USMP UNIVERSIDAD DE SAN MARTÍN DE PORRES

INSTITUTO PARA LA CALIDAD DE LA EDUCACIÓN

CERTIFICACIÓN ACSUG

MATERIAL DE ESTUDIO DE LA ASIGNATURA DE SEMINARIO DE INVESTIGACIÓN

SEMINARIO DE INVESTIGACIÓN

Dra. Patricia Edith Guillén Aparicio

MAESTRÍA EN EDUCACIÓN:
DOCENCIA E INVESTIGACIÓN UNIVERSITARIA
CICLO II

Semestre Académico 2018 - I

VALIDACIÓN DEL INSTRUMENTO DE RECOLECCIÓN DE DATOS

Prueba piloto

Es un ensayo experimental, cuyas conclusiones pueden resultar interesantes para avanzar con el desarrollo de algo.

Alfa de Cronbach y consistencia interna de los ítems de un instrumento de medida

Se trata de un índice de consistencia interna que toma valores entre 0 y 1 y que sirve para comprobar si el instrumento que se está evaluando recopila información defectuosa y por tanto nos llevaría a conclusiones equivocadas o si se trata de un instrumento fiable que hace mediciones estables y consistentes.

Alfa es por tanto un coeficiente de correlación al cuadrado que, a grandes rasgos, mide la homogeneidad de las preguntas promediando todas las correlaciones entre todos los ítems para ver que, efectivamente, se parecen.

Su interpretación será que, cuanto más se acerque el índice al extremo 1, mejor es la fiabilidad, considerando una fiabilidad respetable a partir de 0,80.

Su fórmula estadística es la siguiente:

$$\alpha = \frac{K}{K-1}\left[1-\frac{\sum S_i^2}{S_T^2}\right]$$

Donde:
K: El número de ítems
Si^2: Sumatoria de Varianzas de los Ítems
S_t^2: Varianza de la suma de los Ítems
α: Coeficiente de Alfa de Cronbach
Criterio para evaluar al Coeficiente Alfa de Cronbach

Como criterio general, se sugieren las recomendaciones siguientes para evaluar los resultados del Coeficientes de Alfa de Cronbach:

Coeficiente alfa de Cronbach mayor a 0,9 es Excelente

Coeficiente alfa de Cronbach mayor a 0,8 y menor a 0,9 es Bueno

Coeficiente alfa de Cronbach mayor a 0,7 y menor a 0,8 Aceptable

Coeficiente alfa de Cronbach mayor a 0,6 y menor a 0,7 Cuestionable

Coeficiente alfa de Cronbach mayor a 0,5 y menor a 0,6 Pobre

Coeficiente alfa de Cronbach menor a 0,5 es Inaceptable

Ejemplo de Cálculo del Coeficiente Alfa de Cronbach

Ítems	I	II	III	Suma de Ítems
Sujetos				
1. Campos	3	5	5	13
2. Gómez	5	4	5	14
3. Linares	4	4	5	13
4. Rodas	4	5	3	12
5. Saavedra	1	2	2	5
6. Tafur	4	3	3	10
VARP	**1,58**	**1,14**	**1,47**	**S_T^2 : 9,14**
(Varianza de la Población)		**$\sum S_i^2$:**	**4,19**	

K: El número de ítems = 3
S_i^2 : Sumatoria de Varianzas de los Ítems = 4,19
S_T^2 : Varianza de la suma de los Ítems = 9,14
α : Coeficiente de Alfa de Cronbach

$$\alpha = \frac{3}{3-1}\left[1 - \frac{4.19}{9.14}\right]$$

α : Coeficiente de Alfa de Cronbach = 0,80
Entre más cerca de 1 está α, más alto es el grado de confiabilidad

Variables y tipo de muestreo

La Teoría de Muestreo establece dos tipos de muestreo, el Muestreo Probabilístico y el Muestreo No Probabilístico, en el primero se puede determinar de antemano cuál es la probabilidad de selección de cada una de las muestras que sea posible seleccionar, mientras que en el segundo tipo sucede todo lo contrario, no se aplican criterios ni normas probabilísticas de selección.

Para efecto de este estudio vamos a hacer uso del Muestreo Probabilístico, el cuál está basado en la teoría de la aleatoriedad o del azar, el cual tiene sus principios y bases en la estadística matemática.

Dentro del muestreo probabilístico tenemos cinco tipos que son:

- Muestreo Aleatorio Simple
- Muestreo Aleatorio Estratificado
- Muestreo Sistemático
- Muestreo por Conglomerados

Para la elaboración de esta investigación haremos uso de dos métodos de muestreo, el Muestreo Aleatorio Simple y el Muestreo Aleatorio Estratificado.

