Cátedra ESTADISTICA

TRABAJOS PRÁCTICOS

(SEGUNDA ETAPA)

2016

Facultad de Ingeniería

Universidad Nacional de La Patagonia S. J. B.

Comodoro Rivadavia

TRABAJO PRÁCTICO Nº 4

INFERENCIA ESTADÍSTICA

PRE - REQUISITOS:

Se requiere lectura previa y manejo conceptual de los siguientes conceptos trabajados en la primera etapa:

- ✓ Estadística Descriptiva – Variables, tipos - Medidas de posición y dispersión –
- ✓ Probabilidad – Teoría Axiomática – Teoremas – Sucesos independientes -
- ✓ Distribuciones de probabilidad, discretas y continuas – Propiedades, parámetros, usos, cálculos – Identificación de las variables aleatorias.
- ✓ Esperanza (media) y Varianza de las Distribuciones de probabilidad.
- ✓ Muestreo aleatorio -

Y análisis y comprensión de los contenidos que se están desarrollando en esta segunda etapa:

- ✓ Distribuciones muestrales.
- ✓ Teorema Central de Límite (T. C .L.) y sus consecuencias.
- ✓ Significado de ESTIMAR, diferencia entre estimación puntual y estimación por Intervalo de Confianza. ¿Cuál es realmente útil y por qué?
- ✓ Diferenciación entre intervalo de confianza (I. C.) y prueba de hipótesis (dócima), recordando los diferentes conceptos y razonamientos que incluye cada análisis.

CONSIGNA PARTICULAR

Leer y comprender los conceptos que se presentan a continuación, recurrir a los textos propuestos, analizar los problemas resueltos.
Con esta lectura comprensiva y considerando también los otros pre-requisitos, usted está ya en condiciones de trabajar con sus compañeros.
Recuerde que en esta segunda etapa de la cursada, usted debe ya ser capaz de trabajar en equipo y consultar sólo sus dudas.

Recuerde que es indispensable la **correcta interpretación en términos del problema, aún cuando no se lo señale específicamente en cada ejercicio.**

A continuación se transcriben párrafos del texto recomendado "**PROBABILIDAD Y ESTADÍSTICA PARA INGENIEROS"**, 4/E, de Miller, I. ; Freund, J. E.; Jhonson, R. A., Edición 1992, Prentice-hall Hispanoamericana, S. A. , (capítulos 6 y 7), actualmente citado como: Jhonson, Richard A.; Torre Marina Juan A. , **"Probabilidad y Estadística para Ingenieros de Miller y Frend"**, Prentice-hall Hispanoamericana, S. A., 3ra. Edición, México, 1997 (Clas.: 519.5/J.26, en Biblioteca Central de la U. N. P. S. J. B., Comodoro Rivadavia)

. . . "se supondrá que estamos trabajando con una clase particular de muestra denominada muestra aleatoria" . . . "se debe a que permite generalizaciones válidas o lógicas de . . .",
. . . "veremos como ciertos estadísticos puede esperarse que varíen de muestra a muestra". ..
"El concepto de distribución muestral es . . . "(pág. 187)

"Si una muestra aleatoria de tamaño "n" se elige de una población que tiene media **μ** y varianza **σ²**, entonces **x media** es una variable aleatoria cuya distribución tiene media **μ** y:

- para muestras tomadas de poblaciones infinitas la varianza de esta distribución es:

  σ^2 / n

- para muestras tomadas de poblaciones finitas de tamaño "N" la varianza de esta distribución es:

$$\left(\frac{\sigma^2}{n}\right) * \left(\frac{N-n}{N-1}\right) \qquad \text{(pág. 192)}$$

Si

$$x \approx N(E(x) = \mu, Var(x) = \sigma^2) \Rightarrow \qquad \frac{x-\mu}{\sigma} = z$$

$$\bar{x} \approx N(E(\bar{x}) = \mu, Var(\bar{x}) = \frac{\sigma^2}{n}) \Rightarrow \qquad \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = z$$

n grande

Si p=x/n (proporción)

$$\frac{p-\pi}{\sqrt{\frac{p(1-p)}{n}}} = z$$

Si σ^2 es desconocida

$$\frac{x-\mu}{s} = t_n$$

$$\frac{\bar{x}-\mu}{s/\sqrt{n}} = t_{n-1}$$

Distribución Siempre normal

Y de pág. 210...
"Cuando empleamos una media muestral para estimar la media de una población, sabemos que, aunque estamos utilizando un método de estimación con ciertas propiedades deseables, las posibilidades son escasas y prácticamente nulas de que la estimación en realidad sea igual a μ. En consecuencia, convendría acompañar la estimación puntual de μ con una afirmación de cuán cercana podemos razonablemente esperar que se encuentre la estimación.

El error, $\bar{x} - \mu$, es la diferencia entre la estimación y la cantidad que se supone que estima.

