

PARTE DOS DEL TEMA DOS

2.3 ANÁLISIS DE DISTRACTORES

Los distractores son las respuestas erróneas que tiene una pregunta tipo test para intentar engañar a la gente. Para que un ítem sea bueno todas las alternativas tendrían que tener la misma probabilidad de ser elegida, si el alumno no conoce la respuesta segura. Para comprobarlo se utiliza la prueba chi cuadrado o ji cuadrado , que este año veréis las múltiples utilidades que tiene a parte de cómo prueba de bondad de ajuste.

El chi cuadrado es esta fórmula:

$$\chi^2 = \sum \frac{(f_o - f_c)^2}{f_e}$$

Donde se habla de comparar la frecuencia observada , que es la frecuencia absoluta que realmente tenemos en una opción , o sea el número de personas que han escogido esa respuesta en la pregunta, y compararlo con la fe que es la frecuencia esperada , lo que debería de haber sido, el numero de personas que deberían haber escogido esa opción si todos los distractores lo respondiesen el mismo número de gente.

Y como esta es una prueba de contraste de hipótesis que os debería sonar del año pasado de estadística , son pruebas que servían para saber si se cumple o no una Ho, aquí claramente no hace esa función pero sin embargo si sigue su formato , o sea el resultado de la formula , que será el valor empírico se debe comparar con el valor ideal que es el teórico. Si el ideal es mayor es que los distractores son buenos, si es el empírico el mayor es que los distractores son malos.

Un ejemplo del libro:

Se supone que de 300 individuos contestan a un ítem con cuatro alternativas y que la D es la respuesta correcta. Los valores obtenidos para los 198 individuos que fallan en las distintas alternativas erróneas son los siguientes:

A	B	C
85	40	73

$$\chi^2 = \frac{(85 - 66)^2 + (40 - 66)^2 + (73 - 66)^2}{66} = \frac{1086}{66} = 16,45$$

Para saber si estos distractores son adecuados hacemos este chi cuadrado y comprobamos al compararlo con el resultado de la tabla chi cuadrado si si o no son independientes y las respuestas son iguales según las alternativas , vamos que los distractores funcionan.

En este caso hacemos la prueba chi utiliza como frecuencia esperada la media de gente así debería sumar la gente que escogió cada opción $85+40+73=198$ y dividirlo entre tres que es el número de distractores que hay y da 66 que es lo que entenderemos aquí como frecuencia esperada.

Ahora para saber si es bueno o no debería comparar el resultado 16.45 con el valor teórico que viene en la tabla del chi cuadrado de las tablas donde deberé mirar el nivel de confianza que tiene el investigador o el 95 o el 99 y cortar esa columna en la fila del $k-1$ número de alternativas de respuesta incorrectas menos 1 , así saldría algo como esto:

Dis

Nivel de confianza(NC) = $1 - \alpha = 1 - \text{nivel de significac}$

NC					90%	95%
α	0,995	0,975	0,900	0,500	0,100	0,050
n						
1	0,000	0,000	0,016	0,455	2,706	3,841
2	0,010	0,051	0,211	1,386	4,605	5,991

Y mi valor teórico critico o ideal seria 5.991 mientras el empírico es 16.45

$16.45 > 5.991$ por lo tanto se rechaza la hipótesis de que los distractores son buenos.

2.4 COEFICIENTE DE FIABILIDAD

Siempre tengo que tener en cuenta a la hora de valorar un instrumento no solo como se evaluan las notas sino también hay que tener en cuenta los posibles errores de medición, Esta relación entre la evaluación y los errores de medición es lo que conocemos como la fiabilidad de un ítem. Esta fiabilidad como toda relación se representa por medio de una función lineal.

Donde hay tres supuestos básicos, ideas que si o si se deben cumplir . La primera que las notas que el profesor asigna a cada examen serán igual a las puntuaciones verdaderas. En segundo lugar las puntuaciones verdaderas serán igual a las notas y por tanto los errores de medición es inexistente. Y que la correlación entre los errores del ítem y del examen es también inexistente. Porque el ítem tiene que ser igual que el examen o al menos parecido , igual de buenos ambos. Partiendo de esta idea observaremos el siguiente modelo ideal:

$$X = V + e$$

Así la nota dependerá del verdadero conocimiento de la persona que se denomina puntuación verdadera y el error cometido en la medición.

Cuando no existe el error de medida, el instrumento de medida es fiable. Que un instrumento sea fiable nos indica que el instrumento es merecedor de confianza, libre de errores de medida. Y es básico para hacer una medición, aunque no la única. Y se calcula mediante el coeficiente de fiabilidad que es la fiabilidad relativa y la fiabilidad absoluta que es el error típico de medida. El coeficiente de fiabilidad es la forma de conocer la fiabilidad relativa, que es la correlación entre las puntuaciones obtenidas por la persona y otras puntuaciones, hay muchas formas diversas de calcularlo.