2.2.1 Muestreo Aleatorio Simple

Previo a la definición del Muestreo Aleatorio Simple, es necesario recalcar que la población de esta investigación es finita; ya que conocemos el total de elementos o entes a investigar.

Una muestra aleatoria simple tomada de una población finita, es seleccionada de tal manera que cada muestra posible del mismo tamaño tiene igual probabilidad de ser seleccionada de la población. Para obtener una muestra aleatoria simple, cada elemento en la población debe tener la misma probabilidad de ser seleccionado.

El objetivo fundamental de este tipo de muestreo es tratar de eliminar la predisposición con la que los elementos de la muestra podrían ser elegidos. El método más fácil que se utiliza para lograr extraer la muestra es enumerar todos los **N** elementos, luego fijamos el tamaño **n** de la muestra y empezamos a tomar al azar los **n** números.

2.2.2 Muestreo Aleatorio Estratificado

Para obtener una muestra aleatoria estratificada, primero se divide la población de **N** individuos en **H** grupos, llamados estratos, cada uno de ellos con sus propias características, esto es son heterogéneos entre sí, pero en su interior son lo más homogéneos posible y en conjunto abarcan en su totalidad a la población, esto es:

$$N = N_1 + N_2 + N_3 + N_4 + \ldots + N_H$$

Donde $N_1, N_2, N_3, N_4, \ldots, N_H$ son los tamaños de cada uno de los estratos.

Definido cada estrato y cumpliendo la homogeneidad en el interior de ellos, se extrae una muestra aleatoria simple de cada uno de ellos, dichas extracciones deben realizarse en forma independiente en cada uno de los estratos. El tamaño de cada uno de los estratos son representados por:

$$n = n_1 + n_2 + n_3 + n_4 + \ldots + n_H$$

El tamaño de las muestras de los **H** estratos pueden ser de igual o diferente tamaño, esto varía de acuerdo al tamaño de cada estrato. Este tipo de asignación se lo conoce con el nombre de afijación proporcional.

2.2.3 El muestreo sistemático

Es un tipo de muestreo probabilístico donde se hace una selección aleatoria del primer elemento para la muestra, y luego se seleccionan los elementos posteriores utilizando intervalos fijos o sistemáticos hasta alcanzar el tamaño de la muestra deseado.

Pasos para la selección de un muestreo sistemático

1. Definir la población objetivo.
2. Determinar el tamaño deseado de la muestra (*n*).
3. Identificar el marco muestreo existente o desarrollar un marco de muestreo de la población objetivo.
4. Evaluar el marco muestral por falta de cobertura, cobertura excesiva, múltiple cobertura, agrupación, periodicidad, y hacer los ajustes cuando sea necesario. Idealmente, la lista estará en un orden aleatorio con respecto al estudio variable o, mejor aún, ordenados en función de la variable de interés o su correlación, creando así estratificación implícita.
5. Determinar el número de elementos en el marco de la muestra (*N*).
6. Calcular el intervalo de muestreo (*i*) dividiendo el número de elementos en el marco de muestreo (*N*) por el tamaño de la muestra específica (*n*). Uno debería ignorar el resto y redondear o terminar en el número entero más próximo. El redondeo hacia abajo y truncando puede hacer que el tamaño de la muestra sea más grande de lo deseado. Si es así, se puede eliminar de forma aleatoria las selecciones adicionales. Si no se conoce el tamaño exacto, o es poco práctico determinar, se puede fijar una fracción de muestreo.
7. Seleccionar al azar un número, *r*, de “1” mediante *i*.
8. Selecciona para la muestra, *r*, *r* + *i*, *r* + 2i, *r* ,+3i, y así sucesivamente, hasta agotar el marco.

A nivel técnico, el muestreo sistemático no crea una muestra verdaderamente aleatoria. Sólo la selección del primer elemento de muestreo sistemático es una selección de probabilidad. Una vez que el primer elemento es seleccionado, algunos de los elementos tendrán una probabilidad cero de selección. Además, cierta combinación de elementos, como los elementos que son adyacentes entre sí en el marco de muestreo, pueden no ser seleccionados. Muestreos sistemáticos repetidos pueden utilizarse para abordar este problema

Fortalezas y debilidades del muestreo sistemático

Al resaltar los puntos fuertes y débiles del muestreo sistemático, podemos compararlo con el muestreo aleatorio simple. El muestreo sistemático se utiliza a menudo cuando es imposible o poco práctico utilizar un muestreo aleatorio simple.

