

MUESTRA Y MUESTREO
VALIDEZ Y CONFIABILIDAD
ESP. VICTOR VEGAS

INDICE DE CONTENIDO

- **ESTIMACIÓN DEL TAMAÑO DE UNA MUESTRA**
- **EL MUESTREO**
- **DISTRIBUCIÓN DE LA MUESTRA**
- **REQUISITOS QUE DEBE REUNIR UN INSTRUMENTO DE MEDICIÓN**
- **CONFIABILIDAD DE UN INSTRUMENTO**
- **VALIDEZ DE UN INSTRUMENTO**
- **OTROS ELEMENTOS RELACIONADOS CON LA VALIDEZ**

ESTIMACION DEL TAMAÑO DE UNA MUESTRA

En estadística el tamaño de una muestra no es otra cosa que el número de unidades de observación (sujetos u objetos) extraídos de la población objeto de estudio de forma tal que el tamaño muestral sea representativo de dicha población.

Ahora bien hay que tener presente que existen diferentes formulas para determinar o calcular el tamaño de una muestra, y esto va a depender primero de los parámetros que se conozcan, y segundo del tipo de estudio o investigación que se está realizando.

FORMULAS DE CALCULO:

(Cuando el estudio es de la media aritmética y la población es infinita y se conoce la desviación estándar)

$$n = \frac{(Z)^2 * S^2}{e^2}$$

(Cuando el estudio es de la media y la población es finita y se conoce la desviación estándar)

$$n = \frac{(Z)^2 * S^2 * NP}{(NP - 1) * e^2 + (Z)^2 * S^2}$$

(Cuando el estudio es de la proporción y la población es infinita y se conoce o P o Q)

$$n = \frac{Z^2 * P * Q}{e^2}$$

(Cuando el estudio es de la proporción y la población es finita y se conoce o P o Q)

$$n = \frac{(Z)^2 * P * Q * NP}{(NP - 1) * (e)^2 + (Z)^2 * P * Q}$$

EJEMPLO DE CALCULO PARA EL ESTUDIO DE LA MEDIA ARITMETICA Y LA POBLACION ES INFINITA

Se quiere estimar la estatura media de un grupo numeroso de estudiantes con un error no mayor de 2 centímetros y un 95% de confianza en la estimación.

Por estudios anteriores se conoce que la desviación típica de la distribución de estaturas es de aproximadamente 10 Cm. ¿Cuál será el tamaño muestral ideal para realizar el estudio en cuestión?

FORMULA A UTILIZAR:

Cuando el estudio es de la media aritmetica y la poblacion es infinita y se conoce la desviación estándar)

$$n = \frac{(Z)^2 * S^2}{e^2}$$

DATOS:

$$Zc^2 = (1,96)^2 = 3,84$$

$$S^2 = (10)^2 = 100$$

$$e^2 = (2)^2 = 4$$

$$n = \frac{(3.84) * (100)}{(4)} = 96$$

SOLUCION:

(96 Estudiantes sería el tamaño mínimo de la muestra para el estudio)

VEAMOS EL MISMO EJEMPLO DE CALCULO PERO PARA UNA POBLACIÓN FINITA

Se quiere estimar la estatura media de una población de 600 estudiantes de una ciudad determinada con un error no mayor de 2 centímetros y un 95% de confianza en la estimación.

Por estudios anteriores se conoce que la desviación típica de la distribución de estaturas es de aproximadamente 10 Cm. ¿Cuál será el tamaño de muestra ideal para realizar el estudio en cuestión?. **(NOTESE QUE EN ESTE CASO LA POBLACIÓN ES FINITA 600 ESTUDIANTES)**

FORMULA A UTILIZAR:

(Cuando el estudio es de la media y la población es finita y se conoce la desviación estándar)

$$n = \frac{(Z)^2 * S^2 * NP}{(NP - 1) * e^2 + (Z)^2 * S^2}$$

DATOS:

$$Zc^2 = (1,96)^2 = 3,84$$

$$s^2 = (10)^2 = 100$$

$$e^2 = (2)^2 = 4$$

$$NP = 600$$

$$NP - 1 = 599$$

SOLUCION:

$$n = \frac{(3.84) * (100) * (600)}{(599) * (4) + (3.84) * (100)} = 83$$

(83 Estudiantes sería el tamaño mínimo de la muestra para el estudio)

EJEMPLO DE CALCULO PARA EL ESTUDIO DE UNA PROPORCION O PORCENTAJE Y LA POBLACION ES INFINITA

Se desea estimar el porcentaje de estudiantes que estaría a favor de un determinado candidato rectoral de una universidad del país, con un error no mayor de 5% y con un 95% de confianza. Determine el tamaño muestral que se debe escoger para el estudio.