A fin de examinar este error, utilizaremos el hecho de que

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

es un valor de una variable aleatoria que tiene aproximadamente la distribución normal estándar. En consecuencia, podemos asegurar con una probabilidad de 1-α que la desigualdad

$$-z_{\alpha/2} \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}$$

será satisfecha o que

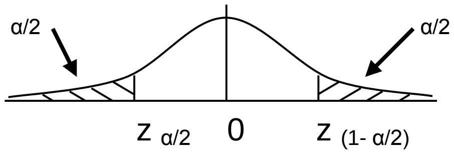
$$\frac{\left|\bar{x}-\mu\right|}{\sigma/\sqrt{n}} \leq z_{\alpha/2}$$

donde $z_{\alpha/2}$ es tal que el área bajo la curva normal a su derecha es α/2. Si ahora indicamos con E el valor máximo de $\left|\bar{x}-\mu\right|$, o sea el error máximo de estimación, tenemos

$$\boxed{E = z_{\alpha/2} * \sigma/\sqrt{n}}$$

con una probabilidad (1- α) . En otras palabras, si intentamos estimar μ con la media de una muestra aleatoria, podemos afirmar con una probabilidad de (1- α) que el error $\left|\bar{x}-\mu\right|$, será a lo sumo $z_{\alpha/2} * \sigma/\sqrt{n}$. Los valores de mayor uso para (1- α) son 0.95 y 0.99 y los valores correspondientes de $z_{\alpha/2}$ son $z_{0.025}$ = 1.96 y $z_{0.005}$ = 2.575.

De acuerdo con el teorema central del límite, es de esperarse que la distribución muestral de $\bar{X}$ tenga una distribución aproximadamente normal con media $\mu_{\bar{x}} = \mu$ y desviación estándar $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ .Al escribir $z_{\alpha/2}$ para el valor z, sobre el cual se encuentra un área de α/2, a partir de la figura siguiente se advierte que: $P\left(-z_{\alpha/2}. < z < z_{\alpha/2}\right) = 1-\alpha$



donde,

$$z = \frac{\left|\bar{x}-\mu\right|}{\sigma/\sqrt{n}}.$$

De aquí que

$$P\left(-z_{\alpha/2} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha \,.$$

Al multiplicar cada término de la desigualdad por $\sigma/\sqrt{n}$, y después restar de cada término y multiplicar por –1 (para invertir el sentido de las desigualdades), se obtiene:

$$P\left(\bar{x}-z_{\alpha/2} * \sigma/\sqrt{n} < \mu < \bar{x}+z_{\alpha/2} * \sigma/\sqrt{n}\right) = 1-\alpha$$

Se selecciona una muestra aleatoria de tamaño n de una población cuya variancia σ^2 se conoce y la media se calcula para obtener el intervalo de confianza (1- α)100%.

Intervalo de confianza de μ, conociendo σ^2

Si es la media de una muestra aleatoria de tamaño n de una población con una variancia conocida σ^2 el intervalo de confianza de (1- α)100% para μ es,

$$\bar{x}-z_{\alpha/2} * \sigma/\sqrt{n} < \mu < \bar{x}+z_{\alpha/2} * \sigma/\sqrt{n}$$

donde $z_{\alpha/2}$ es el valor de z a la derecha del cual se tiene un área de α/2.

Ejemplo:

Una muestra aleatoria de tamaño n = 100 se extrae de una población con $\sigma^2 = (5.1)^2$. Dado que la media muestral es $\bar{x} = 21.6$, construir un intervalo de confianza del 95% para la media de la población μ .

<u>*Solución:*</u> Sustituyendo los valores dados de n, $\bar{x}$, σ^2 y $z_{0.025} = 1.96$ en la fórmula para el intervalo de confianza, obtenemos

$$21.6 - 1.96 * \frac{5.1}{\sqrt{100}} < \mu < 21.6 + 1.96 * \frac{5.1}{\sqrt{100}}$$

ó $20.6 < \mu < 22.6$. Por supuesto, el intervalo de 20.6 a 22.6 contiene la media de la población μ o no, pero tenemos una confianza del 95% de que así sea. Esto significa que el método con que se obtuvo el intervalo "funciona" el 95% de las veces. En otras palabras, en repetidas aplicaciones de la fórmula del intervalo de confianza cabe esperar que el 95% de los intervalos contenga las medias de las poblaciones respectivas.

Ejemplo:

Un investigador quiere determinar el tiempo promedio que un mecánico tarda en rotar los neumáticos de un automóvil, y además desea poder asegurar con una confianza del 95% que la media de su muestra sea a lo sumo de 0,50 minutos. Se puede asumir por experiencia que $\sigma^2 = (1{,}6)^2$ minutos^2. ¿Qué tamaño debe tener la muestra?

<u>*Solución*</u>: Sustituyendo E = 0,50, $\sigma^2 = (1{,}6)^2$, y $z_{0,025} = 1{,}96$ en la fórmula para n obtenemos

$$n = \left[\frac{1.96 * 1.6}{0.5}\right]^2 = 39.3$$

ó 40, redondeado al entero más cercano. Así, el investigador tendrá que observar 40 operaciones de la tarea de rotar los neumáticos de un automóvil.

