

DICCIONARIO DE PSICOMETRIA

"A"

ABCISA. En el eje de las abscisas se representa la variable de observación (en general, variable independiente// Eje horizontal de las coordenadas (Cortada de Kohan, Carro, 1978; Gonzalvo Mainar, 1978)

ADAPTACION DE TECNICAS. En la utilización de técnicas psicométricas provenientes de países extranjeros, tanto para investigación cuanto para diagnóstico y evaluación psicológica general, tienen que considerarse tres cuestiones 1) una comprensión adecuada de las diferencias entre lo émico y lo ético 2) El establecimiento de equivalencias conceptuales (implica que los constructos que se pretenden evaluar tengan el mismo significado en las diversas culturas), lingüísticas (que las palabras utilizadas tengan el mismo significado connotativo y denotativo) y métricas (estudios locales de validez y confiabilidad) 3) El control de las variables relacionadas con la cultura subjetiva del individuo que pueden afectar los patrones de respuesta (por ejemplo, deseabilidad social, autopercepción) (Casullo, 1992; Marín, 1986)

ADMINISTRACION (Tipos de)

- Administración Individual. Se trata de la aplicación de una o más técnicas a un solo sujeto. Pueden ser técnicas autoadministrables o que requieran la intervención directa del entrevistador.
- Administración Colectiva. Consiste en la aplicación de una o más técnicas a un grupo de sujetos en el mismo tiempo y espacio.
- Autoadministración. El/los sujeto/s puede/n contestar o completar el/los protocolo/s de la/s técnica/s sin la necesidad de la intervención del entrevistador.

AMPLITUD. Es una medida de variabilidad o dispersión.

- Amplitud total o extensión. Es la diferencia entre el valor más bajo y el más alto de una serie; o sea, la diferencia entre el límite superior real y el límite inferior real de una distribución de frecuencias. Es la medida más simple y se usa cuando sólo queremos una comparación rápida entre dos distribuciones. No nos informa nada acerca de sus formas. No es muy confiable, sobre todo cuando se trabaja con pocos casos o cuando existen intervalos de frecuencia nula. Sólo se basa en dos observaciones extremas; el resto de las observaciones no intervienen en su determinación.// Es la primera medida de dispersión que se obtiene en cuanto han sido ordenadas las puntuaciones. Se denomina también rango total o gama.

- Amplitud intercuartil. Distancia entre los cuartiles(Q) Q3 y Q1. Incluye el 50 % de los casos. Fórmula: $AI = Q3 - Q1$. Equivale a desviación cuartil (DQ) (Cortada de Kohan, Carro, 1978; Gonzalvo Mainar, 1978) ANALISIS CUALITATIVO. Se refiere al tipo de análisis que toma en cuenta tanto el comportamiento del entrevistado durante la administración de una técnica como su desempeño en la misma. Los siguientes elementos, entre otros, son pasibles de análisis: captación de la consigna, modo de responder a la misma, actitudes, tiempo de respuesta, tolerancia o no al fracaso, modo de formulación de las respuestas, estrategias utilizadas, verbalizaciones espontáneas, tipo de relación que establece con el entrevistador, significación psicológica de los puntajes. Se hace hincapié especialmente en el modo en que ese sujeto se comporta durante la totalidad del proceso de prueba y su singular forma de organizar un cúmulo de respuestas ante diversos estímulos, con la finalidad de describir diferentes conductas para aclarar facetas múltiples del sujeto (Cook, Reichardt, 1986) // Los significados psicológicos atribuidos a los puntajes cuantitativos obtenidos por un sujeto (por ejemplo, CI = 119, clasificación de inteligencia como normal brillante)

DICCIONARIO DE PSICOMETRIA

ANALISIS CUANTITATIVO. Se trata de la puntuación de los datos recogidos a través de diferentes procedimientos (tales como cuestionarios, escalas, tests, entrevistas estructuradas, inventarios, etc.) concebidos para captar las respuestas de los sujetos ante estímulos (verbales o no verbales) prefijados y con opciones establecidas de respuesta. Debe considerarse el puntaje bruto o directo y el puntaje transformado. Para analizar la información se emplean procedimientos estadísticos de diversa complejidad, se asignan valores numéricos a las observaciones y respuestas del sujeto. Dichos valores numéricos pueden, en algunos tests, brindar una lectura diagnóstica a través de criterios preestablecidos o de normas poblacionales (Cook, Reichardt, 1986)

ANALISIS DE ÍTEMS. (Ingl. *ítem analysis*): Comprobación empírica de la utilidad de los ítems componentes de un test. Deben cumplir las condiciones siguientes:

- 1) En su formulación han de expresar claramente la tarea que debe realizarse, sin lugar a dudas, de modo que pueda llevarse a cabo objetivamente (objetividad)
- 2) Las soluciones dadas deben ser independientes del lugar y del tiempo, siendo iguales las de un mismo probando en presencia de las mismas condiciones (estabilidad)
- 3) La tarea debe corresponder a un aspecto esencial de lo que se quiere investigar (validez)
- 4) Cada ítem o tarea debe ser independiente de los demás, evaluable por sí solo (independencia)
- 5) La tarea debe ser clara, concreta, enunciada de un modo enteramente comprensible (concreción)

Estos criterios se aplican en un examen previo de evaluación (ingl. *rating*) del test. Los criterios que se indican a continuación imponen la necesidad de un estudio estadístico basado en la aplicación del test a un número de personas suficientemente grande (de 200 a 500)

6) Grado de dificultad. Los ítems que no son realizados con éxito por ningún probando o que lo son por todos, por ser demasiado difíciles o demasiado fáciles, no sirven para diferenciar. Es apropiada una tarea cuando tiene un grado de dificultad medio, de aproximadamente 50. Significa esto que, en un grupo de examinados, un 50 % dan la solución acertada. Si el grado de dificultad es, por ej. , 96, la tarea es excesivamente fácil. También el promedio de los valores p (grado de dificultad) del total de ítems de un test debe ser, aproximadamente, 50 para obtener una buena distribución de los valores obtenidos en los examinados.

7) Sensibilidad, capacidad de apreciar y medir diferencias pequeñas. Debe existir correlación (de 0,3 a 0,8) entre el resultado global y los resultados de cada test, apreciada con un coeficiente de correlación biserial.

- Análisis de ejercicios, procedimiento para la obtención de valores característicos en ejercicios y para la selección de ítems con el fin de mejorar los criterios de fiabilidad y validez. Se lleva a cabo, según esto, en la fase de construcción de los tests. Para hacer que los resultados del análisis sean efectivos para el empleo posterior de los tests, es necesaria la representatividad de la muestra y de la población a la que se aplica. En principio, valen para los ítems los mismos criterios de calidad que para los tests en general; por ej. , la objetividad, la fiabilidad, la validez. Mientras que la objetividad y la fiabilidad sólo han de medirse raramente, la validez se determina con la selectividad del ítem. Un importante valor característico para los ítems es su dificultad. Con el índice de dificultad se determina el porcentaje de sujetos que con relación a la muestra total ha resultado positivamente el test. En las técnicas de personalidad el desenlace de los ejercicios debe entenderse de modo que la respuesta vaya en dirección de la dimensión a medir.

DICCIONARIO DE PSICOMETRIA

Respecto a los criterios de calidad de un test, puede afirmarse que cuanto mayor sea la objetividad de un ejercicio o un ítem tanto más elevada será su fiabilidad. La dificultad y la selectividad, en cambio, se encuentran en la relación de una curva en U invertida. Esto significa que los ejercicios muy "difíciles" y muy "fáciles" son poco selectivos. Los valores característicos obtenidos en el análisis de ítems se emplean para una selección o una revisión de ítems.// Proceso consistente en evaluar ítems de prueba aisladas con arreglo a ciertas características. Suele implicar la determinación del valor de dificultad y del poder de discriminación del ítem, y muchas veces su correlación con algún criterio externo (Dorsch, 1991, Machrens, Schmann, 1982)

ANALISIS FACTORIAL. Cualquiera de los varios métodos que se utilizan para analizar las intercorrelaciones existentes entre una serie de variables como las puntuaciones alcanzadas en tests psicológicos. El análisis factorial intenta explicar estas intercorrelaciones en términos de algunos "factores" subyacentes, cuyo número es de preferencia menor que el de las variables originales, y revela qué proporción de la variación observada en cada una de las medidas originales se debe a, o se asocia con, cada uno de los factores hipotéticos. El análisis factorial ha contribuido a una mejor comprensión de la organización de los componentes de la inteligencia, aptitudes y personalidad: también marcó el camino hacia la elaboración de pruebas más "puras" de los diferentes componentes.// Designación de diversos métodos mediante los cuales se estudian datos empíricos y correlaciones entre datos aislando factores que intervienen en ellos. El material empírico debe ser de la misma clase o proceder de un área común (por ej., resultados de pruebas de inteligencia). Cada observación ha de ponerse en una relación matemática con cada una de las restantes observaciones. Se efectúa esto en la mayor parte de casos mediante correlaciones y se obtiene una matriz de correlación. Algunos investigadores, basándose en determinadas premisas matemáticas, efectúan una construcción geométrica de dos ejes cruzados perpendicularmente en los que pueden representarse óptimamente todos los valores hallados en el cálculo. Los ejes (factores) de este sistema no se conciben, por lo general, únicamente como un medio descriptivo, sino que se les atribuye una cierta realidad. Así, por ej., los ejes de referencia deducidos de tests de inteligencia se interpretan como capacidades que corresponden originariamente a actividades aisladas. La principal área de aplicación es la psicología diferencial. Pero también se investigan con métodos de análisis factorial la percepción (Thurstone) y los datos de fisiología cerebral, entre otros temas. Se habla del análisis factorial como escuela psicológica cuando se emplea este método como medio de investigación básico en todos los ámbitos de la psicología. Pero el campo de aplicación predilecto del análisis factorial es el estudio del número y clase de capacidades y cualidades generales no derivadas de otras. Con el análisis factorial puede darse a esta cuestión una respuesta puramente empírica y objetiva, en lugar de presentar clasificaciones apriorísticas y arbitrarias. El material consiste en una gran cantidad de resultados de tests representativos de una determinada área, obtenidos con muestras también representativas. Se correlacionan los resultados de cada test con los resultados de cada uno de los restantes y se "deducen" mediante ciertos procedimientos matemáticos cierto número de factores, respecto a los cuales se buscará el establecimiento de correlaciones. Dar a estas magnitudes matemáticas un sentido psicológico es otra cuestión. Qué capacidades o cualidades generales resulten del análisis factorial depende de los procedimientos matemáticos utilizados y de los tests y las muestras que han servido para el estudio. El objetivo ideal del análisis factorial no podrá alcanzarse hasta que concuerden todos los resultados relativos a número y clase de factores considerados como generales. Por ahora sólo parcialmente se ha conseguido esto.

DICCIONARIO DE PSICOMETRIA

En cuanto a la historia del análisis factorial, señalemos que se inició con las investigaciones de Spearman (1904) sobre la inteligencia general. Con el método, por este investigador elaborado, de las diferencias tetrádicas, mostró que en todas las tareas de actividad de la inteligencia puede aislarse un factor de inteligencia general (factor general, factor *g*). Si se elimina este factor, queda un factor específico (factor *e*) para cada tarea. Esta concepción de una organización que va de un factor general a factores de creciente especificación ha recibido la denominación de "teoría jerárquica de las capacidades". Aparecen en ella, además, los factores *c* y *w*, de los que el primero representa la capacidad de orientación, de saber buscar diversos modos de resolución de un problema (en ingl., *cleverness*), y el segundo corresponde a la voluntad (en ingl. *will*). Todavía hoy cuenta esta teoría con defensores en la llamada escuela inglesa, aunque con algunas diferencias respecto a factores aislados, basándose siempre en el método de Spearman. Se ha revisado especialmente el supuesto de que después de excluir el factor *g* quedan solamente factores específicos, válidos únicamente para tareas determinadas. Se ha comprobado que pueden distinguirse por lo menos dos factores de grupo. Un factor, que se relaciona especialmente con tareas del tipo de las que se hacen en la escuela, correspondiente al pensamiento teórico y al verbal, y otro factor que aparece más bien en tareas de tipo práctico, principalmente mecánicas y físicas (Vernon). Una variación de este método es el análisis factorial múltiple, elaborado por Thurstone, que se basa en el supuesto de la igualdad de valor de todos los factores. No se deducen los factores sucesivamente sólo por correlaciones según el orden de su generalidad, sino que se parte de todas las correlaciones simultáneamente (matriz de correlación) y se tratan como sistema cerrado. Las correlaciones se expresan geoméricamente por el ángulo que forman partiendo de un centro u origen común (método centroide). En esta construcción, todas las realizaciones de tests análogas aparecen con distancias angulares pequeñas, mientras que las disímiles tienen distancias mayores. La construcción espacial que resulta es concebida por Thurstone como "espacio matemático" o *supra espace*, cuyas dimensiones corresponden al número de factores que intervienen. Los ejes de referencia de este espacio deben situarse de tal modo (rotación de ejes) que presenten el sistema de resultados de tests en condiciones de óptimo estudio y bajo criterios determinados (principio mínimo). Estos ejes de referencia corresponden por su número y situación al número y clase de capacidades (de modo que se equivalgan factor, vector y dimensión) que han intervenido en todos los tests. Si se parte de tareas aisladas, se describe cada una por su posición en este sistema de coordenadas. Thurstone, basándose en sus investigaciones sobre las capacidades humanas, llegó a la conclusión de que deben distinguirse ante todo los siguientes factores fundamentales: N, factor numérico, habilidad en cálculos; V, comprensión del lenguaje; F, fluidez en la expresión verbal de los pensamientos; S1, S2, S3, factores visuales diversos, espaciales y otros. Meili, por su parte, postula la existencia de cuatro factores que intervienen en proporciones distintas en todo trabajo de la inteligencia. Son los factores de plasticidad, de complejidad, de totalidad y de fluidez. Se corresponden con variables esenciales elaboradas por la psicología experimental del pensamiento y pueden interpretarse, por otra parte, como una síntesis de los factores de Thurstone. Para las pruebas de validez de tests psicológicos, ofrece el análisis factorial posibilidades de substituir la concordancia con otros criterios empíricos (por ej., rendimiento en una asignatura escolar), para determinar la participación del factor que mejor se corresponde con la capacidad que se prueba. Dar una denominación psicológicamente significativa a los factores, que representan una base irreductible para todas las actividades es muy difícil. Puede darse nombre a algunos porque se observa que intervienen en grado especialmente elevado en ciertos tests (análisis de *cluster* o agrupamientos), pero otros no pueden interpretarse psicológicamente porque intervienen con igual "carga" o participación en las más diversas tareas. Como se

DICCIONARIO DE PSICOMETRIA

mencionó antes, Thurstone aisló diferentes factores respecto de la inteligencia. Cattell, por su parte, transfiere el análisis factorial a la psicología de la personalidad y halla doce cualidades caracteriales primarias, independientes entre sí, que tienen aproximadamente las significaciones siguientes:

- 1) Calor afectivo - frialdad afectiva.
- 2) Inteligencia general - necedad.
- 3) Estabilidad emocional - labilidad emocional.
- 4) Afirmación de la personalidad propia - subordinación.
- 5) Optimismo - depresión.
- 6) Sensibilidad - prosaica indiferencia.
- 7) Sentido de finura - tosquedad.
- 8) Sentido de responsabilidad, actuación consciente - acción subalterna, sin responsabilidad propia.
- 9) Amabilidad, exenta de suspicacias - actitud retenida, suspicaz.
- 10) Energía - flaqueza.
- 11) Excitabilidad emocional - flema.
- 12) Confianza - desconfianza.

En general, estos factores son difícilmente definibles con expresiones del lenguaje corriente y no concuerdan bien con los tipos designados con los términos indicados. En los casos ideales, a cada cualidad en el sentido del análisis factorial corresponde una base biológica irreductible. La falta de concordancia de las consideraciones teóricas expuestas por diversos autores, junto con la diversidad de las bases matemáticas, es a menudo causa de diversidad en las denominaciones. Así, pudo comprobar Eysenck que la mayor parte de los estudios de análisis factorial conducen a resultados análogos cuando la semejanza entre los factores se vuelve a someter al análisis factorial para obtener factores de segundo orden. Estos concuerdan en general con los factores complejos hallados por él (neuroticismo, introversión, extraversión).

El análisis factorial orientado a la deducción de capacidades y cualidades generales se basa en la correlación de características en el grupo de personas investigadas.

Este modo de proceder se designa ordinariamente como "técnica R". Se emplea la "técnica Q" o "análisis factorial inverso" (Stephenson, 1939) cuando se correlacionan personas con un grupo de características.

Los factores de esta manera obtenidos se fundamentan en la semejanza entre personas y constituyen una tipología.//

Problemas y presupuestos en la aplicación del análisis factorial:

- 1) Antes del análisis debe delimitarse lo más claramente posible el ámbito de variables que hay que analizar.
- 2) Elección representativa de todas las posibles variables del ámbito total. Cada factor de un ámbito debe estar representado al menos por dos variables con cargas considerables.
- 3) Examen de la validez de los presupuestos para el análisis factorial: Las variables se distribuyen con una distribución normal que rara vez es exacta; las regresiones de las variables son con frecuencia no lineales (pero el coeficiente producto-momento comprende sólo los componentes lineales); y cabe dudar seriamente, muchas veces, de que las dimensiones subyacentes en las variables, en tanto no se trate sólo de unidades meramente operacionales, como en el análisis de *cluster*, actúen conjuntamente en forma aditiva. Todas estas dificultades hacen sospechar que el presupuesto para el análisis factorial no suele cumplirse y por ello su aplicación no será legítima ni sus resultados serán válidos. En una serie de análisis de matrices de correlación construidas, cuya composición factorial se establecía a priori (así, Thurstone), se obtuvieron sin embargo, a pesar del deliberado no cumplimiento de los

DICCIONARIO DE PSICOMETRIA

presupuestos, unos resultados en los que las dimensiones subyacentes podían aún reconocerse claramente en cuanto al número y la clase. Esto justifica una cautelosa aplicación del análisis factorial (en su forma actual) a objetos complejos de investigación, como los que se dan en la psicología y ciencias afines.

4) La interpretación de los resultados apunta en primer término al esclarecimiento de la importancia de los factores obtenidos; sólo sobre la base de una representación provisional de la naturaleza de un factor pueden elaborarse hipótesis fecundas para ulteriores investigaciones causales (observaciones experimentales y sistemáticas). El procedimiento usual para la verificación de una hipótesis sobre la importancia de un factor es la consideración de las variables que en un factor ofrecen cargas elevadas o nulas.

Entre los críticos del análisis factorial, pero también entre los que hacen uso de él, no existe aún acuerdo y claridad sobre el alcance científico de este método, y particularmente sobre el problema de si mediante él se obtienen datos "reales" o simples síntesis meramente operacionales de relaciones observadas entre procesos o estados reales. No puede tratarse de una alternativa fundamental entre factores y dimensiones reales, de una parte, y operacionales o convencionales de otra; pero hay muchos resultados obtenidos mediante análisis factorial que a consecuencia de errores de experimentación en el sentido más amplio del término sólo pueden interpretarse en forma deficiente y, por tanto, de modo provisional.

Ver Yela, Mariano: La técnica del análisis factorial. Madrid, Biblioteca Nueva, 1957, 198 págs. Se recomienda esta obra para los interesados en el tema (Dorsch, 1991; W. Maehrens, S. Sehmman, 1982; Gonzalvo Mainar., 1978; Yela, 1957)

AQUIESCENCIA. En los tests, la tendencia a reaccionar con una actitud, espontánea, e irreflexiva de conformidad, que se observa, sobre todo, en la perseveración de un tipo de respuesta. Por ejemplo, contestar a todos los ítems de un inventario como "verdadero", "falso", o de un cuestionario como "sí", "no", etc. Altera y falsea los resultados. (Dorsch, 1991)

ANSIEDAD DEL TEST. Sentimiento del sujeto antes y durante la ejecución del test (especialmente tests psicológicos y de rendimiento escolar), que en casos extremos puede llegar a deformar la "verdadera" capacidad. Para la teoría y la construcción de tests, esta ansiedad se presenta como problema de validez suficiente (la fiabilidad no debería quedar afectada por la ansiedad del test si las condiciones de ejecución no varían en relación con la ansiedad)

AZAR. Causa probable de un hecho aislado imprevisible. Sucesos al azar. Sin. sucesos aleatorios.

