SOLUCIÓN EJERCICIOS DE REPASO 2º PARCIAL

1.- El tiempo de respuesta de un sistema informático sigue una distribución normal de media m=5. Se realizan N=20 pruebas. Suponiendo que la desviación típica muestral resulta s=2,5, calcula la probabilidad de que la media muestral sea mayor que 6,17.

m=5 N=20 s=2,5

$$P(\bar{X} > 6{,}17) = P\left(\frac{\bar{X}-5}{2{,}5/\sqrt{20}} > \frac{6{,}17-5}{2{,}5/\sqrt{20}}\right) = P(t_{19} > 2{,}09) = 0{,}025$$

2.-La temperatura que soporta un tipo determinado de equipo informático sigue una distribución normal. Si extrae dos muestras de tamaños $N_1=9$ y $N_2=16$, calcula la probabilidad de que la primera varianza muestral resulte más de 4 veces el valor de la segunda varianza muestral.

$N_1=9$ $N_2=16$

$P(s_1^2 > 4s_2^2) = P(\frac{s_1^2}{s_2^2}>4)=P(F_{8,15}>4)=0{,}01$

3.-Se ha registrado el número de errores en 14 programas. La media muestral es 9 y la desviación típica muestral 1,6.

a) Calcula el intervalo de confianza para el número medio de errores (nivel de confianza 95%). ¿Se puede admitir que el número medio de errores en la población de programas es m=7?

b) Calcula el intervalo de confianza para la desviación típica del número de errores (nivel de confianza 90%). ¿Se puede admitir σ=3?

a) N=14 $\bar{X} = 9$ $s = 1{,}6$

$$t_{13}^{\alpha/2=0{,}025} = 2{,}16$$

$[9-2{,}16 \frac{1{,}6}{\sqrt{14}}, 9+2{,}16 \frac{1{,}6}{\sqrt{14}}] =[8{,}076, 9{,}92]$

No se puede admitir que m=7 porque no está en el intervalo

b) $\left[\sqrt{\frac{(14-1)1{,}6^2}{22{,}362}}, \sqrt{\frac{(14-1)1{,}6^2}{5{,}892}}\right]$=[1,22, 2,38] No se puede admitir σ=3 porque no está en el intervalo.

4.-En un estudio de comparación de dos algoritmos de inversión de matrices se ha tomado una muestra de N_1=7 matrices y se han invertido con el algoritmo 1, obteniéndose una desviación típica s_1=1,25 para el tiempo de inversión. Con el algoritmo 2 se ha invertido una muestra de N_2=10 matrices y los resultados han sido s_2=2,4.

a)¿Se puede admitir que las varianzas son iguales? (α=10%)

b)La SCTotal=61,215 y la SCalgoritmo=18,16, con α=1%, ¿hay efecto significativo de algoritmo sobre la media?

a) $F_{6,9}^{\alpha=0,1} \rightarrow f_1$=0,24

f_2= 3,37

Intervalo de confianza para el ratio de varianzas

$\left[\frac{1,25^2}{2,4^2 x\ 3,37}, \frac{1,25^2}{2,4^2 x\ 0,24}\right]$= [0,08, 1,13]

Se puede admitir que las varianzas son iguales porque el cociente 1 está en el intervalo de confianza.

b)ANOVA sobre el efecto de algoritmo

Origen variabilidad	Suma de cuadrados	Grados de libertad	Cuadrado Medio	F-ratio
Algoritmo	18,16	1	18,16	6,48
Residual	43,055	15	2,8703	
Total	61,215	16		

$$F - tabla_{1,15}^{\alpha=0,01} = 4{,}54$$

Como F-ratio es mayor que F-tabla hay efecto significativo de algoritmo sobre la media.

5.- Se consideran dos tipos de distribución de ficheros (A y B) en unidades de almacenamiento. Se realiza un experimento con 13 repeticiones con la distribución A y 15 con la B. Se registra el tiempo de acceso en cada prueba. Los resultados correspondientes son:

Distrib. A: varianza=3,27

Distrib. B: varianza=21,8

SCTotal=422,707 SCResidual=345,159

a) Analiza si hay diferencias significativas entre las medias (α=5%)

b) Analiza si hay diferencias significativas entre las varianzas (α=5%).

a) ANOVA para estudiar la diferencia de medias

Origen de variabilidad	Suma de cuadrados	Grados de libertad	Cuadrado medio	F-ratio
Distribución	77,548	1	77,548	5,84
Residual	345,159	26	13,275	
Total	422,707	27		

$$F - tabla_{1,26}^{\alpha=0,05} = 4{,}23$$

Como F-ratio es mayor que F-tabla si que hay diferencia significativa entre las medias.

