Bioestadística – Laboratorio

**MBIO2401: Laboratorio (1 crédito)**

*Profesora del Laboratorio:* Luisa A. Castellanos.

*Profesor del Curso:* Andrew J. Crawford.

*Monitora:* Valentina Muñoz

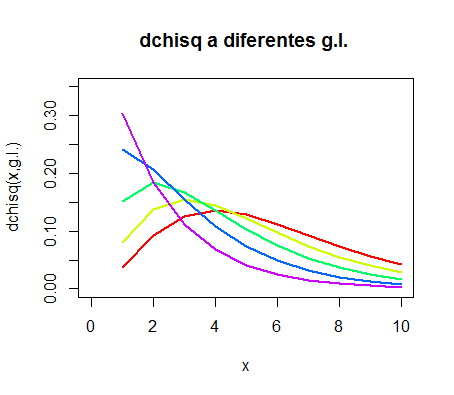
Departamento de Ciencias Biológicas, Universidad de los Andes

**Lab 11 – Prueba de bondad de ajuste y Análisis de contingencia**

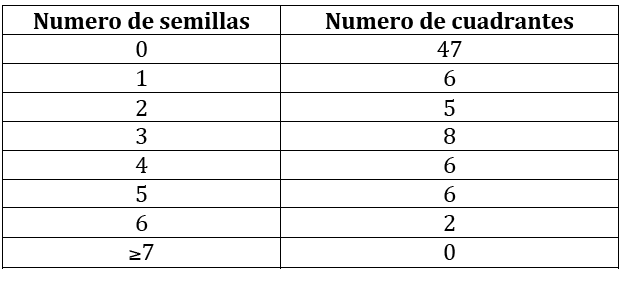
**La asistencia al laboratorio es obligatoria y por lo tanto un requisito para presentar este taller**

Recuerde plantear tanto su hipótesis científica como las hipótesis estadísticas (H0 y H1) en cada caso. Las respuestas de los ejercicios deben incluir el estadístico de prueba así como el Valor-P y el nivel de significancia que se va a usar.

1. [1 punto] Realice una gráfica donde muestre como cambia la distribución de probabilidad para valores de chi-cuadrado desde 1 hasta 10 a medida que van disminuyendo los grados de libertad desde 6 hasta 2. Puede incluir todas las distribuciones en la misma gráfica (una línea para cada distribución de probabilidad) o hacer unagráfica con varios paneles donde se muestren todas las distribuciones.



Notemos como a medida que aumentan los grados de libertad, la distribución se aplana.

1. [2 puntos] El estudio de la distribución espacial de vegetación frecuentemente hace uso de muestras aleatorias de “cuadrantes”, los cuales tienen un tamaño fijo y son colocados al azar en la región de muestreo. En uno de estos estudios, un investigador contó el numero de semillas de pino blanco que crecían en 80 cuadrantes de 10m\*10m, el quería saber si la distribución de las semillas de pino en el bosque era aleatoria, agrupada o dispersa. El investigador obtuvo los siguientes resultados:
2. Si la hipótesis nula de distribución aleatoria de semillas de pino a través del bosque es correcta, ¿Cuál es la distribución de probabilidad teórica que debería explicar las frecuencias en los cuadrantes?

Si la distribución de las semillas fuera verdaderamente aleatoria, habría la misma probabilidad de obtener cada uno de los números de las semillas. La distribución de probabilidad sería la siguiente:

|  |  |  |
| --- | --- | --- |
| **Numero de Semillas** | **Pr(x)** | **Numero de cuadrantes Esperados** |
| 0 | 1/8 | 10 |
| 1 | 1/8 | 10 |
| 2 | 1/8 | 10 |
| 3 | 1/8 | 10 |
| 4 | 1/8 | 10 |
| 5 | 1/8 | 10 |
| 6 | 1/8 | 10 |
| >=7 | 1/8 | 10 |

1. Pruebe si la hipótesis nula anterior es correcta.

Como se cumplen todos los supuestos de la prueba de chi cuadrado, podemos hacer una prueba de chi cuadrado:

> data <-scan()

1: 47

2: 6

3: 5

4: 8

5: 6

6: 6

7: 2

8: 0

9:

Read 8 items

> probs=rep(1/8,8)

> chisq.test(data,p=probs)

Chi-squared test for given probabilities

data: data

X-squared = 161, df = 7, p-value < 2.2e-16

Como alfa<0.05, rechazamos H0 y afirmamos que el numero de semillas NO esta distribuido de manera aleatoria.