"B"

BAREMO: Mediante investigaciones sistemáticas en un gran número de personas se establecen los valores centrales o medios y valores de dispersión para determinadas prácticas (intelectivas, por ej.) contenidas en un test psicológico. Estas normas o valores normales sirven para evaluar las observaciones efectuadas en un individuo particular o en grupos de individuos. Véase "Normas Intragrupo"

"C"

CARACTERISTICA. Es una propiedad o una cualidad de un individuo (Botella, León y San Martín, 1992) // Fenómeno, categoría o clase que pueden ser objeto de

DICCIONARIO DE PSICOMETRIA

tratamiento y estudio estadísticos. Los fenómenos *cualitativos* se denominan *atributos* (por ejemplo, sexo); los *cuantitativos variables*, siendo su expresión numérica (Dorsch, 1991)

CARACTERISTICAS DE DEMANDA. Aquellos indicios asequibles a un sujeto en un experimento que le pueden permitir determinar el "propósito" del experimento; los indicios que le permiten a un sujeto inferir lo que el experimentador "espera" de él; un tipo de variable de confusión Anastasi, 1998)

CENTIL. Valor de la distribución ordenada bajo el cual se halla un determinado porcentaje de casos, p. ej.: el centil 80 significa que es el punto bajo el cual están el 80% de las puntuaciones. En la práctica se habla de percentiles (percentil 90, percentil 80, etcétera). Cuando hablamos solo de centil, se hace referencia, en general, a un rango centil, o porcentaje bajo el punto mencionado. (Ej.: centil 80 = 80% bajo el centil 80). Véase "Percentil"

COEFICIENTE DE CONFIABILIDAD. Coeficiente de correlación entre dos formas de una prueba, entre las puntuaciones obtenidas en dos administraciones de la misma prueba, o entre las dos mitades de una prueba, debidamente corregida. Todos éstos son coeficientes de confiabilidad propiamente dichos, a pesar de medir aspectos un tanto diferentes de la confiabilidad (Anastasi, 1996). // Nombre de diversas medidas propuestas por varios autores para medir la consistencia interna de un test, su grado de homogeneidad respecto a la característica que se quiere medir con el test (Test alfa de Cronbach, Hoyt, Richardson, Rulon). La fiabilidad de un test se define como la variación de la puntuación verdadera con respecto a la puntuación observada, calculada a través de la razón entre las respectivas varianzas. Por lo tanto, la relación sd^2_v / sd^2_x , es por definición la medida de la fiabilidad. Si se tiene en cuenta que esta relación entre varianzas es formalmente igual al cuadrado del coeficiente de correlación lineal obtenido entre las puntuaciones observadas y las verdaderas, y la significación dada a dichos índices, se define el coeficiente de fiabilidad como: $r^2_{xv} = sd^2_v / sd^2_x$. Dando al coeficiente de correlación r_{xv} la denominación es la de índice de fiabilidad. La relación dada anteriormente, en la que se expresa el coeficiente de fiabilidad de un test como la relación entre la varianza de las puntuaciones verdaderas y la de las observadas, así como su expresión alternativa de un coeficiente de correlación lineal entre las puntuaciones observadas y las verdaderas, no son plausibles para ser utilizadas en la práctica, puesto que las puntuaciones verdaderas de los sujetos son inobservables. Por lo tanto, es necesario buscar otras expresiones equivalentes en las que se pueda operar con valores observables, y a través de las cuales, se calculen dichos coeficientes o índices. La ausencia de fiabilidad se identifica con el valor $r_{xx'} = 0$. La fiabilidad perfecta se identifica con el valor $r_{xx'} = 1$, y supone teóricamente que, no existe varianza de error y que toda la varianza de la puntuación observada se debe a la puntuación verdadera. Todos los demás casos son casos intermedios, diciendo que hay una escasa fiabilidad a medida que nos aproximamos al valor cero, y que la fiabilidad es elevada, cuando obtenemos para $r_{xx'}$ valores próximos a uno. Un test de fiabilidad moderada podría tener asociado un coeficiente de fiabilidad con valor $r_{xx'} = 0,62$, por ejemplo, u otro similar, no pudiéndose dar normas fijas ni valores en el intervalo de $r_{xx'}$ que establezcan, con exactitud, si una prueba debe ser o no admitida como fiable, salvo que estos valores no sean significativamente diferentes de cero. Esta significación se obtiene, a determinados niveles de probabilidad, mediante el correspondiente contraste de hipótesis, bajo la hipótesis nula $r_{xx'} = 0$. El nivel de precisión exigible a la prueba está en función del tipo de prueba y del objetivo, u objetivos, para el que esa prueba se construye.

DICCIONARIO DE PSICOMETRIA

Una forma rigurosa de proceder es la siguiente. Una vez decidido cuál es el nivel mínimo de recisión exigible a la prueba, se materializa esa exigencia fijando un cierto valor k (por ejemplo, 0,80) para el coeficiente de fiabilidad, por debajo del cual una menor precisión no sería admisible. Si el valor experimentalmente obtenido para $r_{xx'}$ es un valor mucho menor que el fijado k , se rechaza como admisible esa prueba, puesto que no alcanza el nivel mínimo de precisión deseado. Sin embargo, la apreciación de las diferencias entre el valor observado y el esperado puede ser equívoca. En el caso de que el valor experimentalmente obtenido para $r_{xx'}$ sea mayor o menor que el fijado k , se debe hacer un contraste de hipótesis estadístico que nos revele a qué nivel de confianza podemos decir que el valor obtenido es significativamente diferente del valor dado k , y si el test debe ser admitido o no como fiable, bajo las condiciones previamente establecidas. La forma práctica de realizar esa correlación, que va a proporcionar el valor del coeficiente de fiabilidad $r_{xx'}$, depende de la forma en que se obtengan las medidas supuestamente paralelas.

- Condiciones para estimar empíricamente un coeficiente de fiabilidad (Martínez Arias, 1996)

1) Tamaño muestral. Aunque la forma matemática exacta de la distribución muestral de un estimador de fiabilidad es desconocida, se espera que el valor del error típico de estimación de un parámetro esté inversamente relacionado a la raíz cuadrada del tamaño muestral; es decir, para reducir a la mitad el error típico del coeficiente de fiabilidad, necesitamos cuadruplicar el tamaño de la muestra

2) Representatividad de la muestra. En lo posible, los examinados serán elegidos de modo que representen a una población bien definida

3) Independencia experimental de las medidas. Esto implica que la medida de un examinado no debe influir ni estar influida sobre la medida de otro examinado. Es un tema particularmente relevante cuando las administraciones son colectivas

4) Condiciones estándar de administración. Las circunstancias en que se recogen los datos deben ser idénticas a aquellas con las que se administra el test en la práctica (tiempo, forma de aplicación, instrucciones, nivel de ruido ambiental, etc.)

5) Disponer de dos o más medidas. Las puntuaciones pueden proceder de dos o más aplicaciones del mismo test, de formas alternativas o de la división del test en partes Véase también "Confiabilidad"

COEFICIENTE DE CORRELACION. El grado de relación puede ser representado numéricamente por el coeficiente de correlación. Un coeficiente de correlación perfecta es + 1,00. Un coeficiente de correlación negativa perfecta es - 1,00. Una relación debida puramente al azar es cero. Raramente se encuentran los 7 coeficientes de correlación perfecta de + 1,00 o de - 1,00, particularmente cuando se trata de rasgos humanos. Aunque algunas relaciones tienden a mostrarse bastante consistentes, existen excepciones que reducen el coeficiente medido desde - 1,00 o + 1,00 hacia cero.

- Interpretación de los coeficientes de correlación. Existen falsas interpretaciones y límites que deberían tenerse en cuenta al considerar el significado de un coeficiente de correlación. Un coeficiente no implica una relación de causa - efecto entre las variables. Coeficientes de correlación altamente positivos, estadísticamente auténticos, pueden ser explicados por un factor común (por ejemplo, crecimiento de la población). Antes de que pueda captarse una relación causal entre variables, debe ser establecida por un análisis lógico. El proceso de cálculo del coeficiente de correlación cuantifica solamente la relación que ha sido previamente establecida. Una relación de un tipo determinado sólo puede existir dentro de ciertos límites, y debe ser comprobada para determinar si es lineal o curvilínea. El análisis lógico debe ser también aplicado para

DICCIONARIO DE PSICOMETRIA

determinar qué otros factores pueden participar significativamente en el proceso normal.

- Significado de los coeficientes de correlación. Suele admitirse como criterio general lo siguiente para la evaluación del significado de los coeficientes de correlación:

Valor del coeficiente Relación

00 a 0,20 despreciable

0,20 a 0,40 baja o ligera

0,40 a 0,60 moderada

0,60 a 0,80 sustancial o marcada

0,80 a 1,00 alta o muy alta

Lo anterior es un análisis básico. El significado de un coeficiente de correlación depende de la naturaleza de los factores relacionados, del número de casos implicados, de la amplitud de las puntuaciones y de los propósitos de la aplicación de la medida.

COEFICIENTE DE PRODUCTO-MOMENTO DE PEARSON. Coeficiente de correlación lineal. Para dos variables X e Y este coeficiente se define como el cociente entre la covarianza de las dos variables y el producto de sus desviaciones típicas (Para su fórmula y cálculo consúltese Santisteban, Cortada de Kohan, Pearson) Valores y significado. El coeficiente de correlación lineal r sólo puede tomar valores entre -1 y 1.

Valores:

El valor $r = 0$ indica ausencia de correlación.

El valor $r = +1$ indica correlación máxima positiva y dependencia lineal.

El valor $r = -1$ da la correlación máxima negativa y dependencia lineal.

La proximidad al valor cero indica débil correlación.

La proximidad a los valores extremos, ya sea a -1 ó a +1 indica fuerte correlación.

Signo: El signo indica el sentido directo (positivo) o inverso (negativo) de la correlación. Cuando el signo es positivo ambas magnitudes varían en el mismo sentido, ya sea creciente o decreciente. Cuando el signo es negativo, aun cuando exista muy alta correlación entre ambas variables ésta se hace de manera que cuando una aumenta la otra disminuye, es decir que covarían de forma inversa. Ejemplo:

Entre las variables X = saldos en cuenta e Y = pagos realizados durante el mes, se ha obtenido un coeficiente $r = -0,92$. Este coeficiente indica alta correlación negativa. Ambas magnitudes están muy correlacionadas negativamente, es decir, covarían en sentido inverso.

Consecuencia: El que el coeficiente de correlación lineal r sólo pueda tomar valores entre -1 y +1, indica que la covarianza entre dos variables nunca puede ser mayor que el producto de las desviaciones típicas de ambas variables.

COEFICIENTE DE VALIDEZ. Véase "Validez"

CONFIABILIDAD. La fiabilidad de un test se refiere a la *consistencia* de las puntuaciones obtenidas por los mismos individuos cuando son examinados con el mismo test en diferentes ocasiones, con conjuntos distintos de elementos equivalentes o bajo otras condiciones variables de examen. Este concepto de fiabilidad es la base del cálculo del *error de medida* de una puntuación sencilla, por lo que podemos predecir la amplitud probable de la fluctuación en la puntuación de un solo individuo como resultado de factores casuales.// El grado hasta donde una prueba es consistente en la medición de lo que pretende medir; seguridad, estabilidad, integridad y relativa ausencia de errores de medición. La confiabilidad suela expresarse mediante algún tipo de coeficiente de confiabilidad o a través del error tipo de medición que se deriva de ella. El concepto de fiabilidad del test se ha usado para abarcar varios aspectos de la

DICCIONARIO DE PSICOMETRIA

consistencia de la puntuación. En su sentido más amplio, la fiabilidad del test indica hasta qué punto pueden atribuirse a errores de medida las diferencias individuales en las puntuaciones del test y hasta qué punto cabe atribuir las diferencias verdaderas de la característica que estamos sometiendo a consideración. Para expresarlo en términos más técnicos, toda medida de la fiabilidad del test denota qué proporción de la *varianza total* de las puntuaciones es *varianza de error*. La importancia del asunto reside, sin embargo, en la definición de la varianza de error. Factores que con una determinada finalidad pudieran considerarse varianza de error, se clasificarían como verdadera varianza con otra finalidad diferente; así, p. Ej., si deseamos medir las fluctuaciones del humor, los cambios diarios que surjan en las puntuaciones de un test de animación - depresión serían adecuados al propósito del test, y por ello constituirían una parte de la *varianza verdadera* de las puntuaciones. Si, por el contrario, el test está destinado a medir características de personalidad más permanentes, las mismas fluctuaciones diarias figurarían bajo el encabezamiento de varianza de error.

En esencia, cualquier condición que no sea afín al propósito del test representa una varianza de error. Así, cuando el examinador trata de mantener uniformes las condiciones de aplicación del test, controlando ambiente, instrucciones, límites de tiempo, *rapport* y otros factores similares, está reduciendo la varianza de error y haciendo más fiables las puntuaciones. Sin embargo, aún dándose las condiciones óptimas, ningún test es un instrumento perfectamente fiable. De aquí que todo test deba ir acompañado del enunciado de su fiabilidad. Esta medida de fiabilidad caracteriza el test que se aplica en condiciones tipificadas y a sujetos análogos a los que formaban el grupo normativo. Deben, por tanto, especificarse las características de dicha muestra, además del tipo de fiabilidad que se midió.// El concepto de fiabilidad, aparecido en la terminología propia de la literatura psicométrica, es un concepto análogo al utilizado en otras ciencias bajo la denominación de precisión. Al hablar de la fiabilidad de un test nos estamos refiriendo a la precisión de ese test, considerado como instrumento de medida, en correspondencia con el concepto de precisión que se tiene acerca de los instrumentos utilizados para las medidas físicas. Teniendo en cuenta que el objetivo de un test es el de poder inferir magnitudes que reflejen los inobservables psíquicos a través de sus manifestaciones observables, hay que ser cuidadosos con la interpretación dada a la fiabilidad pues, a veces, es confundida con otros conceptos como el de adecuación o de validez. La fiabilidad de una prueba o test se refiere, exclusivamente, a la precisión de esa prueba utilizada como instrumento de medida y nunca a si ese instrumento es idóneo para la medida de la aptitud, actitud, destreza o cualquier otro tipo de rasgo que se quiere evaluar a través de esa prueba.

La cuestión de la idoneidad del instrumento de medida con respecto al rasgo que se desea evaluar es un problema de validez, que es otra de las características necesariamente exigibles a las pruebas psicométricas, cuya problemática se trata en otro apartado.

El problema de cómo medir la fiabilidad de los tests se aborda conceptualmente desde la misma perspectiva con la que otras ciencias plantean el problema de la precisión de sus instrumentos de medida, si bien, la metodología de la que la Psicometría hace uso suele ser diferente, llevando a veces a tener que construir otro instrumento análogo e intercambiable al que se denomina test paralelo. El control de la fiabilidad se realiza partiendo del mismo supuesto que en las medidas físicas. Si el instrumento de medida es preciso, medidas repetidas serán semejantes y, salvo errores accidentales, se dará la constancia en las medidas.

La constancia o estabilidad de las puntuaciones se evalúa mediante un coeficiente de correlación lineal entre dos series de puntuaciones, consideradas como paralelas, que usualmente se obtienen de la aplicación de dos formas del test a los mismos sujetos.

DICCIONARIO DE PSICOMETRIA

A los coeficientes de correlación e indicadores de la fiabilidad, se les da la denominación de coeficiente e índice de fiabilidad, respectivamente, expresando el primero la correlación entre dos series de medidas paralelas observadas y, el segundo, el coeficiente de correlación entre las puntuaciones observadas y sus correspondientes puntuaciones verdaderas (Anastasi, 1998)

CONFIABILIDAD, FORMAS DE OBTENCION DE LA. Puesto que la fiabilidad se determina formalmente mediante el coeficiente de correlación lineal entre medidas repetidas, realizadas con el mismo test y en las mismas condiciones, o con tests considerados equivalentes por proporciones medidas paralelas, existen procedimientos, que han venido a ser clásicos en el diseño de las pruebas, para la obtención de la fiabilidad. Estos son los denominados test-retest, formas paralelas y división del test en dos mitades. Cada una de las modalidades tiene asociada una problemática distinta y no responden exactamente al mismo concepto de fiabilidad, pues en tanto que unos diseños nos llevan a su medida en el sentido de la estabilidad de las puntuaciones (test – retest), otros nos indican cual es la consistencia interna de la prueba (división por mitades; formas equivalentes en una sola aplicación). Otra de las anotaciones que debemos de hacer aquí, es la de que las “medidas repetidas”; entre las que son usuales el test-retest, formas paralelas y división en dos mitades, deben cumplir las condiciones formales de paralelismo para que se les utilice como “medidas paralelas”. Si las medidas y las varianzas de las dos secuencias de medidas repetidas no son significativamente iguales, no se puede hacer uso de todas aquellas propiedades de las medidas que fueron deducidas a través de las condiciones de paralelismo.

- División en dos mitades. Se trata en este caso de poner en relación las puntuaciones entre dos partes del mismo test, actuando ambas como si fuesen dos formas paralelas. La división del test en dos mitades se puede llevar a cabo de muy distintas formas, sin embargo, para realizarla correctamente con el propósito de calcular la fiabilidad, hay que tener en cuenta cómo ese test ha sido diseñado. En un test con ítems muy homogéneos, cualquier división en dos mitades se puede aceptar como adecuada. Por el contrario, en aquellos casos en los que el test va creciendo en dificultad a medida que se avanza en su ejecución o bien, en aquellos otros en que sin ser de dificultad creciente, se les ha introducido al principio un pequeño número de ítems de pequeña dificultad con el fin de motivar al sujeto, no sería conveniente el realizar una división arbitraria en dos mitades sin tener en cuenta dichas características. Una de las divisiones en dos mitades, que tiende a paliar las dificultades anteriormente mencionadas, es aquella en que el criterio es el de elegir por un lado los elementos que ocupan los lugares pares del test y por otro los impares, para utilizarlos como dos formas paralelas. El coeficiente de confiabilidad se obtiene correlacionando las puntuaciones logradas en una mitad de una prueba con las alcanzadas en la otra mitad de la misma, y aplicando la fórmula Spearman – Brown para compensar la doble longitud de toda la prueba. Los coeficientes de confiabilidad de división en dos mitades a veces se llaman medidas de consistencia interna de una prueba; sólo se refieren al muestreo de contenido y no a la estabilidad de la prueba a través del tiempo. Este tipo de coeficiente de confiabilidad es impropio para las pruebas en que la rapidez es importante. Este procedimiento de división en dos mitades es adecuado en cuanto a lo que la forma de obtener la correlación se refiere, si bien, hay que hacer notar que la fiabilidad así calculada no tiene el mismo sentido que en ese otro caso de utilización de dos test diferentes como formas paralelas.

- Test-retest. Bajo esta denominación se indica que las puntuaciones repetidas X y X' se han obtenido aplicando el mismo test a los mismos individuos y en las mismas condiciones en dos ocasiones diferentes en el tiempo. Es evidente que, si se determina que el test debe ser aplicado a los mismos individuos y en las mismas condiciones, hay que ser extremadamente cuidadosos en la verificación de que esas premisas se

cumplen. Uno de los problemas a los que con mayor frecuencia se hace alusión en la literatura es al posible efecto de aprendizaje del sujeto, y que influirá en sus resultados, cuando realiza por segunda vez la prueba. Por lo tanto una fuente de varianza de error a controlar es el muestreo del tiempo. Este efecto se elimina, o al menos queda paliado, si se deja transcurrir un tiempo suficiente entre ambas aplicaciones de la prueba. Esta solución sin embargo no siempre es plausible, bien porque no se pueda tener disponibles a los mismos sujetos en un tiempo posterior conveniente, o bien porque el rasgo que se está evaluando pueda sufrir también variaciones en ese intervalo de tiempo, por ser función del tiempo (es el caso de las variables evolutivas) o porque se produzca alguna intervención que lo modifique.

Si la prueba es fiable, y las condiciones de aplicación de la prueba en las dos ocasiones son esencialmente iguales, se observará una significativa estabilidad en las puntuaciones. En general, se utiliza el coeficiente de correlación producto – momento de Pearson para obtener el coeficiente de confiabilidad test – retest. Actualmente, la expresión “test - retest” se utiliza también para describir la administración de diferentes formas de una misma prueba, en cuyo caso, este coeficiente de confiabilidad se combina con un coeficiente de formas alternas. En cualquier caso, intervienen (1) fluctuaciones en el tiempo y en la situación de prueba, así como (2) cualquier efecto que ejerza la primera prueba sobre la segunda. Cuando es considerable el intervalo de tiempo transcurrido entre las dos administraciones, como por ejemplo, un intervalo de varios meses, un coeficiente test-retest de confiabilidad no sólo reflejaría la consistencia de medición inherente a la prueba, sino también la estabilidad de la característica medida en los examinados.

- Formas alternas, equivalentes o paralelas. Se entiende por formas alternas, equivalentes o paralelas de un mismo test a dos tests contruidos de forma que resulten ser prácticamente iguales, de manera que las puntuaciones de los sujetos en ambos sean comparables ítem a ítem. Se trata, entonces, de la correspondencia o correlación entre los resultados obtenidos en formas alternas de una prueba; por lo tanto, es una medida del grado hasta donde estas dos formas son congruentes o confiables en la medición. Si el método utilizado es el de formas alternas inmediata, los intervalos de tiempo que separen a las dos administraciones deben ser relativamente cortos, de forma tal que la fuente de varianza de error, esto es el muestreo de contenido, quede relativamente menguada. También, habría que tener en cuenta la longitud de las pruebas, por si deben de ser aplicadas en una misma sesión o en varias, para que las respuestas en alguna de ellas no estén contaminadas por un efecto de fatiga en los sujetos. En cambio, si se utiliza el de formas alternas demorada, los intervalos de tiempo serán más largos y aquí habrá que controlar como posible fuente de varianza de error tanto el muestreo de contenido cuanto el de tiempo. El gran inconveniente que tiene esta modalidad es la de que multiplica el esfuerzo en el proceso de construcción y selección de ítems, si bien, en muchos casos, ya se han construido éstos con anterioridad y se tiene disponible un gran banco de ítems de los que se conocen sus características.

- Confiabilidad de Kuder – Richardson y coeficiente alfa. Las fórmulas deducidas por Kuder y Richardson están destinadas a estimar la confiabilidad de una prueba. Estas fórmulas se basan en la consistencia inter – ítem y sólo requieren una administración de la prueba. La más usada de ellas, la fórmula 20, requiere información basada en el número de ítems de la prueba, la desviación estándar de la puntuación total y la proporción de examinados que aprobaron cada ítem. Las fórmulas de Kuder-Richardson no deben utilizarse con pruebas en donde la velocidad de resolución es importante. El coeficiente alfa de Cronbach, por su parte, suele utilizarse para estimar la consistencia interna de un test y se trata de un método basado en la covarianza de los ítems. Los procedimientos descriptos están basados en el supuesto de que las puntuaciones de las partes son esencialmente equivalentes.