b) $F_{12,14}^{\alpha=0,05} \rightarrow f_1 = 0{,}31 \quad f_2 = 3{,}05$

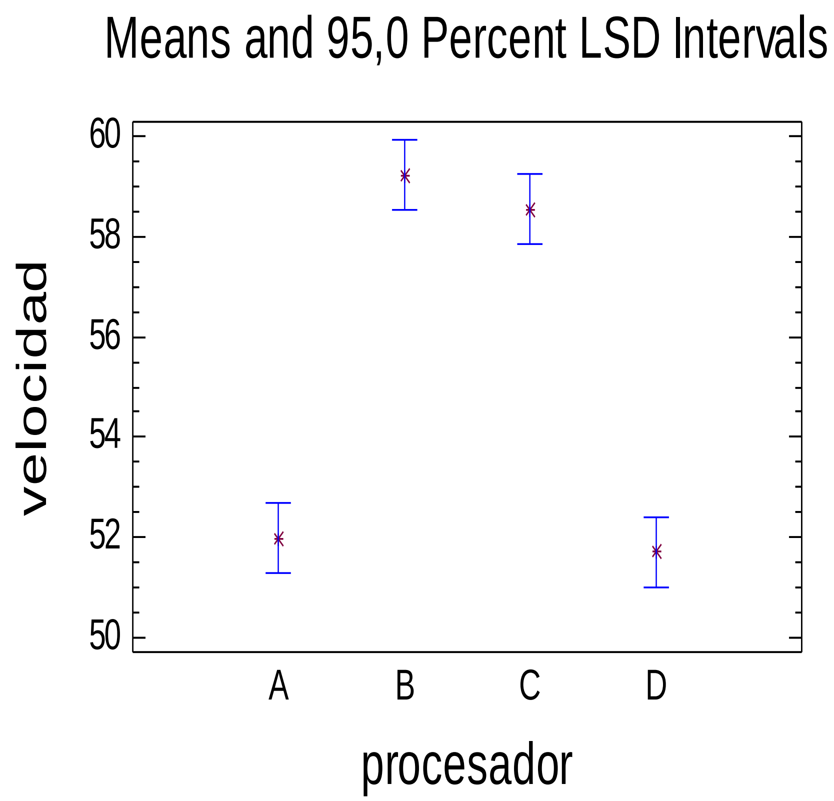
Intervalo de confianza para el ratio de varianzas

$\left[\frac{3{,}27}{21{,}8x3{,}05}, \frac{3{,}27}{21{,}8x0{,}31}\right]$=[0,049,0,48] Como no contiene el cociente 1 las varianzas difieren significativamente, siendo mayor la de distribución B.

6.- Se ha realizado una experiencia para comparar cuatro tipos de procesadores (A, B, C y D). Se ha medido la velocidad de ejecución de procesos, con 4 repeticiones por cada nivel del factor.

a) ¿Influye el procesador en la velocidad media de ejecución? Calcula la tabla del ANOVA (SCTotal=209,837, SCprocesador=199,872). Utiliza α=1%

b) Interpreta el efecto de procesador a partir del gráfico siguiente:



a) ANOVA para estudiar el efecto de procesador

Origen de variabilidad	Suma de cuadrados	Grados de libertad	Cuadrado medio	F-ratio
Procesador	199,872	3	66,624	80,23
Residual	9,965	12	0,8304	
Total	209,837	15		

$F - tabla_{3,12}^{\alpha=0,01} = 5,95$ Como F-ratio es mayor que F-tabla existe efecto significativo de procesador sobre la media.

b) La velocidad de ejecución media difiere entre procesador A y B , y A y C. Entre B y C no hay diferencia significativa. Tampoco la hay entre A y D. D difiere de B y de C. La velocidad de ejecución media es significativamente menor con los procesadores A y D.

7.- Se desea estudiar la influencia que la configuración (tres posibles A, B y C) y el tamaño de memoria Caché (3 niveles: 0, 1 y 2) tienen sobre el rendimiento medio de un sistema informático. Se ha diseñado un experimento en el que cada uno de los nueve posibles tratamientos fue ensayado cuatro veces.

a)Estudia qué efectos son significativos a partir del cuadro resumen del ANOVA (utiliza α=0,05). SCTotal=22078,4. SCconfig=1208,94. SCcaché=17780,8. SCconfig*caché=1708,65.

b)Interpreta el efecto del factor configuración a partir de los intervalos L.S.D.

Los intervalos resultan:

Config. A [41,925; 47,915] Config. B [28,43; 34,42] Config. C [38,997; 44,987]

c)Interpreta los gráficos de medias para el efecto de memoria Caché y para la interacción ConfigxCaché.

a) ANOVA para estudiar la significación de los efectos:

Origen de variabilidad	Suma de Cuadrados	Grados de libertad	Cuadrado medio	F-ratio
Configuración	1208,94	2	604,47	11,83
Caché	17780,8	2	8890,4	173,95
Interacción	1708,65	4	427,16	8,36
Residual	1380,01	27	427,16	
Total	23078,4	35		

$$F - tabla_{2,27}^{\alpha=0,05} = 3,35$$

$$F - tabla_{4,27}^{\alpha=0,05} = 2,73$$

Como F-ratio de configuración es mayor que F-tabla, este factor influye significativamente. F-ratio de caché también es mayor que F-tabla por lo que este factor también tiene efecto significativo. Finalmente la interacción también tiene un F-ratio mayor que la de tabla, por lo que es un efecto significativo.

b) El intervalo de Configuración A se solapa con el de C, por lo que no hay diferencia en rendimiento medio entre ambas configuraciones. Sin embargo ninguno de los dos se solapa con el de configuración B, teniendo esta un rendimiento medio inferior.
c) En el gráfico de medias de memoria caché se aprecia que el efecto de ésta sobre el rendimiento medio es lineal positivo, no existiendo efceto cuadrático.