Fortalezas A diferencia de un muestreo aleatorio simple:	**Debilidades A diferencia de un muestreo aleatorio simple:**
Si el proceso de selección es manual, el muestreo sistemático es más fácil, más simple, menos tiempo, y más económico.	Si el intervalo de muestreo se relaciona con el orden periódico de los elementos en el marco de muestreo, puede resultar una mayor variabilidad.
La población objetivo no tiene porqué ser numerada y se compila un marco si hay representación física.	Elementos combinados tienen diferentes probabilidades de ser seleccionados.
Si el orden de los elementos en el muestreo se asignaron al azar, el muestreo sistemático puede producir resultados similares al muestreo aleatorio simple.	Técnicamente, sólo la selección del primer elemento es una selección de probabilidad ya que para las selecciones posteriores, habrán elementos de la población objetivo que tendrán cero probabilidad de ser seleccionados.
El muestreo sistemático elimina la posibilidad de autocorrelación.	La estimación de las variaciones es más complejo que en el muestreo aleatorio simple.

Como lleves a cabo tu muestreo será determinante para los resultados que vas a obtener. En conclusión, considera la elección de un muestreo sistematico si:

- Es difícil identificar elementos utilizando un método de muestreo aleatorio simple.
- Es importante utilizar un procedimiento de muestreo de probabilidad que pueda ser fácilmente implementado.

2.2.4 El muestreo por conglomerados:

Nos ayuda cuando es imposible o poco práctico crear un marco de muestreo de una población objetivo debido a que está muy dispersa geográficamente y el costo de la recopilación de datos es relativamente alto.

El muestreo por conglomerados es un procedimiento de muestreo probabilístico en que los elementos de la población son seleccionados al azar en forma natural por agrupaciones (clusters). Los elementos del muestreo se seleccionan de la población de manera individual, uno a la vez.

Las unidades de muestreo o grupos pueden ser espaciados, tal como ocurre naturalmente en las unidades geográficas o físicas (por ejemplo: estados, delegaciones o distritos); en base a una organización como escuelas, grado escolar; o servicio telefónico tales como códigos de área o el cambio de las claves lada de los números de teléfono. La heterogeneidad del grupo es fundamental para un buen diseño del muestreo por conglomerados. Por otra parte, los elementos dentro de cada grupo debe ser tan heterogéneos como la población objetivo

Pasos para seleccionar un muestreo por conglomerados

1. Definir la población objetivo.
2. Determinar el tamaño de la muestra deseada.

3. Identificar un marco de muestreo existente o desarrollar un nuevo marco de muestreo de grupos de la población objetivo.
4. Evaluar el marco de muestreo para la falta de cobertura, cobertura excesiva, múltiple cobertura, y la agrupación, y hacer los ajustes cuando sea necesario. Idealmente, los grupos serían tan heterogéneos como la población, mutuamente excluyentes, y colectivamente exhaustivos. La duplicación de elementos de la muestra puede aparecer si elementos de la población pertenecen a más de un grupo. La omisión dará lugar a un sesgo de cobertura.
5. Determinar el número de grupos que se seleccione. Esto se puede hacer dividiendo el tamaño de la muestra por el número promedio estimado de elementos de la población en cada grupo. En la medida en que la homogeneidad y la heterogeneidad de los grupos sean diferentes a la de la población, el número del grupo aumenta e incrementa la precisión. Por otra parte, si las diferencias aumentan, la precisión disminuye.
6. Seleccionar al azar el número previsto de las agrupaciones.

Fortalezas y debilidades del muestreo por conglomerados en comparación con el muestreo aleatorio simple

Fortalezas en comparación con el muestreo aleatorio simple:	Debilidades en comparación con el muestreo aleatorio simple:
Si los grupos están definidos geográficamente, el muestreo por conglomerados requiere menos tiempo, dinero y mano de obra.	Una muestra por conglomerados puede no ser la más representativa de la población como es el caso de una muestra aleatoria simple del mismo tamaño de la muestra.
El muestreo por conglomerados permite muestreos posteriores debido a que los grupos incluidos en la muestra son elementos agregados.	Las variaciones en las muestras tienden a ser mucho más altas que en las del muestreo aleatorio simple.

Uno puede estimar características de los grupos como el de la población.	El muestreo por conglomerados hace aún más complejo el análisis de datos e interpretación de los resultados.
El muestreo por conglomerados no requiere de un marco de muestreo de todos los elementos de la población objetivo.	El muestreo por conglomerados produce errores de muestreo más grandes para muestras de tamaño comparable que otras muestras de probabilidad.