Se supone que el porcentaje a favor del candidato estará alrededor del 40%

FORMULA A UTILIZAR:

(Cuando el estudio es de la proporción y la población es infinita)

$$n = \frac{Z^2 * P * Q}{e^2}$$

DATOS:

$$Zc^2 = (1,96)^2 = 3,84$$

$$P = (40\%)$$

$$Q = (100-40=60\%)$$

$$e = (5\%)^2 = 25$$

SOLUCION:

$$n = \frac{(3.84) * (40) * (60)}{(25)} = 368.64 \cong 369$$

(369 estudiantes sería el tamaño muestral para el estudio)

EJEMPLO DE CALCULO PARA EL ESTUDIO DE UNA PROPORCION O PORCENTAJE Y LA POBLACION ES FINITA

Se desea estimar el porcentaje de estudiantes que estaría a favor de un determinado candidato rectoral de una universidad del país, con un error no mayor de 5% y con un 95% de confianza. Se sabe que la población total de votantes es de 1200 estudiantes. Determine el tamaño muestral que se debe escoger para el estudio.

Se supone que el porcentaje a favor del candidato estará alrededor del 40%

FORMULA A UTILIZAR:

(Cuando el estudio es de la proporción y la población es finita y se conoce P o Q)

$$n = \frac{(Z)^2 * P * Q * NP}{(NP - 1) * (e)^2 + (Z)^2 * P * Q}$$

DATOS:

$$NP = 1200$$

$$Zc^2 = (1,96)^2 = 3,84$$

$$P = (40\%)$$

$$Q = (100-40=60\%)$$

$$e = (5\%)^2 = 25$$

$$NP-1 = 1199$$

SOLUCION:

$$n = \frac{(3.84) * (40) * (60) * (1200)}{(1199) * (25) + (3.84) * (40) * (60)} = 282,18 \cong 282$$

(282 estudiantes sería el tamaño muestral para el estudio)

EL MUESTREO

En estadística se entiende por muestreo, la forma como son escogidas las unidades de observación de una muestra, de manera tal que existen varios procedimientos o tipos de muestreo.

El muestreo es por lo tanto una herramienta de la investigación científica, cuya función básica es determinar como y que parte de una población debe examinarse, con la finalidad de hacer inferencias sobre dicha población.

Los errores más comunes que se pueden cometer son:

- 1.- Hacer conclusiones muy generales a partir de la observación de sólo una parte de la Población, se denomina error de muestreo.**
- 2.- Hacer conclusiones hacia una Población mucho más grandes de la que originalmente se tomo la muestra. Error de Inferencia.**

TIPOS DE MUESTREO

MUESTREO ALEATORIO SIMPLE:

El procedimiento empleado es el siguiente: 1) se asigna un número a cada individuo de la población y 2) a través de algún medio mecánico (bolas dentro de una bolsa, tablas de números aleatorios, números aleatorios generados con una calculadora u ordenador, etc.) se eligen tantos sujetos como sea necesario para completar el tamaño de muestra requerido.

MUESTREO ALEATORIO ESTRATIFICADO:

Consiste en considerar categorías típicas diferentes entre sí (estratos) que poseen gran homogeneidad respecto a alguna característica (se puede estratificar, por ejemplo, según la profesión, el municipio de residencia, el sexo, el estado civil, etc.). Lo que se pretende con este tipo de muestreo es asegurarse de que todos los estratos de interés estarán representados adecuadamente en la muestra.

DISTRIBUCIÓN DE LA MUESTRA

La distribución de la muestra en función de los diferentes estratos se denomina afijación, y puede ser de diferentes tipos:

AFIJACIÓN SIMPLE:

A cada estrato le corresponde igual número de elementos muestrales.

AFIJACIÓN PROPORCIONAL:

La distribución se hace de acuerdo con el peso (tamaño) de la población en cada estrato.

EJEMPLOS DE DISTRIBUCIÓN DE LA MUESTRA

AFIJACIÓN SIMPLE: Supóngase que se ha calculado el tamaño muestral de un estudio acerca del rendimiento académico de los estudiantes de una universidad que tiene 4 carreras, cuyo resultado fue de 120 estudiantes.

El procedimiento es simplemente dividir el tamaño muestral entre los cuatro estratos que en este caso son las carreras $120/4 = 30$. Luego se toman muestras de 30 estudiantes al azar en cada estrato

CARRERAS	ALUMNOS	MUESTRA
ADMINISTRACIÓN	250	30
CONTADURÍA	240	30
EDUCACIÓN	351	30
TURISMO	125	30
TOTAL	966	120

EJEMPLOS DE DISTRIBUCIÓN DE LA MUESTRA

AFIJACIÓN PROPORCIONAL: Supóngase que se ha calculado el tamaño muestral de un estudio acerca del rendimiento académico de los estudiantes de una universidad que tiene 4 carreras, cuyo resultado fue de 120 estudiantes.

