Universidad Nacional de La Plata

Facultad de Ciencias Agrarias y Forestales

CÁLCULO ESTADÍSTICO Y BIOMETRÍA

Prueba de Hipótesis

Hasta ahora hemos visto pruebas de hipótesis utilizando las estadísticos "Z", "t", χ^2 y en todos los casos, las pruebas han coincidido en las siguientes características:

- se han postulado hipótesis referidas a parámetros.

- se ha establecido el cumplimiento de supuestos con relación a las distribuciones de las poblaciones originales de donde se han extraído las muestras.

- para establecer las hipótesis se propuso una distribución teórica subyacente para explicar el comportamiento del estadístico de prueba.

$\Downarrow$

Pruebas paramétricas

Prueba de Hipótesis

Hay otro grupo de pruebas de hipótesis que se caracterizan por las siguientes características:

- las hipótesis hacen referencia a la forma de la distribución de los datos.

- se utilizan cuando disponemos de datos expresados en frecuencias.

- no requieren el cumplimiento de supuestos con relación a las distribuciones de las poblaciones de donde se han extraído las muestras.

- el estadístico de prueba es χ^2

⬇

Pruebas no paramétricas

Pruebas de Chi-cuadrado

Las pruebas conocidas como chi-cuadrado utilizan el estadístico que diseñó Karl Pearson en 1899. Es un índice que mide la desviación de las frecuencias observadas en una muestra respecto a las esperadas bajo una hipótesis.

i	$f_{observada}$	$f_{esperada}$
1	f_1	$\hat{f}_1$
2	f_2	$\hat{f}_2$
3	f_3	$\hat{f}_3$
.	.	.
.	.	.
.	.	.
.	.	.
k	f_k	$\hat{f}_k$

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(f_i - \hat{f}_i\right)^2}{\hat{f}_i}$$

$$\boxed{\chi^2 = \sum_{i=1}^{k} \frac{\left(o_i - e_i\right)^2}{e_i}}$$

Otra fórmula de cálculo: $\chi^2 = \sum_{i=1}^{k} \frac{o_i^2}{e_i} - n$

Pruebas de Chi-cuadrado

Pearson estudió la distribución teórica bajo H_0 verdadera. Algunas características son:

1. Tiene distribución Chi-cuadrado con k-1 grados de libertad

2. Si el valor del estadístico es 0 (cero) significa que no hay diferencias entre las frecuencias observadas y esperadas. Cuanto más se aleja de 0 mayor la discrepancia entre las frecuencias

3. Si las frecuencias esperadas < 5 el estadístico se aleja de la distribución Chi-Cuadrado por lo que se deben reagrupar intervalos o aumentar el numero de individuos estudiados.

4. Si la frecuencia total es muy pequeña (sobre todo para el caso de 1 grado de libertad) es aconsejable introducir el factor de corrección de Yates al calcular el estadístico.

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(\left|o_i - e_i\right| - 0{,}5\right)^2}{e_i}$$

Clasificación de Pruebas no paramétricas

I) Pruebas para la bondad de ajuste

- Para determinar si los datos provienen de una población que tiene una **distribución de probabilidad (proporciones) específica**, o

- Para determinar si los datos provienen de una población que tiene una **distribución de probabilidad conocida**.

II) Pruebas de independencia

III) Pruebas de homogeneidad de muestras

I. Pruebas para la bondad de ajuste

Se parte de una distribución de frecuencias empíricas (frecuencias observadas) para una muestra aleatoria, y luego se procede a ajustar un modelo probabilístico que se selecciona pensando que explica el comportamiento de la variable de interés. En este caso la distribución de frecuencia esperadas (teóricas) se calcula como:

$$\boxed{\hat{f}_i = f_{esperada} = P_{x_i} \times N}$$

Siendo: P_{xi} = la probabilidad bajo el supuesto de que H_o es verdadera

N = el tamaño muestral o total de frecuencias observadas

El objetivo de aplicar este tipo de prueba es comprobar que no existen discrepancias significativas entre las frecuencias esperadas y las observadas, o sea que se trata de probar que el ajuste realizado resulta apropiado.