Los métodos expuestos hasta ahora exigen conocer σ ó que pueda ser aproximada mediante la desviación estándar muestral S, requiriendo así que n sea relativamente grande. Sin embargo, dado que es razonable suponer que estamos muestreando en una población normal, podemos fundamentar nuestro argumento en el teorema central del límite, es decir, en el hecho de que

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

es un valor de una variable aleatoria que tiene la distribución *t* con n-1 grados de libertad. Con un razonamiento similar al empleado para la variable *z,* llegamos al resultado de que podemos asegurar con una probabilidad de (1-α) que el error en que incurrimos al emplear para estimar μ será a lo sumo de

$$\boxed{E = t_{\alpha/2} * S / \sqrt{n}}$$

llamado *Error máximo de estimación*

y obtenemos la fórmula del intervalo con un nivel de confianza del (1-α) por ciento cuando no se conoce σ^2 .

$$\bar{x} - t_{\alpha/2} * S / \sqrt{n} < \mu < \bar{x} + t_{\alpha/2} * S / \sqrt{n}$$

INFERENCIA ESTADÍSTICA: TEOREMA CENTRAL DEL LÍMITE y DISTRIBUCIONES MUESTRALES

1. - Cierta población tiene media y varianza iguales a 500 y 900 respectivamente. Si se seleccionan muchas muestras de tamaño 36 y se calculan las medias muestrales:

a) ¿Qué valor es de esperar que tenga la media de todas esas medias?

b) ¿Qué valor se espera para la varianza y la desviación estándar de todas las medias muestrales?

c) ¿Qué forma es de esperar que tenga la distribución de todas las medias muestrales? Si es necesario, plantee los supuestos adecuados.

2. - Considere el experimento de tomar una prueba estandarizada sobre matemática. La variable X es la calificación recibida. Este examen tiene una calificación media igual a 720 puntos y una varianza de 360; un grupo de 40 estudiantes presenta el examen, siendo la media muestral del grupo de 725,6 puntos. Se forma una distribución de muestreo de medias con los promedios de todos esos grupos de 40 estudiantes.
Determine:

a) La calificación media de esta distribución de medias muestrales. Interprete.

b) La varianza y desviación estándar para esta distribución de medias muestrales. Interprete.

c) Calcule y luego comente, analizando los resultados obtenidos.

c1) P (X > 725) c2) P (X> 720)
c3) P ($\bar{x}$ > 725) c4) P ($\bar{x}$ > 720).

INFERENCIA ESTADÍSTICA: INTERVALOS DE CONFIANZA

3. - ¿Cuál es el error máximo que puede esperarse con una confianza del 90% cuando utilizamos la media de una muestra aleatoria de tamaño n = 64 para estimar la media de una población con σ^2=2,56?

4. - Si un intervalo de confianza para μ está dado por 13,2 ± 1,0; con un coeficiente de confianza del 75% ¿Cuál es el coeficiente de confianza para que el intervalo sea 13,2 ± 2,0?

5 - Si un intervalo de confianza para la media de una población dio 8,12 ± 2,5 para lo cual se tomó una muestra de tamaño 100. ¿Qué intervalo puede esperar de una muestra de tamaño 400 con la misma confianza y siendo la nueva media igual a la anterior?

6 - De una población normal con σ = 2.5, se extrae una muestra aleatoria simple de tamaño 9 cuyos valores se presentan en la siguiente tabla:

165	162	166	164	165	170	169	165	168

(a) Obtenga el intervalo de confianza para la media poblacional con un nivel de confianza del 95%.
(b) Repita (a) suponiendo que la varianza poblacional σ^2 es desconocida.

7.- En la construcción de una represa hidroeléctrica, se efectúan rigurosos controles para determinar la calidad del hormigón utilizado. Con el propósito de hallar la resistencia a la compresión promedio, se seleccionan 25 probetas, obteniéndose una media muestral de 450 Kg/cm², con una varianza (75 Kg/cm²)^2.
Admitiendo que la población de probetas moldeadas para efectuar los ensayos, sigue una distribución normal en cuanto a su resistencia a la compresión:

a) Construya un intervalo de confianza del 95% para la verdadera media.
b) Construya un intervalo de confianza del 90% para la varianza.
c) Interprete y concluya en términos del problema en ambos casos.

8. - Una cadena de hipermercados adquiere mil contenedores con 100 televisores cada uno. Se toma una muestra de 20 contenedores y se encuentra que el 15% de los aparatos es defectuoso. Calcule un intervalo de confianza del 90% para la proporción de televisores defectuosos en la población.

9. A.- Al estimar la media poblacional mediante un intervalo de confianza del 99% se obtiene: Límite inferior y superior 270 y 330 respectivamente. Si dichos valores se calcularon en función de 41 observaciones, ¿cuál es el valor de la media muestral y de la varianza muestral utilizados para determinar el intervalo?