DICCIONARIO DE PSICOMETRIA

- Confiabilidad entre calificadores. Una fuente importante de varianza de error, sobre todo en los instrumentos clínicos, es la varianza del examinador. La confiabilidad del calificador se obtiene con una muestra de tests calificados en forma independiente por dos o más examinadores. Los dos (o más) resultados se correlacionan de acuerdo a la forma usual y el consecuente coeficiente de correlación es la medida de la confiabilidad del calificador. Esta clase de fiabilidad suele calcularse en la investigación de instrumentos de calificación subjetiva. (Anastasi, 1998; Martínez Arias, 1996; Machrens, Sehmman, 1982)

CONFIANZA, LIMITES DE. Superior e inferior, derivados de un ejemplo y utilizados para efectuar deducciones acerca de una población; p. ej.: a partir de la media y de su error típico en una distribución, son determinables los límites que permitan afirmar que existe la probabilidad de un 95% de que la media de esta población cae dentro de estos límites.

CONSISTENCIA INTERNA. Grado de relación existente entre los ítems de una prueba; consistencia en un muestreo de contenido. Véase también "Confiabilidad"

CONSTANTE. Un término en una fórmula matemática que no varía // Característica que puede manifestarse solamente bajo una única modalidad.

CONSTRUCCION DE TESTS. Por lo general en la construcción de tests con una aproximación centrada en el sujeto, suelen seguirse los siguientes pasos 1) Identificar el propósito o finalidad para el que se utilizará el test 2) Especificar a priori las principales restricciones con las que deberá operar el test (tiempo, medio, situaciones, etc.) 3) Identificar las conductas observables representativas del constructo (dominio o atributo de interés; análisis de contenido, revisión de las investigaciones publicadas, incidentes críticos, observaciones directas, juicio de expertos) 4) Especificar el formato de los ítems (elección múltiple, selección o identificación, ordenación, sustitución/corrección, completamiento, etc.) 5) Estudio piloto y tratamiento de los datos del estudio piloto 6) Planificar la tipificación del test (grupos normativos a partir de muestras representativas, tipos de escalas en las que se establecerán las normas – percentiles, puntuaciones típicas, cocientes intelectuales, etc. – , estimación de las propiedades psicométricas del test – confiabilidad y validez) 7) Elaboración del manual del test y otros materiales de apoyo y 8) Considerar cuidadosamente los criterios con consenso internacional para la valoración de un test (Martínez Arias, 1996)

CONSTRUCTO. En psicometría, descripción conceptual hipotética acerca de algún atributo o rasgo psicológico. En términos generales, un concepto o constructo teórico es un contenido figurativo designado por un término o una combinación verbal. Un concepto no es idéntico a los contenidos a los que se refiere su contenido figurativo; tampoco podemos derivar de un mero concepto un enunciado sobre el fenómeno concreto. A través de nuestros conceptos (persona, muerte, deseo, miedo) ordenamos el mundo empírico, pero esta ordenación no necesariamente corresponde a una estructura objetiva de la realidad. La diferenciación entre consciente e inconsciente, entre neurosis y psicosis, son ejemplos de esquemas conceptuales de ordenación cuya deficiencia reconoce hoy quien hace o intenta hacer ciencia. La selección de los conceptos en relación con los fenómenos concretos que intentamos estudiar no es casual ni arbitraria sino que está condicionada por nuestros intereses. Tendemos a percibir de manera preferencial lo que de alguna manera nos resulta importante o significativo. Todo concepto implica elementos descriptivos y valorativos. Hay constructos que no tienen valoración alguna (edad de ingreso a la escuela, por ejemplo). La valoración contenida en un concepto está determinada cultural y

DICCIONARIO DE PSICOMETRIA

socialmente, ideológicamente. Así, por ejemplo, suponen valoraciones socioculturales los conceptos relativos a enfermedad orgánica, enfermedad psíquica, la aceptación de la existencia de lo psíquico independiente de lo orgánico o social. Los elementos valorativos pueden ser factores que perturben la tarea de investigación de determinados fenómenos mientras permanecen implícitos, cuando no los explicitamos. ¿Qué función cumplen los conceptos? Mayntz y col. (1985), en su trabajo sobre métodos de investigación en Sociología, se refieren a cuatro funciones básicas:

1. Ordenar nuestras percepciones.
2. Evaluar lo que percibimos.
3. Guiar nuestras acciones.
4. Facilitar la comunicación.

Un concepto y su definición operacional pueden ser distintos en su extensión y en su contenido semántico. Lo que una definición operacional (como en los constructos) permite conocer puede ser menos extenso que el contenido del concepto.

CONTROL, LISTA DE. La lista de control es un cuestionario breve que comprende cierto número de rasgos y sirve de ayuda para la sistematización y delimitación de la observación. Por ejemplo:

Entre los rasgos de conducta especificados en la siguiente lista de control (ver cuadro) indique marcando mediante una + el grado correspondiente que determina, de acuerdo con su observación del alumno.

RASGO DE CONDUCTA OBSERVADO	SI	NO	DUDOSO
Participación en los juegos			
Ayuda a los compañeros			
Colaboración en trabajos de equipo			
Puntualidad			
Confianza hacia el maestro			
Orden en el aula			
Trabajo cuidadoso			
Alegría			
Veracidad			
Delicadeza en el trato			

CORRECCION PARA EL FACTOR ADIVINATORIO (para el factor azar). Reducción de la puntuación para respuestas equivocadas, a veces aplicada en la calificación de preguntas de falso – verdadero o de elección múltiple. Estas fórmulas de calificación están destinadas a desalentar la costumbre de adivinar las respuestas y a hacer que sean más justos los rangos asignados a los examinandos en términos de sus conocimientos reales. Actualmente, se utilizan mucho menos que cuando se inició la evaluación mediante pruebas (Machrens y Sehmman, 1982)

CORRELACION. La correlación es la relación concomitante entre dos o más variables, o sea, entre dos o más series de datos. El grado de relación puede ser medido y representado por el coeficiente de correlación designado por la letra griega *rho* o por *r*. Pero esta relación no aparece siempre del mismo modo, puede ser que a un cambio de uno de los fenómenos corresponda siempre un cambio exactamente determinado en el otro. Entonces se dice que existe correlación perfecta entre ambos. Pero también se da la correlación imperfecta cuando la variación de un fenómeno suele producir variación en el otro sin que esto ocurra siempre. Cuando no existe relación alguna entre dos fenómenos se dice que la correlación es nula o que no hay correlación.

DICCIONARIO DE PSICOMETRIA

La correlación, sea perfecta o imperfecta, puede producirse de dos maneras: a) Correlación directa o positiva: cuando a un incremento positivo o negativo de un fenómeno corresponde un incremento, positivo o negativo respectivamente, en el otro. b) Correlación inversa o negativa: cuando a un incremento positivo de un fenómeno corresponde un incremento negativo del otro, y a la inversa. Si se observa que los alumnos con elevado CI tienden a obtener puntuaciones altas en tests de cálculo, mientras que los de bajo CI tienden a puntuaciones inferiores, se dice que los factores de inteligencia medidos y las puntuaciones en los tests de cálculo se hallan en una correlación positiva. Los rasgos pueden hallarse en correlación negativa cuando, comparando dos variables, la una aumenta mientras la otra disminuye. Cuando las relaciones entre dos variables son solamente por azar, decimos que no existe ninguna correlación. Los siguientes pares se hallan en correlación positiva:

Costo de la tierra – Productividad por Ha.
Edad – Costo del seguro de vida.
Ingresos familiares – Nivel de Vida.
Inteligencia – Notas escolares.

Otros se hallan corrientemente en correlación negativa. Cuando el uno aumenta, el otro tiene tendencia a disminuir.

Antigüedad de un automóvil – Precio en el mercado.
Tiempo de práctica – Número de errores mecanográficos.
Existen otros rasgos que probablemente no tienen una relación mensurable:
Peso corporal – Inteligencia.
Número del calzado – Sueldo mensual.

Véase también “Coeficiente de correlación”

CRITERIO. Un estándar usado para evaluar la validez predictiva de una prueba.// Una serie de puntuaciones con arreglo a las cuales se comprueba el éxito de un test predictivo. Algunos de estos criterios son: 1) Condiciones, prácticas o rendimientos dominantes de una o varias unidades de trabajo diferentes, pero comparables. 2) Lo que los expertos consideran como las mejores condiciones o procedimientos. 3) Lo que determina como normas apropiadas un grupo profesional, un departamento que concede títulos, una comisión o un cuerpo científico deliberativo. 4) Leyes o normas que han sido estudiadas o promulgadas por un cuerpo legislativo o cuasi-legislativo. 5) La demostración obtenida en la investigación. 6) La opinión pública.// Un modelo elegido como la meta que ha de ser alcanzada en la tarea de aprendizaje, p. ej.: número de recorridos sin error en un laberinto.

CUARTIL (Q). Medida de puntuación relativa o de dispersión que divide la serie de valores en cuatro partes o grupos. Los cuartiles dividen en cuatro partes el 100% total de la muestra. La distancia intercuartil es siempre de un 25%. Una puntuación que se corresponda con el cuartil primero (Q1) deja por debajo de sí un 25% de los casos. La que se equipara con el Q2 un 50%, coincidiendo con la media aritmética. El Q3 deja por debajo de sí un 75%.

CUESTIONARIO. Los términos cuestionarios e inventarios se suelen utilizar como sinónimos a pesar de que hay diferencias marcadas en la forma de su presentación. Un cuestionario supone un conjunto de preguntas que el sujeto tiene que responder en forma positiva o negativa (Casullo, 1992). Por ejemplo, el ítem 8 del Cuestionario de Personalidad de Eysenck: “¿Algunas veces, se siente mal sin motivos?”, al que el sujeto debe contestar en función de “Sí” / “NO”

DICCIONARIO DE PSICOMETRIA

CULTURA, TEST DE AUTENTICA. Tipo de pruebas de inteligencia ideado de modo que reduzca al mínimo las inclinaciones debidas a las diferentes experiencias de los niños educados en una cultura rural que urbana o en una clase media o superior (sin.: tests de cultura libre).

CURVA NORMAL. La forma prevista de la distribución normal. Los análisis matemáticos más antiguos de la teoría de la probabilidad datan del siglo XVIII. Abraham De Moivre, un matemático francés, descubrió que un principio matemático explicaba la probabilidad asociada con diversos juegos de azar. Elaboró la ecuación y la pauta curvilínea que la describe. Durante el siglo XIX, un astrónomo francés, Laplace, y un matemático alemán, Gauss, llegaron independientemente al mismo principio y lo aplicaron con mayor extensión en áreas de medida de las ciencias físicas. Partiendo de las aplicaciones limitadas hechas por esos astrónomos - matemáticos, la teoría de la probabilidad ha sido aplicada a los datos obtenidos en los campos de Biología, Psicología, Sociología y Economía. Se ha encontrado también que describe las fluctuaciones o posibilidades de error en la observación y en la medición.

- Aplicaciones prácticas. Entre las aplicaciones prácticas de la distribución normal, se encuentran:

1. Evaluar la normalidad de una distribución de frecuencias conocidas. Esto puede realizarse superponiendo sobre el polígono de frecuencias dadas una curva de distribución teóricamente normal, con frecuencia, media y desviación típica equivalentes.
2. Ordenar datos cualitativos transformándolos en puntuaciones numéricas. Los datos que han de calificarse pueden ser contestaciones a cuestionarios o a formularios de opiniones, juicios, evaluaciones o escalas de jerarquización.
3. Comprobar la fiabilidad o significación de valores obtenidos sobre muestras. Estos pueden hallarse relacionados con las posibilidades que podrían haber resultado de las fluctuaciones del azar al escoger la muestra. Esta comparación proporcionaría un nivel de confianza al suponer que los valores obtenidos son estadísticamente significativos.
4. Convertir puntuaciones directas en puntuaciones típicas.
5. Normalizar una distribución de frecuencias. Ordinariamente ésta es una parte importante en el proceso de tipificación de un test psicológico.

- Áreas bajo la curva normal, uso de la tabla z. Esta tabla presenta con cada puntuación típica z el porcentaje o áreas comprendidos entre esta puntuación y la media. Si en un problema hemos de calcular el porcentaje de sujetos que presentan igual o menor puntuación típica a una z positiva, p. ej. 1,35; calcularemos el tanto por ciento que hay de esa z a la media, y le sumaremos el 50% de sujetos con puntuación inferior a la media.

Si lo que hemos de hallar es el porcentaje de sujetos con z igual o menor a una z negativa, p. ej. 1,4, calcularemos el % que hay desde la z a la media y el resultado se lo restaremos a 50 (porcentaje de sujetos con puntuaciones inferiores a la media)

COCIENTE INTELECTUAL: el cociente es una relación, una razón, un verdadero cociente o número índice. Por ejemplo, en el cálculo del cociente intelectual de Stern, tomado por Terman $CI = EM/EC \times 100$ (donde EM = Edad Mental y EC = Edad Cronológica) (Cortada de Cohen, N., 1977). Pronto este modo tropezó con un problema técnico importante: que los CI correspondientes a diferentes niveles de edad no resultaban comparables. Por esta razón, este modo de calcular el CI se reemplazó por el CI de desviación, una de las puntuaciones transformadas más usadas en psicología. Estos CI no se obtienen mediante cocientes. La justificación para el nombre de "cociente" estriba en la familiaridad general del término Cociente Intelectual (CI).

DICCIONARIO DE PSICOMETRIA

Uno de los primeros en expresar las calificaciones en términos de CI de desviación fue Wechsler, proponiendo una $m = 100$ y un $DE = 15$.

"D"

DATOS. Las puntuaciones obtenidas en una variable dependiente o medición de rendimiento.

DATOS ESTADISTICOS. Número de unidades correspondientes a los casos estudiados.

DATOS, ORDENACION DE LOS. Los datos se organizan en ordenaciones jerárquicas ascendentes o descendentes por orden numérico. Con frecuencia, se agrupan en intervalos de clase, de modo que su análisis resulta simplificado, y las características son más sobresalientes. Las medidas de la tendencia central (media, mediana, modo) describen los datos en términos de algún tipo de criterio. Las medidas de posición, amplitud o dispersión describen los datos en términos de relación con un punto de la tendencia central. La variación, la desviación, la desviación típica, el percentil, la puntuación típica y la puntuación derivada son medidas útiles de posición, amplitud o dispersión. Las medidas de relaciones describen la relación de variables apareadas cuantificadas por un coeficiente de correlación. El coeficiente resulta útil, en la investigación psicosociopedagógica, para la sistematización de tests, y para hacer predicciones cuando sólo se dispone de algunos de los datos. Debe señalarse que un elevado coeficiente no implica una relación de causa-efecto, sino que sólo cuantifica una relación establecida lógicamente antes de su mensura. La estadística es la servidora -no la dueña- de la lógica; un medio -no un fin- de la investigación. A menos que los supuestos sean válidos, los datos idóneos, cuidadosamente recogidos, registrados y tabulados y el análisis e interpretación sean lógicos, la estadística no puede contribuir a la búsqueda de la verdad.

DECISION ESTADISTICA. Las decisiones estadísticas, basadas sobre la evidencia observada en muestras, implican siempre la posibilidad de error. Los estadísticos no tratan con decisiones fundamentadas en la certeza, sino que, simplemente, estiman la probabilidad o improbabilidad de ocurrencia de los hechos.

El rechazo de la hipótesis nula cuando es realmente verdadera se conoce como Error Tipo I. El nivel de significación seleccionado determina la probabilidad de un error del Tipo I. Por ejemplo, cuando el investigador rechaza la hipótesis nula al nivel 0,05, está corriendo un riesgo del 5% de rechazar lo que pudiera ser una explicación de error de muestra cuando es probablemente verdadera. Aceptar una hipótesis nula cuando es realmente falsa, constituye lo que se conoce como Error Tipo II. Esta decisión se equivoca al aceptar una explicación de error de muestra cuando es probablemente falsa. El nivel de significación más alto, tal como al 0,01 minimiza el riesgo de error de un Tipo I. Pero este alto nivel de significación es tan cauteloso que incrementa el riesgo de un error de Tipo II. Las decisiones estadísticas no son hechas con certeza, pero se basan sobre estimaciones de probabilidad: teorema del límite central, error de varianza, hipótesis nula, niveles de confianza, tests direccionales, no direccionales.

DEDUCCIONES ESTADISTICAS. Las medidas estadísticas descriptivas (como la media o la desviación típica) de una muestra elegida al azar son conocidas como estimaciones. Al valor real de la medida estadística correspondiente de toda la población se le denomina parámetro. Esto debe ser deducido de las estimaciones de la muestra elegida al azar que pueden ser medidas.

DICCIONARIO DE PSICOMETRIA

DEFINICION OPERACIONAL. Designación que se aplica a las definiciones que incluyen la "elaboración" o la " procedencia" de las condiciones correspondientes al fenómeno del que se trata. Es también la designación de un objeto que no apunta a su esencia sino a su manipulación empírica, es decir, se consideran en él ciertas operaciones con las que puede determinarse un constructo// Una definición de un concepto en términos de las operaciones que deben ejecutarse a fin de demostrar el concepto.

DESEABILIDAD SOCIAL. Las pruebas autodescriptivas son especialmente susceptibles a la posibilidad de una descripción falsa. Para la mayor parte de los ítems de un test de este tipo hay respuestas que pueden reconocerse como socialmente más aceptables o deseables que otras. Esta tendencia a elegir respuestas socialmente deseables en los inventarios de personalidad se conoce con el nombre de "deseabilidad social". Edwards (1957) fue el primero en investigar esta variable y la explicó como un efecto de apariencias donde se tiende a falsear para causar una buena impresión. Este falseamiento puede ser consciente o no. Paulhus y Reid (1991) han subrayado esta diferencia distinguiendo entre las nociones de "manejo de impresiones" y de "autoengaño".

DESVIACION. 1) Desviación, referida en general a cualquier campo de la vida 2) en especial, el alejamiento de la conducta normal de la sociedad, quedando el individuo que la practica desacreditado y/o discriminado frente a aquella. // Diferencia entre un valor y otro valor promedio o típico. Sin.: desvío; puntuación diferencial. Notación: x .

DESVIACION, MEDIDAS DE. Las medidas de tendencia central describen ciertas notas de las masas de datos. Hay otras características que requieren un análisis estadístico, en el que la media, la mediana o el modo deben ser complementados por medidas adicionales. (Sin.: medidas de amplitud o dispersión). Ejemplo: Las siguientes puntuaciones fueron obtenidas por dos grupos de estudiantes:

Grupo 1 Grupo 2

Alumnos X Alumnos X

A 100 A 90

B 80 B 90

C 60 C 60

D 40 D 30

E 20 E 30

ÓX 300 300

N 5 5

m 60 60

md 60 60

donde,

ÓX = suma de las puntuaciones

N = número de puntuaciones

m = puntuación media

md = puntuación mediana

Es evidente que las medidas de tendencia central (media y mediana), aunque sean iguales en ambos grupos, no describen las diferencias de rendimiento entre los estudiantes del Grupo I y los del Grupo II. Es necesario emplear una medida de dispersión al mismo tiempo que las medidas de tendencia central para comparar sus rendimientos. El Grupo I es decididamente heterogéneo, con grandes variaciones en los rendimientos. El Grupo II es bastante homogéneo con poca diferencia entre las puntuaciones contiguas y entre las puntuaciones más elevadas y más bajas. Las

DICCIONARIO DE PSICOMETRIA

medidas de dispersión más utilizadas son la amplitud intercuartil, el desvío estándar y la varianza. Véase cada uno de estos términos.

DESVIACION TIPICA O DESVIACION ESTANDAR. Medida de la dispersión de una distribución. Es la raíz cuadrada del promedio de las desviaciones cuadráticas de los índices respecto del valor medio// Es una de los métodos más usados, puesto que permite interpretar los valores en función de su media y de su dispersión o variabilidad. El espacio de la curva normal comprendido entre la media y más menos un desvío estándar es equivalente al 68 por 100 medio de los casos. El espacio de la curva entre la media y más menos dos veces el desvío estándar, comprende el 95 por 100 medio de los casos. El espacio comprendido entre la media y más menos dos veces y media el desvío estándar, tiene el 99 por 100 de los casos. El 68 por 100 es la norma de la distribución. El desvío estándar separa la población normal de la superior e inferior. Se representa por SD (Standard Deviation), DE (desvío estándar), DT (desviación típica) o por la letra griega sigma. Se denomina también desviación media cuadrática. Mide, aproximadamente, las dos terceras partes de la distribución. Se computa partiendo de la media y es la raíz cuadrada del promedio del cuadrado de las desviaciones tomadas desde la media de las puntuaciones obtenidas. Entre la media y cada una de las SD (positiva y negativa) hay un 34,13 por 100, o ligeramente más de un tercio de los casos totales en una distribución normal. Entre más una sigma y menos una sigma se distribuyen, aproximadamente, las dos terceras partes de los casos. La curva normal de frecuencia tiene forma simétrica y posee ciertas propiedades. Cuando un gran número de individuos es sometido a medida respecto a cualquier rasgo mental o físico, existe la tendencia, al representar gráficamente las puntuaciones obtenidas, de aproximación a la curva normal. Debe observarse que hay una tendencia a la acumulación de casos en la zona media de distribución o en las cercanías de la misma. Por ejemplo: una carrera universitaria ofrecerá grandes dificultades a quien no se halla en la zona más allá de la sigma superior desde la media. La curva normal es asintótica, nunca se encuentran sus extremos con el eje de abscisas; ello permite hablar de cuatro sigma, algo muy apartado del centro o media. Un cuatro sigma superior sería un talentoso o genio. En un conjunto de 100, 68 serían "normales"; el resto, "especiales", ya superdotados, ya infradotados. La curva normal teórica es una línea que se aproxima indefinidamente a la abscisa, pero sin encontrarla.