El procedimiento en este caso implica calcular el porcentaje que tiene cada estrato o carrera con respecto al total y luego distribuir la muestra de modo proporcional, de la siguiente forma el 26% de 120 = $120 \times 26 / 100 = 31$ estudiantes de la carrera de administración, luego 25% de 120 = $120 \times 25 / 100 = 30$ de la carrera de contaduría y así sucesivamente para todos los estratos o carreras.

CARRERAS	ALUMNOS	%	MUESTRA
ADMINISTRACIÓN	250	26	31
CONTADURÍA	240	25	30
EDUCACIÓN	351	36	43
TURISMO	125	13	16
TOTAL	966	100	120

REQUISITOS DE DEBE REUNIR UN INSTRUMENTO DE MEDICIÓN

Toda medición o instrumento de recolección de los datos debe reunir dos requisitos esenciales: **CONFIABILIDAD Y VALIDEZ**.

QUÉ ES LA VALIDEZ

La validez, en términos generales, se refiere al grado en que un instrumento realmente mide la variable que pretende medir. Por ejemplo, un instrumento para medir la inteligencia es válido si realmente mide la inteligencia y no la memoria. Una prueba sobre conocimientos de Historia debe medir esto y no por ejemplo conocimientos de autores que escriben de historia.

QUÉ ES LA CONFIABILIDAD

La *confiabilidad* de un instrumento de medición se refiere al grado en que su aplicación repetida al mismo sujeto u objeto, en diferentes momentos produce resultados similares.

CONFIABILIDAD

Por ejemplo, si usted mide hoy el recorrido que hace su vehículo a 60Kph durante tres minutos, verá que será de 3 kilómetros. Si hace la misma prueba una semana después en las mismas condiciones, el resultado será el mismo.

Ahora bien, si aplicamos un test para medir el stress a un sujeto hoy, y luego una semana después le aplicamos el mismo test en las mismas condiciones, lo más probable es que los resultados no sean los mismos, puede ser que los niveles de stress sean mayores o menores.

Tradicionalmente, la fiabilidad de un test puede entenderse de tres maneras diferentes:

- a) Aludiendo a la estabilidad temporal de las medidas que proporciona.**
- b) Haciendo referencia al grado en que diferentes partes del test miden un rasgo de manera consistente.**
- c) Enfatizando el grado de equivalencia entre dos formas paralelas.**

CONFIABILIDAD

❑ CONFIABILIDAD COMO ESTABILIDAD TEMPORAL

Si disponemos de las puntuaciones de N personas en un test y, después de transcurrido un tiempo, volvemos a medir a las mismas personas en el mismo test, cabe suponer que siendo el test altamente fiable, deberíamos obtener una magnitud de correlación muy cerca de (1) entre ambas mediciones. Dicha correlación entre la evaluación test y el retest (r_{xx}) se denomina coeficiente de fiabilidad test-retest, y cuanto más está cerca de 1 mayor confiabilidad tendrá el instrumento. Este método es conocido como el modelo clásico para evaluar la confiabilidad de un instrumento.

CONFIABILIDAD



Ejemplo: A una muestra de 6 estudiantes de la UNEG se le aplica un cuestionario de hábitos de estudio. Transcurridos dos meses, se vuelve a aplicar el mismo test a los mismos estudiantes bajo las mismas condiciones. Sus puntuaciones en las dos aplicaciones fueron las siguientes:

$$r = \frac{(n) * (\Sigma X * Y) - (\Sigma X) * (\Sigma Y)}{\sqrt{[(n * \Sigma X^2) - (\Sigma X)^2] * [(N * \Sigma Y^2) - (\Sigma Y)^2]}}$$

NOMBRES	TEST	RETEST
LAURA	10	9
MARÍA	8	8
PEDRO	10	10
JOSÉ	7	7
ALBERTO	9	9
JUAN	8	7
CORRELACIÓN	0,90 MUY ALTA	

CONFIABILIDAD

Si por el contrario, los resultados fueran los que se muestran en la tabla, esto indicaría que la correlación entre los resultados es muy baja y por lo tanto el instrumento carece de confiabilidad.

$$r = \frac{(n) * (\Sigma X * Y) - (\Sigma X) * (\Sigma Y)}{\sqrt{[(n * \Sigma X^2) - (\Sigma X)^2] * [(N * \Sigma Y^2) - (\Sigma Y)^2]}}$$

NOMBRES	TEST	RETEST
LAURA	10	6
MARÍA	8	4
PEDRO	10	5
JOSÉ	7	6
ALBERTO	9	5
JUAN	8	4
CORRELACIÓN	0,18 MUY BAJA	

CONFIABILIDAD

❑ CONFIABILIDAD COMO CONSISTENCIA INTERNA

La precisión o fiabilidad de un test se puede entender también como el grado en que diferentes subconjuntos de items miden un rasgo o comportamiento homogéneo; es decir, el grado en que covarían, correlacionan o son consistentes entre sí diferentes partes del cuestionario. Lo más usual es obtener la consistencia entre dos mitades del test (método de dos mitades)

