

# **Cátedra ESTADISTICA**

## **TRABAJOS PRÁCTICOS**

(SEGUNDA ETAPA)

**2020**

**Facultad de Ingeniería**

**Universidad Nacional de La Patagonia S. J. B.**

**Comodoro Rivadavia**



## **TRABAJO PRÁCTICO Nº 6**

**INFERENCIA ESTADÍSTICA-Prueba de Hipótesis no paramétricas**

### **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.

### **INFERENCIA ESTADÍSTICA Prueba de hipótesis: Bondad de ajuste**

**Ejemplo de bondad de ajuste para variable discreta<sup>1</sup>:** 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

<sup>1</sup> 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 se ajusta a una distribución de Poisson.

$H_1$ : La forma de la distribución de los defectos no se ajusta a una distribución de 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.

$$\lambda = \frac{(32)(0) + (15)(1) + (9)(2) + (4)(3)}{60} = 0,75$$

A partir de la distribución Poisson con parámetro 0.75, pueden calcularse las probabilidades asociadas con cada 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:  $E(0) = n \cdot p(0) = 60 * 0,472 = 28,32$

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

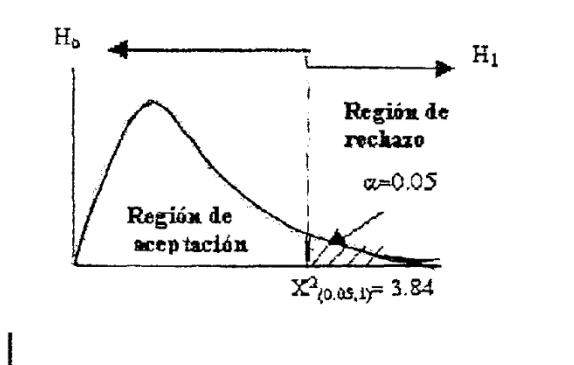
Observando la tabla de distribución de probabilidad de la variable, se tiene que la suma de las frecuencias esperadas coincide con la suma de las frecuencias observadas. (Analice por qué).

Si la frecuencia esperada en alguna celda es menor que 5, se combinan celdas hasta 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 teniendo 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  $K - r - 1 = 3 - 1 - 1 = 1$ , 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 ubica allí.



Regla de decisión:

Si  $X_R^2 < 3,84$  no se rechaza  $H_0$ .

Si  $X_R^2 \geq 3,84$  se rechaza  $H_0$

Para este ejemplo se tiene

$$\chi^2 = \sum_{j=1}^n \frac{(o_j - e_j)^2}{e_j} = 2,94$$

Justificación y decisión: Como 2,94 no es mayor que 3,84, no es posible rechazar la hipótesis nula y se concluye con un  $\alpha = 0,05$  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 $\chi^2$

### CONSIGNA PARTICULAR:

Lea cada uno de los siguientes enunciados y en todos y cada uno explique cuál es la prueba que hará y por qué.

✓ Resuelva los ejercicios que se le señalen en clase.

✓ En todos los casos

- ❑ Desarrolle los pasos de la dócima
- ❑ 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?)

**Ejemplo de bondad de ajuste para variable continua** <sup>2</sup> 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	$x'_i$	$x'_i f_i$	$x_i^2 \cdot f_i$
[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

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

<sup>2</sup> Extraído de Inferencia Estadística y diseño de experimentos. Garcia, Roberto Mariano

$$s^2 = \frac{1}{n-1} \left( \sum x^2 f - \frac{(\sum x'f)^2}{n} \right) = \frac{1}{599} \left( 1525400 - \frac{29660^2}{600} \right) = 98,84$$

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:

$$P(20 \leq x \leq 30) = P\left(\frac{30 - 49,4}{9,94}\right) - P\left(\frac{20 - 49,4}{9,94}\right) = 0,0237$$

Se tiene así la siguiente tabla, en la que se han agregado los intervalos extremos  $(-\infty, 20)$  y  $(90, \infty)$  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 <sub>i</sub>
$(-\infty, 20)$	$P(X < 20) = 0,0015$	0,9
$[20, 30)$	$P(20 \leq X \leq 30) = 0,0237$	14,22
$[30, 40)$	$P(30 \leq X \leq 40) = 0,1459$	87,54
$[40, 50)$	$P(40 \leq X \leq 50) = 0,3516$	210,96
$[50, 60)$	$P(50 \leq X \leq 60) = 0,3335$	200,1
$[60, 70)$	$P(60 \leq X \leq 70) = 0,1245$	74,7
$[70, 80)$	$P(70 \leq X \leq 80) = 0,0182$	10,92
$[80, 90)$	$P(80 \leq X \leq 90) = 0,0010$	0,6
$[90, +\infty)$	$P(X \geq 90) = 0,0001$	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:

X (miles de litros)	F	E <sub>i</sub>
$(-\infty, 30)$	6	15,12
$[30, 40)$	90	87,54
$[40, 50)$	232	210,96
$[50, 60)$	192	200,1
$[60, 70)$	66	74,7
$[70, +\infty)$	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  $\alpha=0.05$ , 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.

1.- 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
<b>Aceptable</b>	144	152	140
<b>Inaceptable</b>	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.

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	Lubricante A	Lubricante B	Lubricante C	Total
1				
2	144	152	140	436
3	56	48	60	164
4	200	200	200	600
5				
6				
7	145.3333333	145.3333333	145.3333333	436
8	54.6666667	54.6666667	54.6666667	164
9	200	200	200	600

The dialog box 'PRUEBA, CHI' displays the following information:

- Rango\_actual: B2:D3 = {144;152;140;56;48;}
- Rango\_esperado: B7:D8 = {145.333333333333;}
- Devuelve la prueba de independencia: el valor de distribución chi cuadrado para la estadística y los grados de libertad apropiados.
- Rango\_esperado es el rango de datos que contiene el resultado del producto de los totales de filas y columnas con el total general.
- La fórmula = 0.390703019

2.- El gerente de una empresa de elaboración de productos químicos desea determinar si el número de empleados que asisten al consultorio médico de la planta se encuentra distribuido en forma equitativa durante los 5 días de trabajo de la semana. Para ello toma una muestra aleatoria de 4 semanas completas de trabajo y obtiene el siguiente número de consultas:

Lunes	Martes	Miércoles	Jueves	Viernes
49	35	32	39	45

¿Existe alguna razón para creer que el número de empleados que asisten al consultorio médico, no se encuentra distribuido de forma equitativa durante los días de la semana? Use  $\alpha=0,05$

3.-Se busca establecer si la calidad de productos obtenidos por tres métodos es similar, o si alguno de los métodos produce calidad diferente. La calidad se califica como “Alta”, “Media” y “Rechazo”. Se toman tres muestras, una por cada método, y se obtienen los siguientes resultados:

	ALTA	MEDIA	RECHAZO	TOTAL
METODO A	180	15	5	
METODO B	152	45	3	
METODO C	125	72	3	
TOTAL				

a) Realice la prueba adecuada para el objetivo del problema. Decida considerando  $\chi^2$  calculado = 47,61 y  $\alpha=0,05$ .

b) Dibuje la distribución, ubique el valor crítico, señale la región crítica y fundamente en palabras por qué la ubica allí.

c) Concluya en términos estadísticos y en términos del problema.

d) Diga con palabras y con fórmulas cómo se calcula el valor esperado **“METODO B, CALIDAD MEDIA”**.

**4.-** Durante un período de 100 semanas, se observó el número semanal de averías de una máquina y se registraron los datos en la siguiente tabla. Sabiendo que el número semanal medio de averías es de 2,1. ¿Con que distribución de probabilidad discreta se podrían ajustar los datos? Realice la dócima con  $\alpha = 0,05$ . Concluya en términos del problema.

Numero de averías	0	1	2	3	4	5 o más
Numero de semanas	10	24	32	23	6	5

**5.-** En un estudio socioeconómico se divide un grupo de mujeres de acuerdo a su estado civil y luego se les consulta sobre su inserción laboral. Los resultados son los siguientes:

ESTADO CIVIL	INSERCIÓN LABORAL		TOTAL
	TRABAJA	NO TRABAJA	
Casada	45	131	176
Soltera	83	127	210
TOTAL	128	258	386

- ¿Qué tipo de prueba elegiría en esta situación? Fundamente su elección.
- Realice la prueba, mostrando todos sus pasos. Decida considerando  $\chi^2_{\text{calculado}}=8,41$ . Use nivel de significación de 0,05.
- ¿Cómo encuentra un valor esperado? Ejemplifique para Soltera y Trabaja. Muestre el desarrollo teórico y sus cálculos.
- Explique en palabras dónde ubicó la región crítica y por qué.