DESVIACION TIPICA, VARIACION DE LA. Cuanto mayor es el valor de la desviación típica, mayor resulta la dispersión o amplitud; cuanto menor sea el valor de la desviación típica, menor es la dispersión. Las puntuaciones por encima de la media en una distribución normal tienen un valor positivo; las que se hallan por debajo, un valor negativo. Como la cantidad de desviación más que la dirección es lo importante como medida de dispersión desde la media, las desviaciones son elevadas al cuadrado para eliminar los signos. Así, la desviación típica fue elaborada por Karl Pearson en 1893 como una medida de la dispersión -a partir de la media de las puntuaciones de una distribución.

DESUDIO ESTANDAR. Véase "desviación típica"

DETERMINISMO. Una premisa de la ciencia que sostiene que los sucesos tienen un número finito de causas que pueden descubrirse.

DIAGRAMA DE DISPERSION. Un mapeo de las puntuaciones logradas por los mismos individuos en dos variables diferentes, que proporciona una representación gráfica del grado de correlación entre dos variables.

DICCIONARIO DE PSICOMETRIA

DIFERENCIAL SEMANTICO. En sentido genérico, cualquier conjunto de escalas de evaluación que tienen como puntos de referencia adjetivos bipolares. El diferencial semántico es un enfoque muy flexible para obtener medidas de actitudes y rasgos. Osgood y sus colegas llevaron a cabo estudios decisivos acerca de las escalas de diferencial semántico. Los fundamentos lógicos se basan en la presunción de que en los lenguajes hablado y escrito, las características de las ideas y de los objetos reales se comunican sobre todo por medio de adjetivos. La mayor parte de los adjetivos tienen opuestos como lo ejemplifican los pares dulce – agrio; oscuro – claro y alto – bajo. En la construcción de un diferencial semántico deben tenerse en cuenta los siguientes pasos:

- 1) preparar una lista de tales pares de adjetivos
- 2) utilizarlos como referencias en las escalas de evaluación (por ej., *Débil* _ _ _ _ _ *Fuerte*)
- 3) buscar factores comunes en las escalas.

Los análisis factoriales realizados con las escalas de diferencial semántico por Osgood y otros investigadores han redundado en la aparición sistemática de tres factores con contenido semejante. El factor que se encuentra con mayor frecuencia es el de *calificación*, definido por pares como bueno – malo; agradable – desagradable; justo – injusto, entre otros. El segundo factor es el de *potencia*, ejemplificado por los siguientes: fuerte – débil; duro – blando y tosco – delicado y el tercero es el de *actividad*, como en activo – pasivo, tenso – distendido y excitable – calmo. Nunally por su parte, halló un cuarto al que denominó factor de *familiaridad*, definido en los pares común – no común; comprensible – misterioso y predecible – impredecible (Nunally, 1970).

DIMENSION. Una escala entre dos extremos a lo largo de la que se producen variaciones ordenadas; p. ej.: la agudeza es una dimensión del tono, la brillantez, una dimensión de la luz, la largura, una dimensión del tamaño, etc.

DISTRIBUCION. Un conjunto de valores para un atributo o variable.

- Distribución de frecuencia. Un conjunto de puntuaciones dispuestas en orden ascendente o descendente. El número de veces en que ocurre cada puntuación.
- Distribución normal. Una curva simétrica, mesocúrtica, en forma de campana. Si las puntuaciones sobre una medición se distribuyen normalmente, mientras más grande sea la diferencia entre una puntuación y la media, menor será la probabilidad de obtener la puntuación.// La de frecuencias -tipo, simétrica, o bien conformada simétricamente- cuyas propiedades sirven para obtener deducciones estadísticas a partir de las medidas derivadas de los casos.
- Distribución, normalización de la. Al sistematizar los tests, se supone que el rasgo de conducta a medir se halla distribuido normalmente en la población. Para su aplicación a grupos, con el propósito de establecer normas, se eligen los ítems del test de modo que las frecuencias de puntuaciones obtenidas se aproximen a la distribución normal. Este es un proceso importante en la tipificación de un test. El concepto se aplica a veces para graduar el rendimiento de los estudiantes en una carrera universitaria determinada. Se ha utilizado mucho el siguiente procedimiento:

Curso Desviación de la media
(en DE) Porcentaje de casos (%)
A por encima de +1,5 7
B de +0,5 a +1,5 24
C de - 0,5 a +0,5 38

DICCIONARIO DE PSICOMETRIA

D de - 1,5 a - 0,5 24

E por debajo de - 1,5 7

Este sistema de graduación supone que los cursos se hallan normalmente distribuidos en una determinada carrera. Es un intento de aplicar un concepto, que es válido para grandes grupos sin seleccionar, a un grupo muy seleccionado en el cual se ha distribuido la normalidad de la distribución. Por ejemplo, en una clase universitaria de 1º o último año, donde las medidas y procesos de selección han eliminado a muchos que hubieran ocupado los puestos inferiores en una distribución no seleccionada, la aplicación no es válida. El forzar las mediciones de rendimiento académico en esta pauta desfigura el concepto básico de la distribución normal.

- Distribución sesgada. Una distribución asimétrica. En una distribución sesgada negativamente, las puntuaciones extremas son inferiores a la media. En una distribución sesgada positivamente, las puntuaciones extremas son superiores a la media.

- Distribución simétrica. Distribución de frecuencia cuyos casos se hallan situados de un modo igual en los intervalos de clase a ambos lados del punto medio; coinciden, en ella, media, mediana y modo.

"E"

EDAD. En psicología, tiempo transcurrido desde el nacimiento para humanos y animales

- Edad cronológica. Tiempo transcurrido desde el nacimiento.

- Edad educacional. La edad educacional expresa el conjunto de conocimientos que corresponden a los alumnos normales que tienen una edad cronológica determinada. La educación es un concepto amplio que hace referencia a los hábitos y a todo lo relacionado con las facultades específicamente humanas. Como el conjunto de conocimientos de esta edad educacional es más restringido que la educación en sí, se cree más correcto hablar de edad de instrucción.

- Edad fisiológica. Tiempo referente a la maduración de tejidos, glándulas, etc.

- Edad mental. Determinada por la puntuación de un test de inteligencia comparada con la media de una determinada edad cronológica. Los tests que determinan aproximadamente la relación entre la edad mental y la cronológica de los sujetos examinados dan así el cociente intelectual (CI). Esta medida es muy relativa, por la imperfección de los tests de nivel, y también por las reacciones afectivas del sujeto en el examen. Sin embargo, los resultados obtenidos sobre un gran número tienen un valor de comparación. Por ejemplo, de 1000 casos, se registran: 19 con un CI inferior a 70; 60 con CI comprendido entre 70 y 90; 291, con CI entre 90 y 110; 550, con CI entre 111 y 130 y 80 alumnos con CI de más de 130. Se admite que un cociente intelectual de 100 es normal, se ve, pues, que el 80 por 100 de los sujetos examinados tienen un nivel mental normal o superior a la media.

- Edad prenatal. Período comprendido desde 280 días antes y hasta el nacimiento (estados germinal, embrionario y fetal).

- Edad real o verdadera. Tiempo transcurrido desde la concepción (EC + EP)

EDUCCION. Intuición inmediata de relaciones en presencia de datos o características.

EFICIENCIA, PREDICCION DE. Calculado un coeficiente de correlación, es interesante saber qué significa, lo íntima que es la relación. El coeficiente no representa un porcentaje de algo, un r de 0,50 no es cualitativamente la mitad e un r de 1-. Suponiendo el problema de predecir notas escolares sobre la base de tests de inteligencia, es preferible que un test nos proporcione una correlación de 0,60 que no 16 utilizar ninguno o que éste fuese cero. Si un estudiante obtiene 7 en el test y 5,7 en

DICCIONARIO DE PSICOMETRIA

la ecuación regresiva, podemos buscar la cuantía de nuestro error de predicción; en un grupo, puede calcularse el error promedio. El error determinado en el caso de que r sea cero nos permitiría predecir que cualquier sujeto puede alcanzar la puntuación media.

Entonces es posible comparar los errores medios producidos con las correlaciones de 0,60 y de cero y ver en qué porcentaje es menor el primero, aunque es algo complicado esperar predecir las notas escolares por las comparaciones de los errores, siendo preferible aplicar la fórmula siguiente que da el error de predicción (E), que es fácil de computar para cualquier correlación obtenida entre tests y notas escolares. El resultado suele denominarse predicción de eficiencia del coeficiente de correlación. Seguidamente damos algunos datos relativos a predicción de eficiencia:

Coeficiente de Predicción de correlación eficiencia (%)

0,10	½
0,20	2
0,30	5
0,40	8
0,50	13
0,60	20
0,70	29
0,80	40
0,90	56
0,95	69
0,98	80
1,00	100

El coeficiente de 0,60 citado reduce el error de predicción sólo en un 20 por 100. Por tanto, incluso utilizando los mejores tests, si obtenemos correlaciones entre 0,60 y 0,70, cuando predecimos calificaciones escolares o el éxito en una tarea todavía podemos cometer algunos errores.

EJES. En las distribuciones de frecuencias medidas con escalas de intervalos o de razones, tenemos dos variaciones isomórficas con las longitudes. Una es la variable de observación y la otra es la frecuencia. Por ello, se pueden utilizar dos direcciones para representarlas. Lo más común es tomar las dos direcciones en forma perpendicular, llamando a la dirección horizontal "eje de las abscisas" y a la dirección vertical "eje de las ordenadas". En el eje de las abscisas se representa la variable de observación (en general, variable independiente) y en el eje de las ordenadas la frecuencia de las observaciones (en general, variable dependiente).// Eje horizontal de las coordenadas (Cortada de Kohan, Carro, 1978; Gonzalvo Mainar, 1978)

EMICO y ETICO. Conceptos postulados por el misionero y lingüista Pike (1954) por analogía con fonémica (*emic*) y fonética (*etic*), quien subrayó que el análisis *etic* no puede llegar a resultados estructurales porque no es posible un sistema universal de las diferencias de sonidos. En psicología, los constructos *emic* se refieren a sistemas lógico – empíricos cuyas distinciones fenoménicas se basan en discriminaciones y contrastes que los propios actores sociales consideran significativos, con sentido, reales, verdaderos o de algún modo apropiados. En cambio, los constructos *etic* dependen de distinciones fenoménicas consideradas adecuadas por la comunidad científica. Algunas variables *etic* pueden analizarse desde una perspectiva *emic*; por ejemplo, aprendizaje, inteligencia, reglas de crianza, atribuciones de causalidad, percepción de enfermedad, salud psíquica, enfermedad, ansiedad.

DICCIONARIO DE PSICOMETRIA

ERROR ESTANDAR DE MEDICION O ERROR TIPICO DE MEDICION (EEM). La desviación estándar de la distribución de muestreo. // La desviación típica de la distribución respecto de la media y de algunos otros estadísticos derivados. Es una medida en especial adecuada para la interpretación de las puntuaciones individuales, por lo que resulta más útil que el coeficiente de confiabilidad para muchos propósitos de investigación. Por ejemplo, si es poco probable que el resultado de un individuo se desvíe a más de 2,58 de EEM de su verdadera calificación, podemos argumentar que su *puntuación verdadera* debe estar dentro de 2,58 EEM de la *puntuación obtenida*. Gulliksen (1950) propuso que el EEM se utilice para estimar los límites razonables de la puntuación verdadera con cualquier resultado obtenido.

ERROR DE ESTIMACION (EE). Así como el EEM indica el margen de error que puede esperarse de la puntuación de un sujeto como resultado de la poca confiabilidad de la prueba, el EE muestra el margen de error que puede esperarse al predecir la puntuación del sujeto debido a la validez imperfecta del test. El cálculo del EE sirve para indicar el tamaño del error en relación con el error que resultaría de una mera conjetura (es decir, con validez 0). Debido a lo dificultoso de obtener la validez de un test, una prueba puede mejorar considerablemente su eficacia predictiva si demuestra alguna relación significativa con el criterio, aunque sea baja. En algunos casos, una validez tan baja como 0,30 puede justificar la inclusión del tests en una batería de evaluación ya que esto es a todas luces mejor que elegir otro al azar (Anastasi, 1998)

ERROR DE MEDICION. Cálculo que permite predecir la probable fluctuación en la calificación de un solo individuo debida a factores aleatorios irrelevantes o desconocidos (Anastasi, 1998). Véase "error estándar de medición"

- Error, fuentes de. Las más frecuentes son:

- 1) Confusión de afirmaciones con hechos
- 2) Incapacidad de reconocer las limitaciones
- 3) Tabulación descuidada o incompetente
- 4) Procedimientos estadísticos inadecuados
- 5) Errores de cálculo
- 6) Defectos de lógica
- 7) Parcialidad inconsciente del investigador
- 8) Falta de imaginación
- 9) Condiciones ambientales inadecuadas en el testeo

ERROR TIPICO. Véase "Error estándar"

ERRORES ALEATORIOS. Se denominan de este modo a las variaciones cuya causa es el azar.

ESCALAMIENTO. Desarrollo de reglas sistemáticas y de unidades significativas de medida para cuantificar las observaciones empíricas. El problema del escalamiento o construcción de escalas ha recibido una gran atención desde la publicación de los trabajos de Stevens, siendo la Teoría Representacional de la Medición una de las teorías más consensuadas en la actualidad. Este modelo trata el tema de la medida en torno a tres grandes áreas: el problema de la representación, el de la unicidad y el de la significación. El primero, implica que la medición supone encontrar un sistema relacional numérico con una estructura semejante al sistema relacional empírico que se pretende medir. Los números pueden representar distinto tipo de relación: de identidad o de diferencia; de orden; de igualdad de diferencias o de razón. Estas relaciones se detallan en "escalas de medición". El segundo problema, el de la

DICCIONARIO DE PSICOMETRIA

unicidad, analiza cómo puede cambiarse un conjunto concreto de números por otro sin que se modifiquen las relaciones numéricas que representan a las relaciones empíricas. En la escala nominal, los números de un conjunto de objetos pueden cambiarse por cualquier otro conjunto que preserve la relación de igualdad vs. diferencia. Son transformaciones admisibles todas las inyectivas (por ej. pertenencia – no pertenencia). En la ordinal, las modificaciones son posibles si el orden relativo de los objetos permanece invariable. Las transformaciones de este tipo se llaman monótonas crecientes. Las escalas de intervalo admiten transformaciones lineales positivas. En cambio, en las de razón, se preserva el homomorfismo. El tercer problema, el de la significación, hace referencia a la validez de la conclusión empírica inferida a partir de un conclusión numérica. Stevens plantea la solución del problema en términos de los estadísticos admisibles para cada tipo de escala. Así, por ejemplo, la media y otros estadísticos basados en ella (como la varianza o el coeficiente de correlación lineal) no deberían utilizarse con escalas ordinales (Martínez Arias, 1996)

ESCALAS DE ACTITUD. En todas las escalas de actitud, los individuos indican su acuerdo o desacuerdo con una serie de afirmaciones sobre el objeto en estudio. Por lo general, arrojan una calificación total que indica la dirección e intensidad de la actitud del individuo hacia una categoría de estímulos (por ejemplo, hacia un estilo de vida, organización de valores, política, etc.). Se han elaborado procedimientos especiales para alcanzar la unidimensionalidad o homogeneidad de los reactivos, la igualdad de las distancias entre las unidades de la escala y la comparabilidad de las calificaciones. Existen tres acercamientos a la construcción de este tipo de escalas:

- Escala tipo Thurstone. La construcción de esta escala se inicia con acopio de afirmaciones que expresan una amplia gama de actitudes hacia el objeto considerado (por ejemplo, actitudes hacia la censura, la pena capital, etc.). Luego, se solicita a numerosos jueces que clasifiquen las afirmaciones de acuerdo con el grado en que resultan favorables. Se eligen las que presenten una variabilidad mínima y una gran extensión de valores de escala (en general, 11 puntos). En la escala final, las afirmaciones se presentan en orden aleatorio sin indicaciones de sus valores de escala y la puntuación que recibe el individuo es la media del valor escala de todas las afirmaciones que aprueba
- Escala tipo Guttman. Se desarrolló como una técnica para determinar si un conjunto de afirmaciones de actitud era unidimensional. Según Guttman, la escala es perfecta si el individuo que concuerda con cierta afirmación de una actitud en particular también lo hace con otras más moderadas de esa misma actitud. El procedimiento para el desarrollo de una escala de este tipo consiste en identificar un conjunto de reactivos que caigan en una secuencia ordenada en función de la aprobación que reciben de los sujetos que las contestan. Los ítems que no cumplen con este requisito son eliminados. Esta misma noción de ordinalidad o progresión uniforme es la que sustenta las escalas piagetianas
- Escala tipo Likert. Método ideado por R.A. Likert para medir la actitud. El sujeto debe expresar su elección ante un cierto número de determinaciones (ítems), elaboradas de modo que pueda apreciarse el grado de conformidad con las actitudes representadas. Las escalas tipo Likert requieren una respuesta graduada que se expresa en función de las siguientes categorías: Totalmente de acuerdo, de acuerdo, indeciso, en desacuerdo, totalmente en desacuerdo. Para calificar la escala, las opciones de respuesta se acreditan como 5, 4, 3, 2 ó 1, del extremo favorable al desfavorable. La puntuación total es la suma de las calificaciones a cada ítem, que debe interpretarse en términos de normas establecidas empíricamente.

DICCIONARIO DE PSICOMETRIA

ESCALAS DE MEDICION. En psicología, se ha hablado mucho en los años recientes de los diferentes tipos posibles de escalas de medición y se ha investigado cuidadosamente la clase de escala apropiada para diversos tipos de medición psicológica. La medida surge de una comparación por medio de un criterio. El criterio de comparación puede ser de diversos tipos. Con mayor frecuencia se utilizan: 1. Criterio de similitud. - Establece una relación de igualdad, o equivalencia. Esta equivalencia puede tener carácter cualitativo o cuantitativo. Cuando la similitud es cualitativa, la escala adopta el tipo de escala apreciativa. Es la graduación teñida de subjetividad (bien, regular, mal). En carácter cuantitativo se presenta mediante la puntuación numérica nominal 2. Criterio de ordenación. - En la faceta subjetiva puede servir de ejemplo la clásica ordenación de los alumnos en clase, según el concepto que de cada uno tiene su profesor (desde el niño primero en la clase, hasta el último). Cuantitativamente, mediante la ordenación jerárquica de las puntuaciones obtenidas en un test o en una prueba objetiva (ordenación del rango o amplitud total). Suele hacerse de mayor a menor.

3. Criterio de agrupación. - Cuando los datos de una distribución son numerosos, lo cual dificultaría las operaciones, se agrupan en intervalos. En cada uno de ellos existe una determinada frecuencia de puntuaciones. A este tipo corresponden las escalas de graduación (termómetro). 4. Criterio de relación. - Establece una distribución por equivalencia relativa (con una unidad, m. cm.), o entre dos datos (razón o cociente de los mismos).

El primero criterio nos proporciona las escalas nominales. En ellas se opera mediante la relación de equivalencia. Sus estadísticos son la frecuencia, el modo y el coeficiente de contingencia. El segundo, nos eleva a las escalas ordinales, jerarquizadas en orden descendente. Sus estadísticos son media, mediana, desviación típica, percentiles y correlación ordinal. El tercero, las de intervalos, cuya unidad se basa en una igualdad de diferencia, y nos lleva a las puntuaciones típicas. El cuarto, produce las escalas de razones.

Los tipos diferentes de escalas de medición entrañan usos distintos de los números. En los apartados siguientes se especifican algunos de estos empleos.