Igualmente hay que señalar, que se pueden aplicar los métodos de formulas directas, tales como el ALFA DE CRONBACH, EL KR20 DE KUDER RICHARDSON, EL PUNTO BISERIAL, entre otros

CONFIABILIDAD



Ejemplo: Se ha aplicado una prueba de habilidad verbal a un grupo de estudiantes de la UNEG y se han obtenido los siguientes resultados

TEST	PUNTOS	PARES	PUNTOS	IMPARES	PUNTOS
1	10	2	9	1	10
2	9	4	8	3	8
3	8	6	7	5	7
4	8	8	8	7	7
5	7	10	8	9	7
6	7	12	7	11	7
7	7	14	8	13	8
8	8	16	6	15	5
9	7	0,899			
10	8				
11	7				
12	7				
13	8				
14	8				
15	5				
16	6				

CONFIABILIDAD

❑ CONFIABILIDAD COMO CORRELACIÓN ENTRE FORMAS PARALELAS

A veces, por razones de índole práctico o investigativo, se diseña un test y una segunda versión del mismo, denominada forma paralela, que intenta evaluar o medir lo mismo que el test original pero con diferentes ítems. Como ya hemos explicado, dos versiones o formas se consideran paralelas si, aplicadas a una misma muestra de personas, obtienen medias y varianzas probabilísticamente similares.

La correlación de Pearson entre las puntuaciones obtenidas en una misma muestra en dos formas paralelas se considera el coeficiente de fiabilidad de cualquiera de ellas, e indicará el grado en que pueden considerarse equivalentes.

CONFIABILIDAD



EJEMPLO: Supóngase que se ha aplicado un una encuesta a un grupo de trabajadores de una determinada empresa para determinar el grado de satisfacción laboral , y posteriormente el investigador realiza un reajuste de la encuesta o forma 1 y aplica, una segunda forma, obteniendo los siguientes resultados:

SUJETOS	PUNTUACIÓN FORMA 1	PUNTUACIÓN FORMA 2
1	12	13
2	14	15
3	10	8
4	14	14
5	12	13

0,9289 INDICA ALTA CORRELACIÓN

No es común diseñar una forma paralela de un test para obtener datos sobre su fiabilidad. Cuando se diseñan (tarea por otra parte difícil) es porque van a utilizarse en determinados trabajos que requieren 2 aplicaciones sucesivas de un test que se puede recordar con facilidad.

MODELO ALFA DE CRONBACH

El modelo de confiabilidad ALFA DE CRONBACH, se suele utilizar cuando el instrumento piloto utilizado para su validación es cerrado, y está conformado en su mayoría por ítems que presentan varias alternativas de respuesta en cada caso.

Para su aplicación efectiva, se suele tomar una muestra piloto a la cual se le aplica el instrumento diseñado, una vez aplicado se vacían las respuestas emitidas por los encuestados en una matriz de datos, y a partir de ella se calcula el coeficiente de confiabilidad aplicando la formula siguiente:

$$\alpha = \frac{K}{K-1} * \left(1 - \frac{\sum S^2_j}{\sum S^2_t} \right)$$

DONDE:

(K) = Número de Items

S^2_j = Sumatoria de la Varianza del Instrumento

S^2_t = Varianza de la suma de los ítems

EJEMPLO DEL CALCULO PARA EL COEFICIENTE ALFA DE CRONBACH

Supongamos que se ha aplicado una prueba piloto de 12 ítems cerrados con 3 alternativas de respuesta (3, SIEMPRE), (2, ALGUNAS VECES), y (1, NUNCA) a una muestra de 10 participantes en una investigación acerca de los hábitos de estudio de los alumnos de la UNEG y los resultados son los siguientes:

ITEMS	SUJETOS									
	1	2	3	4	5	6	7	8	9	10
1	3	2	3	3	1	2	3	3	3	3
2	3	2	3	2	3	1	3	3	3	2
3	3	2	3	1	3	1	3	3	3	1
4	2	1	3	1	3	1	2	2	2	1
5	3	1	2	2	3	3	3	3	3	2
6	3	2	2	3	3	3	3	3	3	3
7	3	2	1	3	2	3	2	2	2	3
8	3	1	2	3	3	3	2	2	2	3
9	2	2	2	2	3	1	1	1	1	2
10	2	2	3	2	3	2	2	2	2	2
11	2	1	3	2	2	1	1	1	1	2
12	3	2	3	3	2	2	3	3	3	3

