Nombre de alumno: Maria Cristel Cajija González

Nombre del profesor: Rosario Gómez Lujano

Nombre del trabajo:

Ensayo: Prueba de hipótesis con una muestra, justificación de hipótesis, hipótesis nula y alternativa, error tipo I y II, contraste de hipótesis bilateral para media, hipótesis y prueba de hipótesis, procedimientos sistemáticos para prueba de hipótesis, prueba para proporciones.

Mapa conceptual: Distribución normal y T de estudent, prueba de una y dos colas, regresión y correlación, correlación por ajustes de una recta con el criterio de mínimos cuadrados, errores de la pendiente, y ordenada en el origen de la recta de regresión, regresión lineal, vertiente descriptiva o correlación, vertiente inferencial o regresión.

Ejercicios: Determinar los valores de Z

Materia: Estadística Inferencial

Grado: Cuarto Cuatrimestre

Grupo: Único

Pichucalco, Chiapas, a 02 de diciembre de 2020.

ENSAYO

ESTADISTICA INFERENCIAL

Prueba de hipótesis para una muestra

Una prueba de hipótesis es una regla que especifica si se puede aceptar o rechazar una afirmación acerca de una población dependiendo de la evidencia proporcionada por una muestra de datos.

Una prueba de hipótesis examina dos hipótesis opuestas sobre una población: la hipótesis nula y la hipótesis alternativa. La hipótesis nula es el enunciado que se probará. Por lo general, la hipótesis nula es un enunciado de que "no hay efecto" o "no hay diferencia". La hipótesis alternativa es el enunciado que se desea poder concluir que es verdadero de acuerdo con la evidencia proporcionada por los datos de la muestra.

Con base en los datos de muestra, la prueba determina si se puede rechazar la hipótesis nula. Usted utiliza el valor p para tomar esa decisión. Si el valor p es menor que el nivel de significancia (denotado como α o alfa), entonces puede rechazar la hipótesis nula.

Justificación de la hipótesis

Son el reverso de las hipótesis de investigación, es decir, refutan o niegan la relación entre variables. ... Indica el porqué de la investigación exponiendo sus razones. Por medio de la justificación debemos demostrar que el estudio es necesario e importante.

En las bases de esta propuesta, se plantea un nuevo equilibrio entre SABER, SABER-HACER y SABER SER. Es decir, la preocupación de la formación estará centrada tanto en los procesos cognitivos del APRENDER A APRENDER, como, asimismo, en los conocimientos prácticos o competencias del SABER-HACER, los conocimientos sociales de la convivencia y el conocimiento personal de sí mismo (SABER SER).

Un elemento que incide significativamente en la fundamentación de un proyecto es la identificación de su carácter estratégico respecto de una visión global del desarrollo local.

Hipótesis nula y alternativa

Las hipótesis nula y alternativa son dos enunciados mutuamente excluyentes acerca de una población. Una prueba de hipótesis utiliza los datos de la muestra para determinar si se puede rechazar la hipótesis nula.

Hipótesis nula (H_0)

La hipótesis nula indica que un parámetro de población (tal como la media, la desviación estándar, etc.) es igual a un valor hipotético. La hipótesis nula suele ser una afirmación inicial que se basa en análisis previos o en conocimiento especializado.

Hipótesis alternativa (H_1)

La hipótesis alternativa indica que un parámetro de población es más pequeño, más grande o diferente del valor hipotético de la hipótesis nula. La hipótesis alternativa es lo que usted podría pensar que es cierto o espera probar que es cierto.

Ejemplos de la hipótesis nula

Un investigador puede postular una hipótesis:

H_1: las plantas de tomate exhiben una mayor tasa de crecimiento cuando se plantan en compost en lugar del suelo.

Y una hipótesis nula:

H_0: las plantas de tomate no presentan una mayor tasa de crecimiento cuando se plantan en el compost en lugar del suelo.

Es importante seleccionar cuidadosamente el texto de la nula y asegurarse de que sea lo más específico posible. Por ejemplo, el investigador puede postular una hipótesis nula:

H_0: las plantas de tomate no muestran ninguna diferencia en sus tasas de crecimiento cuando se plantan en compost en lugar del suelo.

Hay un gran defecto con esta H_0. Si las plantas realmente crecen más lentamente en el compost que en el suelo, se llega a un callejón sin salida. H_1 no está respaldada y tampoco la H_0, ya que existe una diferencia en las tasas de crecimiento.

Ejemplo de hipótesis alternativa

Supongamos que un investigador ha realizado un estudio acerca del salario medio mensual en un determinado barrio de una ciudad. Imaginemos que de la población de ese barrio, el investigador ha encuestado a 1.000 personas llegando a la conclusión de que el salario medio mensual por habitante es de 1.500 u.m. (unidades monetarias).

Por tanto, el investigador quiere contrastar si ese salario medio mensual es igual a 1.500 u.m. (conclusión del estudio y por ende hipótesis alternativa) o si por el contrario el salario medio mensual es distinto a 1.500 u.m. (conclusión contraria a la del estudio que se pretende negar y por ende hipótesis nula)

El contraste a realizar sería el siguiente:

H0: El salario medio mensual es distinto a 1.500 u.m.

H1: El salario mensual es igual a 1.500 u.m.

$$(H0: \bar{X} \neq 1.500\ u.m.)$$

$$(H1: \bar{X} = 1.500\ u.m.)$$

Como vemos, la hipótesis alternativa (H1), es la conclusión alcanzada por el investigador. Para demostrarla el investigador va a tratar de probar que lo contrario a su hipótesis alternativa (hipótesis nula, H0), no es cierto. Como conclusión, podemos deducir que la formulación de la hipótesis alternativa es la que nos va a conducir a la formulación de la hipótesis nula.

Error tipo I (α)

El error tipo I, también llamado alfa (α), **se comete al rechazar la hipótesis nula (H0) siendo esta verdadera**. Así, la probabilidad de cometer un error de tipo I es α, que es el nivel de significación que hemos establecido para nuestra prueba de hipótesis.