- Escalas Nominales. Rótulos. A veces se usan los números para individualizar objetos sin pretensión alguna de someter el número a análisis matemáticos. Por ejemplo, el geólogo puede distinguir sus muestras de rocas como 1, 2, 3, etc. En este caso los números se usarían exclusivamente como etiquetas y carecerían de valor para el análisis matemático. No tendría sentido sumar los números que representan la primera y segunda roca y formar una cierta ecuación con el 3 que se refiere a la tercera. Otros ejemplos de números usados como rótulos son los que utilizan los jugadores de fútbol en la espalda, los números de los carteles de las carreteras y los números de los elementos atómicos. Debe destacarse que cada escala de medición está concebida para un cierto empleo de los números. Una de las posibles utilizaciones de ellos es el de designar cosas. En este caso no existe la intención de realizar análisis matemático de los números ni se considera que éstos representen cantidades de atributos. Sin embargo, puede suceder que los rótulos guarden correlación con cantidades de atributos. Así, en el ejemplo de las rocas del geólogo, podría suceder que a medida que su bolsa de minerales se hiciera más pesada, él eligiera muestras cada vez más livianas. En consecuencia, los números usados como rótulos se relacionarían incidentalmente con el peso de las rocas. En forma semejante, las carreteras de números altos pueden tener alguna diferencia cuantitativa con las de números bajos y los elementos atómicos que ocupan una posición avanzada en el esquema de numeración pueden ser cuantitativamente diferentes de los que ocupan posiciones anteriores. El punto esencial es que en los casos particulares debe justificarse el uso de los números al examinar la naturaleza de las escalas de medición. El hecho de que haya o no una correlación cuantitativa incidental con una serie particular de números

DICCIONARIO DE PSICOMETRIA

no tiene importancia en un análisis de la legitimidad de su utilización. Puesto que los rótulos no están concebidos para indicar cantidades de atributos, no se requiere justificación alguna para emplear números de rótulos. Categorías. La utilización de los números para representar grupos de objetos guarda estrecha relación con su uso como rótulos. Por ejemplo, el geólogo podría clasificar sus rocas en las siguientes categorías: sedimentarias, ígneas y metamórficas y referirse a estas categorías como 1, 2 y 3. Otros esquemas de clasificación incluyen las diferentes profesiones, los dos sexos, el tener una lesión cerebral o ser una persona normal. Las únicas diferencias que existen entre emplear números como rótulos y emplearlos como categorías son que en este último caso: 1) a cada número le corresponde más de un objeto; 2) todos los objetos a los que se les asigna el mismo número son semejantes con respecto al mismo atributo. Al igual que los números que se usan como rótulos, los que se usan para representar categorías en un esquema de clasificación carecen de connotación cuantitativa. De modo similar, los números usados para identificar categorías pueden presentar muchas correlaciones incidentales con cantidades de atributos. Así, en promedio, los varones y las mujeres difieren en altura, habilidad atlética y otros muchos atributos, pero esto no guarda relación alguna con el hecho de justificar el uso de los números 1 y 2 para representar respectivamente a los varones y a las mujeres. Ya que los números empleados para representar categorías no tienen connotación cuantitativa, el usarlos para tal fin no exige justificación alguna. Al examinar este punto se debe hacer una distinción entre el uso de los números para representar categorías y el emplearlos para indicar la frecuencia con que aparecen los objetos en las diferentes categorías. Así, el geólogo puede incluir 22 de sus rocas en la categoría de rocas ígneas. En tales casos, se dice a veces que uno "mide" el número de casos en las diferentes categorías, pero de acuerdo con la definición previa de "medición", éste sería un uso poco apropiado del término. Sería más apropiado decir que uno enumera o cuenta los objetos de cada categoría.

- Escalas Ordinales. Una escala ordinal es aquella en la cual: 1) una serie de objetos o personas está ordenada de "mayor" a "menor" con respecto a un atributo; 2) no hay indicación de "cuánto" posee en un sentido absoluto cualquiera de los objetos, y 3) no existe indicación de cuánto se diferencian los objetos en lo que respecta al atributo. El orden por rangos es la manera más primitiva de la medición si se excluyen de ésta los rótulos y las categorías. Es primitiva en el sentido de que es básica para todas las formas superiores de medición y que sólo proporciona poca información. Se obtiene una escala ordinal, por ejemplo, cuando se ordena un grupo de jóvenes desde el más bajo hasta el más alto. Esta escala no dará indicación de la altura promedio: como grupo, los jóvenes podrían ser relativamente altos o comparativamente bajos. La escala no proporciona información alguna acerca de cuánto varían los jóvenes en altura. Con respecto a los métodos de análisis, no tiene sentido computar la desviación media y estándar de una serie de rangos. A menudo no se tiene en cuenta que los números que se emplean con escalas ordinales sólo proporcionan un sistema abreviado conveniente para designar las posiciones relativas de los objetos. Se obtiene una escala de orden por rangos cuando de N personas cualesquiera (SS) se sabe que $S_i > S_j > S_k > S_n$ con respecto a un atributo. // Aquella en la que los elementos están colocados por orden de tamaño pero sin unidades iguales ni cero verdadero (absoluto)

- Escalas de Intervalos. Una escala de intervalos es aquella en la cual:

- 1) se conoce el orden por rangos de los objetos con respecto a un atributo;
- 2) se sabe cuánto se diferencian los objetos uno de otro, pero
- 3) no se dispone de ninguna información acerca de la cantidad absoluta del atributo con respecto a ninguno de los objetos. Se obtendrá una escala de intervalos para la altura de un grupo de jóvenes si, en lugar de medirlos directamente, se midiera la altura de cada uno con respecto al más bajo del grupo. De ese modo, el más bajo obtendría un puntaje de cero, el que tiene dos centímetros más, un puntaje de 2, un joven con tres centímetros más, un puntaje de 3 y así sucesivamente. Especificar los

DICCIONARIO DE PSICOMETRIA

intervalos en función de la distancia de cada sujeto con respecto a la media aritmética guarda una relación más directa con lo que se hace habitualmente en las mediciones psicológicas. Así, un joven cuya altura estuviese dos centímetros por encima de la media recibiría un puntaje de +2 y otro que está 2 centímetros por debajo de la media, recibiría un puntaje de -2. Pueden calcularse los intervalos con respecto a la altura media o con respecto a cualquier otro atributo sin saber realmente cuánto se aparta una persona del punto cero (por ejemplo, altura cero o inteligencia cero). Hay una información potencialmente importante que las escalas de intervalos no proporcionan: la magnitud absoluta del atributo para cualquier persona u objeto específicos. Así, que el joven más alto tenga un puntaje sobre la media de +6 (6 centímetros por encima del promedio) nada nos dirá acerca de su altura en sentido absoluto. Podría ser el más alto de un grupo de pigmeos. Puesto que a veces se habla de las escalas de intervalos como escalas de "intervalos iguales" es fácil caer en el error de suponer que tales escalas requieren un número igual de personas u objetos en cada punto del continuo, una distribución rectangular de la puntuación. Lo que realmente se quiere decir por "igual" es que los intervalos en la escala son iguales, independientemente del número de personas u objetos que existan en puntos diferentes de ella. Así, en una escala de intervalos para la medición de la inteligencia, la diferencia entre los CI 100 y 105 se presupondrá igual a la diferencia entre los CI 120 y 125. Naturalmente, la mayor parte de las veces las consecuencias prácticas de tales diferencias iguales serán distintas, pero estrictamente hablando, esto nada tiene que ver con el carácter del intervalo de la escala. En forma semejante, si tres automóviles marchan respectivamente a 45, 90 y 135 kilómetros por hora, el intervalo entre los dos primeros es igual al intervalo entre los dos últimos, pero evidentemente estos intervalos tienen importancia diferente en lo que respecta a la seguridad en el tránsito, al rendimiento de la nafta (consumo de nafta por kilómetro) y al desgaste del automóvil. Por lo tanto, es necesario establecer una distinción cuidadosa entre el carácter de una escala de medición cualquiera y los significados prácticos de los puntos de la escala.

- Escalas de cocientes. Se obtiene una escala de cocientes cuando: 1) se conoce el orden por rangos de las personas con respecto al atributo; 2) se conocen los intervalos entre las personas, y 3) para cada persona se conoce además la distancia desde un cero racional. En otras palabras, una escala de cocientes es un tipo particular de escala de intervalos en el cual se expresan las distancias con respecto a un cero racional y no, por ejemplo, con respecto a la altura del joven más alto o del más bajo, o a la altura media. Evidentemente, si se dispone de una escala de intervalos de altura y se conoce además la altura absoluta (distancia con respecto a cero) de un joven cualquiera del grupo, pueden calcularse las alturas absolutas de todos los demás.

- Escalas de Medición en Psicología, aplicación. Teniendo en cuenta las consideraciones de los párrafos anteriores, resulta conveniente estudiar el tipo de escalas que se han establecido para las mediciones psicológicas. Estas utilizan todos los tipos de escalas que hemos examinado. Algunas investigaciones manejan datos de categorías como los números de pacientes mentales de diferentes clases, las comparaciones de personas con lesiones cerebrales con respecto a las normales y la comparación entre fumadores y no fumadores. En algunos casos, los datos psicológicos se ordenan por rangos. Este es el caso, por ejemplo, cuando los examinados ordenan diez pedazos de papel gris desde el más oscuro hasta el más claro. Este caso se presenta también cuando los examinados ordenan los nombres de veinte grupos nacionales desde los que tienen más aceptación hasta los que tienen menos. Cuando se está ante datos que, como éstos, están claramente ordenados por rangos, se debe ser muy cauto al aplicar análisis matemáticos que presupongan escalas de intervalos o de cocientes. Dejemos de lado, por el momento, las escalas de intervalos. Hay algunas ordenaciones de datos psicológicos correctamente consideradas como escalas de cocientes. Por ejemplo, este

DICCIONARIO DE PSICOMETRIA

caso se presenta cuando se expresan algunas ordenaciones en términos de tiempo, como el intervalo que requiere una rata para atravesar un laberinto o el tiempo de reacción en los sujetos humanos. En tales casos tiene sentido hablar de cantidad cero de tiempo y también lo tiene, por ejemplo, decir que un sujeto requiere el doble de tiempo que otro. Otro ejemplo en el cual los datos psicológicos pueden ordenarse en términos de escalas de cocientes se presenta en ciertos tipos de estudios de aprendizaje. Si, por ejemplo, deben recordarse veinte pares de palabras, tiene sentido hablar de una cantidad cero de aprendizaje y también lo tiene formar cocientes entre los números de pares que el mismo sujeto ha aprendido en etapas diferentes del aprendizaje.

Los ejemplos precedentes constituyen más la excepción que la regla, pues la mayoría de los datos psicológicos se expresan legítimamente sólo como escalas de intervalos y no como escalas de cocientes. Por ejemplo, no tiene sentido hablar de inteligencia cero o de autoestima cero y tampoco lo tiene decir que una persona es el doble de ansiosa que otra o que tiene una vez y media la capacidad de razonamiento de otra. Que para la mayor parte de los atributos se utilicen sólo escalas de intervalo en lugar de escalas de cocientes no representa un problema serio. La mayoría de los métodos de análisis matemático y estadístico que se necesitan en la investigación psicológica requieren sólo escalas de intervalos, por ejemplo, análisis correlacional y análisis de la varianza. Hay quienes discuten la utilización de escalas de intervalos para la mayoría de las mediciones psicológicas, pero aparentemente el número de quienes lo hacen disminuye de año en año. De admitirse esta crítica, la psicología tendría que abandonar muchos de sus métodos de análisis estadístico y matemático más eficaces, lo cual provocaría efectos desastrosos en este campo. Podría argumentarse también que para ninguna escala de medición existen intervalos "reales" o "correctos" sino que se los establece por convención. Así, si los científicos hallan que una determinada gradación de los intervalos tiene sentido y es científicamente útil, aceptan estas gradaciones como una convención de medida. Este es, por ejemplo, el caso de la medición de la dureza de las rocas y de la intensidad de los terremotos. Podrían usarse otros métodos de medición y ésto daría como resultado diferentes intervalos de medición. A la larga se impondrá la gradación que conduzca a relaciones matemáticas más sencillas, pero no habría error alguno en tomar como ciertos los intervalos de cada tipo de medición cuando se realiza el análisis matemático. El problema real no se refiere a si se utilizan o no los intervalos de la escala de medición en el análisis matemático sino cuál de las gradaciones de intervalo se mostrará más útil a la larga. En consecuencia, nada tiene de malo realizar operaciones matemáticas que tomen como ciertos los intervalos de medición cuando se investigan los tests de inteligencia, los tests de personalidad y otras técnicas de medición empleadas en psicología. Teniendo en cuenta las consideraciones anteriores, sería una torpeza que los psicólogos no emplearan estos eficaces métodos de análisis matemático y estadístico que presuponen escalas de intervalos, a menos que los datos hubiesen sido obtenidos originariamente por métodos originales.

ESCALA T. Puntaje transformado intragrupo. Comprende una amplitud, en una distribución normal, de 10 desviaciones típicas, desde -5 hasta +5, y 100 unidades. Supone una $m = 50$ y un $DE = 10$. (Véase también "puntajes transformados")

ESCALA TIPO LIKERT. Véase "Escalas de actitud"

ESTADISTICA. Técnica o proceso matemático de recogida, descripción, organización, análisis e interpretación de datos numéricos. Como la investigación busca -del mismo modo- una expresión cuantitativa, la estadística constituye un instrumento fundamental de medida y de investigación.

DICCIONARIO DE PSICOMETRIA

Los resultados de la elaboración de datos producto de una muestra suelen denominarse estadísticos o estadígrafos. Describen la conducta relativa a las características de un grupo población deducidas de cierto número de casos individuales. El estadístico se refiere a la muestra, y no tiene validez hasta que se combinan cierto número de observaciones y medidas individuales, posibilitando generalizaciones, reglas, principios o leyes. Cuando el estadístico adquiere validez, y es reflejo de la población en estudio, se denomina parámetro.

Todo el mundo se halla familiarizado con las expresiones alumno medio, adolescente típico y ciudad representativa. Estos son conceptos estadísticos y, como características de grupo, pueden ser expresados en medidas de edad, tamaño u otro rasgo capaz de ser descrito numéricamente. Cuando decimos que el alumno medio de quinto nivel tiene diez años, describimos todo el grupo, no un niño especial de la escuela. La media estadística (aquí la media) es, pues, una abstracción que puede ser usada en lugar de la gran masa de las medidas individuales.

- Estadística deductiva. Comprende el proceso de muestreo o selección, para su estudio, de un pequeño grupo, al que se supone como representativo de otro más numeroso y del cual procede. El pequeño grupo se denomina muestra; el grupo mayor población, conjunto o universo. Antes de poder suponer la representatividad, es esencial que los individuos seleccionados sean elegidos de modo que el grupo pequeño, o muestra, se aproxime al grupo mayor, o población. Con un margen de error, que siempre se halla presente, y mediante el uso de técnicas apropiadas, esta aproximación puede considerarse supuesta, haciendo posible la estimación de las características del conjunto mediante un análisis de la muestra accesible.

Estudiar ciertas poblaciones directamente sería imposible. Algunos conjuntos son tan extensos que sus características nunca podrían ser medidas completamente; la tarea no tendría fin. Otros estudios, aunque posibles, emplearían demasiado tiempo, dinero y personal para ser practicables. Las autoridades sanitarias encontrarían caro, innecesario y virtualmente imposible, medir directamente la tensión arterial de todas las mujeres que asisten a las universidades. La presión arterial medida en unas pocas muestras, cuidadosamente elegidas, obtendría datos razonablemente exactos para toda la población. La tarea de seleccionar una buena muestra representativa no es fácil. De los muchos métodos empleados describiremos dos tipos muestreo aleatorio y muestreo estratificado.

- Estadística descriptiva. Métodos para resumir, organizar y comunicar datos.// Compendia o simplifica afirmaciones referentes a medidas realizadas sobre una población. Se aplica en sentido amplio a muestras que representan poblaciones a las que debería referirse en sentido estricto. Sirve para la descripción numérica de un grupo particular. Ninguna conclusión va más allá del grupo descrito, y cualquier similitud con los de fuera del grupo no puede ser garantizada. Los datos describen este grupo y solamente éste.

Gran parte de la investigación en psicología, pedagogía y ciencias sociales implica la estadística descriptiva, y proporciona información valiosa acerca de la naturaleza de un particular grupo o clase. Este tipo de datos se utiliza como base para la investigación aplicada. En este apartado entran las medidas y escalas estadísticas.

- Estadística no paramétrica. La que es calculada sin suponer la concurrencia de una distribución subyacente, las fórmulas correspondientes suelen hacer uso de datos ordenados (como en la correlación de diferencias ordenadas). Sin.: estadística de distribución libre.

DICCIONARIO DE PSICOMETRIA

ESTADISTICOS. Valores obtenidos por tratamiento de los elementos muestrales (media, mediana, sigma, etc.). También suelen ser denominados estadígrafos, por su carácter descriptivo de la muestra. Ver "estadística"

ESTIMACION. Valor estadístico para cálculo del parámetro de la población muestral. El tratamiento estadístico paramétrico de los datos se basa sobre ciertos supuestos acerca de la naturaleza de la distribución y los tipos de medidas utilizados. El tratamiento estadístico no paramétrico posibilita decisiones útiles sin supuestos de la naturaleza de la distribución de los datos. Los tests paramétricos deben usarse cuando es lícito establecer supuestos, ya que son más sensibles y oscurecen menos las relaciones significativas de los datos.

ESTRATEGIA SECUENCIAL. Según Cronbach y Gleser (1957), procedimiento de selección por el que se deciden gradualmente y sobre la marcha las nuevas informaciones a recabar sobre los examinados. Mientras en el procedimiento convencional (battery procedure) se admite a aquellos sujetos que han obtenido la puntuación más alta y los demás son rechazados, en el procedimiento secuencial se forman grupos de los admitidos, los rechazados y los que deben someterse a nuevas pruebas. Sólo después de efectuadas las nuevas pruebas se toman ulteriores decisiones.

ESTRATOS. Subdivisiones del total de la población en subpoblaciones, clases o grupos, de los cuales se toma un cierto número de elementos que se consideran representativos.

ESTUDIO LONGITUDINAL. Un método en el cual los individuos son estudiados en el transcurso del tiempo y se obtienen mediciones de los individuos en varios intervalos.

EVALUACION, ESCALAS DE. Mediante la clasificación en varias categorías, permiten la apreciación de rasgos, aspectos, cualidades, de un modo más objetivo. Este medio puede ser un auxiliar útil cuando se procede a evaluar un examen escrito del tipo convencional, y se desea la mayor objetividad posible. Se destacan las características más importantes y se procede a elaborar la escala, disponiendo así de una norma y de unos criterios aplicables a todos los examinados equitativamente.

EVALUACION PSICOLOGICA. Según Cronbach y Gleser (1957) es un proceso de toma de decisiones// Proceso mediante el cual, a partir de algún modelo o enfoque teórico se intenta comprender o analizar un fenómeno concreto, ya sea real o simbólico, en sus aspectos manifiestos y/o latentes. Toda evaluación tiene como objetivos el poder categorizar, comparar, analizar y contrastar tanto datos de tipo cualitativo como cuantitativo, obtenidos mediante la aplicación de técnicas diversas. Las tareas de evaluación pueden referirse tanto a procesos como a productos o resultados. Puede tener como objetivo la evaluación de aspectos deficitarios o patológicos (psicodiagnóstico) o en la evaluación de capacidades y potencialidades de una persona o de un grupo de personas (orientación vocacional, selección de personal, evaluación de psicoterapias y de la eficiencia de técnicas de enseñanza, valoración de planes y programas, etc.) (Casullo, 1996)

EXPLICACION. Una especificación de la condición antecedente necesaria para demostrar un fenómeno.

DICCIONARIO DE PSICOMETRIA

"F"

FACTOR. En general, magnitud influyente, magnitud determinante. En matemáticas, magnitud cognitiva. En genética, sinónimo de gen. En planificación de pruebas, condición que se hace variar. En análisis factorial, magnitud matemática que se obtiene de la matriz de las intercorrelaciones. El factor representa un constructo hipotético que no es observable.

FACTOR GENERAL. 1) Una capacidad general comprendida en las puntuaciones de tests, especialmente en los de inteligencia, distinta de las especiales y únicas para cada prueba (Spearman); 2) capacidad general con la que se correlaciona cada uno de los factores primarios (Thurstone).

FENOMENO. Un hecho o evento que es observable.

FIABILIDAD. Consistencia o constancia de una prueba como instrumento de medida. Véase "confiabilidad"

FRECUENCIA. Número de veces que se repite un valor. Hay tres tipos de frecuencias: a) frecuencia absoluta; b) frecuencia relativa; c) frecuencias acumuladas.

- Frecuencia absoluta es el número entero de observaciones correspondientes al mismo.

- Frecuencia relativa es la frecuencia absoluta correspondiente dividida por el número total de casos u observaciones.

- Frecuencia acumulada. La distribución de frecuencias acumuladas se puede obtener cuando los valores de la variable están ordenados. Significa una ordenación de frecuencias en la que se acumulan a las frecuencias correspondientes a cada valor o clase de variable las de los valores o clases inferiores. La frecuencia acumulada relativa es igual a la absoluta dividida por el número de casos.

- Frecuencia, distribución de. Serie de puntuaciones ordenadas según su valor y agrupadas en intervalos de clase.

- Frecuencias teóricas (f_1). Número de veces que se repite una variable en un intervalo determinado, según una distribución normal.

"G"

GRADUACION, ESCALA GRAFICA DE. Medio de calificar a un sujeto por el que el examinador registra su juicio colocando una marca en algún punto de una línea impresa, uno de cuyos extremos indica el grado más bajo del rasgo y el otro el más elevado.

GRANDES NUMEROS, INERCIA DE LOS, LEY DE. Los grupos grandes tienden a mayor estabilidad de características.

"H"

HEURISTICO, METODO. Método no riguroso para descubrir la solución correcta de un problema, mediante la obtención de aproximaciones a la respuesta adecuada con el empleo de analogías y otros procedimientos de investigación, pero sin la labor de explorar todas las posibilidades.

"I"

INFERENCIA. Inducción de características de un grupo reducido a toda la población, lo que permite extraer generalizaciones del proceso de muestreo.

DICCIONARIO DE PSICOMETRIA

INTERPRETACION DE TECNICAS (Integración de los análisis cuantitativos y cualitativos de una técnica). Se refiere a la integración, en la evaluación de una técnica, de los análisis cuantitativo y cualitativo efectuados, teniendo en cuenta que la misma posee, por lo general, múltiples propósitos a ser entendidos. La integración de ambos tipos de análisis puede brindarnos percepciones de mayor riqueza interpretativa que tomando uno u otro de manera excluyente.