9. B.- Si un valor de 52 tiene un margen de error de ± 5, ¿cómo están relacionados estos números con:

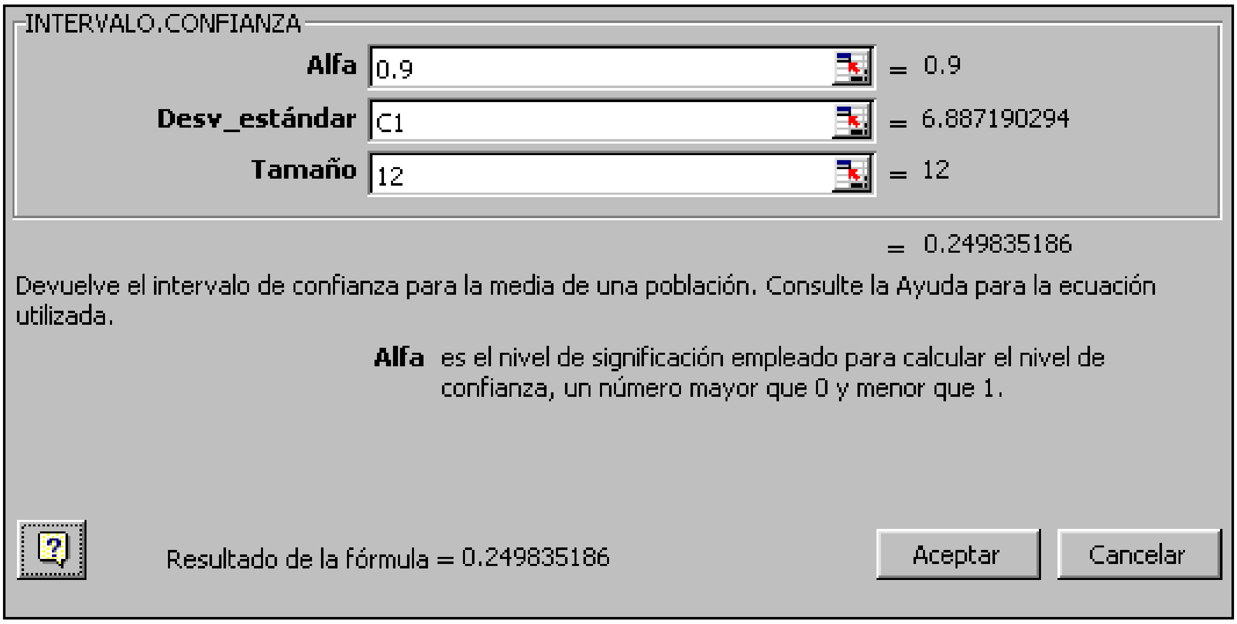
-La estimación puntual?
-El ancho del intervalo?
-Los límites del intervalo?

SALIDA DE PC

10. - Se sabe que la producción diaria de una refinería de petróleo está normalmente distribuida. De año en año la producción diaria usualmente se incrementa, pero la desviación estándar permanece constante.
La siguiente información representa la producción diaria de días seleccionados aleatoriamente durante 2001.

74.45	85.20	88.20	86.10	82.60	90.10	93.90	75.30	79.80	73.20	77.90	89.80

Se sabe de años anteriores que la desviación estándar es 0.9. Se encontró un intervalo de confianza del 90% para la producción media diaria del año 2001.
Calcule por su cuenta, compare con la salida siguiente e interprete.



PRUEBA DE CONCEPTO

1) Coloque una cruz en la variable que se emplea para construir un intervalo de confianza para la varianza poblacional σ^2.

☐ $= \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ ☐ $= \dfrac{X - \mu}{\sigma}$ ☐ $= \dfrac{(n-1)s^2}{\sigma^2}$ ☐ $= \dfrac{(n-1)\sigma^2}{\sigma^2}$

☐Ninguna de las anteriores

2) Señale la expresión correcta para referirse a la interpretación de un intervalo de confianza.
a) El parámetro está dentro del intervalo..........
b) El intervalo encierra al valor del parámetro.
c) El parámetro varía dentro del intervalo................

3) Indique la expresión correcta.
a) $P(10 < \mu < 25) = 0,95$
b) $P (\mu = 25) = 0,95$
c) $10 < \mu < 25$, con coeficiente de confianza 0,95.
d) $10 < \mu < 25$, con nivel de significación 0,95.
e) Ninguna de las anteriores.

4) Seleccione la respuesta correcta:

¿Cuál intervalo de confianza preferiría

$4,32 \pm 0,30$? $4,32 \pm 0,13$?

<u>**Nota**</u>: ambos intervalos están calculados con el mismo coeficiente de confianza.

5) Los límites de confianza del 95% para la media de una población son 20 y 30 kg. ¿Cuál de las siguientes afirmaciones es correcta y por qué?

1.-de 100 medias poblaciones extraídas al azar de la población considerada por Io menos 95 estarán entre 20 y 30 kg.-
.........porque..
2.-de 100 medias muestrales extraídas al azar de la población considerada por lo menos 95 estarán entre 20 y 30 kg.
.........porque..
3.-en el largo plazo el intervalo 20 a 30 kg. incluirá por lo menos en el 95% de los casos a la verdadera media poblacional
.........porque..

INFERENCIA ESTADÍSTICA : <u>PRUEBAS DE HIPÓTESIS (dócimas)</u>

Para poder resolver los problemas que siguen usted debe haber leído los temas desarrollados en clase y ser capaz de responder a la mayoría de las siguientes preguntas.

1) ¿Qué diferencias hay entre una hipótesis estadística y una hipótesis matemática?
2) Describa cómo se prueba una hipótesis estadística referida a un parámetro θ paso a paso.
3) ¿Qué diferencias hay entre el <u>**coeficiente de confianza**</u> utilizado en la estimación de parámetros por intervalos de confianza, y el <u>**nivel de significación**</u> ¿En qué se parecen (conceptualmente)?.
4) ¿Se puede “aceptar” una hipótesis estadística? Explique claramente su respuesta.
5) ¿Se puede “rechazar” una hipótesis estadística? Explique claramente su respuesta.
6) ¿Qué tipo de errores se encuentran presentes al plantear una hipótesis estadística? ¿Qué relación existe entre ellos?
7) ¿Cuál es la hipótesis a partir de la cual se toma la decisión, H_o ó H_1.?
8) ¿Cuándo se dice “ se tiene evidencias para …….” Se refiere a rechazar H_o y H_1.?