Si por ejemplo el α que habíamos establecido es de 0.05, esto indicaría que estamos dispuestos a aceptar una probabilidad del 5% de equivocarnos al rechazar la hipótesis nula.

Error tipo II (β)

El error tipo II o beta (β), se comete al aceptar la hipótesis nula (H0) siendo esta falsa. Es decir, la probabilidad de cometer un error tipo II es beta (β), y depende de la potencia de la prueba (1-β).

Para reducir el riesgo de cometer un error tipo II, podemos optar por asegurarnos de que la prueba tiene suficiente potencia. Para ello, deberemos asegurarnos de que el tamaño de la muestra es lo suficientemente grande como para detectar una diferencia cuando ésta realmente exista.

Contraste de hipótesis bilateral para la media

Queremos contrastar una hipótesis acerca del valor de la media poblacional a partir de los resultados de una muestra. El proceso que seguimos es:

Contraste bilateral $H_0 : \mu = \mu_0 \quad H_1 : \mu \neq \mu_0$	*1) Establecer la hipótesis*	Contraste unilateral $H_0 : \mu \leq \mu_0 \quad H_1 : \mu > \mu_0$
buscamos $z_{\alpha/2}$ tal que $P(-z_{\alpha/2} \leq z \leq z_{\alpha/2}) = 1 - \alpha$  $\left(\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$	Las medias muestrales se distribuyen: $N(\mu, \frac{\sigma}{\sqrt{n}})$ *2) Elegir el nivel de significación α y determinar la zona de aceptación a partir del* Intervalo de confianza	buscamos z_α tal que $P(z \leq z_\alpha) = 1 - \alpha$  $\left(-\infty, \mu_0 + z_\alpha\frac{\sigma}{\sqrt{n}}\right)$
$\bar{x} \in \left(\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$ aceptamos H_0 $\bar{x} \notin \left(\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$ rechazamos H_0	*3) Verificación* *4) Decisión*	$\bar{x} \in \left(-\infty, \mu_0 + z_\alpha\frac{\sigma}{\sqrt{n}}\right)$ aceptamos H_0 $\bar{x} \notin \left(-\infty, \mu_0 + z_\alpha\frac{\sigma}{\sqrt{n}}\right)$ rechazamos H_0

Hipótesis y prueba de hipótesis

Una prueba de hipótesis es una regla que especifica si se puede aceptar o rechazar una afirmación acerca de una población dependiendo de la evidencia proporcionada por una muestra de datos.

Una prueba de hipótesis examina dos hipótesis opuestas sobre una población: la hipótesis nula y la hipótesis alternativa. La hipótesis nula es el enunciado que se probará. Por lo general, la hipótesis nula es un enunciado de que "no hay efecto" o "no hay diferencia". La hipótesis alternativa es el enunciado que se desea poder concluir que es verdadero de acuerdo con la evidencia proporcionada por los datos de la muestra.

Con base en los datos de muestra, la prueba determina si se puede rechazar la hipótesis nula. Usted utiliza el valor p para tomar esa decisión. Si el valor p es menor que el nivel de significancia (denotado como α o alfa), entonces puede rechazar la hipótesis nula.

Un error común de percepción es que las pruebas estadísticas de hipótesis están diseñadas para seleccionar la más probable de dos hipótesis. Sin embargo, al diseñar una prueba de hipótesis, establecemos la hipótesis nula como lo que queremos desaprobar. Puesto que establecemos el nivel de significancia para que sea pequeño antes del análisis (por lo general, un valor de 0.05 funciona adecuadamente), cuando rechazamos la hipótesis nula, tenemos prueba estadística de que la alternativa es verdadera. En cambio, si no podemos rechazar la hipótesis nula, no tenemos prueba estadística de que la hipótesis nula sea verdadera. Esto se debe a que no establecimos la probabilidad de aceptar equivocadamente la hipótesis nula para que fuera pequeña.

Entre las preguntas que se pueden contestar con una prueba de hipótesis están las siguientes:

- ¿Tienen las estudiantes de pregrado una estatura media diferente de 66 pulgadas?
- ¿Es la desviación estándar de su estatura igual a o menor que 5 pulgadas?
- ¿Es diferente la estatura de las estudiantes y los estudiantes de pregrado en promedio?
- ¿Es la proporción de los estudiantes de pregrado significativamente más alta que la proporción de las estudiantes de pregrado?

PROCEDIMIENTO DE 5 PASOS PARA PROBAR UNA HIPOTESIS

PASO 1.-PLANTEAMIENTO DE HIPOTESIS. -Ho: Hipótesis Nula -H1: Hipótesis Alternativo Hipótesis Nula. -Una afirmación o enunciado tentativo que se realiza acerca del valor de un parámetro poblacional. Por lo común en una afirmación de que el parámetro de población tiene valor especifico. Hipótesis Alternativa. -Una afirmación o enunciado que se aceptara si los datos muéstrales proporcionan amplia evidencia de que la hipótesis nula es falsa PASO 2.-NIVELES DE SIGNIFICACION. El riesgo que se acerca en rechazar la hipótesis nula cuando en realidad deben asemejarse por ser verdadera. El nivel de significación se denota mediante la letra griega sigma. No hay un nivel de significación que se aplique a todos los estudios que implican muestreo. Deben tomarse una decisión de usar el nivel 0.05, el nivel 0.01, el 0.10 o cualquier otro nivel entre 0 y 1 Tradicionalmente se relaciona el nivel 0.05 para proyectos de investigación sobre consumo, el 0.01 para control de calidad y el 0.10 para encuesta políticas. Como investigador debe decidir el nivel de significación antes de formular una regla de decisión y recopilar datos muéstrales. ERROR TIPO 1.-La probabilidad de rechazar la hipótesis nula cuando en realidad es verdadera. ERROR TIPO 2.-L probabilidad de aceptar la hipótesis nula cuando en realidad es falsa. PASO 3.-ESTADISTICO DE PRUEBA Un valor, determinado a partir de la información muestral, que se utiliza para aceptar o rechazar la hipótesis nula.