"...es innecesaria una dicotomía entre los métodos cuantitativos y los cualitativos y existen todas las razones para emplearlos conjuntamente con el objeto de satisfacer las exigencias de la evaluación del modo más eficaz posible..." (Cook y Reichardt, 1986)

INSTRUCCION, AÑO DE. Está constituido teóricamente por la cantidad de conocimientos que se adquieren en doce meses. La determinación del año de instrucción no es posible, en realidad, porque el desenvolvimiento mental del hombre no sigue un ritmo uniforme. Por otra parte, en la tarea docente se sigue un orden exigido por la naturaleza misma de las materias de enseñanza y, en virtud de este orden, unos conocimientos se enseñan normalmente a una edad y otros en otra diferente.

INTERFERENCIA SOCIAL. Facilitación e inhibición social. Efecto de interferencia de las opiniones ajenas sobre el individuo y su conducta.

INVENTARIO. Conjunto de frases o palabras a las que el sujeto tiene que responder eligiendo o marcando su posición ante ellas (por ejemplo, al ítem "Me gustan las revistas de mecánica", contestar "verdadero" o "falso").

ÍTEM. Cualquier unidad simple o elemento, de las pruebas o de los materiales experimentales; p. ej.: una sola pregunta en un test que se compone de muchas preguntas, o una sola sílaba sin sentido en una lista de sílabas que han de aprenderse de memoria. . En todo ítem es preciso considerar tres factores diferenciados: dificultad, homogeneidad, validez.

- Ítem, dificultad relativa: La expresión cuantitativa de la dificultad relativa del ítem se puede hacer de dos formas:

- 1) Expresando el porcentaje de sujetos que lo aciertan.
- 2) Expresando la z_i que deja por debajo de sí un número de sujetos igual al que ha superado la prueba.

Por ejemplo $z_i = 0,84$, puntuación que deja por debajo de sí el 80% de los casos// La estimación de la dificultad de un ítem se realiza sólo en el caso en que se puede hablar de error o acierto de la respuesta.

Se puede determinar el grado de dificultad de un ítem restándole a la unidad el cociente resultante de dividir el número de aciertos entre el número de sujetos que realizaron el ítem. Este cociente es el índice de facilidad. Si se aplica un ítem a 100 sujetos y los 100 fallan, el ítem no sirve porque no tiene poder de discriminación. El ítem que más discrimina es el que es superado por la mitad de los sujetos que lo intentan. La dificultad de los ítems en un test de aprendizaje ha de ser, en general: 10% de ítems muy difíciles; 20 % de ítems difíciles; 40 % de ítems de dificultad media; 20 % de ítems fáciles; 10 % de ítems muy fáciles.

- Ítem, homogeneidad de un. Todos los ítems deben intentar medir lo mismo; pero deben ser lo suficientemente homogéneos para que midan lo mismo, y lo suficientemente heterogéneos para que midan los diferentes tipos que presenta la característica a tratar.

Se halla la homogeneidad o heterogeneidad de un ítem obtenido los coeficientes de correlación entre cada dos: a mayor correlación, mayor homogeneidad.

DICCIONARIO DE PSICOMETRIA

- Item, validez de un. Un ítem es válido cuando es capaz de discriminar entre los sujetos superiores e inferiores de la característica que quiera medir. Experimentalmente, para ver la validez del ítem, se obtiene la relación entre éste y el total del test. Cuanto mayor sea la correlación, más válido será el ítem. Si un ítem puede discriminar un 27% superior del grupo de un 27% inferior del mismo grupo al que ha sido aplicado, se dice que es válido. Esto se determina por el siguiente proceder: Sabemos, p. ej., que, en un grupo de 300, en el ítem k han contestado bien 56 y mal 18. El 27% de 300 es 81. Habrá, pues, en el grupo, 81 sujetos mejores y 81 peores. Tras esto, hay que obtener el tanto por ciento que supone 56 en el grupo de 81 y el que supone 18: 56 es el 69% de 81; 18 es el 22% de 81. Buscamos estos porcentajes en la tabla correspondiente y nos da una correlación de 0,475. La línea curva de la tabla separa los coeficientes de correlación que son mayores a 0,32 de los que son menores. Los resultados que se hallan entre la línea curva y la diagonal tienen poder de discriminación, pero este es menor que los que se encuentran por encima de dicha línea curva. La tabla nos ofrece también la dificultad del ítem. A partir de la diagonal hay en la tabla unos signos negativos, porque el porcentaje de la parte superior del grupo es menor que el de la parte inferior, por lo que no es válido el ítem.
- Items pesados. Si un elemento (o un test en una batería de tests) tiene una capacidad de predicción mejor que otro, se le asigna una calificación mayor; así, influirá en una puntuación compuesta más que el elemento (o test) de un valor predictivo inferior.
- Items, ponderación de los. A veces, las categorías de respuesta no son mutuamente exclusivas. Las preferencias para ciertas cosas o razones de un acto son explicadas, en general, en términos de un cierto número de factores, raramente simples. No sería realista esperar, de uno que responde, el que indique su tipo favorito de distracción o la razón simple que le decide a matricularse en una universidad determinada. En tales casos sería más adecuado preguntarle de modo que formulara dos o tres respuestas en orden de importancia o de preferencia. Esta ordenación de ítems posibilita un método útil de análisis. Los elementos son ponderables en orden inverso. Por ejemplo, si han de ordenarse tres ítems, sería adecuado asignar ponderaciones como sigue:
1ª elección 3 puntos.
2ª elección 2 puntos.
3ª elección 1 punto.

Mediante la ponderación de los promedios totales de los sujetos se puede determinar, entonces, un juicio compuesto de la importancia de los ítems. Sería conveniente recordar que cuando los ítems están relacionados en orden, las diferencias entre ellos pueden no ser iguales. La ordenación no es el método más refinado para escalar. Véase también "Análisis de Items"

"L"

LIMITACIONES ESTADISTICAS. Existen cierto número de limitaciones que deberían tenerse presentes al usar los procedimientos estadísticos, al interpretarlos y al derivar conclusiones de sus resultados: 1) El procedimiento estadístico es servidor de la lógica y sólo tiene valor si comprueba, aclara y mide relaciones que han sido establecidas por un análisis claro y lógico. La estadística es un medio, nunca un fin, de la investigación. 2) Un procedimiento estadístico no debe ser empleado en el análisis de los datos a menos que se comprendan claramente los supuestos y limitaciones básicos que le sirven de fundamento. 3) Las conclusiones derivadas del análisis estadístico no serán más exactas o válidas que los datos originales. Ningún refinamiento de la manipulación estadística complicada obtendrá verdades importantes si los datos proceden de medidas toscas o inexactas. 4) Todo tratamiento de los datos debe ser comprobado y

DICCIONARIO DE PSICOMETRIA

controlado con frecuencia, para reducir al mínimo la probabilidad de errores de medida, registro, tabulación y análisis. 5) Existe un constante margen de error dondequiera que se halla implicada una medida realizada por el ser humano. Este error se incrementa cuando se someten a mensura cualidades o características de la personalidad humana, o cuando se establecen deducciones sobre la población a partir de medidas de muestras estadísticas.

Cuando se realizan comparaciones o contrastes, una mera diferencia numérica no es en sí misma una base válida para ninguna conclusión. Debería emplearse un test de significado estadístico al ponderar la posibilidad del azar en la selección de la muestra para producir la diferencia aparente. Aplicar estas medidas de significado estadístico es eliminar de las conclusiones algunas de las dudas. 6) Con frecuencia se equiparan humorísticamente los estadísticos y los embusteros. No hay duda de que los procedimientos estadísticos pueden usarse para demostrar cualquier cosa que uno desee. Partiendo de supuestos falsos, usando procedimientos inadecuados u omitiendo datos importantes, el investigador parcial puede llegar a conclusiones falsas. Esas conclusiones son con frecuencia particularmente peligrosas por la autenticidad que parecen conferir los procedimientos estadísticos. La tergiversación puede ser deliberada o no intencionada. El omitir ciertos hechos o elegir sólo los favorables a la posición de uno es tan culpable como la verdadera deformación. En realidad representa un tipo de falseamiento que no cabe en la investigación. El lector debe siempre controlar los datos, particularmente cuando el propósito del informe parece ser el persuadir.

Una observación final referente al uso de los procedimientos estadísticos en la investigación. Los informes, particularmente en las ciencias de la conducta, incluyen a veces complicados procedimientos estadísticos que no sirven para un propósito útil, sino que se incluyen, al parecer, para prestar un aire científico y digno a un estudio estéril o superficial. No hay que decir que no debería emplearse ninguna técnica estadística a menos que añada claridad o significado al análisis de los datos. Nunca debería utilizarse como pantalla para impresionar al lector.

"M"

MEDIA (x). Una medición descriptiva de la tendencia central apropiada para datos de intervalo; la suma de todas las puntuaciones dividida entre el número de puntuaciones.

MEDIA ARITMETICA, SIGNIFICADO DE LA. Para ver si una muestra posee un número suficiente de sujetos, de modo que los resultados obtenidos con pruebas experimentales, realizadas con dicha muestra, puedan ser inferidos a la población, el investigador tiene que hallar la Razón Crítica. Si la Razón Crítica es inferior a 1,96, el número de sujetos de la muestra es insuficiente, por lo que el nivel de probabilidad de esta muestra no llega ni a un 95 %. Será necesario aumentar el número de sujetos. Si la Razón Crítica fuera igual o superior a 1,96, la prueba tendría una probabilidad del 95% por lo que el número de sujetos sería suficiente.

MEDIANA. Una medición descriptiva de la tendencia central; el valor que divide a la distribución por la mitad. // Es el punto (no necesariamente una puntuación) situado en el centro de una distribución de valores o datos. Por encima de ella se encuentra el 50% de éstos y por debajo el otro 50%. Esta medida es más de posición que de magnitud.

El procedimiento para computarla con datos no agrupados es:

1. Ordenar las puntuaciones jerárquicamente (de mayor a menor).

DICCIONARIO DE PSICOMETRIA

2. Si el número de puntuaciones es impar, la mediana es la puntuación que se halla en el lugar medio de la ordenación.

3. Si el número de puntuaciones es par, la mediana es un punto medio entre las dos puntuaciones centrales. Ejemplo: 9, 8, 7, 4; 8 y 7 son las puntuaciones centrales, su punto medio es 7,5. La Md. es, pues 7,5.

4. La mediana no se halla afectada por los extremos de la serie. En ciertos tipos de análisis, la mediana puede reflejar con más exactitud que cualquier otra medida la posición de la tendencia central. La mediana se emplea para describir todos los casos de la distribución, pues la representa en forma razonablemente típica, ya que es correcto suponer que el caso o calificación medio es típico del grupo. MEDIANA, USO DE LA. Si no es utilizable la media; en distribuciones asimétricas; en distribuciones incompletas; cuando la unidad no es constante (escala ordinal).

MEDICION, ESCALAS DE. Véase "Escala de Medición"

MEDICION NO REACTIVA. Una medición de la conducta en la que el sujeto no es consciente de ser observado y donde la conducta del sujeto no cambia debido al proceso de observación.

MEDIDA. Asignación de números a realidades que pueden recibirlos de acuerdo con determinaciones específicas. Véase "Escala de Medición"

METODO ESTADISTICO. El método estadístico necesita, de un lado, la consideración retrospectiva de los principios fundamentales; de otro implica la respuesta a las siguientes preguntas: 1) ¿Cuáles son las cuestiones significativas que han de ser contestadas? Esta etapa muestra las directrices para recogida de datos. 2) ¿Qué hechos han de ser reunidos para obtener la información suficiente y necesaria que posibilite las respuestas? 3) ¿Cómo esos datos han de ser recogidos, organizados y analizados para que arrojen luz sobre el problema? 4) ¿Qué supuestos se hallan bajo la metodología estadística que se emplea? 5) ¿Qué conclusiones pueden ser derivadas lógicamente del análisis de los datos? En sentido amplio, el campo de la estadística abarca todas estas fases y ninguna investigación obtendrá conclusiones lógicas a menos que cada etapa haya sido cuidadosa y completamente realizada. La investigación pedagógica puede tratar con dos tipos de datos estadísticos: el descriptivo y el deductivo.

METODO EXPERIMENTAL. Un método usado para estudiar fenómenos en el que una o más variables independientes se manipulan y donde se mide el rendimiento en una o más variables dependientes.

METODOS NO PARAMETRICOS. Son aquellas técnicas que nos permiten comparar muestras y hacer inferencias o pruebas de significación, sin tener que suponer normalidad en las poblaciones. Entre estos métodos se encuentran:

- La prueba χ^2 , puesto que su significación depende sólo de los grados de libertad de la tabla, y no hay por qué hacer suposición alguna acerca de la forma de distribución de las variables clasificadas en las categorías de la tabla.
- El coeficiente de correlación para la diferencia de rango
- La prueba del signo es la prueba no paramétrica más sencilla. Se utiliza para hallar la significación de las diferencias entre los sujetos de los grupos, siendo calificadas estas diferencias como + o -.
- La prueba de la mediana que compara los rendimientos de dos grupos independientes. Primeramente se unen los dos grupos y se establece una mediana común. La mitad de las puntuaciones de cada grupo tendría que estar por encima y la

DICCIONARIO DE PSICOMETRIA

otra mitad por debajo de la mediana común, y esta hipótesis es la que habrá que verificar.

- La prueba de la suma de rangos se utiliza para verificar la hipótesis de que dos grupos independientes de observaciones han sido sacados al azar de la misma población, por lo que no habrá ninguna diferencia verdadera entre ellos.

MODA. Una medición descriptiva de la tendencia central; el valor que ocurre con mayor frecuencia.// La moda como medida de tendencia central (M_o), es el caso que más se repite en una distribución. Cuando se trata de una serie sencilla, basta buscar a simple vista cuál es el más repetido de los valores. Ej.: en la serie 3, 4, 5, 5, 5, 5, 5, 6, 6, 7, 7, 8, 8, 9, 9 sería 5. Cuando tenemos los valores agrupados en una tabla de frecuencias, podemos tomar como M_o el punto medio del intervalo que tiene mayor número de frecuencias. La moda se usa muy rara vez. Es el promedio que más fácilmente se halla y tiene valor aproximado. Es el valor más denso. Si una distribución tiene una moda, se llama unimodal, si tiene dos, bimodal; Sin.: modo, módulo.

MUESTRA. Grupo de sujetos seleccionados de la población.// Conjunto de elementos, que constituyendo una pequeña proporción de la población, son seleccionados del todo o de una parte (estrato), con carácter homogéneo, con la finalidad de su análisis. Mediante la observación de la muestra pueden hacerse inferencias acerca de la población.

Muestra aleatoria. La muestra al azar. Significa que los sujetos han sido seleccionados en la forma arbitraria que se emplea para obtener una sección representativa de la población sin tendencias predeterminadas, y que se considera libre de influencias de cualquiera de las variables implicadas. **Muestra estratificada.** A veces conviene subdividir la población en grupos homogéneos más pequeños, al objeto de lograr una representación más exacta. Por ejemplo, al realizar un estudio sobre los ingresos laborales en una determinada comunidad, una verdadera muestra nos daría aproximadamente el mismo número relativo desde cada nivel socioeconómico de la comunidad completa. Si la proporción fuera 15% de trabajadores profesionales, 10% de directivos, 20% de trabajadores especializados y 55% de no especializados, la muestra incluiría aproximadamente la misma proporción en orden a ser considerada representativa. Dentro de cada subgrupo, se emplea generalmente un proceso de selección por probabilidad. Este proceso proporciona al investigador una muestra más representativa que otra seleccionada a partir de la comunidad completa, que podría estar, sin duda, desequilibrada por una preponderancia de trabajadores no especializados. Es evidente que las características de la población completa deben ser consideradas con gran cuidado, junto con los propósitos del estudio antes de decidir sobre una muestra estratificada.

Muestra representativa. Significa que la distribución de las puntuaciones en la misma es casi paralela a la de la población, y contiene, aproximadamente, la misma proporción de sujetos pertenecientes a cada región geográfica, grupo cronológico, grupo ocupacional, etc., que la población completa de la que se tomó la muestra.

Muestra sesgada. Cuando por incorrección del procedimiento existe una causa de error, como cuando no todos los sujetos elegidos responden a un cuestionario por correo. Pueden faltar sectores esenciales.

Muestra, tamaños de la. La confiabilidad de una media o de una desviación típica depende del tamaño de la muestra en que se basa el error típico. Los errores típicos varían en proporción inversa a la raíz cuadrada del tamaño de la muestra, de modo que, cuanto mayor sea N (el número de sujetos), tanto menor es, por lo general, el error típico. Cuanto mayor sea el número de sujetos (correctamente seleccionados), tanto mayor será la desviación típica de la muestra y tanto más representará ésta a la población.

DICCIONARIO DE PSICOMETRIA

MUESTREO. Es el proceso de verificar la selección de un conjunto que se denomina población, para destacar parte de éste considerada más o menos representativa constituyendo la muestra. Hace la investigación más rápida, fácil y económica.

- Muestreo aleatorio. Un procedimiento usado para obtener muestras representativas de una población.

En el muestreo completamente aleatorio, cada sujeto en la población debe tener una oportunidad igual de ser seleccionado y la selección o no selección de un sujeto no puede influir en la selección o no selección de cualquier otro sujeto.

- Muestreo incidental. Este término ha de aplicarse a aquellos grupos a los que se recurre principalmente porque son fácilmente accesibles. Esos grupos casuales pocas veces constituyen muestras al azar de ninguna población definible. Las fórmulas de error típico son aplicables sólo con un alto grado de aproximación a las muestras incidentales. Las generalizaciones basadas sobre tales datos son a menudo engañosas.

"N"

NIVEL DE SIGNIFICACION. La probabilidad usada para definir un resultado experimental como un evento raro si sólo opera el azar y para definir lo que significa grupos equivalentes.

NIVELES DE MEDICION. Véase "Escalas de Medición"

NORMAS DE DESARROLLO. Se trata de sistemas de puntuación que describen cualitativamente

funciones específicas de la conducta. Relación media, común o tipo, bajo condiciones determinadas: p. ej.: la puntuación del test medio de instrucción de los niños de 9 años, o el peso medio al nacer de los varones. Las puntuaciones basadas en normas de desarrollo tienden a ser burdas en términos de la psicometría ya que no se prestan a un tratamiento estadístico preciso. No obstante resultan atractivas para propósitos descriptivos. Las más conocidas son la "Edad Mental" (procedimiento por el cual los reactivos se agrupan en niveles cronológicos); los "Grados Equivalentes" (usadas en ciertas pruebas de aprovechamiento escolar) y las "Escalas Ordinales de Desarrollo" provenientes del campo de la psicología infantil). Por ej. los programas de desarrollo de Gesell presentan, en meses, el nivel aproximado que ha alcanzado un niño en cuatro áreas principales de conducta: motora, adaptativa, lenguaje y personal – social, mientras que los estadios de desarrollo de Piaget señalan la descripción de los procesos cognitivos de la infancia (Anastasi, 1998)

NORMAS INTRAGRUPPO. La mayoría de los tests proporcionan alguna forma de normas intragrupo, en las que el puntaje obtenido por un sujeto se evalúa en términos referidos al grupo de estandarización adecuado. Las puntuaciones intragrupo tienen un significado cuantitativo uniforme y definido y pueden emplearse casi todos los análisis estadísticos. Las más utilizadas son los siguientes: - Percentil: las calificaciones percentiles se expresan en términos del porcentaje de sujetos de la muestra de estandarización que caen bajo determinada puntuación directa. Por ejemplo, si el 28% resuelve correctamente menos de 15 problemas aritméticos, entonces el puntaje directo 15 corresponde al percentil 28. Los percentiles no deben confundirse con los porcentajes, que son las puntuaciones directas que se expresan en términos del porcentaje de los ítems correctos y el total sino que son puntuaciones transformadas no lineales que se expresan en términos del porcentaje de individuos. El P 50 corresponde a la mediana. Los percentiles 25 y 75 se conocen como el primer y el tercer puntos cuartiles (Q1 y Q2) porque separan los cuarto inferior y superior de la distribución. - Puntajes Estándar o Puntuaciones Típicas (puntaje z): estas

DICCIONARIO DE PSICOMETRIA

puntuaciones se representan con la letra z y expresan la distancia del individuo de la media en términos del desvío estándar de la distribución normal. La transformación de puntuaciones brutas en puntajes z no cambia la distribución de los puntajes. Si la distribución original es asimétrica, también lo será la de z . Esta transformación es lineal, lo cual significa que hay una relación en línea recta entre las puntuaciones directas y las puntuaciones transformadas (z). El puntaje z se calcula restando la puntuación bruta a la media del grupo normativo y dividiendo luego por el desvío estándar ($(x - M) / DE$). La media es de 0 y el DE y la varianza de 1, por ello buena parte de las puntuaciones suelen ser negativas y casi todas decimales. Esto hace que resulte incómodo su tratamiento y que muchas veces se busquen procedimientos para superar esta dificultad. A estas transformaciones se las conoce con el nombre de escalas o puntuaciones derivadas. - Escalas Derivadas: Un procedimiento utilizado para salvar la dificultad recién mencionada es transformar las puntuaciones típicas (z) en otras que retengan todas las relaciones que manifiestan los puntajes originales pero evitando la dificultad operativa. Para ello, se parte de las puntuaciones directas, se tipifican y por último se transforman linealmente. El siguiente cuadro (Botella, León y San Martín, 1996) indica el esquema de transformación en una escala derivada

Puntuaciones directas (x_i)

Puntuaciones típicas (z)

Puntuaciones transformadas (T)

media media: 0 media: b

DE / V_z

DE / V_z : 1

DE / V_z : a^2

Tipificación

$[z = (x - m) / DE]$

Transformación lineal

en escala derivada

$[T = a \cdot z + b]$

Las puntuaciones T tienen media 50 y desvío estándar 10 y constituyen, junto con el cociente intelectual (CI), las transformaciones más conocidas y usadas en psicología. En el CI, la media es 100 y el DE 15. El término cociente intelectual se ha mantenido por tradición histórica más que por cómo mide la inteligencia. En ambos casos, existe una equivalencia entre las puntuaciones de las escalas derivadas.