Prueba de hipótesis Ejercicios

11 - Dados los siguientes enunciados plantee las hipótesis nula y alternativa:
a.- Cierto tipo de cultivo de cereal siempre ha crecido hasta una altura promedio de 516 mm, con una desviación típica de 38 mm. Se siembra una muestra de 100 semillas y se aplican fertilizantes para tratar de mejorar su altura promedio. Plantee las hipótesis nula y alternativa.

b.- La duración media de las baterías producidas por una compañía nunca ha sido inferior a 1120 horas, con una desviación típica de 125 horas. Una muestra de 16 baterías arrojó una media de 1070 horas, con lo cual se sospecha que la duración de las baterías en el último año ya no es la misma. Plantee las hipótesis nula y alternativa.

c.- Existía una vez un país en el cual la mortalidad infantil nunca había superado el 0,2 % de la tasa de natalidad. Sin embargo políticos inescrupulosos lo llevaron a la bancarrota, con lo cual se sospecha que esto ha incidido desfavorablemente en el sistema de salud. Plantee las hipótesis nula y alternativa.

d.-. Un metalúrgico realizó 4 determinaciones del punto de fusión del manganeso y quiere probar si se está de acuerdo con el valor hipotético de 1260 grados. Plantee las hipótesis nula y alternativa.

12. - Se pide a un médico que practique un examen físico a un deportista a fin de comprobar que está en condiciones de continuar con la carrera que está disputando. Indique en qué condiciones el facultativo cometería un error de tipo I y en qué condiciones un error de tipo II. Analice.

13. - Históricamente la intención de voto de los electores para cierto partido político superaba el 40%. Una encuesta reciente reflejó que entre 1000 votantes, 342 lo harían por dicha agrupación política.
Plantee las hipótesis nula y alternativa.

14. - Nota: resuelva inciso por inciso en el orden que se pide **sin desarrollar los seis pasos de la dócima** en cada uno.
En una prueba de laboratorio sobre la medición de la tensión a la ruptura de una cuerda de cáñamo, se conoce que la varianza de dichas tensiones es de 25 (libras por pulgada cuadrada).
Tomando una muestra al azar de 4 elementos y a un nivel de significación del 5% se desea probar la hipótesis nula de que la verdadera tensión promedio de la cuerda es de 50 libras por pulgada cuadrada, contra la alternativa de que es menor.
a) Encuentre el punto crítico y las regiones de aceptación y de rechazo.
b) Suponiendo $\mu_1 = 45$, encuentre β y grafique α y β.
c) Suponga ahora que no conoce α, determínelo considerando como punto crítico el valor x_1 barra = 47 libras por pulgada cuadrada.
d) Relacione los tamaños de α y β, observe el gráfico y obtenga conclusiones.

15.- Se instala una máquina para construir rulemanes con 1 cm de radio. Una muestra de 10 rulemanes producidos por ésta máquina tiene un radio medio de 1,004 cm, con

$S^2=(0{,}003)^2$. ¿Hay alguna razón para sospechar que la máquina está produciendo rulemanes con un radio medio mayor de 1 cm?

16. - La desviación estándar de una cierto proceso de producción es de 4 pulgadas. Se sospecha que la varianza se ha hecho demasiado grande. Se toma una muestra de 7 partes producidas en dicho proceso y sus medidas son 5, 7, 2, 4, 8, 6 y 5 pulgadas.
Probar la hipótesis adecuada al nivel del 1%. Luego repetir el desarrollo usando α= 5%. Comparar, analizar.

17.- Una fábrica automotriz asegura que de cada 1000 vehículos que produce, 800 llegan a superar los 100.000 km recorridos sin necesidad de efectuar ajuste alguno en el motor.
Controladores del proceso de fabricación, sospechan que esto no es así. Toman una muestra aleatoria de 100 de estos automóviles y encuentran que 72 llegaron a cumplir con este requisito.
Con un nivel de significación del 1% ¿daría Ud. la razón al fabricante?

18.- Informes meteorológicos, indican que históricamente la media anual de precipitaciones caídas en cierta región europea, nunca había sido inferior a 255 mm/año. Sin embargo, sobre una muestra de 25 poblados de dicha región, se obtuvo que en el último año la precipitación media anual fue de 235 mm/año con una desviación típica de 35 mm/año.
Docime la hipótesis de interés al 1% de significación, e interprete en términos del problema.

19.- Dada la información siguiente ¿Cuál sería la conclusión al probar la hipótesis H_o: μ =10, contra cada una de las alternativas indicadas?