PASO 4.-REGLA DE DECISION Es una regla simple la cual es una afirmación de las condiciones bajo las que se acepta la hipótesis nula. PASO 5.-TOMA DE DECISION Es la toma de decisión si se debe aceptar o rechazar la hipótesis nula Nota: Los conceptos que están en amarillo, son conceptos que deben investigar para resolver el experimento.

Pruebas de proporciones

Las pruebas de proporciones son adecuadas cuando los datos que se están analizando constan de cuentas o frecuencias de elementos de dos o más clases. El objetivo de estas pruebas es evaluar las afirmaciones con respecto a una proporción (o Porcentaje) de población. Las pruebas se basan en la premisa de que una proporción muestral (es decir, x ocurrencias en n observaciones, o x/n) será igual a la proporción verdadera de la población si se toman márgenes o tolerancias para la variabilidad muestral. Las pruebas suelen enfocarse en la diferencia entre un número esperado de ocurrencias, suponiendo que una afirmación es verdadera, y el número observado realmente. La diferencia se compara con la variabilidad prescrita mediante una distribución de muestreo que tiene como base el supuesto de que H_0 es realmente verdadera.

En muchos aspectos, las pruebas de proporciones se parecen a las pruebas de medias, excepto que, en el caso de las primeras, los datos muestrales se consideran como cuentas en lugar de como mediciones. Por ejemplo, las pruebas para medias y proporciones se pueden utilizar para evaluar afirmaciones con respecto a:

1) Un parámetro de población único (prueba de una muestra)

2) La igualdad de parámetros de dos poblaciones (prueba de dos muestras), y

3) La igualdad de parámetros de más de dos poblaciones (prueba de k muestras). Además, para tamaños grandes de muestras, la distribución de muestreo adecuada para pruebas de proporciones de una y dos muestras es aproximadamente normal, justo como sucede en el caso de pruebas de medias de una y dos muestras.

Prueba de proporciones de una muestra

Cuando el objetivo del muestreo es evaluar la validez de una afirmación con respecto a la proporción de una población, es adecuado utilizar una prueba de una muestra. La metodología de prueba depende de si el número de observaciones de la muestra es grande o pequeño.

Como se habrá observado anteriormente, las pruebas de grandes muestras de medias y proporciones son bastante semejantes. De este modo, los valores estadísticos de prueba miden la desviación de un valor estadístico de muestra a partir de un valor propuesto. Y ambas pruebas se basan en la distribución normal estándar para valores críticos. Quizá la única diferencia real entre las ambas radica en la forma corno se obtiene la desviación estándar de la distribución de muestreo.

Esta prueba comprende el cálculo del valor estadístico de prueba Z

$$Z_{prueba} = \frac{\frac{x}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Donde:

$x = ocurrencias$

$n = observaciones$

$\frac{x}{n} = proporción\ de\ la\ muestra$

$p_0 = proporción\ propuesta$

$$\sqrt{\frac{p_0(1-p_0)}{n}} = desviación\ estándar\ de\ la\ proporción$$

Si se muestrea a partir de una población finita

$$\frac{n}{N} \cdot 100\% > 5\%$$

Se debe utilizar el factor finito de corrección

$$Z_{prueba} = \frac{\frac{x}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$$

Posteriormente este valor es comparado con el valor de Z, obtenido a partir de una tabla normal a un nivel de significación seleccionado.

Como ocurrió con la prueba de medias de una muestra, las pruebas de proporciones pueden ser de una o dos colas.

El tipo de prueba refleja H_1. Por ejemplo, hay tres posibilidades para H_1:

H_1: $p > p_0$ H_1: $p < p_0$ H_1: $p \neq p_0$

La hipótesis nula es: H_0: $p = p_0$

La primera alternativa establece una prueba de cola derecha, la segunda, izquierda y la tercera, una prueba de dos colas.

Ejemplo ilustrativo

En un estudio se afirma que 3 de 10 estudiantes universitarios trabajan. Pruebe esta aseveración, a un nivel de significación de 0,025, respecto a la alternativa de que la proporción real de los estudiantes universitarios trabajan es mayor de lo que se afirma, si una muestra aleatoria de 600 estudiantes universitarios revela que 200 de ellos trabajan. La muestra fue tomada de 10000 estudiantes.

Los datos son:

$$p_0 = \frac{3}{10} = 0{,}333$$

$\alpha = 0{,}025$
$n = 600$
$X = 200$
$N = 10000$

Las hipótesis son:

$H_0: p = p_0$

$H_1: p > p_0$

Con lectura en la tabla para un área de 0,025 le corresponde un valor $Z_{tabla} = 1{,}96$. Se toma en cuenta el valor positivo porque se trata de una prueba de hipótesis a cola derecha.

Como en los datos aparece el tamaño de la población, se debe verificar si el tamaño de la nuestra es mayor que el 5%. Se remplaza valores en la siguiente fórmula:

$$\frac{n}{N} \cdot 100\% > 5\%$$

$$\frac{600}{10000} \cdot 100\% = 6\%$$

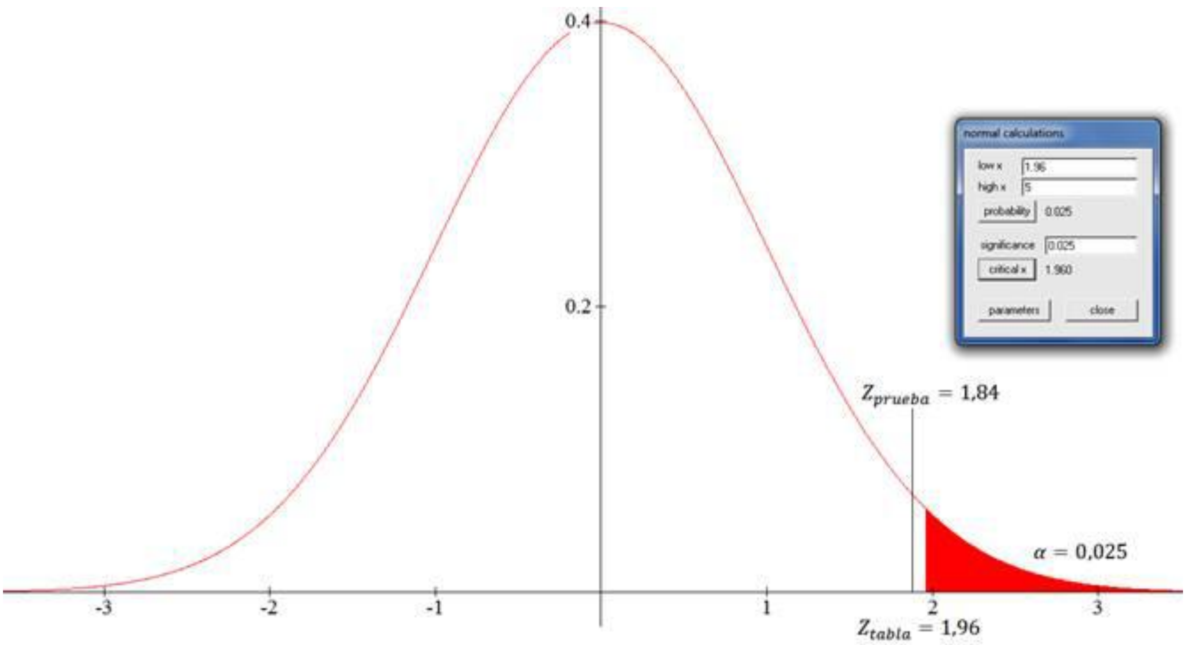
Por lo tanto se debe utilizar la fórmula con el factor finito de corrección.

$$Z_{prueba} = \frac{\frac{x}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}} = \frac{\frac{200}{600} - 0{,}333}{\sqrt{\frac{0{,}333(1-0{,}333)}{600}} \cdot \sqrt{\frac{10000-6000}{10000-1}}} = 1{,}84$$

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G	H	I
1	p_o	0,3	=3/10						
2	n	600							
3	x	200							
4	α	0,025							
5	N	10000	$\frac{n}{N}\cdot 100\% > 5\%$		6	=(B2/B5)*100			
6	$H_O: p = p_0$								
7	$H_1: p > p_0$								
8	z tabla	-1,96	=DISTR.NORM.ESTAND.INV(B4)						
9		1,96	=B8*-1						
10									
11	$z_{prueba} = \dfrac{\frac{x}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}\sqrt{\frac{N-n}{N-1}}}$			1,84	=(B3/B2-B1)/(RAIZ(B1*(1-B1)/B2)*RAIZ((B5-B2)/(B5-1)))				
12									
13									

El gráfico elaborado en Winstats y Paint se muestra a continuación:



Decisión:

H_0 es aceptada, ya que $Z_{prueba} = 1{,}84$ es menor que $Z_{tabla} = 1{,}96$, por lo tanto es cierto que 3 de 10 estudiantes universitarios trabajan.

Prueba de proporciones de dos muestras

El objetivo de una prueba de dos muestras es determinar si las dos muestras independientes fueron tomadas de dos poblaciones, las cuales presentan la misma proporción de elementos con determinada característica. La prueba se concentra en la diferencia relativa (diferencia dividida entre la desviación estándar de la distribución de muestreo) entre las dos proporciones muestrales. Diferencias pequeñas denotan únicamente la variación casual producto del muestreo (se acepta H0), en tanto que grandes diferencias significan lo contrario (se rechaza H0). El valor estadístico de prueba (diferencia relativa) es comparado con un valor tabular de la distribución normal, a fin de decidir si H0 es aceptada o rechazada. Una vez más, esta prueba se asemeja considerablemente a la prueba de medias de dos muestras.

La hipótesis nula en una prueba de dos muestras es

$$H_0: p_1 = p_2$$

Las hipótesis alternativas posibles son

$$H_1: p_1 \neq p_2 \quad H_1: p_1 > p_2 \quad H_1: p_1 < p_2$$

La estimación combinada de p se puede calcular de la siguiente manera:

$$p = \frac{x_1 + x_2}{n_1 + n_2}$$

Donde:
p = proporción muestral
x_1 = número de aciertos en la muestra 1
x_2 = número de aciertos en la muestra 2
n_1 = número de observaciones de la muestra 1
n_2 = número de observaciones de la muestra 2

Este valor de p se utiliza para calcular el valor estadístico de prueba

$$z_{prueba} = \frac{\frac{x_1}{n_1} - \frac{x_2}{n_2}}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Ejemplo ilustrativo

Se ponen a prueba la enseñanza de la estadística empleando Excel y Winstats. Para determinar si los estudiantes difieren en términos de estar a favor de la nueva enseñanza se toma una muestra de 20 estudiantes de dos paralelos. De paralelo A 18 están a favor, en tanto que del paralelo B están a favor 14. ¿Es posible concluir con un nivel de significación de 0,05 que los estudiantes que están a favor de la nueva enseñanza de la Estadística es la misma en los dos paralelos?

Los datos son:

$n_1 = 20$

$n_2 = 20$

$x_1 = 18$

$x_2 = 14$

$\alpha = 0{,}05$

Las hipótesis son

$H_0: p_1 = p_2$

$H_1: p_1 \neq p_2$

Como se trata de una prueba de hipótesis a dos colas se debe calcular

$$\frac{\alpha}{2} = \frac{0{,}05}{2} = 0{,}025$$

Con lectura en la tabla para un área de 0,025 le corresponde un valor $Z_{tabla} = \pm 1{,}96$.