-2DE -1DE m +1DE - 2

DE Puntaje z :

-2 -1 0 1 2

Puntaje T :

30 40 50 60 70

Puntaje CI:

70 85 100 115 130

- Puntajes estándar normalizados: son puntajes estándar expresados en términos de una distribución que ha sido transformada para ajustarse a una curva normal y que pueden adoptar cualquier forma conveniente. Por ejemplo, en las escalas clínicas del MMPI-2 la distribución de los puntajes directos no son iguales por lo que se utilizan puntuaciones T normalizadas, obtenidas a partir de una distribución centil, donde los porcentajes que quedan por encima y por debajo de una puntuación T determinada sean los mismos para cada escala.

DICCIONARIO DE PSICOMETRIA

"O"

OJIVA. Curva que reúne las frecuencias acumuladas o convertidas en porcentajes. El eje vertical de la gráfica se divide en 100 partes iguales. El valor principal de la ojiva se halla en el hecho de que pueden leerse directamente los percentiles. La representación gráfica que tiende a obtener una curva, sea la campana de Gaus o la ojiva de Galton, es realmente la representación de una relación entre dos series de datos numéricos. Estos se hallan unos en función de otros, y el problema se centra en la representación gráfica de una función dada en la curva correspondiente.

La solución se halla mediante la determinación convencional de un sistema de coordenadas rectangular, cartesiano u ortogonal. El punto cero es el origen de las coordenadas. El eje horizontal es el de las abscisas (x), y el eje vertical el de las ordenadas (y). Los dos ejes, en relación, se denominan coordenadas. Los cuadrantes son -cada uno de ellos- la cuarta parte de un plano infinito. Se numeran en sentido opuesto al giro de las agujas de un reloj. El cuadrante que nos interesa ahora es el primero bipoisitivo. El segundo es negativo-positivo; el tercero es binegativo; el cuarto es positivo-negativo

ORDEN. Una premisa de la ciencia que afirma que los eventos se suceden unos a otros en secuencias regulares.

ORDENADA. El eje Y o vertical de una gráfica, en general, variable dependiente

"p"

PARAMETRO. Estadístico respecto de la población, aceptado por su validez demostrada. Los parámetros son el correlato de los estadísticos en la población, que son más o menos representativos de ésta según el grado de aproximación y fidelidad de la muestra, de la cual se deducen las características del conjunto (población o universo). Cuanto menor es la diferencia entre el estadístico y el parámetro más representativa es la muestra. El valor correspondiente a la muestra en la población constituye el parámetro.

PEARSON, COEFICIENTE DE CORRELACION (r). Es el índice más significativo para expresar la correlación existente entre dos variables cuantitativas. (Véase "Coeficiente de correlación producto – momento de Pearson")

PERCENTIL. Describe una puntuación en relación a la posición de otras ordenadas, siendo el punto de la serie jerárquica por debajo del cual se encuentra un porcentaje dado de puntuaciones. El percentil 70 coincidiendo con la puntuación 65 indica que el 70% de las puntuaciones se halla por debajo de los 65 puntos. El sujeto situado en dicho percentil se hallaría en el 30% superior de la distribución y superaría al 70% de los incluidos en la misma. Véase "Normas intragrupo"

PERCENTIL, PUNTUACION DEL. Conociendo el percentil, si se desea hallar la puntuación correspondiente se procede así:

1. Multiplicar el número total de alumnos por el percentil dado, y dividir el resultado por 100. Restarle 1.
2. Contar desde abajo en la distribución tantos lugares como indique el resultado anterior.
3. El lugar hallado será correspondiente a la puntuación buscada.

PERCENTILAR. Distribuir una serie en 100 intervalos, cada uno de los cuales incluye exactamente el 1 por 100 del número total de casos.

DICCIONARIO DE PSICOMETRIA

Las notaciones percentiles facilitan:

- a) La interpretación significativa de las puntuaciones individuales.
- b) La comparación de puntuaciones obtenidas por un mismo sujeto en dos o más tests de una materia, o en pruebas distintas (ortografía, vocabulario).

PERSONALIDAD, MEDIDAS DE LA. Las llamadas escalas de personalidad siguen el tipo de autoexploración. El individuo responde a ciertas cuestiones o afirmaciones. Esos instrumentos proporcionan puntuaciones, las cuales se suponen que miden ciertos rasgos o tendencias de la personalidad. Pueden ser considerados como tests en la medida en que han sido cuidadosamente tipificados y obtienen puntuaciones cuantitativas. Pero tienen muchas de las características de los cuestionarios o escalas de autoevaluación.

A causa de la dificultad, incapacidad o resistencia de los individuos a manifestar objetiva o exactamente sus propias reacciones, esos instrumentos son de valor limitado. Parte de esta limitación puede ser debida a las inadecuadas teorías de la personalidad de acuerdo con las que se han elaborado. Aún así, proporcionan datos útiles sobre los cuales basar la necesidad de análisis posteriores. Algunos tienen una razonable validez empírica en grupos especiales de individuos, pero resultan inadecuados cuando se aplican a otros.

POBLACION. Las unidades potenciales para observación de donde se extrae la muestra para ser observada.// Se considera formada por todos los elementos o unidades correspondientes a un tipo homogéneo o, en caso contrario, agrupable por una o varias características definidas (atributos), sin.: universo o conjunto de todos los casos posibles de los cuales se elige una muestra. Las fórmulas estadísticas usuales para obtener deducciones a partir de las muestras se aplican cuando la población supera apreciablemente a la muestra (p. ej.: de cinco a diez veces mayor).

PORCENTAJE, FIABILIDAD DEL. Se multiplica por 100 la fiabilidad de la proporción.

POSICION RELATIVA, MEDIDAS DE. Una puntuación es plenamente significativa si se expresa en relación con la medida del grupo o conjunto a que pertenece el sujeto. Para ello tenemos los cuantiles, que dividen la serie de valores en partes iguales. Los más usados son: cuartiles, deciles, percentiles, centiles.

PROBABILIDAD. Un cálculo de la probabilidad de que ocurra un resultado particular; la razón del número de resultados favorables con el número total de resultados posibles si sólo opera el azar.

- Probabilidad a posteriori. Probabilidad estadística.
- Probabilidad a priori. Estudio de la relación entre casos reales y posibles.
- Probabilidad elemental. Si sometemos a medida, conteo y puntaje una cierta característica en parte de los elementos de una población, encontramos generalmente una frecuencia determinada entre ellos (p. ej.: entre los 100 muchachos de un colegio de 11 años de edad existen ciertas frecuencias de estaturas, es entonces presumible que entre los siguientes 100 muchachos de 11 años de ese colegio hallaré las mismas frecuencias de estaturas). Dispondré entonces de datos para definir la probabilidad a partir de una frecuencia relativa. En una estimación de la probabilidad, si en los primeros 100 muchachos de 11 años 30 miden 1,40 m de estatura (30/100) no hay razón fuerte que se oponga a que en el grupo siguiente no encuentre también 30/100, o sea, que es probable que la segunda muestra de la población de muchachos no difiera mucho de la primera. En el caso de lanzamiento de monedas, en el que cara y cruz son dos sucesos mutuamente excluyentes, la probabilidad de que aparezca una de entre las 2 es 1/2. Que salga cara en la moneda es $p = 1 / 2$, que no salga es $1 - 1 / 2$.

DICCIONARIO DE PSICOMETRIA

2. Si denotamos $1 - p = q$, entonces $p + q = 1$. Si tenemos n sucesos, H_i , p_i probabilidades ($i = 1, 2, 3, \dots k$), la probabilidad de H_1 , o H_2 , o H_3 , ... o H_k sería $p(H_1 + H_2 + H_3, \dots + H_k) = p_1 + p_2 + p_3 + \dots p_k$; siempre y cuando los sucesos sean excluyentes entre sí.

La probabilidad de que dos sucesos a y b se verifiquen simultáneamente, o sucesivamente, la denotamos $P_{a \text{ y } b}$, y es $P_{a \text{ y } b} = p_a \cdot p_b$.

Si tenemos en una bolsa 3 dados con puntos y 2 dados en blanco en todas sus caras, la probabilidad de obtener un dado con puntos en la primera extracción es $3 / 5$, la probabilidad de obtener un dado en blanco en la segunda es $2 / 4$, a la tercera extracción el que sea con puntos tendrá la probabilidad $2 / 3$, y así sucesivamente.

PROMEDIO. Valor representativo de una distribución de frecuencias, alrededor del cual se encuentran diversos otros valores; p. ej.: media, mediana, modo de una distribución de puntuaciones. Un promedio debe reunir las siguientes características:

- 1) Estar perfectamente definido, para que se pueda calcular en cualquier distribución.
- 2) Basarse en todas las observaciones, para que efectivamente represente a la distribución. Los promedios más usados en investigación pedagógica son: media aritmética, media geométrica, media armónica, moda, mediana.

PSICOMETRIA. Esta denominación abarca el conjunto de modelos formales que posibilitan la medición de variables psicológicas, centrándose en las condiciones que permitan llevar a cabo todo proceso de medición en Psicología.

PUNTAJE OBSERVADO. La puntuación obtenida por un sujeto es un test psicológico. Puntuación empírica de un test compuesta por una parte denominada *puntuación o puntaje verdadera/o* y de otra que es el *error de medida*

PUNTAJE TRANSFORMADO. Transformación de las puntuaciones brutas en cualquier escala derivada a fin de hacer comparables los resultados. Los más usuales son los puntajes de normas intragrupo (percentil, puntaje T, Cociente Intelectual por desviación). Véase "Normas Intragrupo"

PUNTAJE VERDADERO. La puntuación verdadera es constante, pero las puntuaciones observadas y los errores son aleatorios. Si, por ejemplo, suponemos que el *puntaje verdadero* de una persona en inteligencia es de $CI = 110$ no podemos predecir con certeza cuál será la puntuación observada o empírica que obtendrá esa persona en aplicaciones estandarizadas y repetidas del mismo test, dada la influencia del error de medida (Martínez Arias, 1996)

"R"

RANGO. Para algún grupo específico, la diferencia entre la más alta y la más baja de las puntuaciones obtenidas en una prueba; por lo tanto, una medida muy burda de expansión o variabilidad, ya que solamente se basa en dos puntuaciones extremas. El término rango se utiliza igualmente en lo que se refiere a la posible dispersión de la medición que permite una prueba, que en la mayoría de los casos, corresponde al número de ítems de la prueba.

RELACION LINEAL. Una relación entre dos variables en la que un incremento en una variable siempre va acompañada de un incremento constante o siempre va acompañado de una disminución constante en la otra.

DICCIONARIO DE PSICOMETRIA

RELACION MONOTONICA. Una relación entre dos variables en la que un incremento en una variable siempre va acompañado por un incremento o siempre va acompañado por una disminución en la otra variable.

ROTACION. Transformación analítica o gráfica de una matriz factorial con fines de interpretación material de los factores extraídos. Un criterio de rotación bastante empleado es la estructura factorial simple de L.

L. Thurstone. Requiere que los factores roten, para que, sobre o cerca de ellos, se encuentre el mayor número posible de puntos extremos de vectores de la característica. En la rotación gráfica o visual, esto se produce por evaluación subjetiva de la estructura factorial; y en la rotación analítica, por la definición matemática de valores óptimos máximos y mínimos. Los métodos de rotación se dividen, según el ángulo formado por los factores entre sí, en ortogonales y oblicuos. En las rotaciones ortogonales se mantiene la independencia o ausencia de correlación entre los factores, mientras que en las oblicuas se admite que los factores resultantes estén correlacionados.

ROTACION DE CRITERIO. Procedimiento propuesto por Eysenck (1950) para la rotación o la interpretación de un factor. En el análisis de los tests se acepta un criterio extremo que muestre una elevada correlación con el factor. Véase "Análisis Factorial"

"S"

SIGNIFICACION. La significación estadística se refiere a si los resultados obtenidos son un suceso raro o común si sólo opera al azar. La significación psicológica se refiere a la calidad de las ideas, la suficiencia de la prueba de la idea, y la claridad de los resultados.

SUPUESTO. Un principio básico de una teoría que se da por sentado y de donde se derivan otros principios; condición que debe satisfacerse antes de que se aprueben ciertas conclusiones (como los supuestos de una prueba estadística).

"T"

TECNICAS DE OBSERVACION. Un método de investigación basado en la observación de la conducta sin la manipulación sistemática de una variable independiente.

TEORIA. Una explicación tentativa de un fenómeno o conjunto de fenómenos.

TEORIA CLASICA DE LOS TESTS (TCT). Modelo psicométrico que describe la influencia de los errores de medida en las puntuaciones observadas y sus relaciones con las puntuaciones verdaderas. El error de medida es una desviación no sistemática o aleatoria de la puntuación verdadera. Los supuestos de la TCT son los siguientes (Martínez Arias, 1996):

- 1) Modelo lineal para las puntuaciones. La puntuación observada es igual a la suma de la puntuación verdadera y el error de medida (relación aditiva).
- 2) Para una población de personas medidas con el test o para una repetición infinita de medidas realizadas sobre la misma persona, el valor esperado de la variable error de medida (variable aleatoria) es igual a 0.
- 3) Las puntuaciones verdaderas y error no están correlacionadas. Esto implica que los errores no están correlacionados con atributo alguno ya que de lo contrario no sería aleatorio.
- 4) La puntuación de error de un test no está correlacionada con la puntuación de error de otro test. Esto significa que si un sujeto tiene un error positivo en el test 1, su error

DICCIONARIO DE PSICOMETRIA

en el test 2 podrá ser positivo, negativo o nulo, por lo que no puede predecirse a partir del tipo de error en un test qué error cometerá el sujeto en otro test.

5) Los errores de medida de un test no están correlacionados con las puntuaciones verdaderas de otro test. Este supuesto sería violado si el test 2 mide alguna dimensión que influye en los errores del test 1. Para una explicación detallada de la teoría, recomendamos consultar: Martínez Arias, R. (1996): "Psicometría: Teoría de los tests psicológicos y educativos" Madrid: Síntesis Psicología.

TEORIA DE LA RESPUESTA A ITEM O DE LAS CURVAS CARACTERISTICAS DE ITEM (TRI). Modelo psicométrico que considera al ítem como unidad de análisis básica del test en lugar de las puntuaciones totales como ocurre en la TCT. Partiendo de supuestos o hipótesis fuertes, intenta dar una fundamentación probabilística al problema de la medición de rasgos y constructos no observables (*rasgos latentes*). Es importante señalar que la medida básica que se utiliza en la TRI es la probabilidad de que un sujeto con determinada habilidad (rasgo latente) acierte en un ítem de dificultad conocida. Sin embargo, ello no implica que existan rasgos o habilidades subyacentes en sentido físico o fisiológico. Los rasgos latentes son constructos estadísticos derivados matemáticamente de relaciones empíricas observadas entre las respuestas a los tests. Las *curvas características del ítem* (CCI) se grafican según funciones matemáticas derivadas. Hay diferentes modelos TRI que emplean funciones distintas. Por ejemplo, en el modelo triparamétrico, las CCI son descritas por tres parámetros (discriminación; dificultad; adivinación) que se derivan de los datos empíricos (Anastasi, 1998). Los supuestos más importantes de la TRI se describen a continuación (Martínez Arias, 1996):

1) Unidimensionalidad: se exige que el rendimiento sea explicado por un factor dominante o aptitud medida por el test. Los modelos que se establecen sobre esta base se denominan unidimensionales; en cambio, aquellos en los que se asume la necesidad de más de un factor para explicar el rendimiento en el test son los multidimensionales.

2) Independencia local: significa que si se mantienen constantes las medidas que explican el rendimiento en el test, las respuestas de los examinados a un par de ítems cualesquiera son estadísticamente independientes.

Para una explicación detallada de la teoría, recomendamos consultar: Martínez Arias, R. (1996): "Psicometría: Teoría de los tests psicológicos y educativos" Madrid: Síntesis Psicología.

TEORIA DE LA GENERALIZABILIDAD. Modelo psicométrico propuesto por Cronbach y col (1963) que introdujo las siguientes modificaciones en la TCT:

1) utiliza el concepto de *muestreo de fuentes de variación múltiples*, lo que permite tratar cada característica de la situación de medición como una faceta de un plan de medición concebido como un diseño experimental. Cada una de estas facetas tiene su propia variabilidad, por ende, se consideran varias fuentes de error

2) sustituye el concepto de medidas paralelas por el de *medidas aleatoriamente paralelas*, considerando que las distintas condiciones de un procedimiento de evaluación son una muestra de un universo más amplio

3) reemplaza el concepto de *puntuación verdadera* por el de *puntuación de dominio o del universo*, definida como la esperanza matemática de todas las mediciones posibles

4) analiza la fiabilidad de los tests por medio del análisis de varianza en vez del análisis de la correlación entre puntuaciones verdaderas y observadas como en la TCT

Para una explicación detallada de la teoría, recomendamos consultar: Martínez Arias, R. (1996): "Psicometría: Teoría de los tests psicológicos y educativos" Madrid: Síntesis Psicología

DICCIONARIO DE PSICOMETRIA

TEST. Prueba psicológica para estudiar alguna función.

TESTS PSICOMETRICOS. Medida objetiva y estandarizada de una muestra de conducta. Las respuestas se evalúan según normas cuantitativas, todos sus elementos se valoran de forma numérica e independiente, el resultado final es una puntuación cuantitativa y suelen referirse a características psicológicas concebidas como más o menos independientes (rasgos / atributos).

- Tests de despistaje o *screening*: Técnicas psicométricas utilizadas para distinguir entre potenciales casos o no casos objeto de estudio. Su finalidad es realizar un tamizado básico respecto de la conducta o rasgo a medir por medio de algún puntaje de corte.

- Tests de ejecución. Se acostumbra llamar a los instrumentos tests de "ejecución" cuando le quitan importancia a los requisitos verbales, emplean materiales tridimensionales y requieren respuestas manipulativas. Debido a estos componentes, los tests de ejecución miden generalmente la coordinación motora, la velocidad, los factores perceptuales y los factores espaciales.

- Tests de ejecución máxima: la situación de cada ítem plantea un problema que el sujeto ha de resolver poniendo en funcionamiento su capacidad en alto grado. Son habituales en la evaluación del rendimiento, aptitudes, inteligencia, etc.

- Tests de ejecución típica: los ítems plantean situaciones habituales de la vida corriente y/o son una muestra del comportamiento típico o más frecuente del sujeto en determinadas situaciones. Son los tests más comunes en la evaluación de la personalidad, actitudes, intereses, etc.

- Tests de investigación. Se usan para obtener medidas, analizar diferencias individuales, calcular estadísticos y estimadores, pruebas de hipótesis, etc. Con frecuencia son operacionalizaciones de variables en psicología experimental y de las diferencias individuales entre otras.

- Tests de diagnóstico. Se usan para evaluar la calidad o el grado de algún rasgo o de la conducta de un sujeto.

- Tests referidos a normas. Técnicas psicométricas que utilizan como referencia para la comparación normas o baremos, derivados empíricamente en función de medidas estadísticas de tendencia central y de dispersión. La finalidad es describir al sujeto en el continuo de algún rasgo, expresando su posición relativa respecto al grupo de sujetos de la muestra normativa.

- Tests referidos a criterios. Técnicas psicométricas que se utilizan para evaluar el estatus absoluto de un sujeto con respecto a algún dominio de conductas bien definido. Algunos autores (Hambleton y Rogers, 1991) enfatizan la analogía con los llamados "tests de diagnóstico"

El siguiente cuadro muestra las diferencias entre ambos tipos de tests (Martínez Arias, 1996)

Aspecto	Tests referidos a normas	Tests referidos a criterios
Finalidad de la evaluación		

Poner de relieve diferencias interindividuales en la conducta o rasgo que mide el test

Estimar el rendimiento o conducta del sujeto en los objetivos que mide el test

Construcción del test / Contenido.

Los ítems suelen derivarse de alguna teoría de rasgos o constructos. Suelen ser limitados, ambiguos y sin delimitación clara del dominio de contenidos

Se comienza con una clara especificación del dominio de contenidos o conductas y del uso pretendido del test.

Selección de los ítems

Para poner de relieve las diferencias individuales, se maximiza la varianza del test, seleccionando ítems de dificultad media y alto poder discriminante.

Los ítems se seleccionan en función de los objetivos y del uso pretendido del test

Significado de las puntuaciones

DICCIONARIO DE PSICOMETRIA

La puntuación se considera un indicador de la puntuación verdadera en un "rasgo latente"

La puntuación representa el estimador muestral de la conducta o rendimiento del sujeto en el dominio.

Interpretación de las puntuaciones

Tiene significado sólo en relación a los resultados del grupo normativo

La puntuación tiene significado en términos absolutos (en relación al dominio)

- Tests verbales. Los instrumentos se denominan tests verbales si son formas impresas, acentúan la comprensión verbal y exigen respuestas simbólicas. Debido a la facilidad con la cual ciertas clases de materiales de tests pueden ser dados en formas impresas, los tests verbales tienden a medir la comprensión verbal, el cálculo numérico y los factores de razonamiento. No existe una separación neta entre los factores que se encuentran en los tests verbales y de ejecución; por el contrario, diferentes factores tienen tendencia a aparecer en las dos clases de materiales.

