Facultad de Ciencias **Bioestadística**
Universidad de la Repúblicas **Curso 2019**

Nombre:	**CI:**	**Año en que cursó**

Examen - 5 de Febrero de 2020 - Soluciones

Problema 1: Se tiene una urna con 6 bolillas de las cuales 4 son azules y 2 son rojas. Se extran (sin reposición) 3.

1. Calcular la probabilidad de las 3 bolillas extraidas sean de color azul. $\frac{C_3^4}{C_3^6} = \frac{1}{5}$
2. Calcular la probabilidad de que las 3 bolillas extraidas sean de color rojo. 0
3. Calcular la probabilidad de extraer 2 azules y 1 roja. $\frac{C_2^4 C_1^2}{C_3^6} = \frac{3}{5}$
4. Se tiene ahora una caja que tiene unicamente 1 bolilla azul y una roja. De la primera caja se extrae una y se coloca en la segunda caja. Luego se realiza una extracción de la segunda caja. Calcular la probabilidad de que la bolilla extraida de la segunda caja sea roja. $\frac{2}{6}\frac{2}{3} + \frac{4}{6}\frac{1}{3} = \frac{4}{9}$

Problema 2: Para estudiar la relación lineal entre el viento y la temperatura se ejecutó el siguiente comando:

```
datos<-airquality[complete.cases(airquality),]
lin<-lm(Wind~Temp,datos)
summary(lin)

##
## Call:
## lm(formula = Wind ~ Temp, data = datos)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.9443 -2.3584 -0.3005  1.6136  9.6852
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.37877    2.43137  10.027  < 2e-16 ***
## Temp        -0.18561    0.03102  -5.983 2.84e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.101 on 109 degrees of freedom
## Multiple R-squared:  0.2472,Adjusted R-squared:  0.2403
## F-statistic: 35.79 on 1 and 109 DF,  p-value: 2.842e-08
```

1) Plantear el modelo teórico correspondiente, indicando quienes son las variables y que hipótesis debe cumplir el error. El modelo teórico correspondiente es $W = aT + b + \epsilon$, siendo W la variable viento, T la temperatura y ϵ el error con distribución $N(0, \sigma^2)$.

2) Según el resultado del comando anterior, ¿es razonable suponer que se cumplen las hipótesis sobre los errores? Si lo es, pues Median $= -0.3005$.

3) Plantear y realizar una prueba de hipótesis (a nivel 0.05) para decidir si existe o no dependencia lineal entre las variables. La prueba de hipótesis a realizar es

$$\begin{cases} H_0: & a = 0 \\ H_1: & a \neq 0 \end{cases}$$

Dado que el p-valor $2.842e^{-08} < 0.05$ se rechaza H_0, es decir hay una relación lineal entre la variable viento y la varible temperatura.

4) Dar un valor del coeficiente de determinación e interpretar. El coeficiente de determinación corresponde a $R^2 = 0.2472$ lo que indica que la variación explicada por el modelo es 1/4 de la variación total, por lo cual el ajuste no es muy bueno.

Problema 3: Considere los datos

$$2.55 \quad 1.14 \quad 2.69 \quad 0.50 \quad 2.14 \quad -0.97 \quad 0.23 \quad 1.35 \quad -0.21 \quad 1.46$$

1) Realizar la prueba de Spearman a nivel $\alpha = 0.05$ para decidir si los mismos son iid. La prueba a realizar es

$$\begin{cases} H_0: & \text{Los datos son iid} \\ \\ H_1: & \text{los datos no son iid} \end{cases}$$

Calculamos el estadístico $\rho_S = 1 - \frac{6D}{n(n^2-1)}$, aquí $D = 234$ y $n = 10$ por lo tanto $\rho_S = -0.418$. Usando la tabla de Spearman se obtiene que el p-valor es $2 * 0.116$ y como es mayor a 0.05 se tiene que no rechazamos H_0 a dicho nivel.

2) Hallar un valor estimado para la esperanza, la mediana y el desvío de la muestra. Unos estimadores de los valores pedidos son los siguientes:

- Para la Esperanza: $\bar{X}_{10} = \frac{X_1+\ldots+X_{10}}{10} = 1.088$.
- Para el desvío: $S_{10} = \sqrt{\frac{1}{9}\left(\sum X_i^2 - 10(\bar{X}_{10})^2\right)} = 1.2029$
- Para la mediana: 1.245

3) Considere la siguiente salida de R

```
a=c(2.55,1.14,2.69,0.50,2.14,-0.97,0.23,1.35,-0.21,1.46)
ks.test(a,"pnorm",1,1)

##
##  One-sample Kolmogorov-Smirnov test
##
## data:  a
## D = 0.17286, p-value = 0.8784
## alternative hypothesis: two-sided
```

¿Que se puede concluir de la misma? El p-valor es bastante grande, por lo cual no se rechaza (sin importar el nivel) que los datos correspondan a $N(1,1)$.

4) Considere la siguiente salida de R

```
a=c(2.55,1.14,2.69,0.50,2.14,-0.97,0.23,1.35,-0.21,1.46)
ks.test(a,"pnorm",2,1)

##
##  One-sample Kolmogorov-Smirnov test
##
## data:  a
## D = 0.4054, p-value = 0.05359
## alternative hypothesis: two-sided
```

¿Que se puede concluir de la misma? Teniendo en cuenta el resultado de los dos test anteriores,¿que distribución es mas razonable suponer para los datos?

En esta situación si trabajamos con un nivel $\alpha > 0.05359$ se rechaza que los datos correspondan a una $N(2,1)$. Claramente es mas razonable suponer que los datos corresponden a una $N(1,1)$.