Tamaño de la muestra

2.6. Determinación del Tamaño de la Muestra

El tamaño de la muestra a través de proporciones se determinó con los siguientes datos y parámetros:

1. Error del diseño = 0.04
2. Nivel de Confianza (1-α)100% con α =0.05, se obtiene entonces un 95% de confianza.
3. $z_{\alpha/2}$=1.96, dicho valor es obtenido de la Tabla de distribución Normal.
4. Tamaño de la población, **N**=154,281

$\hat{p} = 0.19$

$\hat{q} = 0.81$ 5.

$\hat{\sigma} = \hat{p} * \hat{q} = 0.19 * 0.81 = 0.1539$ 6.

7. El tamaño de la muestra para el muestreo aleatorio simple se calcula con la siguiente ecuación:

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

El valor de n_0 proviene de:

$$n_0 = \frac{Z^2{}_{\alpha/2}}{e^2}\hat{p}\hat{q}$$

Al reemplazar los valores en ambas fórmulas tenemos:

$$n_0 = \frac{1.96^2}{0.04^2}(0.19 * 0.81) = 369.514$$

$$n = \frac{369.514}{1 + \frac{369.514}{154281}} = 368.63 \approx 369$$

Lo cual nos indica que el tamaño de la muestra para este estudio es de 369 profesionales de tercer nivel.

MEDIDAS DE TENDENCIA CENTRAL

Ejemplo: De la producción diaria de una máquina se eligió una muestra de 100 baterías que se probaron para ver cuanto tiempo operarían en una lámpara medida en horas y los resultados fueron los siguientes.

28	14	30	47	33	21	17	22	31	20
36	16	13	22	34	27	11	17	43	41
31	39	48	40	41	11	20	23	27	29
12	12	32	43	35	31	40	31	17	29
17	14	19	23	46	40	27	28	31	35
20	17	39	42	41	50	30	17	46	31
11	33	36	37	19	17	22	36	47	17
49	16	37	43	42	41	22	17	19	20
35	17	25	36	39	30	40	36	36	38
23	23	13	16	46	40	22	23	21	39

Para hallar la tabla de distribución de frecuencias tenemos que encontrar el número de intervalos y el tamaño del intervalo.

- El número de intervalos es igual a la raíz cuadrada de todos los datos. Frecuencias de datos agrupados

$$\sqrt{100} = 10$$

Recuerda que el resultado es un número no exacto se debe de aproximar.

- Hallamos el tamaño del intervalo Frecuencias de datos agrupados

Tamaño del intervalo = 50 – 11
10

Tamaño del Intervalo = 39 = 3,9 Como el resultado es un número decimal 10 debemos de aproximar.

Por lo tanto el tamaño del intervalo es de 4.

- A partir del dato menor comenzamos a sumar el tamaño del intervalo

Intervalo (Horas)
11-15
16-20
21-25
26-30
31-35
36-40
41-45
46-50
51-55
56-60
TOTAL

- Hallamos la frecuencia absoluta (ni) contando cuantas veces se repiten los números para cada intervalo.
-

Intervalo (Horas)	***Frecuencia Absoluta (ni)***
11-15	9
16-20	20
21-25	13
26-30	10
31-35	13
36-40	18
41-45	9
46-50	8
51-55	0
56-60	0
TOTAL	100

- Hallamos la frecuencia relativa hi, dividiendo cada una de las frecuencias absolutas entre el total de datos.

Intervalo (Horas)	***Frecuencia Absoluta (ni)***	***Frecuencia Relativa (hi)***
11-15	9	0,09
16-20	20	0,20
21-25	13	0,13
26-30	10	0,10
31-35	13	0,13
36-40	18	0,18
41-45	9	0,09
46-50	8	0,08
51-55	0	0,0
56-60	0	0,0
TOTAL	100	1

- Hallamos el porcentaje multiplicando la frecuencia relativa por 100

Intervalo (Horas)	*Frecuencia Absoluta (ni)*	*Frecuencia Relativa (hi)*	*Porcentaje %*
11-15	9	0,09	9
16-20	20	0,20	20
21-25	13	0,13	13
26-30	10	0,10	10
31-35	13	0,13	13
36-40	18	0,18	18
41-45	9	0,09	9
46-50	8	0,08	8
51-55	0	0,0	0
56-60	0	0,0	0
TOTAL	100	1	100%

- Hallamos la marca de clase Mi, sacando el promedio de cada uno de los intervalos, es decir, sumando los dos intervalos y luego dividiéndolos entre dos.