TESTS PROYECTIVOS. Técnicas de evaluación psicológica, especialmente referidas a características generales y globales de la personalidad. Para su evaluación se siguen criterios cualitativos donde las diferentes respuestas suelen valorarse con relación al resto de la población.

"V"

VALIDACION. Estudio tendiente a la comprobación de la validez de un test.

VALIDACION CRUZADA. 1) Procedimiento para controlar la validez de los resultados de tests, de creciente uso en la actualidad porque se ha observado que, al variar las muestras de validación, varían también los coeficientes de validación. Los coeficientes de validación hallados en una muestra vuelven a determinarse en otra muestra. Puede indagarse, por ej., si una serie de ítems de un test que han permitido diferenciar en un grupo diversos grados de expresión de una cualidad diferencian igualmente en otro grupo iguales grados. 2) Replicación independiente de la validación. Este método fue propuesto por Mosier en 1947 sobre la base de su observación acerca de la deficiente replicabilidad de los coeficientes de validez del test de Rorschach.

VALIDACION DIFERENCIAL. Designa la validación específica de tests para determinadas muestras. Durante mucho tiempo se entendió que las relaciones entre test y criterio son homogéneas en muestras grandes y sólo sufren alteración por errores de medida; pero ha podido demostrarse experimentalmente que los coeficientes de validación pueden ser muy diferentes para subgrupos específicos. Un test puede no ser válido, por ej., para una muestra de variables heterogéneas; y puede ser válido para subgrupos constituidos en relación a edad, sexo o características personales. Esta observación ha inducido a la construcción de tests especiales que predicen la validez de tests en determinados grupos de examinados. Cronbach ha propuesto una fórmula para el cálculo de la validación diferencial (Ghiselli, 1963)

VALIDEZ. Uno de los principales criterios de bondad de un test, que expresa la exactitud de éste en relación con un criterio. La validez indica el grado de exactitud con que un test "mide efectivamente aquella característica de personalidad o aquel comportamiento que debe o se propone medir" (Lienert). Según las directrices de la APA para tests pedagógicos y psicológicos, hay que distinguir entre validez de contenido, validez de criterio y validez de constructo.

Hay validez de contenido cuando el test o las tareas contienen la característica o las cualidades a medir. Se habla entonces, también, de validez lógica. Si el valor obtenido

DICCIONARIO DE PSICOMETRIA

en un test se pone en relación con un valor-criterio, hay validez de criterio. Aunque otras especies de validez se obtienen también o se controlan empíricamente, algunos autores designan la validez de criterio como validez empírica. Según los datos de los que se extraiga el criterio, se habla de validez (de criterio) interna o externa. Si se extrae como criterio un valor de test que mide constructos o cualidades semejantes, o constructos o cualidades desemejantes, se habla de validez interna. Si se extrae un criterio fuera de los valores del test (por ej., midiendo un rendimiento concreto o mediante juicio estimativo), se habla de validez externa. El punto temporal de la obtención del valor del test y del criterio hace razonable distinguir entre validez de coincidencia y validez de predicción. Si un valor de test ha de predecir valores de criterio que pertenecen al futuro, se habla de validez de predicción. En la validez de coincidencia no se exige este pronóstico. En la validez de constructo se trata principalmente de la clarificación teórica de un valor de test o de una variable. Indica la importancia psicológica que posee un valor o una variable. La determinación de la validez de un test es costosa en tiempo y en recursos económicos. Los resultados de tal análisis los da a conocer el autor del test a los usuarios. El tipo de valores de la validez de un test resulta del método estadístico empleado. Si se utiliza para la validación el método de los grupos extremos, la medida de validez resultante es la diferencia entre valores medios de los grupos. Si se utiliza el método de los grupos representativos, la medida son los coeficientes de correlación según la clase de datos existentes.

- Validez de Concepto. Denominación que da Cattell a la validez de constructo.
- Validez de Contenido. Expresa el grado en que el contenido de un test constituye una muestra representativa de los elementos del constructo que pretende evaluar. Vale decir si resulta una muestra representativa y relevante del dominio comportamental que se pretende medir. En la validación de contenido deben seguirse las siguientes fases.

- 1) Definición del universo de observaciones admisibles
- 2) Identificación de expertos en dicho universo
- 3) Juicio de los expertos acerca del contenido del test
- 4) Procedimiento para resumir los datos resultantes

En el caso de las pruebas de rendimiento, la validez es el grado hasta donde el contenido de la prueba representa un muestreo equilibrado y adecuado de los resultados (conocimientos, destrezas, etc.) del curso o del programa de enseñanza que pretende abarcar. Se demuestra mejor mediante una comparación del contenido de la prueba con los cursos impartidos, los materiales de enseñanza y las formulaciones de los objetivos de enseñanza y, muchas veces, mediante un análisis de los procesos requeridos para proporcionar las respuestas correctas a los ítems. La validez aparente, referente a una observación de lo que *parece* medir una prueba, es un tipo extratécnico de evidencia que, en algunos tipos de tests, puede resultar conveniente

- Validez de Constructo. Designación propuesta por Cronbach y Meehl (1955) para un tipo de análisis de validez en el que el significado del valor del test se verifica mediante conceptos o constructos psicológicos. Con la validez de constructo se intenta expresar algo sobre el significado psicológico o la base teórica de un test.

Se trata de un procedimiento muy genérico. Mientras en la validez orientada hacia el criterio no se plantea el problema de la causa del grado de correlación entre valor del test y valor del criterio, esta cuestión es fundamental en la validez de constructo. Esta incluye algunos pasos principales. En el primer paso el investigador debe formular, sobre la base de la teoría por él escogida, hipótesis acerca del grado mayor o menor de valores del test. Luego, seleccionar ítems o tests que representen manifestaciones del constructo y realizar la recolección de datos empíricos para poner a prueba las hipótesis. Finalmente, establecer la consistencia entre los datos y las hipótesis y examinar hasta qué punto podrían explicarse mediante modelos alternativos o

DICCIONARIO DE PSICOMETRIA

hipótesis rivales. La validez de constructo es esencial para realizar inferencias acerca de conductas o atributos que pueden agruparse bajo un constructo psicológico particular. Integra toda la evidencia que permite la interpretación de las puntuaciones de los tests.

A continuación se citan los procedimientos utilizados con más frecuencia en la validación de constructo, solos o en combinación.

1) Diferenciación entre grupos: se trata de establecer la validez a través de la diferenciación entre grupos de edades, en el caso de que la variable sea susceptible de ello. Ejemplo de esto son las habilidades cognitivas que aumentan con la edad. Este método es también apropiado cuando se busca la validación de constructos en escalas ordinales, como en el caso de la Teoría de los estadios de Piaget o los Estadios del desarrollo moral de Kohlberg.

2) Correlaciones con otras medidas del constructo: se trata del análisis de correlación entre un nuevo test y otros tests ya validados como evidencia de que el nuevo test mide el mismo constructo.

3) Análisis factorial: Véase "Análisis factorial"

4) Matrices Multimétodo – multirasgo: Para aplicar este procedimiento deben disponerse de dos o más modos de medir el constructo de interés e identificarse otros constructos diferentes (pero relacionados) que puedan ser medidos por los mismos métodos (Por ejemplo: Métodos: "Autoinforme" y "Evaluación de los compañeros"; Constructos: "Liderazgo"; "Sociabilidad" y "Popularidad"). Se selecciona una muestra de sujetos a los que se aplican todas las medidas de cada constructo y de cada método. Se calculan las correlaciones y se presentan en forma de una matriz denominada Matriz multimétodo – multirasgo. Se obtienen las siguientes datos:

- coeficientes de fiabilidad: correlaciones entre medidas del mismo constructo usando el mismo método de medida.
- coeficientes de validez convergente: correlaciones entre medidas del mismo constructo usando diferentes métodos.
- coeficientes de validez discriminante: correlaciones entre medidas de diferentes constructos usando el mismo método de medida o entre diferentes constructos utilizando distintos métodos.

5) Validación basada en la Teoría de la Generalizabilidad: se estudia si las observaciones para un individuo sobre un constructo son invariantes frente a distintos métodos de medida.

6) Contribuciones de la Psicología Cognitiva: validación mediante diferentes métodos propios de esta corriente. Por ejemplo, simulación de procesos a la exploración de lo que miden los tests de inteligencia; análisis de protocolos verbales, en los que se pide a los sujetos que "piensen en voz alta" mientras realizan una tarea; análisis cronométrico; técnica de correlatos cognitivos; movimientos oculares, etc.

- Validez de Criterio. Expresa las relaciones del constructo con otros constructos, operacionalizada, en general, en términos de correlaciones y regresiones del test con otras medidas//. El grado hasta donde las puntuaciones alcanzadas en una prueba concuerdan con (validez concurrente), o predicen (validez predictiva), alguna medida criterio. La validez predictiva se refiere a la exactitud con que una prueba de aptitud, de pronóstico o de madurez, indica el futuro éxito didáctico en alguna área, demostrados por las correlaciones observadas entre las puntuaciones alcanzadas en la prueba y futuras medidas criterio de semejante éxito (como por ejemplo, la relación existente entre la puntuación alcanzada en una prueba de aptitud académica, administrada durante la segunda enseñanza, y el promedio académico alcanzado en cuatro años de bachillerato). En la validez concurrente, no transcurre ningún intervalo de tiempo significativo entre la administración de la prueba que se valida y la medida

DICCIONARIO DE PSICOMETRIA

criterio. Este tipo de validez puede demostrarse a través de medidas concurrentes de capacidad y rendimiento académico, de la relación existente entre una nueva prueba y otra, generalmente aceptada o reconocida como válida, o mediante la correlación observada entre las puntuaciones obtenidas en una prueba y medidas criterio que son válidas, pero menos objetivas y más difíciles de obtener que una puntuación de prueba.// Este tipo de validez indica cuál es la eficacia del test en la predicción de algún tipo de comportamiento futuro del sujeto en situaciones específicas. Por ejemplo, un test de razonamiento abstracto podría utilizarse para predecir el éxito de un alumno en la disciplina de Matemáticas. Esta predicción será tanto más precisa cuanto mayor sea la validez del test de razonamiento con respecto al rendimiento en Matemáticas.

- Validez, coeficiente de. La validez relativa al criterio se juzga como el grado de relación entre el test que actúa como predictor y otra variable de interés que actúa como criterio.

Si la relación entre las puntuaciones en el test y las de la variable criterio es lineal, entonces el coeficiente de correlación lineal es una medida apropiada de la relación.

Se define el coeficiente de validez de una variable X con respecto a una segunda medida Y como el valor absoluto de la correlación

La variable X, que toma los valores de las puntuaciones en el test, es la variable predictor, y la variable Y que se predice se la denomina variable criterio.

Las medidas en la variable criterio pueden hacerse al mismo tiempo en que se administra el test, o en momentos diferentes. Estas situaciones han dado lugar a que a veces se distinga entre dos tipos de validez, la llamada validez concurrente y la denominada validez predictiva.

En la validez predictiva transcurre un período de tiempo entre la medición del predictor y del criterio y en ese período se produce algún tipo de intervención en los sujetos, o se gana experiencia, o conocimientos etc. La validez concurrente, por el contrario, implica la recogida de ambos tipos de medidas al mismo tiempo, o al menos en momentos tan contiguos como para que no pueda considerarse que se producen cambios ni en los sujetos ni en las condiciones.

Los procedimientos estadísticos utilizados para la validación en forma concurrente o predictiva son los mismos y en las últimas versiones de los "APA Tests Standards" se minimiza la importancia de esa distinción en cuanto a las relaciones temporales; sólo en ciertas situaciones aplicadas se mencionan, especialmente en las dirigidas a la selección de personal, en las que se suele utilizar la validación predictiva, quedando relegada la concurrente a los procesos de sustitución de unos instrumentos por otros. Esta definición dada del coeficiente de validez como la correlación entre la variable predictor X y el criterio Y, se puede relacionar en este contexto de la teoría clásica con el coeficiente de fiabilidad, como correlación entre dos medidas paralelas X y X'. Desde este punto de vista puede considerarse la fiabilidad de un test como su validez con respecto a un test paralelo.

- Validez Ecológica. Término ambiguo, aún no bien aclarado, que a veces no se distingue lo bastante de la "representatividad ecológica". La validez significa aquí (casi siempre) "legitimidad", la proporción de correspondencia o coincidencia. Pero desde la perspectiva de la validez ecológica no se juzgan sólo los tests y otros medios de investigación diagnóstica, sino también pruebas experimentales o cada uno de sus componentes, como tareas, condiciones marco espaciales y temporales, resultados de investigación, teorías, etc. Se busca así la correspondencia con el "biotipo" (Pawlik 1976), con las condiciones en circunstancias de vida "natural" (de ahí, ecológico). Este tipo de cuestiones sobre la validez o correspondencia se presenta especialmente cuando se abordan conocimientos científicos empíricos desde una situación relativamente artificial, construida específicamente, alejada del contexto vital natural (situaciones de test, de laboratorio).

DICCIONARIO DE PSICOMETRIA

VALOR ESTANDAR. (Ingl. Standard score): Expresión numérica de la ejecución de un test por la población utilizada para la estandarización de un test. La diferencia entre la puntuación no elaborada (raw score) de un individuo y el valor promedio de la población que ha servido para el establecimiento de los valores estándar, se expresa en medidas de dispersión (desviación estándar). El valor estándar no solamente sirve para comparar el resultado de individuos con el promedio de un grupo representativo, sino que permite comparar entre sí los resultados en la ejecución de tests diversos. Son valores estándar corrientes, para los tests de escalas, los siguientes: desviación del cociente de inteligencia, normas de notas escolares, normas estandinas, normas T, normas Z., Norma, normas de un test, sistemas de medida.

VARIABILIDAD. En biología y en psicología, grado en que se presentan variaciones, es decir, frecuencia y magnitud de las variaciones de una característica; valor promedio de esta variación. La variabilidad es una característica general de la naturaleza viva. Variabilidad de una figura: Relación entre variabilidad y constancia en una figura.

VARIABLE. Una cosa o evento que puede medirse o manipularse; un símbolo y su conjunto de valores.

- Variable atributiva. Una característica de un sujeto que puede medirse.
- Variable continua. Variable medida sobre una escala continua.
- Variable cualitativa. Característica cualitativa.
- Variable cuantitativa. Característica cuantitativa.
- Variable de confusión. Una variable no manipulada como variable independiente que varía sistemáticamente con una variable independiente de modo que los efectos de la variable de confusión y el efecto de la independiente no pueden separarse.
- Variable dependiente. En la experimentación, se llama "dependiente" la variable que se considera como función de una o más variables independientes y cuyos valores correspondientes pueden predecirse.

Dependencia de variables.

- Variable dicotómica. Una variable con dos y sólo dos categorías mutuamente excluyentes (hombre y mujer, sí y no).
- Variable discreta. Variable medida sobre una escala discreta.
- Variable independiente. La variable que en un experimento se hace variar a voluntad, de acuerdo con el plan.
- Variable interviniente. Variable interpuesta, es decir, magnitud que entra en juego entre observaciones anteriores, que son condiciones antecedentes, y la conducta observada, y a la que, aparte de la cualidad de un elemento funcional definido matemáticamente, no se atribuyen en la función de dependencia otras cualidades suplementarias (*surplus meaning*).
- Variable moderadora. Una variable que ejerce influjo en el grado de relación entre otras dos o más variables. Si se trata de un test, se habla de test moderador. Si se considera la variable moderadora en su función para la validez de un test, se la puede designar también como variable que divide las muestras en grupos. La edad, el sexo o las características de la personalidad son variables moderadoras. Bibl.: Saunders 1956.
- Variable ordinal. Variable cuantitativa, clasificada en una escala ordinal.
- Variable supresora. Variable predictora en una serie de predictores que no está correlacionada con el criterio, pero está altamente correlacionada con otro predictor. El empleo de una variable supresora puede aumentar la validez múltiple de una batería de tests por la predicción de un criterio, por ej., Ya que queda reprimida parte de la varianza sistemática de otros predictores que no están correlacionados con el criterio. Las variables supresoras tienen una elevada influencia (negativa, pero no todas son variables negativas).

DICCIONARIO DE PSICOMETRIA

VARIACION. Desviación respecto a un valor central (promedio, por ej.). Amplitud de la variación: Techo que media entre los valores mínimo y máximo de una característica en el conjunto de variaciones. Sin.: Amplitud de oscilación. Coeficiente de variación: Medida de la dispersión referida al valor medio//. En el estudio de las variaciones en herencia se distinguen: a) Variaciones hereditarias entre individuos pertenecientes a un mismo linaje, resultantes de nuevas combinaciones de factores hereditarios = combinación. b) Variaciones debidas únicamente a influencias recibidas en la vida (del medio ambiente, profesionales, residencia, etc.) = modificaciones no hereditarias. c) Variaciones hereditarias imprevisibles, debidas a causas desconocidas = mutación//. Expresándolo desde la combinatoria, la variación designa la posibilidad de poner n elementos diversos en combinación con k elementos, donde se permite la repetición de elementos. Sobre esta base, las variaciones se designan también a menudo como combinación con repetición.

VARIANZA. Para una distribución, el promedio del cuadrado de las desviaciones a partir de la media; por lo tanto, el cuadrado de la desviación estándar.

- Varianza de error y varianza verdadera. Si se somete cada una de dos o más muestras de sujetos a condiciones experimentales distintas, puede descomponerse la varianza total de los valores de medición resultantes del experimento en dos constituyentes: varianza sistemática y varianza de error (llamada también "error del experimento"). En los experimentos realizados correctamente, la varianza sistemática o verdadera refleja la influencia de las diversas condiciones experimentales y permite predecir su acción. La varianza de error resulta de las diferencias casuales en los objetos de experimentación (en la investigación psicológica, generalmente en sujetos). Puede darse por segura la intervención de las condiciones de experimentación si la varianza sistemática no es varias veces mayor que la varianza de error. Por ello, el examinador procura, mediante una hábil elaboración del plan, mantener la varianza de error lo más baja posible.

- Varianza homogénea de muestras. Existe cuando las varianzas obtenidas para dos o más muestras no son significativamente diversas en su magnitud. La homogeneidad de las varianzas de muestras es un presupuesto para la aplicación de diversos métodos estadísticos y se puede comprobar mediante el test de Bartlett o el test F.

- Varianza residual. Se llama así aquella parte de la varianza total de los datos que no se puede reducir a variación sistemática, experimental. Representa en el análisis de la varianza la magnitud con la que se compara la "intervarianza" cuando no es posible una estimación directa de la varianza de error ("intravarianza").

VERIFICACION. "Demostración científica", medio auxiliar de la investigación. Con arreglo a la lógica, la verificación es la forma progresiva de la conclusión por reducción. La verificación sucede a la formación de hipótesis y permite deducir (casi siempre aduciendo leyes ya seguras) otros hechos nuevos en el terreno de la investigación y en condiciones dadas. Si los hechos se producen, entonces la hipótesis ha sido verificada y queda convertida en ley, siempre que los hechos exigidos por la hipótesis hayan tenido lugar sin excepción. Pero, si no suceden los hechos esperados para la verificación de una hipótesis, hay que hablar de falsación. Teoría falsable, pensamiento.

Bibliografía

- Anastasi, A, Urbina, S. (1998): *Tests Psicológicos*. México: Prentice Hall

DICCIONARIO DE PSICOMETRIA

- Botella, J., León, O.G., San Martín, R. (1996): *Análisis de datos en Psicología I*. Madrid: Pirámide
- Casullo, M.M. (1996). *Evaluación Psicológica y Psicodiagnóstico*. Capítulos 2 y 3. Buenos Aires: Catálogos.
- Casullo, M.M. (1999). La evaluación psicológica: Modelos, Técnicas y Contexto sociocultural. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 1, 97-113
- Casullo, M.M; Figueroa, N; Aszkenazi, M. (1991). *Teoría y técnicas de evaluación psicológica*. Buenos Aires: Psicoteca. Caps. 1y 2.
- Reichardt, Ch. S. y Cook, T.D. (1986): *Métodos Cualitativos y Cuantitativos en Investigación Evaluativa*. Madrid: Morata.
- Cortada de Kohan, N. (1977): *El profesor y la orientación vocacional*. México: Trillas
- Cortada de Kohan, N., Carro, J. M. (1978): *Estadística Aplicada*. Buenos Aires: Eudeba
- Dorsch, F. (1978): *Diccionario de Psicología*. Barcelona: Herder
- Gonzalvo Mainar, G. (1978): *Diccionario de Metodología Estadística*. Madrid: Morata
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley
- Machrens, W., Schmann, S. (1982): *Medición y Evaluación en la Educación y la Psicología*. México: Cecsa
- Marín.G. (1986). Metodología de la Investigación Psicológica. *Acta Psiquiátrica y Psicológica de América Latina*, 32, 3
- Martínez Arias, R. (1996): *Psicometría: Teoría de los tests psicológicos y educativos*. Madrid: Síntesis Psicología
- Yela, M. (1957): *La técnica del análisis factorial*. Madrid: Biblioteca Nueva

Autor: María Elena Brenlla (comp.)

Co – autores: María Eugenia Basile, Paula Blezowsky, Alicia N. Cayssials, María Mercedes Fernández Liporace, Vanesa Góngora, Selva Moretto, Carina Naisberg, Marcelo Pérez, María José Scheinsohn, Adriana Wiater, Fabiana Uriel