

Nombre:

Grupo:

MATEMÁTICAS III. GMAT 11-JUNIO RECUPERACIÓN 2017-2018. CONTROL2. CUESTIONES.

1) **(0.5 puntos)** Calcular la correlación de $x = (2, 1, 0, -1, -2)$ y $y = (0, 1, 4, 6, 8)$

2) **(0.5 puntos.)** Para comprobar la relación entre las notas por islas de Baleares se toman 3 muestras de 50 individuos de cada una de las islas y sus notas numéricas en las Pruebas de Bachillerato para el Acceso a la Universidad ¿Qué tipo de contraste podemos aplicar y por qué?

MATEMÁTICAS III. GINF CONTROL PARTE 2. EJERCICIOS.

1) El data frame `datos_vuelos` contiene información del retraso en minutos de vuelos de varias compañías aéreas diferentes.

```
head(datos_vuelos)
str(datos_vuelos)
aggregate(retraso~compania,data=datos_vuelos,FUN=mean)
aggregate(retraso~compania,data=datos_vuelos,FUN=sd)
var.test(retraso~compania)
boxplot(datos_vuelos$retraso~datos_vuelos$compania,
        main="Diagramas de caja de los retraso por compañía)

## Error: <text>:7:14: inesperado INCOMPLETE_STRING
## 6: boxplot(datos_vuelos$retraso~datos_vuelos$compania,
## 7:         main="Diagramas de caja de los retraso por compañía)
##         ^
```

Contestad a las siguientes cuestiones justificando que parte del código utilizáis

1) Interpretar y poner un título adecuado al diagrama de cajas ¿Qué nos dice el diagrama sobre la igualdad de medias del retraso? (**0.5 puntos**)

2) Escribid hipótesis del contraste de medias y discutid si se cumplen las condiciones necesarias para realizarlo. (**0.5 puntos**)

3) Realizar un contraste de igualdad de medias entre las dos compañías $\alpha = 0.1$ y discutid los resultados obtenidos a partir la salida del código. (**0.5 puntos**)

2) Para estudiar si hay evidencia de que el retraso de un vuelo en la salida aumenta el retraso de su llegada se toma una muestra aleatoria simple de 100 vuelos y se anota para cada vuelo si tuvo retraso en la salida y en la llegada (en minutos). La tabla siguiente resume los resultados:

| Salida / Llegada | No Retraso | Retraso |
|------------------|------------|---------|
| No Retraso | 75 | 15 |
| Retraso | 6 | 4 |

1) Plantear un contraste de igualdad de proporciones entre la proporción de vuelos retrasados en la salida y en la llegada. ¿Qué diseño experimental estamos utilizando? **(0.5 puntos.)**

2) Resolver el contraste al nivel de significación $\alpha = 0.1$ **(0.5 puntos.)**

3) Calcular el p -valor del contraste anterior. **(0.5 puntos.)**

4) Calcular e interpretar un intervalo de confianza para la diferencia de proporciones al nivel 99%. **(0.5 puntos.)**

3) Se piensa que el tiempo en segundos transcurrido entre dos reservas de vuelos de avión en un mismo día podría seguir una distribución exponencial con una reserva cada cinco segundos. Se toma una muestra de 10 tiempos en segundos.

| Vuelo | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---------|------|------|------|------|------|------|------|------|------|------|
| Retraso | 0.50 | 1.40 | 1.60 | 2.20 | 2.40 | 3.70 | 3.90 | 4.50 | 5.20 | 7.10 |

1) ¿Cuál es y qué parámetros tiene la función de distribución teórica propuesta? Escribid correctamente la función de distribución. **(0.5 puntos)**

2) Contrastar la hipótesis del enunciado con el test KS, al nivel de significación $\alpha = 0.1$. **(1 puntos)**

4) La siguiente tabla contiene los valores de `retraso_llegada`, `retraso_salida` y `distancia` del trayecto del vuelo para cuatro vuelos. Las distancias vienen dadas en centenas de kilómetros y los retrasos en decenas de minutos.

```
df=data.frame(retraso_llegada,retraso_salida,distancia)
df
```

```
##   retraso_llegada retraso_salida distancia
## 1             27             3         10
## 2             -9            -1         15
## 3             18             2         20
## 4             46             5          5
```

```
X=cbind(rep(1,4),df$retraso_salida,df$distancia)
X
```

```
##      [,1] [,2] [,3]
## [1,]    1    3   10
## [2,]    1   -1   15
## [3,]    1    2   20
## [4,]    1    5    5
```

```
Y=matrix(df$retraso_llegada,ncol=1)
Y
```

```
##      [,1]
## [1,]   27
## [2,]  -9
## [3,]   18
## [4,]   46
```

```

t(X)%*%X

##          [,1] [,2] [,3]
## [1,]      4    9   50
## [2,]      9   39   80
## [3,]     50   80  750

det(t(X)%*%X)

## [1] 5150

solve(t(X)%*%X)

##          [,1]      [,2]      [,3]
## [1,]  4.4368932 -0.53398058 -0.23883495
## [2,] -0.5339806  0.09708738  0.02524272
## [3,] -0.2388350  0.02524272  0.01456311

solve(t(X)%*%X)%*%t(X)%*%Y

##          [,1]
## [1,]  0.57281553
## [2,]  9.07766990
## [3,] -0.03980583

X%*%solve(t(X)%*%X)%*%t(X)%*%Y

##          [,1]
## [1,] 27.407767
## [2,] -9.101942
## [3,] 17.932039
## [4,] 45.762136

sum((X%*%solve(t(X)%*%X)%*%t(X)%*%Y)^2)

## [1] 3249.762

```

Usad el código anterior cuando pertoque para contestar a las siguientes preguntas.

- 1) Escribid y explicad la ecuación del modelo de regresión lineal múltiple que predice el `retraso_llegada` a partir de las otras dos variables. **(0.5 puntos.)**
- 2) Calcular R^2 y R^2 ajustado de la anterior regresión. **(0.5 puntos.)**
- 3) Calcula el AIC de este modelo. **(0.5 puntos.)**
- 4) Calcular el intervalo de confianza al 95% para el coeficiente de la variable distancia ¿Qué se puede deducir de su presencia en el modelo? **(0.5 puntos.)**