Intervalo (Horas)	*Frecuencia Absoluta (ni)*	*Frecuencia Relativa (hi)*	*Porcentaje %*	*Marca de Clase (Mi)*
11-15	9	0,09	9	13
16-20	20	0,20	20	18
21-25	13	0,13	13	23
26-30	10	0,10	10	28
31-35	13	0,13	13	33
36-40	18	0,18	18	38
41-45	9	0,09	9	43
46-50	8	0,08	8	48
51-55	0	0,0	0	53
56-60	0	0,0	0	58
TOTAL	100	1	100%	

- Multiplicamos la frecuencia absoluta por la marca de clase (ni*Mi), es te dato es necesario para poder luego hallar las medidas de tendencia central.

Intervalo (Horas)	***Frecuencia Absoluta (ni)***	***Frecuencia Relativa (hi)***	***Porcentaje %***	***Marca de Clase (Mi)***	***Frecuencia Absoluta por Marca de clase (ni*Mi)***
11-15	9	0,09	9	13	117
16-20	20	0,20	20	18	360
21-25	13	0,13	13	23	299
26-30	10	0,10	10	28	280
31-35	13	0,13	13	33	429
36-40	18	0,18	18	38	684
41-45	9	0,09	9	43	387
46-50	8	0,08	8	48	384
51-55	0	0,0	0	53	0
56-60	0	0,0	0	58	0
TOTAL	100	1	100%		2940

- Por último para terminar nuestra tabla hallamos la frecuencia absoluta acumulada que es el primer dato e irle sumando sucesivamente cada dato de la frecuencia.

Intervalo (Horas)	***Frecuencia Absoluta (ni)***	***Frecuencia Relativa (hi)***	***Porcentaje %***	***Marca de Clase (Mi)***	***Frecuencia Absoluta por Marca de clase (ni*Mi)***	***Frecuencia Absoluta Acumulada (Ni)***
11-15	9	0,09	9	13	117	9
16-20	20	0,20	20	18	360	29
21-25	13	0,13	13	23	299	42
26-30	10	0,10	10	28	280	52
31-35	13	0,13	13	33	429	65
36-40	18	0,18	18	38	684	83
41-45	9	0,09	9	43	387	92
46-50	8	0,08	8	48	384	100
51-55	0	0,0	0	53	0	-
56-60	0	0,0	0	58	0	-
TOTAL	100	1	100%		2940	

- Para hallar las medidas de tendencia central vamos a realiza los siguientes procedimientos:

MODA: La moda es la clase o dato que más se repite, es decir 20 que corresponde al intervalo 16-20 Mo = 20

MEDIA ARITMÉTICA: La media aritmética es el promedio, es decir la sumatoria del ni*Mi dividido entre el total de datos.

$X = \frac{2940}{100} \qquad X = 29{,}40$

MEDIANA O VALOR CENTRAL: Hallamos la media teórica, es decir el número total de datos dividido entre dos (2), y lo buscamos entre la frecuencia absoluta acumulada, entonces el total de datos es 100 y lo dividimos entre 2, 100/2 = 50, este dato se encuentra aproximadamente en el cuarto intervalo y ubicamos el límite inferior de la clase o sea 23, luego ubicamos en la frecuencia absoluta acumulada el número anterior al dato donde se encuentra la media teórica, ubicamos la frecuencia absoluta que corresponde al intervalo y por último multiplicamos por el tamaño del intervalo.

$$Me = Li + \frac{N/2 - hi - 1}{100} * ai$$

Li: Es el límite inferior de la clase (intervalo) donde se encuentra la media teórica.

N / 2: Es la media teórica, es decir el número total de datos dividido entre dos.

Fi - 1: Es el dato anterior en la frecuencia absoluta acumulada donde encontramos la media teórica.

fi: Es la frecuencia absoluta que corresponde al dato donde se encuentra la media teórica.

ai: Es la amplitud o tamaño del intervalo.

Me = 27 + <u>100/2 – 42</u> *
4 6

$$Me = 27 + \frac{50 - 42}{6} * 4$$

$$Me = 27 + \frac{8}{6} * 4$$

Me = 27 + 1,33 * 4

Me = 27 + 5,33

Me = 28,33

Los resultados son entonces los siguientes:

Mo = 20

X = 29,40

Me = 28,33

MEDIDAS DE DISPERSIÓN

Rango de variación

Se define como la diferencia entre el mayor valor de la variable y el menor valor de la variable.

$$\boxed{\text{Rango de variación} = \text{Máximo} - \text{Mínimo}}$$

La mejor medida de dispersión, y la más generalizada es la varianza, o su raíz cuadrada, la desviación estándar. La varianza se representa con el símbolo σ^2 (sigma cuadrado) para el universo o población y con el símbolo s2 (s cuadrado), cuando se trata de la muestra. La desviación estándar, que es la raíz cuadrada de la varianza, se representa por σ (sigma) cuando pertenece al universo o población y por “s”, cuando pertenece a la muestra. σ^2 y σ son parámetros, constantes para una población particular; s2 y s son estadígrafos, valores que cambian de muestra en muestra dentro de una misma población. La varianza se expresa en unidades de variable al cuadrado y la desviación estándar simplemente en unidades de variable.