Calculando la proporción muestral se obtiene:

$$p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{18 + 14}{20 + 20} = 0{,}8$$

Calculando Z_{prueba} se obtiene:

$$Z_{prueba} = \frac{\frac{x_1}{n_1} - \frac{x_2}{n_2}}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\frac{18}{20} - \frac{14}{20}}{\sqrt{0{,}8(1-0{,}8)\left(\frac{1}{20} + \frac{1}{20}\right)}} = 1{,}58$$

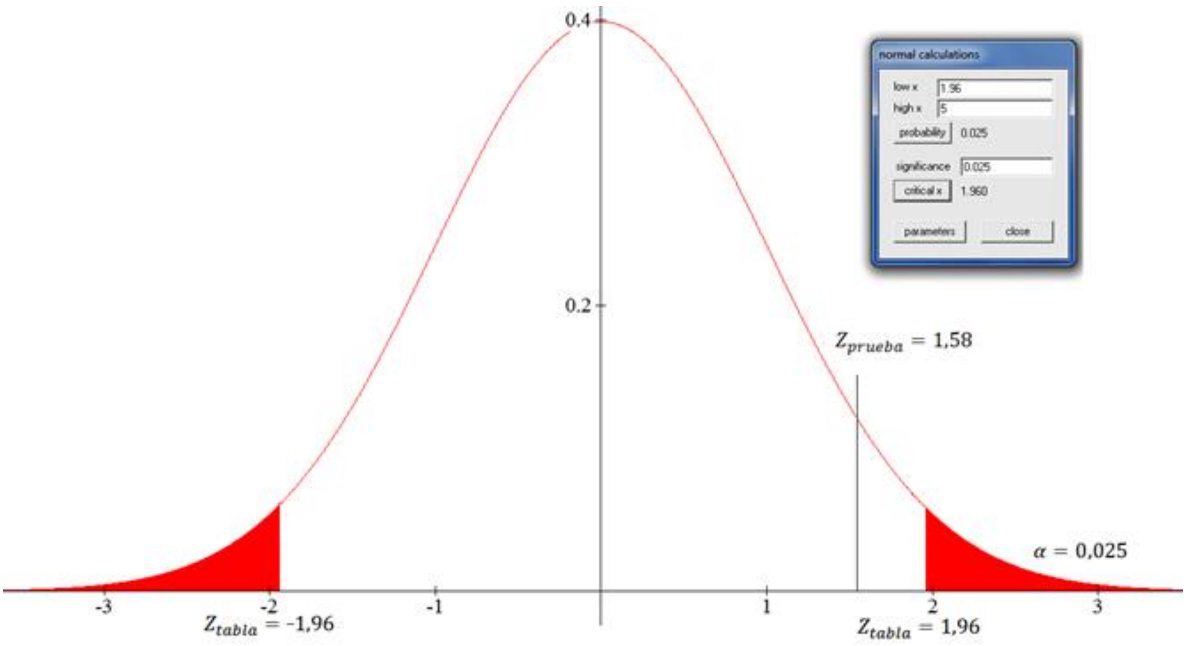
$$Z_{prueba} = \frac{\frac{18}{20} - \frac{14}{20}}{\sqrt{0{,}8(1-0{,}8)\left(\frac{1}{20} + \frac{1}{20}\right)}} = 1{,}58$$

$$Z_{prueba} = 1{,}58$$

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G
1	n_1	20					
2	n_2	20					
3	x_1	18					
4	x_2	14					
5	α	0,05					
6	$H_0: p_1 = p_2$						
7	$H_1: p_1 \neq p_2$						
8	$\frac{\alpha}{2}$	0,025					
9							
10	Z_{tabla}	-1,96	=DISTR.NORM.ESTAND.INV(B8)				
11		1,96	=B10*-1				
12	$p = \frac{x_1 + x_2}{n_1 + n_2}$	0,8	=(B3+B4)/(B1+B2)				
13							
14							
15	$Z_{prueba} = \frac{\frac{x_1}{n_1} - \frac{x_2}{n_2}}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$	1,58	=(B3/B1-B4/B2)/RCUAD(B12*(1-B12)*(1/B1+1/B2))				
16							
17							

El gráfico elaborado en Winstats y Paint se muestra a continuación:



Decisión:

H_0 es aceptada, ya que $Z_{prueba} = 1{,}58$ está en la zona de aceptación $Z_{tabla} = \pm 1{,}96$, entonces, la proporción de los estudiantes que están a favor de la nueva enseñanza de la Estadística es la misma en los dos paralelos.

Prueba de proporciones de k muestras

La finalidad de una prueba de k muestras es evaluar la aseveración que establece que todas las k muestras independientes provienen de poblaciones que presentan la misma proporción de algún elemento. De acuerdo con esto, las hipótesis nula y alternativa son

H_0:Todas las proporciones de la población son iguales.

H_1: No todas las proporciones de la población son iguales.

La estimación combinada de la proporción muestral "p" se calcula de la siguiente manera:

$$p = \frac{x_1 + x_2 + x_3 + \cdots x_n}{n_1 + n_2 + n_3 + \cdots n_n}$$

En una muestra se puede dar un conjunto de sucesos, los cuales ocurren con frecuencias observadas "o"(las que se observa directamente) y frecuencias esperadas o teóricas "e" (las que se calculan de acuerdo a las leyes de probabilidad).

La frecuencia esperada "e" se calcula así: $e = p \cdot o_{total}$

p = proporción muestral

o_{total} = frecuencia total observada

El estadístico de prueba es

$$\chi^2_{prueba} = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \frac{(o_3 - e_3)^2}{e_3} + \cdots \frac{(o_n - e_n)^2}{e_n}$$

$$\chi^2_{prueba} = \sum \frac{(o_i - e_i)^2}{e_i}$$

Donde:
χ es la letra griega ji
χ^2 se lee ji cuadrado

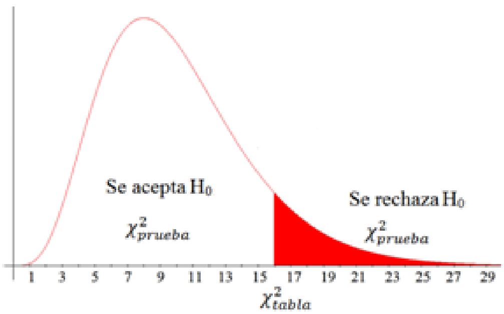
Por lo tanto el valor estadístico de prueba para este caso es la prueba *ji cuadrado* o conocida también como *chi cuadrado*

Como sucede con las distribuciones t y F, la distribución ji cuadrado tiene una forma que depende del número de grados de libertad asociados a un determinado problema.