ITEMS	SUJETOS										VAR ITEMS
	1	2	3	4	5	6	7	8	9	10	
1	3	2	3	3	1	2	2	3	3	3	0,50
2	3	1	3	2	3	1	3	3	1	2	0,84
3	3	1	3	1	3	1	3	1	1	1	1,07
4	2	1	3	1	3	1	2	2	2	1	0,62
5	3	1	2	2	3	3	3	1	3	2	0,68
6	3	1	2	3	3	3	3	3	3	3	0,46
7	3	1	1	3	2	3	2	2	2	3	0,62
8	3	2	1	3	3	3	2	2	2	3	0,49
9	3	1	1	2	3	1	1	1	1	2	0,71
10	3	1	3	2	3	2	2	2	2	2	0,40
11	3	1	3	2	2	1	1	1	1	2	0,68
12	3	2	3	3	2	2	3	3	3	3	0,23
VAR ITEMS	35,00	15,00	28,00	27,00	31,00	23,00	27,00	24,00	24,00	27,00	
SUMA VAR ITEMS	7,30										
VAR INSTRU	27,88										
α	0,81										

$$\alpha = \frac{K}{K-1} * \left(1 - \frac{\sum S^2_j}{\sum S^2_t} \right)$$

$$\alpha = \frac{12}{12-1} * \left(1 - \frac{7,30}{27,88} \right) = 0,81$$

MODELO KR20 DE KUDER RICHARDSON

El modelo de confiabilidad KR20, se suele utilizar cuando el instrumento piloto utilizado para su validación es cerrado, y está conformado por ítems con dos categorías de respuesta, o sea ítems dicotómicos

Para su aplicación efectiva, se suele tomar una muestra piloto a la cual se le aplica el instrumento diseñado, una vez aplicado se vacían las respuestas emitidas por los encuestados en una matriz de datos, y a partir de ella se calcula el coeficiente de confiabilidad aplicando la formula siguiente:

$$Kr20 = \frac{K}{K-1} * \left(1 - \frac{\Sigma P * Q}{S^2} \right)$$

DONDE:

(K) = Número de Items

P = Proporción de Respuestas a favor

Q = Proporción de Respuestas en Contra (Q = 1- P)

S² = Varianza del Instrumento

EJEMPLO DEL CALCULO PARA EL COEFICIENTE KR20 DE KUDER RICHARDSON



ITEMS											
SUJETOS	A	B	C	D	E	F	G	H	I	J	PUNTOS
A	0	1	1	1	0	1	1	1	1	1	8
B	1	1	1	1	0	1	1	1	1	1	9
C	1	1	0	0	1	1	0	1	0	1	6
D	1	0	1	1	1	1	1	1	1	1	9
E	0	0	0	0	0	0	0	0	0	0	0
F	1	1	0	1	1	1	1	1	1	1	9
G	1	1	1	1	1	1	1	1	1	0	9
H	0	0	0	0	0	0	0	0	0	0	0
I	1	1	1	0	1	1	1	1	1	1	9
J	1	1	1	1	1	1	1	1	0	1	9
Σ	7	7	6	6	6	8	7	8	6	7	13,7333333
P	7/10 = 0,7	0,7	0,6	0,6	0,6	0,8	0,7	0,8	0,6	0,7	
Q	1-0,7 = 0,3	0,3	0,4	0,4	0,4	0,2	0,3	0,2	0,4	0,3	
P.Q	0,7*0,3 = 0,21	0,21	0,24	0,24	0,24	0,16	0,21	0,16	0,24	0,21	
ΣP.Q	2,12										0,93959008

$$Kr20 = \frac{K}{K-1} * \left(1 - \frac{\Sigma P * Q}{S^2} \right)$$

$$Kr20 = \frac{10}{10-1} * \left(1 - \frac{2,12}{13,73} \right) = 0,93$$

INTERPRETACIÓN DE LOS COEFICIENTES DE CONSISTENCIA INTERNA.

- **Expresa la proporción de varianza resultante de la relación entre los ítems. Un coeficiente de 0,70 indica que el 70% de la varianza se debe a lo que los ítems tienen en común, y un 30% se debe a errores de medición.**
- **Son indicadores de homogeneidad de los ítems, de que todos midan lo mismo. La homogeneidad conceptual se interpreta como descriptor del rasgo que suponemos presente en todos los ítems.**
- **Son una estimación del coeficiente de correlación que podemos esperar con un test similar, con el mismo número y tipo de ítems. De un universo de posibles ítems hemos escogido una muestra de ítems. Si la fiabilidad es alta, con otra muestra de ítems de la misma población de ítems obtendríamos unos resultados semejantes.**
- **Son indicadores de homogeneidad de los ítems, de que todos midan lo mismo. La homogeneidad conceptual se interpreta como descriptor del rasgo que suponemos presente en todos los ítems.**
- **La raíz cuadrada de un coeficiente de fiabilidad equivale al coeficiente de correlación entre las puntuaciones obtenidas y las puntuaciones verdaderas, a estos e le llama índice de precisión.**

VALIDEZ POR JUICIO DE EXPERTOS

La validez a través del juicio de expertos, es quizás uno de los procedimientos más frecuentemente utilizados para validar instrumentos. Este procedimiento consiste en entregar el instrumento a un grupo de profesionales que se supone tienen suficiente experticia en el tema y estos a través de una **cuidadosa revisión** determinan si el instrumento reúne los requisitos de validez para ser aplicado.