Fórmulas

Donde μ es el promedio de la población.

$$\sigma^2 = \frac{(Y_1-\mu)^2+(Y_2-\mu)^2-\ldots\ldots+(Y_N-\mu)^2}{N}$$

$$\sigma^2 = \frac{\sum_i (Y_i-\mu)^2}{N}$$

Donde Ȳ es el promedio de la muestra.

$$s^2 = \frac{(Y_1-\bar{Y})^2+(Y_2-\bar{Y})^2+\ldots\ldots+(Y_N-\bar{Y})^2}{n-1}$$

$$s^2 = \frac{\sum_i (Y_i-\bar{Y})^2}{n-1}$$

Consideremos a modo de ejemplo una muestra de 4 observaciones

Según la fórmula el promedio calculado es 7, veamos ahora el cálculo de las medidas de dispersión:

Y_i	$(Y_i-\bar{Y})$	$(Y_i-\bar{Y})^2$
3	-4	16
6	-1	1
8	+1	1
11	+4	16
		34

TÉCNICAS E INSTRUMENTOS

TÉCNICA	CONCEPTUALIZACIÓN	PROPIEDADES	INSTRUMENTOS
Observación	Es la inspección analítica y rigurosa de las variables de la investigación, proceso que nos permitirá comprender las características, atributos y propiedades que tiene, posibilitándonos ver su desenvolvimiento en el medio ambiente que se encuentra. La observación debe ser lo más verídico posible, tiene que ser fiel reflejo de la realidad existente.	• Debe ser contemplado en función tanto del problema, los objetivos y las variables de la investigación. El objeto observado se ha de determinar mediante la pregunta ¿qué y por qué se va a observar?, ¿Qué importancia genera su observación? • Debe ser proyectado de forma tal que me brinde información útil, veraz y confiable, para ello se ha de preguntarse ¿para qué se va a realizar la observación?, ¿cuál es su utilidad? • Tiene que ser imparcial y neutral, es decir no poseer ningún criterio que distorsione o modifique los resultados de la observación. • La información que brinda la observación, ha de ser registrada en instrumentos que posibiliten su adecuado tratamiento y análisis para su correcta interpretación y validación.	• Guía de observación de campo. • Guía de observación experimental. • Lista de cotejo. • Registros de anécdotas.
Entrevista	Es el diálogo que se establece entre el investigador (entrevistador) y la persona que posee información valiosa para la investigación (entrevistado), Genéricamente, las entrevistas para las tesis son de profundidad, es decir el entrevistado tiene que tener un alto conocimiento acerca de la investigación. Pero también se puede realizar entrevistas aleatorias, en donde se busca saber la opinión de los entrevistados, es decir que tanto conoce sobre el tema, sin importar mucho el nivel de cultural de ellos.	• La entrevista debe ser previamente estructurada, con preguntas que posibiliten ampliar los saberes del entrevistador para que lo aplique en la tesis. • Para ello el entrevistador ha de formular las preguntas en función al problema de la investigación, de modo tal que el entrevistado, al responder las preguntas, pueda ilustrar, ampliar y profundizar los conocimientos del entrevistador. • El entrevistador debe tener curiosidad, y saber abordar la entrevista en relación a lo que necesita investigar o a las características del entrevistado. • Muchas veces las entrevistas deben de contener preguntas libres o que se formulan según el rumbo que toma la conversación, ya que en todo diálogo siempre surgen considerandos nuevos o imprevistos, que pueden alterar en el acto la estructura de la investigación. Las preguntas libres pueden memar el impacto de las nuevas orientaciones que surjan, sin perjudicar la estructura interna que ha de tener la entrevista.	• Cuestionario de la entrevista estructurada. • Cuestionario de la entrevista semi estructurada. • Cuestionario de la entrevista no estructurada.