La fiabilidad es un indicador de la probabilidad de fallo de nuestro instrumento y cuanto se parece o se diferencia de la nota que de verdad se merecería. Aunque un instrumento de medida no es bueno solo por ser fiable, ser merecedor de confianza y no cometer errores de medida, puede por ejemplo que lo que mide no interese, por eso podemos decir que la fiabilidad es una condición necesaria aunque no suficiente. Aunque hay dos tipos de fiabilidad la fiabilidad relativa que se mediría por el coeficiente de fiabilidad (muestra la fiabilidad relativa a un grupo concreto, y la fiabilidad absoluta que se mide por el error típico. Esta relación entre la fiabilidad, la nota que se merece un individuo y la nota que de verdad obtiene se expresa con la siguiente fórmula

$$\rho_{XX'} = \frac{\sigma_V^2}{\sigma_X^2}$$

La fiabilidad es un indicador de la probabilidad de fallo y mira el grado de discrepancia con el verdadero valor que se hace comparando la nota de un instrumento con otro instrumento que se considere óptimo y que trabaje lo mismo claro está, se le denominaría formas paralelas. Aunque hay que tener en cuenta que las diferencias existentes entre lo que se busca medir antes y después debe ser mínima para medir ese error de estimación.

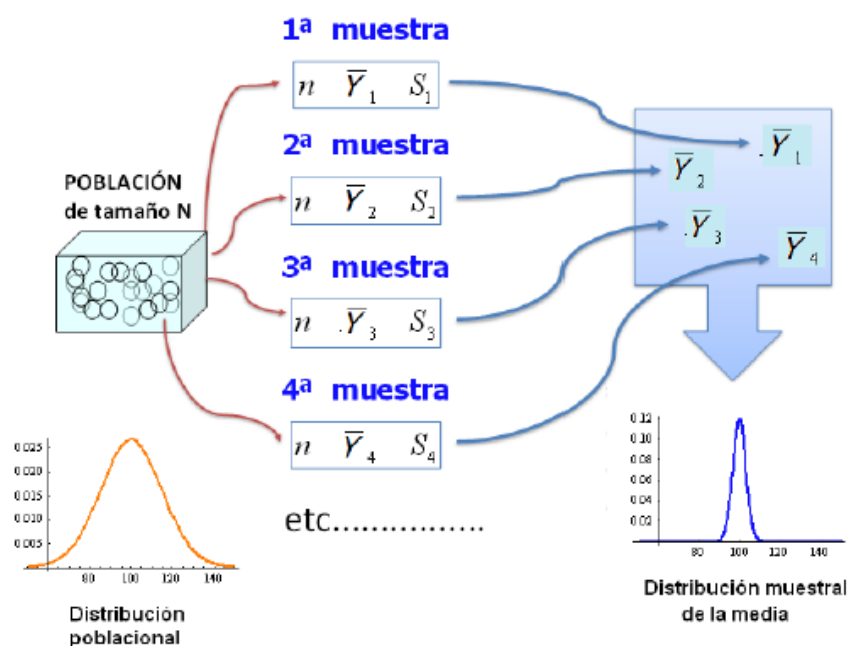
$$\rho_{XX'} = 1 - \frac{\sigma_e^2}{\sigma_X^2}$$

Estas son las formulas que relacionan lo dicho anteriormente relacionan por un lado la puntuación verdadera de cada individuo unida en el calculo de la varianza al hacer la varianza con todas las puntuaciones verdaderas lo que deberían de sacar veremos lo unido que está. Y lo relacionamos con las puntuaciones empíricas lo que realmente pasa. Y relacionándola debería ser igual o muy parecida, que es lo explicado por la fórmula Todo esto mide la fiabilidad como probabilidad de fallo pero si queremos medir la fiabilidad como precisión, tendremos que hablar de la fiabilidad absoluta que es el error típico de medida.

Si a un individuo le hago el examen una y otra vez con las notas que saque que serán muy parecidas siempre la desviación típica que calcularía con estos datos sería el error típico , porque el individuo siempre sabe lo mismo lo que se desvía de lo que debería de saber es exclusivamente un error de medición.