	N	$\bar{x}$	S^2	α	H_1
A	9	12	36	0.05	$\mu > 10$
B	16	13	64	0.05	$\mu \neq 10$
C	16	11	81	0.01	$\mu > 10$
D	25	8	64	0.01	$\mu < 10$
E	25	9	49	0.02	$\mu \neq 10$

INFERENCIA ESTADÍSTICA <u>Prueba de hipótesis : Bondad de ajuste</u>

Ejemplo de bondad de ajuste para variable discreta[1]: Se sabe que el número de defectos en las tarjetas de circuito impreso sigue una distribución de Poisson. Se reúne una muestra aleatoria de 60 tarjetas de circuito impreso y se observa el número de defectos. Los resultados obtenidos son los siguientes:

Número de defectos	Frecuencia observada
0	32
1	15
2	9
3 ó más	4

[1] Extraído de Probabilidad y Estadística Aplicadas a la Ingeniería. Montgomery Runger

¿Muestran estos datos suficiente evidencia para decir que provienen de una distribución Poisson? Haga la prueba con $\alpha = 0.05$.

Solución
H_0: La forma de la distribución de los defectos es Poisson.
H_1: La forma de la distribución de los defectos no es Poisson.

La media de la distribución Poisson propuesta en este ejemplo es desconocida y debe estimarse a partir de los datos contenidos en la muestra.

$$\mu = \lambda = \frac{(32)(0)+(15)(1)+(9)(2)+(4)(3)}{60} = 0.75$$

A partir de la distribución Poisson con parámetro 0.75, pueden calcularse las propiedades asociadas con el valor de X, utilizando:

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-0.75}0.75^x}{x!}$$

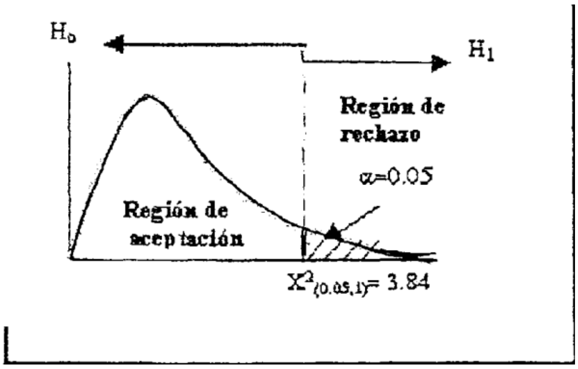
Se calculan las probabilidades y se multiplican por 60 para obtener los valores esperados. Por ejemplo, para encontrar: $E(0)=n.p(0)= 60* .472= 28.32$

Número de defectos	Probabilidad	Frecuencia esperada	Frecuencia observada
0	0.472	28.32	32
1	0.354	21.24	15
2	0.133	7.98	9
3 (o más)	0.041	2.46	4
		60	60

Observando la tabla de distribución de probabilidad de la variable, se tiene que la suma de las frecuencias esperadas coincide con la suma de las frecuencias observadas. (Analice por qué).
Si la frecuencia esperada en alguna celda es menor que 5, se combinan celdashasta lograr “5” o más.

Número de defectos	Frecuencia esperada	Frecuencia observada
0	28.32	32
1	21.24	15
2 ó más	10.44	13

Los grados de libertad se calculan tendiendo en cuenta el número de categorías o de valores de la variable, quitando uno como siempre (K-1). En este caso, los grados de libertad serán K —p — 1 = 3 – 1 – 1= 1, debido a que la media de la distribución Poisson fue estimada a partir de los datos.
Teniendo en cuenta que la región crítica es unilateral derecha (analice por qué) se tiene



Regla de decisión:

Si X_R^2 <3.84 no se rechaza H_0.

Si X_R^2 ≥3.84 se rechaza H_0

Para este ejemplo se tiene

$$\chi^2 = \sum_{j=1}^{n} \frac{\left(o_j - e_j\right)^2}{e_j} = 2.94$$

Justificación y decisión: Como 2.94 no es mayor que 3.84, no es posible rechazar la hipótesis nula y se concluye con un α= 0.05 que no hay evidencias para decir que la distribución de los defectos en las tarjetas de circuito impreso no es Poisson.

INFERENCIA ESTADÍSTICA : PRUEBAS DE HIPÓTESIS: USOS DE χ^2

CONSIGNA PARTICULAR:

- ✓ **PRIMERO:**

1.-Lea cada uno de los siguientes enunciados y en todos y cada uno explique cuál es la prueba que hará y por qué.
2.-Tenga en cuenta con cuántos grados de libertad deberá trabajar en cada caso y por qué. (conoce los parámetros????)

- ✓ **SÓLO A CONTINUACIÓN: Resuelva los ejercicios que se le señalen en clase.**

- ✓ **En todos los casos y sin necesidad de que se lo repitan:**
  - ❑ **Escriba correctamente las hipótesis**
  - ❑ **Desarrolle los pasos**
  - ❑ **RAZONE EN FÓRMULA Y EN PALABRAS UN VALOR ESPERADO**
  - ❑ **EXPLIQUE EN CADA EJERCICIO DÓNDE Y POR QUÉ COLOCA LA REGIÓN CRÍTICA**
  - ❑ **Tenga en cuenta con cuántos grados de libertad deberá trabajar y por qué. (conoce los parámetros????)**

20.- Una caja contiene gran cantidad de tornillos de 4 diferentes tamaños. Se supone que están en la proporción 8:7:3:2.
Una muestra de 400 tornillos contiene: 180 del tamaño 1, 120 del tamaño 2, 40 del tamaño 3 y 60 del tamaño 4.
a) Estime las proporciones de cada tipo de tornillos en la caja.
b) Verifique que los valores observados se asemejan a los estimados en a).