Encuesta	Es aplicada a todos los elementos que conforman el espacio muestral, para saber la opinión, tendencias y actitudes que permitan obtener información relevante a la investigación. Las encuestas deben ser estructuradas a fin de poder determinar la veracidad o falsedad de la hipótesis de la investigación. La información obtenida ha de ser contrastada y analizada para poder formular conclusiones.	• Las encuestas pueden ser de varios tipos, como: a) descriptivas, que buscan describir las condiciones reales en que se hallan los elementos de la muestra; b) analíticas, buscan describir, analizar y explicar las causas que posibilitan el problema de la investigación; c) de respuesta abierta, las que posibilitan al encuestado responder libremente a las preguntas, esto permite obtener datos de mayor diversidad y grado de profundidad; d) de respuesta cerrada, que solamente tienen opciones predeterminadas para seleccionar, permiten obtener mayor grado de cohesión y precisión en las respuestas; e) encuesta telefónica, se dan genéricamente cuando hay mucha distancia entre el encuestador y el encuestado, las preguntas deben ser pocas e ir directamente a lo que se necesita obtener; d) encuestas webs, puestos en moda gracias a las redes sociales como Facebook, Twiter e Instagram, en donde se difunden las encuestas, genéricamente de muy pocas preguntas, ya sea una o dos, con un máximo de diez interrogantes, las cuales son respondidas por grandes cantidades de personas en muy poco tiempo, estas encuestas webs, también pueden ser aplicadas a los medios tanto electrónicos como laptops u ordenadores fijos, como a los teléfonos inteligentes.	• Cuestionario de la encuesta.
Experimento	Es la manipulación de las variables independientes o causales, con el propósito de observar analíticamente los cambios que operan cuando se les somete a modificaciones o alteraciones, y son comparadas frente a otro grupo de control, el cual puede ser de su propio ambiente (pre experimental) o en otro (cuasiexperimental o experimental propiamente dicho), el cual no ha sido alterado.	• Ha de establecerse previamente a la experimentación un fundamento teórico, para poder establecer los presupuestos mínimos que se ha de necesitar para realizar el experimento. • Todo experimento ha de ser realizado sistemáticamente, es decir con orden y un procedimiento que posibilite la observación detallada de cada operación realizada dentro de la experimentación. • La valoración del experimento está en concordancia al grado de efectividad que tenga para poder afirmar o negar la hipótesis de la investigación. • En el proceso de experimentación, el investigador debe tener una actitud reflexiva, analítica y comparativa, para poder establecer conclusiones, las cuales deben de ceñirse estrictamente a los resultados objetivos de la experimentación.	• Programa de experimentación. • Informe de la experimentación.

CORRELACIÓN DE PEARSON-SPEARMAN

Correlación de Pearson

Tengamos las siguientes puntuaciones en las variables V1 (inteligencia) e V2 (rendimiento académico): Calcular el coeficiente de correlación de Pearson:

V1:X:	105	116	103	124	137	126	112	129	118	105
V2: Y:	4	8	2	7	9	9	3	10	7	6

Solución

Configuremos la siguiente tabla:

X	Y	X^2	Y^2	XY
105	4	11025	16	420
116	8	13456	64	928
103	2	10609	4	206
124	7	15376	49	868
137	9	18769	81	1233
126	9	15876	81	1134
112	3	12544	9	336
129	10	16641	100	1290
118	7	13924	49	826
105	6	11025	36	630
1175	65	139245	489	7871

De donde:

$$\overline{X} = \frac{\sum X}{N} = \frac{1175}{10} = 117.5$$

$$\overline{Y} = \frac{\sum Y}{N} = \frac{65}{10} = 6.5$$

$$S_x = \sqrt{\frac{\sum X^2}{N} - \overline{X}^2} = \sqrt{\frac{139245}{10} - 117.5^2} = 10.874$$

$$S_x = \sqrt{\frac{\sum Y^2}{N} - \overline{Y}^2} = \sqrt{\frac{489}{10} - 6.5^2} = 2.579$$

Aplicando (1.9):

$$r_{xy} = \frac{\frac{\sum XY}{N} - \overline{X}\overline{Y}}{S_x S_y} = \frac{\frac{7871}{10} - 117.5 * 6.5}{10.874 * 2.579} = 0.8327$$

Recuerda tener en cuenta lo siguiente:

Valores de r	Tipo y grado de correlación
-1	Negativa perfecta
-1 < r ≤ -0,8	Negativa fuerte
-0,8 < r < -0,5	Negativa moderada
-0,5 ≤ r < 0	Negativa débil
0	No existe
0 < r ≤ 0,5	positiva débil
0,5 < r < 0,8	positiva moderada
0,8 ≤ r < 1	positiva fuerte
1	Positiva perfecta

Spearman – Brown

$$R = \frac{2r}{1+r} \qquad r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$

PRUEBA DE INDEPENDENCIA

En el análisis estadístico es frecuente considerar un análisis bivariado y si queremos conocer si dos variables cualitativas son independientes o dependientes utilizaremos la prueba Chi cuadrado.