Para obtener un valor crítico (valor que deja un determinado porcentaje de área en la cola) a partir de una tabla de ji cuadrado, se debe seleccionar un nivel de significación y determinar los grados de libertad para el problema que se esté resolviendo.

Los grados de libertad son una función del número de casillas en una tabla de $2 \cdot k$. Es decir, los grados de libertad reflejan el tamaño de la tabla. Los grados de libertad de la columna son el número de filas (categorías) menos 1, o bien, $r-1$ .Los grados de libertad de cada fila es igual al número de columnas (muestras) menos 1, o bien, $k-1$. El efecto neto es que el número de grados de libertad para la tabla es el producto de (número de filas -1) por (número de columnas -1), o bien, $(r-1)(k-1)$.Por lo tanto con 2 filas y 4 columnas, los grados de libertad son $(2-1)(4-1)=3$

La prueba ji cuadrado requiere la comparación del χ^2_{prueba} con el χ^2_{tabla}. Si el valor estadístico de prueba es menor que el valor tabular, la hipótesis nula es aceptada, caso contrario, H_0 es rechazada.



Nota: Un valor estadístico de χ^2_{prueba}menor que el valor crítico χ^2_{tabla} o igual a él se considera como prueba de la variación casual en donde H_0 es aceptada.

Ejemplos ilustrativos:

1) El siguiente valor $3 \cdot 4$ representa el tamaño de una tabla $r \cdot k$.

Determine el número de grados de libertad y obtenga el valor crítico en el nivel 0,05 se significación.

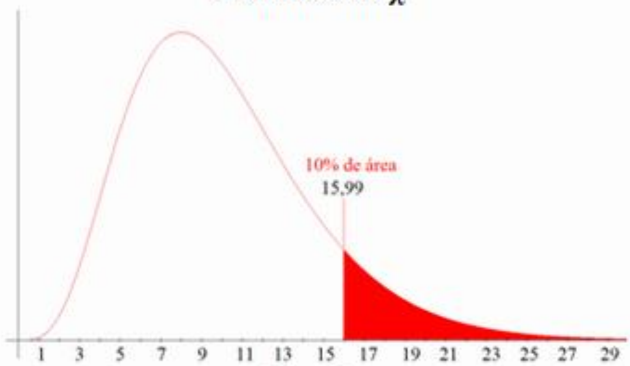
Solución:

Los grados de libertad se calculan aplicando la fórmula:

$$Grados\ de\ libertad = (r-1)(k-1)$$

$$Grados\ de\ libertad = (3-1)(4-1) = 12$$

TABLA Nº 6
DISTRIBUCIÓN χ^2



Ejemplo:
Para 10 grados de libertad
$P(\chi^2 > 15,99) = 0,10 = 10\%$

	0,995	0,990	0,975	0,950	0,900	0,750	0,500	0,250	0,100	**0,050**	0,025	0,010	0,005
1	0,000	0,000	0,001	0,004	0,016	0,102	0,455	1,323	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	0,575	1,386	2,773	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	1,213	2,366	4,108	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	1,923	3,357	5,385	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	2,675	4,351	6,626	9,236	11,070	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	3,455	5,348	7,841	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	4,255	6,346	9,037	12,017	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	3,490	5,071	7,344	10,219	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	5,899	8,343	11,389	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	6,737	9,342	12,549	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	7,584	10,341	13,701	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	8,438	11,340	14,845	18,549	**21,026**	23,337	26,217	28,300

Con lectura en la tabla con 12 grados de libertad y 0,05 de área se obtiene $\chi^2_{tabla} = 21{,}026$

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D
1	r	4		
2	k	5		
3				
4	(r-1)(k-1)	12	=(B1-1)*(B2-1)	
5				
6	α	0,05		
7	χ^2_{tabla}	21,026	=PRUEBA.CHI.INV(B6;B4)	
8				
9	α	0,01		
10	χ^2_{tabla}	26,2170	=PRUEBA.CHI.INV(B9;B4)	

2) La siguiente tabla muestra las frecuencias observadas y las frecuencias esperadas al lanzar un dado 60 veces. Contrastar la hipótesis de que el dado es bueno, con un nivel de significación de 0,01.

Cara del dado	1	2	3	4	5	6
Frecuencia observada	6	8	9	15	14	8
Frecuencia esperada	10	10	10	10	10	10

Solución:

$r = 2$
$k = 6$

$$\alpha = 0{,}01$$

Las hipótesis son:
H_0:Todas las proporciones de la población son iguales.
H_1: No todas las proporciones de la población son iguales.

Los grados de libertad se calculan aplicando la fórmula:
$Grados\ de\ libertad = (2-1)(6-1)$
$Grados\ de\ libertad = 5$

Con lectura en la tabla con 5 grados de libertad y 0,01 de área se obtiene $\chi^2_{tabla} = 15{,}086$

Calculando χ^2_{prueba} se obtiene:

$$\chi^2_{prueba} = \sum \frac{(o_i - e_i)^2}{e_i}$$

$$\chi^2_{prueba} = \frac{(6-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(9-10)^2}{10} + \frac{(15-10)^2}{10} + \frac{(14-10)^2}{10} + \frac{(8-10)^2}{10}$$