21.- Se probaron seis ejemplares, de un metro de longitud, de un cierto alambre aislado para encontrar los lugares de aislamiento insuficiente. El número de tales lugares fue 2, 0, 1, 1, 3, 2. El fabricante afirma que la calidad es tal que hay menos de 120 de dichos defectos por cada 100 metros. ¿Cumple el conjunto del cual proceden los ejemplares probados con lo afirmado por el fabricante para un nivel de significación del 0,05? (Use la distribución de Poisson).

Ejemplo de bondad de ajuste para variable continua [2] Los datos de la tabla siguiente son consumos diarios de agua en una curtiembre. Se desea ensayar la hipótesis de que responden a una distribución Normal.

X (miles de litros)	Días	x'_i	$x'_i f_i$	$x_i^2 * f_i$
[20 – 30)	6	25	150	3750
[30 – 40)	90	35	3150	110250
[40 – 50)	232	45	10440	469800
[50 – 60)	192	55	10560	580800
[60 – 70)	66	65	4290	278850
[70 – 80)	12	75	900	67500
[80 – 90)	2	85	170	14450
Total	600		29660	1525400

En primer lugar estimamos la media y la varianza

$$\bar{x} = \frac{\sum x' f}{n} = \frac{29660}{600} = 49{,}4 \text{ miles de litros}$$

$$s^2 = \frac{1}{n-1}\left(\sum x'^2 f - \frac{(\sum x' f)^2}{n}\right) = \frac{1}{599}\left(1525400 - \frac{29660^2}{600}\right) = 98.84$$

S= 9.94 miles de litros

Para calcular las frecuencias esperadas, conforme a la hipótesis nula, obtenemos primero las probabilidades de los intervalos, suponiendo el modelo Normal, y utilizando las estimaciones de sus parámetros. Por ejemplo:

$$P(20 \le x \le 30) = P\left(\frac{30-49.4}{9.94}\right) - P\left(\frac{20-49.4}{9.94}\right) = 0.0237$$

Se tiene así la siguiente tabla, en la que se han agregado los intervalos extremos (-∞, 20) y (90,∞) para contemplar todo el dominio teórico de la variable Normal. Las frecuencias esperadas se han calculado multiplicando por 600 sus probabilidades.

X (miles de litros)	Probabilidad	E_i
(-∞,20)	P(X<20)=0.0015	0.9
[20, 30)	P(20≤X≤30)=0.0237	14.22
[30, 40)	P(30≤X≤40)=0.1459	87.54
[40, 50)	P(40≤X≤50)=0.3516	210.96
[50, 60)	P(50≤X≤60)=0.3335	200.1
[60, 70)	P(60≤X≤70)=0.1245	74.7
[70, 80)	P(70≤X≤80)=0.0182	10.92
[80, 90)	P(80≤X≤90)=0.0010	0.6
[90, +∞)	P(X≥90)=0.0001	0.06
	1	600

A efectos de cumplir con el requerimiento de que todas las frecuencias esperadas sean mayores o iguales que 5 es necesario amalgamar intervalos, para obtener finalmente la siguiente tabla:

[2] Extraído de Inferencia Estadística y diseño de experimentos. Garcia, Roberto Mariano

X (miles de litros)	F	E_i
(-∞,30)	6	15.12
[30, 40)	90	87.54
[40, 50)	232	210.96
[50, 60)	192	200.1
[60, 70)	66	74.7
[70, +∞)	14	11.58

El nro de grados de libertad =nro de intervalos – nro de parámetros estimados – 1
Nro grados libertad= 6-2-1=3. Si se considera α=0.05, el valor crítico para la prueba es $\chi^2_{v,1-\alpha}=\chi^2_{3,0,95}=7.81$ en tanto que el estadístico de prueba arroja el valor 9.55, por lo tanto, la hipótesis nula se rechaza y podemos asegurar, con probabilidad máxima de equivocarnos del 5%, que esta variable no tiene distribución Normal.

22.-En un estudio sobre dosificación de mezclas de hormigón, se seleccionaron dos muestras de 200 ensayos de asentamiento cada una, para comparar dos métodos empíricos (A yB), donde se midió si el asentamiento era bueno o malo.

Asentamiento	Malo	Bueno
Método A	43	157
Método B	81	119

a) Plantee las hipótesis de interés al 5% de significación. No calcule. El valor calculado es: χ^2_{calc}.= 16,87.
b) Explique cómo calcula un valor esperado (en fórmulas y en palabras)
c) Grafique la región crítica y explique claramente por qué la dibuja así. Relacione con la "forma" de la expresión analítica de la variable chi cuadrado aproximada.

23.- Si en el problema anterior usted hubiera considerado que se tienen 400 ensayos que luego se clasifican de acuerdo con el método empleado y según que su asentamiento sea bueno o malo, ¿en qué hubiera cambiado su razonamiento? ¿Qué se mantiene y qué cambia y por qué?

24.- Se toman muestras aleatorias de 50 operarios de cuatro centros laborales y se confecciona la siguiente tabla de contingencia, donde se diferencia a los obreros que trabajan horas extras:

Centros Laborales	A1	A2	A3	A4
No trabajan horas extras	30	35	40	45
Trabajan horas extras	20	15	10	05

¿Se puede suponer con un nivel de significación del 5% que la proporción de obreros que trabajan horas extra es la misma en los cuatro centros?
Use χ^2_{calc}.= 13,33.