Ahora bien en muchas ocasiones, sólo se entrega una constancia de aprobación firmada por el o los expertos, sin otra evidencia que refleje los criterios que utilizó el experto para validar un instrumento. En este sentido es muy común encontrar opiniones divergentes entre otros expertos luego que el instrumento ha sido aplicado en cuyo caso es muy poco lo que el investigador puede hacer.

VALIDEZ POR JUICIO DE EXPERTOS

En este taller he incluido una forma sencilla y técnica que puede permitir realizar una validación por juicio de expertos operando con seis criterios que pueden ayudar a eliminar la ambigüedad y así evitar opiniones encontradas.

Aquí se trata de establecer 6 criterios: CLARIDAD, OBJETIVIDAD, ACTUALIDAD, ESTRUCTURA, SIFICIENCIA Y PERTINENCIA del instrumento, luego se establece la siguiente escala de puntuación:

(EXCELENTE = 10), (BUENO ENTRE 8 Y 9), (SUFICIENTE ENTRE 6 Y 7), (INSUFICIENTE ENTRE 1 Y 5).

A través de este procedimiento, cada experto coloca la calificación respectiva, por cada criterio, a cada criterio se le calcula el promedio, y la calificación final se obtiene luego de sumar los promedios entre los criterios, y luego de acuerdo con el resultado se dará el veredicto de cada experto.

VALIDEZ POR JUICIO DE EXPERTOS

CLARIDAD DEL INSTRUMENTO: Forma como está redactado o sea si el lenguaje es apropiado al nivel de las personas a las cuales va dirigido.

OBJETIVIDAD DEL INSTRUMENTO: Se refiere a si está expresado en términos de conducta observable

ACTUALIDAD DEL INSTRUMENTO: Se refiere a si el contenido esta acorde con los nuevos avances científicos y tecnológicos

ESTRUCTURA DEL INSTRUMENTO: Se refiere a si el contenido del mismo presenta un estructura lógica acorde con los objetivos de la investigación

SUFICIENCIA DEL INSTRUMENTO: Se refiere a si el contenido del mismo abarca todos los aspectos de los objetivos de la investigación en cantidad y calidad

PERTINENCIA DEL INSTRUMENTO: Se refiere a si es útil y adecuado para la investigación

VER EJEMPLO SIGUIENTE:

VALIDEZ POR JUICIO DE EXPERTOS (EJEMPLO)



Supongamos que se va a validar un instrumento de 8 preguntas a través del juicio de (1) experto y arroja los siguientes resultados

ÍTEMS	CLARIDAD	OBJETIVIDAD	ACTUALIDAD	ESTRUCTURA	SUFICIENCIA	PERTINENCIA
A	8	7	7	9	10	9
B	4	8	7	9	10	8
C	5	7	7	9	10	9
D	8	8	8	9	9	8
E	9	9	8	9	9	9
F	9	9	9	9	9	8
G	7	8	8	9	9	9
H	8	7	8	9	9	8
PROMEDIO DE CALIFICACIÓN	7,25	7,875	7,75	9	9,375	8,5
PROMEDIO ENTRE CRITERIOS	1,21	1,31	1,29	1,50	1,56	1,42
PUNTAJE FINAL	$(1,21+1,31+1,29+1,50+1,56+1,42) = 8,29$					
(EXCELENTE = 10) (BUENO ENTRE 8 Y 9) (SUFICIENTE ENTRE 6 Y 7) (INSUFICIENTE ENTRE 1 Y 5)						

Nótese que el resultado está en la categoría de bueno, indicativo de que el instrumento tiene validez de: calidad, objetividad, actualidad, estructura, suficiencia y pertinencia, por lo tanto es aplicable. Ahora bien si son varios expertos se promedia el puntaje final entre el número de expertos

MÉTODO RECOMENDADO CUANDO SE VALIDA A TRAVÉS DE JUICIO DE EXPERTOS:

Existen muchos métodos utilizados para el consenso en juicio de expertos, sin embargo considero que el más fácil de aplicar es el de consenso grupal.

MÉTODO DEL CONSENSO GRUPAL: Este método consiste en reunir a los expertos en un lugar determinado, se indica al grupo que su tarea consiste en lograr una estimación de la pertinencia y otros aspectos relacionados con la elaboración de los ítemes, que sea satisfactoria para todos los expertos. Con estas instrucciones se maximizan los intercambios de información y opiniones dentro del grupo de expertos, y se puede llegar a un consenso adecuado.

Para evitar discusiones interminables e innecesarias se recomienda que el grupo de expertos no sea muy numeroso, ya que es mas probable que se pongan de acuerdo pocos que muchos. Por lo general se recomiendan 3

ASPECTOS A CONSIDERAR EN LA ELABORACIÓN DE INSTRUMENTOS

EN LA ELABORACIÓN O CONSTRUCCIÓN DEL INSTRUMENTO:

- Se debe tener cuidado en la formulación las preguntas o ítems. Una formulación incorrecta puede dar lugar a interpretaciones diferentes por parte del entrevistado a las que el observador desea.