1. [2 puntos] Las ventanas de los edificios matan más aves que cualquier otro factor relacionado con humanos. Solo en Norteamérica, entre 100 millones y 1 billón de aves mueren por choques contra ventanas. Este escenario representa un poco más del 5% del número total de aves en el área. Una posible solución a este problema es colocar las ventanas en un determinado ángulo, así estas reflejaran el suelo más que una imagen del cielo a las aves en vuelo. Un experimento fue realizado para comparar el numero de aves que murieron como resultado de una ventana vertical, una ventana con un ángulo de inclinación de 20° con respecto a la vertical y una ventana con un ángulo de inclinación de 40° con respecto a la vertical (Klem *et al.*, 2004). Los ángulos fueron asignados aleatoriamente a seis ventanas idénticas y en esta asignación, el tipo de ventana fue variado diariamente de forma aleatoria por cuatro meses. Durante el transcurso del experimento, 30 aves murieron por choque con ventanas en orientación vertical, 15 por choque con ventanas en ángulo de 20° y 8 por choque con ventanas en ángulo de 40°.
2. Establezca la hipótesis nula y alternativa más apropiada para que los investigadores contesten su pregunta.
3. Proponga una prueba estadística para probar esta hipótesis y realícela. ¿A qué conclusiones llegó? ¿Son o no efectivas las ventanas inclinadas?
4. Halle el Valor- P para estos datos. Utilice el método directo, no use chisq.test( ).
5. **Bono: [1 punto]** Creen una función que le permita encontrar los valores esperados para una tabla de contingencia de 2 × 2.

aves

sexo

Tipo\_cruce machos hembras

hibrido 16 10

puro 72 73

> expected<-function(matrix){

+ tj1<-matrix[1]+matrix[3];

+ tj2<-matrix[2]+matrix[4]

+ ti1<-matrix[1]+matrix[2]

+ ti2<-matrix[3]+matrix[4]

+ tot<-matrix[1]+matrix[2]+matrix[3]+matrix[4]

+ x1<-c(tj1\*ti1/tot,ti2\*tj1/tot,tj2\*ti1/tot,ti2\*tj2/tot)

+ return(x1)}

> expected(aves)

[1] 13.38012 12.61988 74.61988 70.38012

1. **[4 puntos]** En la sección de contenidos, en Sicua-Plus encontrará un archivo de texto con el nombre **‘Datos\_contingencia’.** Este archivo corresponde a los resultados de una encuesta realizada a 170 jóvenes entre los 19 y 25 años de edad e incluye respuestas a varias preguntas sobre aspectos de la vida de estos jóvenes.

El taller del día de hoy consiste en plantear, con base a estos datos, **3 preguntas** que puedan ser probabas con un análisis de contingencia.

Deben incluir:

* La hipótesis nula y alternativa para cada una de las preguntas.
* La evaluación de los supuestos con el comando creado en el punto 1 (mínimo para una de las 3 preguntas y para todas las preguntas que usen tablas de contingencia de 2 × 2).
* El comando y la matriz creada en ‘R’.
* El comando y la respuesta de la prueba estadística.

**1ra pregunta**

La hipótesis nula y alternativa para cada una de las preguntas.

H0: El consumo de drogas es independiente al consumo de tabaco

Ha: El consumo de drogas NO es independiente al consumo de tabaco

La evaluación de los supuestos con el comando creado en el punto 1 (mínimo para una de las 3

preguntas y para todas las preguntas que usen tablas de contingencia de 2 × 2).

> expected(fumDrog)

[1] 39.98824 63.01176 26.01176 40.98824

Todos los valores esperados se encuentran por encima de 5.

El comando y la matriz creada en ‘R’.

> vect<-c(length(which(data$Fuma=="SI" & data$Consumo\_drogas=="SI")),length(which(data$Fuma=="SI" & data$Consumo\_drogas=="NO")),length(which(data$Fuma=="NO" & data$Consumo\_drogas=="SI")),length(which(data$Fuma=="NO" & data$Consumo\_drogas=="NO")))

> fumDrog<-matrix(vect, # Juntar todos los datos

+ byrow=T, # Instrucción llenar por filas

+ ncol=2, # Número de columnas

+ nrow=2, # Número de filas

+ dimnames=list( # Poner nombres a las filas y columnas

+ "Fuma"=c("SI","NO"), #Nombres de las filas, usamos concatenar

+ "Consume Drogas"=c("SI","NO"))) # Nombres de las columnas

>

> fumDrog

Consume Drogas

Fuma SI NO

SI 43 60

NO 23 44

El comando y la respuesta de la prueba estadística.

> chisq.test(fumDrog,correct = F)

Pearson's Chi-squared test

data: fumDrog

X-squared = 0.9408, df = 1, p-value = 0.3321

P>alfa, luego no se puede rechazar H0. Es decir, el consumo de drogas es independiente al consumo de tabaco.