25.- Se tomó una muestra al azar de 1000 jóvenes y a cada miembro de la muestra se lo clasificó de acuerdo a su nivel intelectual y según el número de libros leídos durante el último año, obteniéndose:

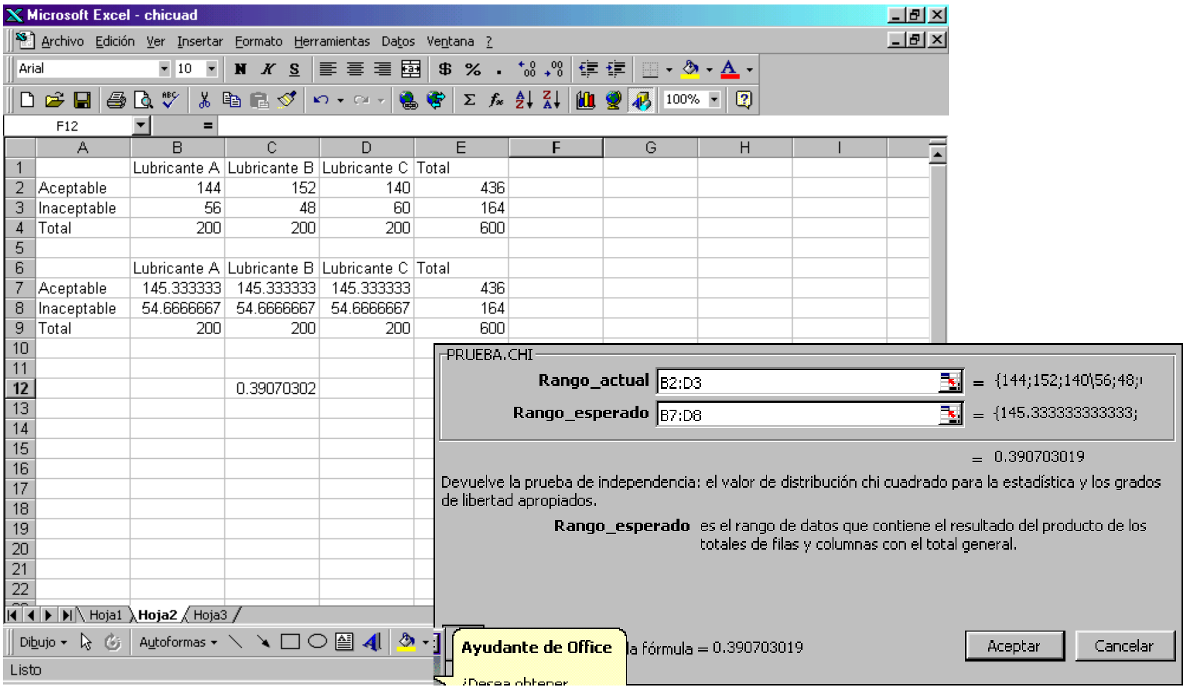
	Nivel intelectual			
Libros leídos	**12 años ó menos**	**más de 12 años pero menos de 16**	**16 años o más**	**Totales**
Ninguno	330	50	20	**400**
1	50	100	50	**200**
2	100	150	50	**300**
3 ó más	20	30	50	**100**
Totales	**500**	**330**	**170**	**1000**

Probar la hipótesis que considere de interés al nivel del 5%.
a) Justifique y plantee la hipótesis seleccionada.
b) Indique la variable pivotal.
c) Indique la región crítica (justifique) y plantee la regla de decisión.
d) Explique cómo obtiene el valor esperado de la celda (1,1).
e) Suponga que el valor calculado es 344,65. Decida y concluya.

26.- Mediante un nuevo proceso se prepararon tres clases de lubricantes.
Cada uno de los lubricantes se prueba con cierto número de máquinas, y el resultado es luego clasificado como aceptable o no.
Los datos representan los resultados de este experimento.

	Lubricante 1	**Lubricante 2**	**Lubricante 3**
Aceptable	144	152	140
Inaceptable	56	48	60

a) Docime la hipótesis de interés al nivel del 5% y compare con la salida de pc que está a continuación.
b) ¿Qué test utilizó?. Justifique.



	A	B	C	D	E
1		Lubricante A	Lubricante B	Lubricante C	Total
2	Aceptable	144	152	140	436
3	Inaceptable	56	48	60	164
4	Total	200	200	200	600
5					
6		Lubricante A	Lubricante B	Lubricante C	Total
7	Aceptable	145.333333	145.333333	145.333333	436
8	Inaceptable	54.6666667	54.6666667	54.6666667	164
9	Total	200	200	200	600
12			0.39070302		

27.- Para poder competir en el mercado; una empresa productora de un nuevo sistema de construcción de viviendas realiza una encuesta piloto sobre la preferencia del público en las distintas zonas del país, tomando muestras proporcionales al número de habitantes. Los resultados son los siguientes

Preferencia	**Zona A**	**Zona B**	**Zona C**	**Zona D**
Nuevo sistema	35	22	28	33
Sist. tradicional	45	28	12	17

¿Se puede concluir que existe la misma relación entre la preferencia del público y las zonas geográficas a un nivel del 5%?