El error típico me servirá así para:

- Estimar el limite entre que nota y que nota se encontrará la puntuación verdadera de un individuo
- Para comparar las puntuaciones de varios sujetos en diferentes instrumentos o en el mismo
- Las medidas o notas que saca la gente se ha de distribuir también de manera normal como si fuese un estadístico (cálculo para la muestra) más, si yo extraigo todas las muestras posibles y de todas ellas saco sus notas. Si pongo todas las notas que hemos sacado de todas las muestras juntas. Entonces podréis ver que todas estas notas juntas se distribuirán de manera normal, o sea que si tomamos estas medidas como si fuesen los números obtenidos de una muestra, y por ejemplo con ellos dibujo lo que sería una gráfica , la de la curva normal , veré que queda exactamente como quedaría la campana de gauss perfecta. Porque a fin de cuentas lo que obtendría es la población y eso ya sabemos que se distribuye siempre de manera normal. A esta distribución de las medidas de todas las muestras posibles las conocemos como distribución muestral de la medida. Y sin lugar a dudas donde mejor se ve es en este dibujo colgado por los profesores.



ESTIMACIONES EMPÍRICAS DEL COEFICIENTE DE FIABILIDAD.

- FIABILIDAD COMO EQUIVALENCIA - FORMAS PARALELAS.

Un instrumento es equivalente a otro cuando miden lo mismo . Los errores de medida que se puedan producir son debidos a falta de equivalencia entre los elementos.

Para medir la fiabilidad teniendo en cuenta esta característica se hará con el coeficiente de equivalencia. Se cogen dos muestras de una misma población , y dos exámenes que sean equivalentes porqu sean similares.

Deben tener el mismo número de ítems

Tienen que estar escritos igual, mismo vocabulario, mismo idioma...

Que sean preguntas del mismo estilo

Trabajen el mismo contenido y objetivo

El índice de dificultad debe ser igual.

No debe haber diferencias significativas entre medias, se cumpla la Ho en el contraste entre grupos

El formato debe ser igual y la presentación también .

Ahora a estasn dos muestras les aplicaré los dos exámenes y comprobare por el coeficiente de equivalencia o el coeficiente de correlación entre las puntuaciones de ambas pruebas

Si son pruebas criterioles, que se compara a un individuo con un modelo de referencia, el coeficiente de fiabilidad mas usado es el coeficiente k kappa

$$K = \frac{F_c - F_a}{N - F_a}$$

Donde Fc es la frecuencia de las coincidencias y Fa es la frecuencias de azar para elementos que coinciden y N nuemro de casos. El resultado que dará será similar a la correlación pearson para los datos dicotómicos . Y como máximo daría 1 y como mínimo 0

EXAMEN DE AÑOS ANTERIORES QUE HAN SALIDO:

EXAMEN 2011-2012 SEPTIEMBRE ORIGINAL

Un profesor está realizando una prueba de rendimiento en lengua. Pasa dicha prueba a un colectivo de 200 estudiantes de un colegio. Desea saber las características métricas de la misma. Se pide:

- b) Calcular el índice de dificultad corregido de un ítem de cuatro alternativas que acierta 80 estudiantes suponiendo que lo han respondido todo el colectivo del colegio.
- c) Del ítem anterior se quiere saber el análisis de distractores del mismo y se sabe que la respuesta correcta es la b y que el reparto de las otras categorías es el siguiente:

a	c	d
60	20	40

Lo primero para hacer el apartado c que es el que no hemos hecho hasta ahora tenemos que entender el número de personas que fallaron antes de nada, en total eso se puede hacer sumando la gente que contestó a , que contesto c y d o al total restarle los aciertos.

$$60+20+40 = 120$$

$$200 - 80 = 120$$

Una vez que tengo eso claro miro que para analizar los distractores deberé coger la fórmula del chi cuadrado.

$$\chi^2 = \sum \frac{(f_o - f_c)^2}{f_e}$$

De donde los fo los se porque son el número de gente que escogió cada respuesta errónea y el fe que es la frecuencia esperada la media de los valores que fallaron $60+20+40 = 120$ y dividido entre el número de opciones que hay $120/3 = 40$. Y con estos valores calcularemos el chi para cada celda.

$$(60-40)^2 / 40 = 10$$

$$(20 - 40)^2 / 40 = 10$$

$$(40 - 40)^2 / 40 = 0$$

Ahora que tenemos el chi cuadrado de cada celda buscaremos el chi cuadrado total que se hallará a partir de la suma del chi de cada celda $10+10+0 = 20$ este será el valor empírico.

Ahora con este valor lo compararé con el valor teórico que se ve por la tabla del formulario , donde me fijaré en los grados de libertad y el número de errores menos 1

Dis

Nivel de confianza(NC) = 1 - α = 1 - nivel de significac

NC					90%	95%
α	0,995	0,975	0,900	0,500	0,100	0,050
n						
1	0,000	0,000	0,016	0,455	2,706	3,841
2	0,010	0,051	0,211	1,386	4,605	5,991

Comparo los resultados:

20 > 5.991 como el valor empírico es mayor que el teórico es que estos distractores no son buenos no distraen bien y para mejorar el ítem debería cambiarlo.