Planteo de Hipótesis

Distribución de proporciones (Modelo genético):

H_o: ‘la segregación de cierto poroto responde a la teoría de Mendel 9:3:3:1’

(Lisos Amarillos – Rugosos Amarilos – Lisos Verdes – Rugosos Verdes)

H_o: ‘la relación de moscas ojos blancos respecto a moscas ojos rojos es 1 a 3’

Distribución de probabilidad conocida:

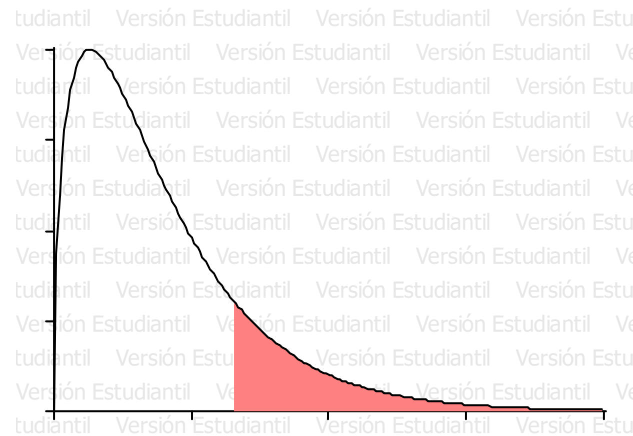
H_o: ‘la variable X sigue una distribución Binomial’

H_o: ‘la variable X sigue una distribución Poisson’

H_o: ‘la variable X sigue una distribución Normal’

Región de rechazo y regla de decisión

Prueba de hipótesis	**gl**
Distr. de proporciones	**n -1**
Distr. de probabilidad conocida	**n-k-1** Binomial y Poisson k=1 Normal k=2



χ^2 critico (1-α;gl)

Cálculo de Chi-cuadrado

X_i	f_{obs}	P_i	f_{esp}
x_1	f_1	p_1	$\hat{f}_1$
x_2	f_2	p_2	$\hat{f}_2$
x_3	f_3	p_3	$\hat{f}_3$
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
x_k	f_k	p_k	$\hat{f}_k$

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

Ejercicio 1

De acuerdo a la teoría mendeliana de la herencia, cuando se cruzan plantas de arveja (*Pisum sativum*), puede esperarse con relación a la herencia de las características textura del tegumento y el color del grano que se presente una interrelación de 9:3:3:1.

En una experiencia se obtuvieron los siguientes datos:

Fenotipo	**Amarilla-lisa**	**Amarilla-rugosa**	**Verde-lisa**	**Verde-rugosa**	**Total**
N° de semillas	315	101	108	32	556

Establecer las hipótesis nula y alternativa y realizar el contraste correspondiente utilizando un $\alpha=0.05$, a fin de verificar si hay concordancia de los valores observados con las leyes de herencia.

Ejercicio 2

De un monte de cerezos atacado por pulgón verde, un técnico fruticultor ha extraído una muestra aleatoria de 100 hojas. Examinado el material recolectado, se han encontrado los siguientes resultados:

N° de pulgones /hoja	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
N° de hojas	39	21	18	9	5	4	3	1

Establecer las hipótesis nula y alternativa y realizar el contraste correspondiente utilizando un $\alpha=0.05$, a fin de verificar si la variable bajo estudio se ajusta a la Distribución de Poisson.

II. Pruebas de independencia de dos factores

Determina si dos factores que clasifican a una población o muestra son estadísticamente independientes. Es decir, los niveles de un determinado factor no afectan a los niveles del otro factor. Es la Hipótesis general de una tabla de contingencia y el principio de análisis de datos categorizados, lo que implicaría que las frecuencias observadas son iguales a las frecuencias esperadas.