RESPECTO AL ORDEN DE LOS REACTIVOS (PREGUNTAS)

- Es conveniente situar los reactivos que sean más difíciles de ser contestado fielmente al final, de esta manera no se desanimará de antemano el entrevistado.
- Otra opción es repetir dos o tres reactivos que posean la misma información pero con diferente redacción. Estos reactivos de control permitirán detectar cuándo el entrevistado está contestando honestamente.

RESPECTO A LA REDACCIÓN DE LOS REACTIVOS:

- La redacción, y el vocabulario, debe estar acorde a la persona observada, tomando en cuenta su edad, nivel cultural, nivel escolar, nivel socio-económico, etcétera.
- Cada uno de los ítems debe contener una y sólo una pregunta.
- En la redacción de la pregunta no debe estar sugerida alguna de las respuestas.
- Tampoco conviene apoyarse o mencionar opiniones o sugerencias ya existentes, como son posiciones de instituciones, de personas, etcétera.

ASPECTOS A CONSIDERAR EN LA ELABORACIÓN DE INSTRUMENTOS

DECISIONES SOBRE EL CONTENIDO DE LAS PREGUNTAS:

- ¿Es necesaria la pregunta? ¿Será útil?
- ¿Se necesitan varias preguntas sobre esta cuestión?
- ¿Cuentan los informantes con los datos necesarios para contestar la pregunta?
- ¿Necesita la pregunta ser más concreta, específica e íntimamente ligada con la experiencia personal del informante?
- ¿Es el contenido de la pregunta lo suficientemente general y está libre de concreciones y especificidades falsas?
- ¿Expresan las preguntas actitudes generales y son tan específicas como suenan?
- ¿Está el contenido de la pregunta polarizado o cargado en una dirección sin preguntas acompañantes que equilibren el énfasis?
- ¿Darán los informantes la información que se les pide?

ASPECTOS A CONSIDERAR EN LA ELABORACIÓN DE INSTRUMENTOS

DECISIONES SOBRE LA FORMA DE RESPUESTA DE LA PREGUNTA:

- ¿Puede contestarse mejor la pregunta con un impreso que exija la contestación por una marca (o contestación corta de una o dos palabras, o un número), de respuesta libre o por una marca con contestación ampliatoria?
- Si se usa la contestación por una marca, ¿cuál es el mejor tipo de cuestión: dicotómica, de elección múltiple, o de escala?
- Si se usa una lista de comprobación, ¿cubre adecuadamente todas las alternativas significativas sin solaparse y en un orden definible? ¿Es de una longitud razonable? ¿Es la redacción de los ítems imparcial y equilibrada?
- ¿Es fácil, definida, uniforme y adecuada para la finalidad, la forma de respuesta?

ASPECTOS A CONSIDERAR EN LA ELABORACIÓN DE INSTRUMENTOS

DECISIONES SOBRE LA UBICACIÓN DE LA PREGUNTA EN LA SECUENCIA:

- ¿Puede verse influida por el contenido de las cuestiones precedentes la contestación a la pregunta?
- ¿La pregunta está dirigida en una forma natural? ¿Está en correcto orden lógico?
- ¿Aparece la pregunta demasiado pronto o demasiado tarde desde el punto de vista de despertar interés y recibir la atención suficiente?

NOTA: Se puede consultar el material, disponible a través de internet, que Cadoche y sus colaboradores han elaborado, y en el cual se trata más a fondo este punto de la elaboración del cuestionario y las preguntas.

FACTORES QUE AFECTAN LA CONSTRUCCIÓN DE LOS INSTRUMENTOS

ALGUNOS DE LOS FACTORES A EVITAR, PARA QUE EL INSTRUMENTO SEA VÁLIDO SON:

- **Instrucciones imprecisas o vagas**
- **Redacción con un lenguaje inadecuado para el nivel de la audiencia o personas que van a responder**
- **Preguntas que sugieren la respuesta**
- **Ambigüedad en la formulación de las preguntas, que lleven a diferentes interpretaciones**
- **Cuestionarios demasiado cortos (no incluyen una muestra adecuada de los indicadores a medir), con pocos reactivos (preguntas)**
- **Cuestionarios demasiado largos y tediosos afectan la fidelidad de las respuestas por cansancio o fastidio.**
- **Preguntas incongruentes con el universo de contenido, sin relación con los rasgos o características a medir**
- **Ordenamiento inadecuado de las preguntas.**

FACTORES QUE AFECTAN LA CONSTRUCCIÓN DE LOS INSTRUMENTOS

ADMINISTRACIÓN Y CALIFICACIÓN DEL INSTRUMENTO.