Tabla de Contingencia.

También llamado cuadro de doble entrada, nos permite hacer un análisis conjunto de dos variables para conocer la ocurrencia o incidencia de las variables a la vez.

Triola (2009) afirma "Una tabla de contingencia o una tabla de frecuencia de dos factores es una tabla en que las frecuencias corresponden a dos variables. Una variable se utiliza para categorizar renglones, y una segunda variable se utiliza para categorizar columnas". (p.606).

Prueba de hipótesis de Independencia

Córdova (2008) afirma: "La prueba de hipótesis de independencia implican dos variables categóricas y lo que se prueba es la suposición de que las dos variables son estadísticamente independientes". (p.344)

Hipótesis

La hipótesis nula H_0 establece que las variables cualitativas son independientes y la hipótesis alterna establece lo contrario es decir que las variables no son independientes.

Formula:

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

Nota1:

Cuando la muestra es pequeña, digamos menos de 50, o cuando algunas o todas las frecuencias de las celdas son menores que 5, o cuando el grado de libertad es igual a 1, se aplicara la corrección de yates se efectúa por:

$$X^2{}_{cal} = \frac{\left(\left|O_{ij} - e_{ij}\right| - 0.5\right)^2}{e_{ij}}$$

Nota 2:

El coeficiente de contingencia es una medida de correlación entre dos variables cuyos valores se registran en una tabla e contingencia.

El índice o coeficiente de contingencia se define por:

$$C = \sqrt{\frac{X^2{}_{cal}}{n + X^2{}_{cal}}}$$

Donde, n es el tamaño de muestra.

Si $Si: \ X^2{}_{cal} \ es\ significativo, también\ lo\ es\ el\ coeficiente\ de\ contingencia$

Ejemplo

Cuatrocientos trabajadores fueron seleccionados y clasificados según el nivel de estrés en el que se encuentran y la productividad como se indica en la tabla que sigue. Al nivel de significancia del 5% ¿se puede inferir que el nivel de estrés y la productividad son independientes? Y ¿calcular el nivel de correlación?

	Desempeño laboral		
Nivel de estrés	Alto	Medio	Bajo
Alto	15	60	70
Medio	10	40	30
Bajo	60	30	10

Tabla 1

Frecuencias observadas del Nivel de estrés y desempeño laboral

	Desempeño laboral			
Nivel de estrés	Bueno	Regular	Malo	Total
Alto	15	60	70	145
Medio	10	40	30	80
Bajo	60	30	10	100
Total	85	130	110	325

Tabla 2

Frecuencias esperadas del Nivel de estrés y desempeño laboral

Desempeño laboral

Nivel de estrés	Bueno	Regular	Malo
Alto	37.9	58	49.1
Medio	20.9	32	27.1
Bajo	26.2	40	33.8

Solución:

1. Plantear hipótesis

   H_0: El nivel de estrés y el desempeño laboral son independientes.

   H_0: El nivel de estrés y el desempeño laboral son dependientes.

2. Especificar el nivel de significancia:

$$\alpha = 0.05$$

3. Seleccionar el estadístico de prueba:

$$X^2 = \sum_{i=1}^{I} \sum_{j01}^{J} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

4. Calcular el estadístico de prueba

$$X^2 = \frac{(15-37.9)^2}{37.9} + \frac{(60-58)^2}{58} + \frac{(70-49.1)^2}{49.1} + \frac{(10-20.9)^2}{20.9} + \frac{(40-32)^2}{32} + \frac{(30-27.1)^2}{27.1} + \frac{(60-26.2)^2}{26.2} + \frac{(30-40)^2}{40} + \frac{(10-33.8)^2}{33.8}$$

$$X^2 = 93.67$$

5. Regla de decisión

Si el estadístico calculado X^2 calculado es mayor al $X2_{Tabla}$ ~~Entonces~~ se rechaza la H_0

$$X^2{}_{calculado} = 93.67$$

$$X^2{}_{tabla} = 9.49$$

En este caso como el ${X^2}_{calculado}\ es\ mayor\ que\ {X^2}_{tabla}$ se rechaza la hipótesis nula.

6. Conclusión

Con un nivel de significancia de 0.05 se rechaza la H_0 y se concluye que el nivel de estrés y el desempeño laboral son dependientes.

Coeficiente de correlación

Para calcular el nivel de correlación utilizaremos el coeficiente de contingencia

$$\boxed{C = \sqrt{\frac{{X^2}_{cal}}{n + {X^2}_{cal}}}}$$

$$C = \sqrt{\frac{93.66}{325 + 93.66}}$$

$$C = 0.47$$