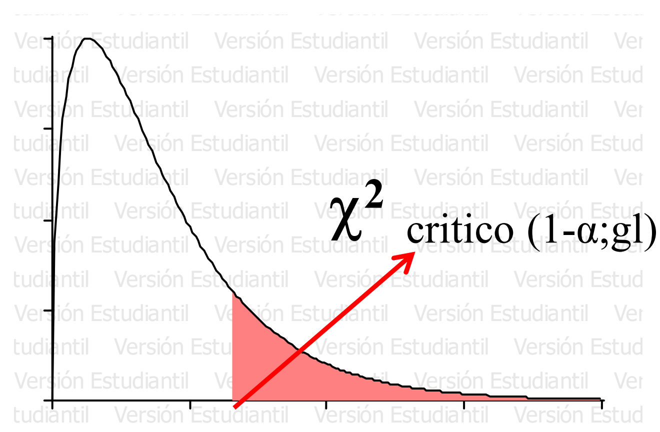
Planteo de Hipótesis

H_o: 'los criterios de clasificación son independientes'

H_1: 'los criterios de clasificación son dependientes'

Región de rechazo y regla de decisión

$$\boxed{gl = (f - 1) \times (c - 1)}$$



Cálculo de Chi-cuadrado

Frecuencias observadas en cada celda por combinación de dos niveles de factores

Factor A

Factor B

	A_1	A_2	...	A_a	
B_1	f_{11}	f_{12}	...	f_{1a}	$f_{1.}$
B_2	f_{21}	f_{22}	...	f_{2a}	$f_{2.}$
...	...	...	f_{ij}	...	...
B_b	f_{b1}	f_{b1}	...	f_{ba}	$f_{b.}$
	$f_{.1}$	$f_{.2}$	...	$f_{.a}$	$f_{..}$

Marginales fila

Total fila

Marginales columna

Total columna

Total General

Cálculo de Chi-cuadrado

	A_1	A_2	...	A_a	
B_1	f_{11} / $\hat{f}_{11}$	f_{12}	...	f_{1a}	$f_{1.}$
B_2	f_{21}	f_{22}	...	f_{2a}	$f_{2.}$
...	...	...	...	...	...
B_b	f_{b1}	f_{b1}	...	f_{ba}	$f_{b.}$
	$f_{.1}$	$f_{.2}$	...	$f_{.a}$	$f_{..}$

Bajo la hipótesis nula, las frecuencias esperadas en una tabla de contingencia se pueden obtener como producto de las frecuencias marginales observadas.

Es decir que bajo independencia de factores se cumple que:

$$\frac{f_{11}}{f_{.1}} = \frac{f_{12}}{f_{.2}} = \frac{f_{1a}}{f_{.a}} = \ldots = \frac{f_{1.}}{f_{..}} \Rightarrow \hat{f}_{11} = \frac{f_{.1} \times f_{1.}}{f_{..}}$$

$$\boxed{\chi^2_{obs} \sim \chi^2_{[(a-1).(b-1)]}}$$

Ejercicio 3

Se quiere evaluar si la germinación o no de semillas está asociada a la condición de haber sido tratadas con un fungicida. Para ello dos lotes de tamaño similar, fueron tratadas con fungicida o dejadas como control no tratadas. Luego las semillas se hicieron germinar y se registró el número de germinadas y no germinadas en cada uno de los grupos: control y tratadas con fungicida. El resultado de este conteo se presenta a continuación:

Condición	no germinó	germinó	Total
Control	245	1190	**1435**
Fungicida	123	1358	**1481**
Total	**368**	**2548**	**2916**

Establecer las hipótesis de trabajo y realizar el contraste correspondiente utilizando un $\alpha=0.05$, a fin de verificar si hay evidencia estadísticamente significativa de que el uso de un fungicida mejora el poder germinativo.