Entre estos factores están:

- **Tiempo insuficiente para responder.**
- **Ayuda adicional de otros sujetos.**
- **Las chuletas o copias**
- **La subjetividad en los puntajes de las preguntas.**

RESPUESTAS DE LOS SUJETOS.

- **Bloqueo de los sujetos al responder debido a situaciones emocionales**
- **Las respuestas al azar**

NATURALEZA DEL GRUPO Y DEL CRITERIO.

El instrumento debe ser aplicado para el grupo al cual fue diseñado, puede ser inapropiado para otros grupos.

EJEMPLO DE INSTRUMENTOS MAL ELABORADOS

¿DONDE ESTÁ EL ERROR?

1.- ¿ES USTED OPUESTO A PEDIR A SUS ALUMNOS QUE NO REALICEN LOS EJERCICIOS PRÁCTICOS LUEGO DE ESTUDIAR LA TEORÍA CORRESPONDIENTE?

2.- USTED ESTÁ SATISFECHO CON EL CAMBIO DE SU SUPERIOR INMEDIATO

SI: ____ NO: ____ TAL VEZ: ____

3.- REPRENDE USTED A SUS ALUMNOS

SIEMPRE: ____

ALGUNAS VECES: ____

NUNCA: ____

4.- ¿CON QUIÉN VIVES ACTUALMENTE?

CON MIS PADRES: ____

CON OTROS FAMILIARES: ____

SÓLO: ____

EN UNA CASA ALQUILADA: ____

5.- ¿CÓMO TE SIENTES EN LA ESCUELA?

A GUSTO: ____

NI BIEN NI MAL : ____

ME TRATAN MAL: ____

OTROS ELEMENTOS RELACIONADOS CON LA VALIDEZ

La validez es un concepto del cual pueden tenerse diferentes tipos de evidencia. (Wiersma, 1986; Gronlund, 1985):

- 1) EVIDENCIA RELACIONADA CON EL CONTENIDO,**
- 2) EVIDENCIA RELACIONADA CON EL CRITERIO**
- 3) EVIDENCIA RELACIONADA CON EL CONSTRUCTO.**

EVIDENCIA RELACIONADA CON EL CONTENIDO

La *validez de contenido* se refiere al grado en que un instrumento refleja un dominio específico de contenido de lo que se mide. Es el grado en que la medición representa al concepto medido (Bohrnstedt, 1976).

Por ejemplo: una prueba de operaciones estadísticas descriptivas no tendría validez si sólo contiene problemas relativos a las medidas de tendencia central, y no se incluyen las medidas de variabilidad, posición y forma.

EVIDENCIA RELACIONADA CON EL CRITERIO

La *validez de criterio* establece la validez de un instrumento de medición comparándola con algún criterio externo. Este criterio es un estándar con el que se juzga la validez del instrumento (Wiersma, 1986).

Mientras los resultados del instrumento de medición se relacionen más al criterio, la validez del criterio será mayor.

Por ejemplo: un cuestionario para detectar las preferencias del electorado por los distintos partidos contendientes, puede validarse aplicándolo tres o cuatro días antes de la elección y sus resultados compararlos con los resultados finales de la elección

Evidencia relacionada con el constructo.

La *validez de constructo* es probablemente la más importante sobre todo desde una perspectiva científica y se refiere al grado en que una medición se relaciona consistentemente con otras mediciones de acuerdo con hipótesis derivadas teóricamente y que conciernen a los conceptos (o constructos) que están siendo medidos. Un *constructo* es una variable medida y que tiene lugar dentro de una teoría o esquema teórico.

Por ejemplo: supongamos que un investigador desea evaluar la *validez de constructo* de una medición particular, digamos una escala de motivación intrínseca: “El Cuestionario de Reacción a Tareas”, versión mexicana (Hernández-Sampieri y Cortés, 1982).

Estos autores sostienen que el nivel de motivación intrínseca hacia una tarea está relacionado positivamente con el grado de persistencia adicional en el desarrollo de la misma (los empleados con mayor motivación intrínseca son los que suelen quedarse más tiempo adicional una vez que concluye su jornada). Consecuentemente, la predicción teórica es que a mayor motivación intrínseca, mayor persistencia adicional en la tarea.

Por ejemplo: supongamos que un investigador desea evaluar la *validez de constructo* de una medición particular, digamos una escala de motivación intrínseca: “El Cuestionario de Reacción a Tareas”, versión mexicana (Hernández-Sampieri y Cortés, 1982).

Estos autores sostienen que el nivel de motivación intrínseca hacia una tarea está relacionado positivamente con el grado de persistencia adicional en el desarrollo de la misma (los empleados con mayor motivación intrínseca son los que suelen quedarse más tiempo adicional una vez que concluye su jornada). Consecuentemente, la predicción teórica es que a mayor motivación intrínseca, mayor persistencia adicional en la tarea.

Coeficiente de Confianza	80%	90%	95%	98%	99%
Valor de Z	1,28	1,65	1,96	2,33	2,58

FIN DE LA PRESENTACIÓN