**6.-** Un ingeniero de control de calidad investiga el cumplimiento de 4 estándares de calidad en sus productos, estableciéndose una probabilidad igual de que cumpla o no cualquiera de los estándares. Desea probar con un nivel de significación de 0,05 que la muestra proviene de una distribución binomial. Para ello toma una muestra aleatoria de 200 productos con los siguientes resultados:

Número de estándares que cumple	Frecuencia observada
0	18
1	51
2	70
3	32
4	9
Total	200

**7.-** La siguiente tabla muestra los resultados de un estudio del tránsito de la ciudad, relacionado al uso del cinturón de seguridad. Con tal fin se toma una muestra de los controles de rutina, y los datos se muestran a continuación.

	Femenino	Masculino
Rara vez o nunca usa cinturón de seguridad	208	324
Usa cinturón de seguridad	1217	1184

Responda:

- ¿Qué tipo de prueba elegiría? Fundamente su elección.
- Realice la prueba, mostrando todos sus pasos. Use nivel de significación de 0,01.

- c- ¿Cómo encuentra un valor esperado? Ejemplifique para Femenino y Rara vez o nunca usa el cinturón de seguridad. Muestre el desarrollo teórico y sus cálculos.
- d- Explique por qué señaló la región crítica donde lo hizo.

**8.-** En una planta de ensamblaje de camiones, la supervisión quiere saber si hay relación entre la calidad de la soldadura y los turnos. Para esto se toma una muestra de 1732 soldadores y obtiene la siguiente tabla:

	Alta	Media	Baja
Turno día	471	178	54
Turno tarde	458	154	32
Turno noche	257	111	17

- a- Diga qué tipo de prueba realizaría. ¿Por qué?
- b- Plante la hipótesis de interés. Use  $\alpha = 0,05$
- c- Explique cómo calcula un valor esperado, en fórmulas y en palabras.
- d- Cuántos grados de libertad usa? Explique cómo los calcula.
- e- Grafique la región crítica y explique por qué la dibuja allí
- f- Use  $X^2 = 9,4668$ ; concluya e interprete en términos del problema.

**9.-** Una fábrica de fertilizantes está desarrollando un nuevo producto y quiere saber si diferentes concentraciones de éste muestran diferencias en la forma de crecimiento de las plantas. Para ello tratan a 378 plantas de álamo y se confecciona la siguiente tabla:

	Crecimiento		Total
	Mayor	Normal	
Concentración alta	138	49	187
Concentración media	39	41	80
Concentración baja	35	76	111
Total	212	166	378

- a) ¿A qué prueba corresponde el enunciado del ejercicio?. Justifique.
- b) Plantee los primeros cuatro pasos teóricos de la prueba.
- c) Razone y explique con palabras cómo obtiene un valor esperado (en fórmulas y palabras).
- d) **NO CALCULE.** Decida considerando  $\chi^2_{\text{calculado}} = 52.736$  y  $\alpha = 0,05$
- e) Concluya en términos del problema.
- f) Explique en palabras dónde ubicó la región crítica y por qué

**10.-** Una fábrica de fertilizantes está desarrollando un nuevo producto y quiere saber si diferentes concentraciones de este tienen distinto efecto en el crecimiento del álamo. Para ello se forman tres grupos de plantas aplicándoles una concentración de fertilizante diferente a cada grupo y se confecciona la misma tabla que en el ejercicio anterior.

- a) Lea el enunciado del ejercicio. ¿A qué prueba corresponde el enunciado del ejercicio? Justifique. ¿Qué diferencia encuentra con el enunciado del ejercicio anterior?
- c) Razone y explique con palabras cómo obtiene un valor esperado (en fórmulas y palabras).

---

**PRUEBA DE CONCEPTOS**

1.- Dadas las hipótesis

$H_0$ : el estado civil de las personas es independiente de la situación laboral.

$H_1$ : el estado civil de las personas no es independiente de la situación laboral.

Represente gráficamente la variable pivotal y marque la región crítica.

2.- En una prueba de bondad de ajuste los grados de libertad de la variable pivotal se calculan como:

☐  $n-1$

☐  $k-1$

☐  $(f-1)(c-1)$

☐  $n.p$

☐  $k-1-r$

3.- Señale la respuesta correcta. ¿Cuántos grados de libertad se requieren en una prueba de independencia basada en una tabla de contingencia de 5 columnas y 4 filas?

☐ 2

☐ 7

☐ 12

☐ 20