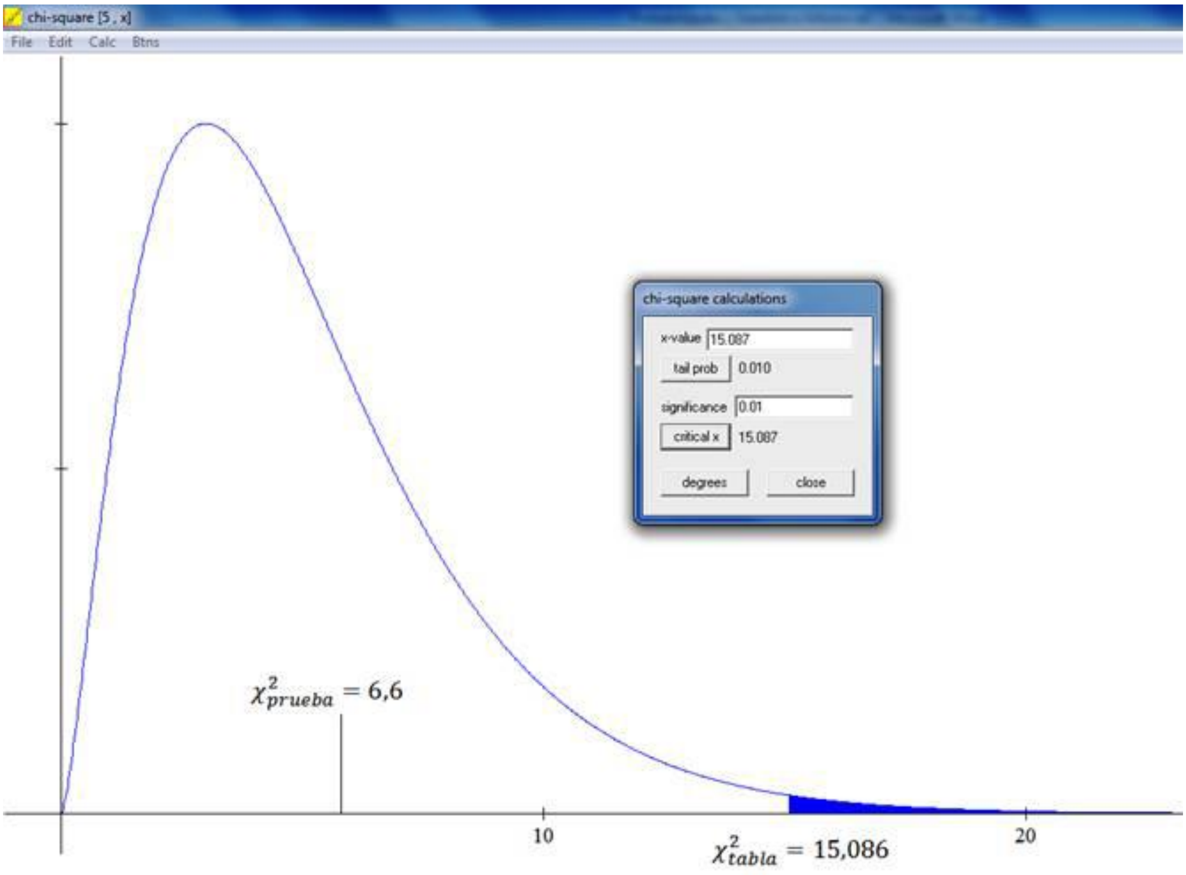
$$\chi^2_{prueba} = 1{,}6 + 0{,}4 + 0{,}1 + 2{,}5 + 1{,}6 + 0{,}4$$

$$\chi^2_{prueba} = 6{,}6$$

Los cálculos en Excel se muestran en la siguiente figura:

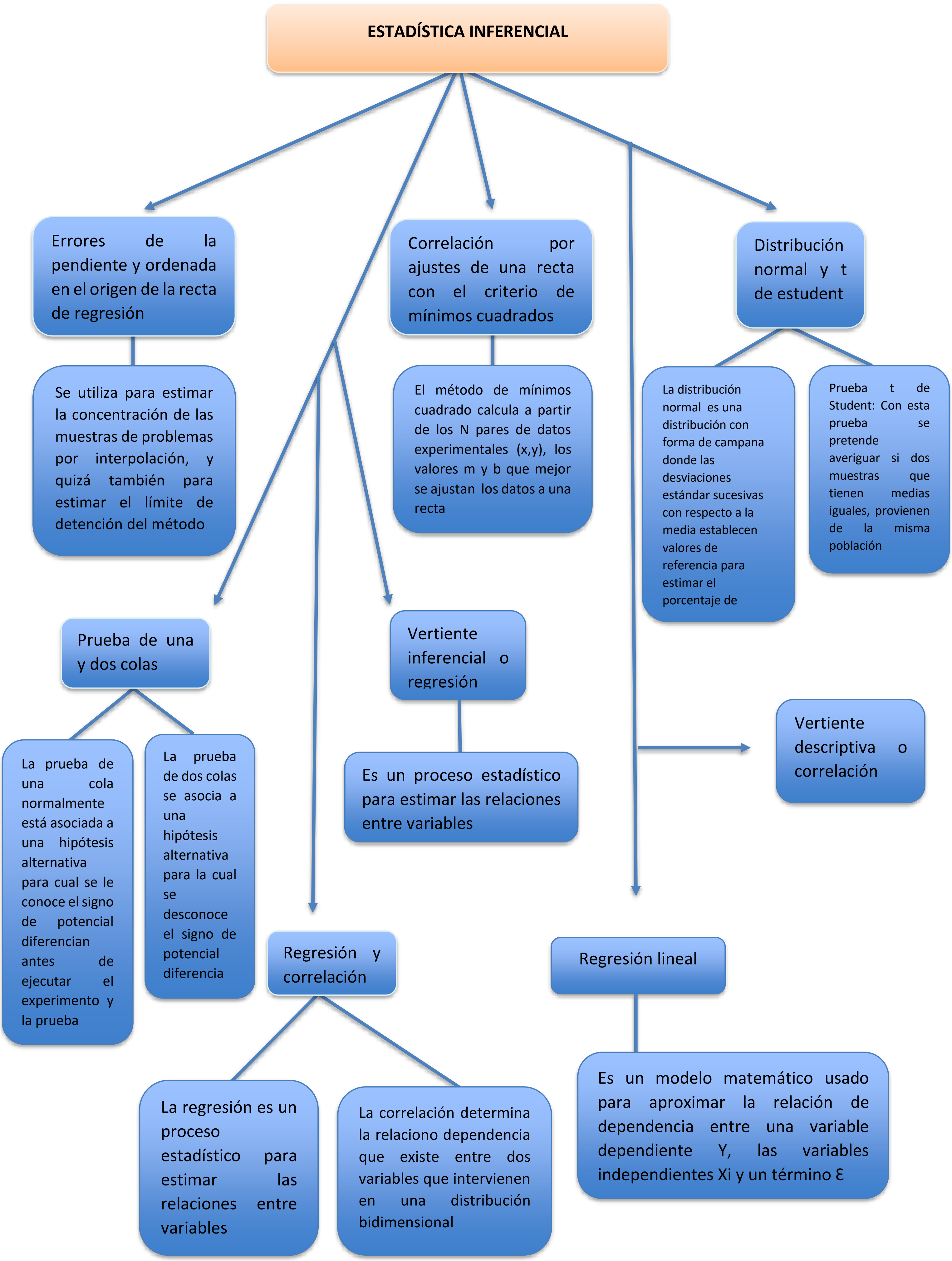
	A	B	C	D	E	F	G
1	Cara del dado	1	2	3	4	5	6
2	Frecuencia observada	6	8	9	15	14	8
3	Frecuencia esperada	10	10	10	10	10	10
4							
5	α	0,01					
6	r	2	=CONTAR(B2:B3)				
7	k	6	=CONTAR(B2:G2)				
8	(r-1)(k-1)	5	=(B6-1)*(B7-1)				
9	χ^2_{tabla}	15,086	=PRUEBA.CHI.INV(B5;B8)				
10							
11	Probabilidad de χ^2_{prueba}	0,2521	=PRUEBA.CHI(B2:G2;B3:G3)				
12	$\chi^2_{prueba} = \sum \frac{(o_i - e_i)^2}{e_i}$	6,6	=PRUEBA.CHI.INV(B11;B8)				
13							

El gráfico elaborado en Winstats y Paint se muestra a continuación:



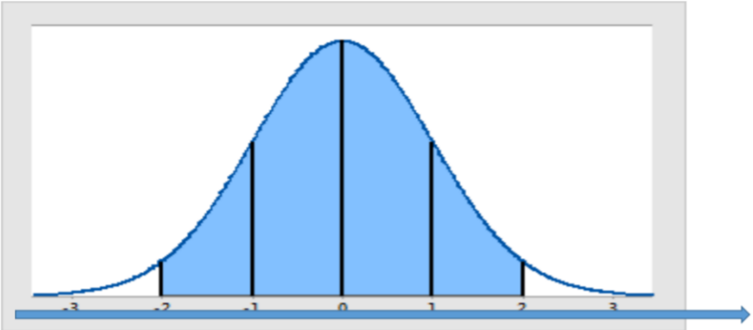
Decisión:

H_0 es aceptada, ya que χ^2_{prueba} (6,6) es menor que χ^2_{tabla}(15,086), por lo tanto, se concluye que todas las proporciones de la población son iguales, es decir, el dado es bueno.


ESTADÍSTICA INFERENCIAL
Errores de la pendiente y ordenada en el origen de la recta de regresión
Se utiliza para estimar la concentración de las muestras de problemas por interpolación, y quizá también para estimar el límite de detención del método
Correlación por ajustes de una recta con el criterio de mínimos cuadrados
El método de mínimos cuadrado calcula a partir de los N pares de datos experimentales (x,y), los valores m y b que mejor se ajustan los datos a una recta
Distribución normal y t de estudent
La distribución normal es una distribución con forma de campana donde las desviaciones estándar sucesivas con respecto a la media establecen valores de referencia para estimar el porcentaje de
Prueba t de Student: Con esta prueba se pretende averiguar si dos muestras que tienen medias iguales, provienen de la misma población
Prueba de una y dos colas
La prueba de una cola normalmente está asociada a una hipótesis alternativa para cual se le conoce el signo de potencial diferencian antes de ejecutar el experimento y la prueba
La prueba de dos colas se asocia a una hipótesis alternativa para la cual se desconoce el signo de potencial diferencia
Vertiente inferencial o regresión
Es un proceso estadístico para estimar las relaciones entre variables
Vertiente descriptiva o correlación
Regresión y correlación
La regresión es un proceso estadístico para estimar las relaciones entre variables
La correlación determina la relaciono dependencia que existe entre dos variables que intervienen en una distribución bidimensional
Regresión lineal
Es un modelo matemático usado para aproximar la relación de dependencia entre una variable dependiente Y, las variables independientes Xi y un término Ɛ

Resuelve los siguientes problemas

- Determine el valor de Z, cuando la media es cinco, sabiendo que la desviación estándar es 2 y x=15



$$Z=\frac{x-M}{\sigma}$$

$$Z=\frac{15-5}{2}=2$$

Z

- Una encuesta nacional reciente determino que los estudiantes de secundaria miraban 6.8 películas al mes con una desviación estándar de 0.5 horas. Una muestra aleatoria de 36 estudiantes revelo que la cantidad media de película que vieron fue de 6.2, con un nivel de significancia de 0.05 ¿Puede concluir que los estudiantes universitarios ven menos película que los estudiantes de secundaria?

H0 : $\mu \geq 6.8$
HI : $\mu < 6.2$

Datos del problema

Media teórica=6.8 películas

n= 36

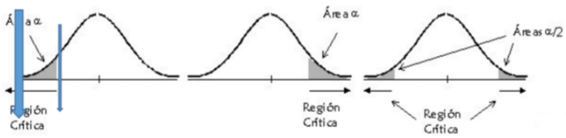
Media muestral=6.2 películas

Desviación estándar=0.5 horas

Nivel de significancia=0.05
0.5 -0.05=0.45

Z=-1.64

Nivel de confianza=95%



$$Z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

$$Z=\frac{6.2-6.8}{0.5/\sqrt{36}}=\frac{-0.6}{0.5/6}=-7.5$$

R: SE RECHAZA LA HIPOTESIS QUE LOS ALUMNOS VEN MAYOR O IGUAL A 6.8 PELICULAS AL MES CON UN NIVEL DE SIGNIFICANCIA DE 0.05.