En el gráfico de la interacción se observa:
Con configuración A el efecto de memoria caché sobre el rendimiento medio es lineal positivo y cuadrático negativo. Con configuración B dicho efecto es lineal positivo sólo. Con configuración C este efecto es lineal positivo y de menor magnitud que con configuración B.

8.- El tiempo (en segundos) que tarda un sistema informático en ejecutar un programa prueba, depende del número de usuarios conectados a él. La tabla siguiente muestra la matriz de varianzas-covarianzas observada después de medir ambas variables durante varios días:

Covariances

	N_USUARIOS	T_EJECUCION
N_USUARIOS	65,9933 (25)	48,6027 (25)
T_EJECUCION	48,6027 (25)	36,116 (25)

a) Calcula el coeficiente de correlación lineal e interpreta el resultado.
b)¿Qué apariencia tendrá el diagrama de dispersión entre ambas variables?
c)¿Qué porcentaje de la varianza observada en el tiempo de ejecución está explicada por el número de usuarios?

a) $r = \frac{48{,}6027}{\sqrt{65{,}9933x36{,}116}} = 0{,}9955$ La relación es lineal positiva y fuerte
b) El diagrama de dispersión presentará los puntos muy cercanos a una recta de pendiente positiva.
c) El porcentaje de la varianza observada en el tiempo de ejecución explicada por el número de usuarios es el valor del coeficiente de correlación al cuadrado $0{,}9955^2$x100=99,11%

9.- Con los mismos datos del ejercicio anterior, la media de N_USUARIOS ha resultado 12,08 y la media de T_EJECUCION vale 9,94.
a) Calcula la recta de regresión que relaciona la media del T_EJECUCION con N_USUARIOS.
b) Calcula la probabilidad de que el tiempo de ejecución esté entre 10 y 11 segundos cuando el número de usuarios es 15.

a) b=r $\frac{s_y}{s_x} = 0{,}9955\sqrt{\frac{36{,}116}{65{,}9933}}$=0,736
a=$\bar{y} - b\bar{x}$=9,94-0,736x12,08=1,049
La recta es:
E(T_EJECUCION/N_USUARIOS)=1,049+0,736N_USUARIOS

b) $s_{residual}=\sqrt{36{,}116(1-0{,}9955^2)}=0{,}569$
$m_{TEJECUCION/N\_USUARIOS=15}=1{,}049+0{,}736X15=12{,}089$
$p(10<N(12{,}089,0{,}569)<11)=P(\frac{10-12{,}089}{0{,}569}<N(0{,}1)<\frac{11-12{,}089}{0{,}569})=$
$=P(-3{,}67<N(0{,}1)<-1{,}91)=P(N(0{,}1)<-1{,}91)-P(N(0{,}1)<-3{,}67)=$
$0{,}0281-0{,}00012=0{,}02798$

10.- En los datos de la encuesta se dispone de 89 observaciones del peso y estatura en chicos. Con dichos datos se ha calculado la recta que relaciona ambas variables, obteniéndose la siguiente tabla:

```
Multiple Regression Analysis

----------------------------------------------------------------------Dependent
variable: PESO

Selection variable: SEXO="chicos"

----------------------------------------------------------------------

                                Standard          T

Parameter        Estimate          Error       Statistic        P-Value

----------------------------------------------------------------------

CONSTANT            65,224        1,65319

ESTATURA-165      0,498687       0,118347

----------------------------------------------------------------------
```

a) Completa la tabla anterior y analiza si los dos parámetros de la recta son significativos (α=5%).
b) Se ha calculado también el ANOVA de la recta. Completa la tabla y explica las conclusiones de la misma (α=1%). Calcula el coeficiente de determinación e interpreta el resultado.

```
                         Analysis of Variance
-----------------------------------------------------------------------
Source      Sum of Squares     Df  Mean Square    F-Ratio      P-Value
-----------------------------------------------------------------------
Model
Residual             4745,9
-----------------------------------------------------------------------
Total (Corr.)        5714,49
```

c) ¿Entre qué valores estará el peso del 95% de los chicos que miden 175 cm?

a)

CONSTANT t-statistics=62,224/1,65319=39,45

ESTATURA-165 t-statistics=0,498687/0,1183347=4,21

$$t-tabla_{89-2}^{\alpha=0{,}05} entre\ 2\ y\ 1{,}98$$

Multiple Regression Analysis

Dependent variable: PESO

Selection variable: SEXO="chicos"

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	65,224	1,65319	39,45	<0,05
ESTATURA-165	0,498687	0,118347	4,21	<0,05

Las dos t-statistics son en valor absoluto mayores que la t de tabla, por lo que tanto la ordenada como la pendiente difieren significativamente de cero.

b) **Analysis of Variance**

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	968,59	1	968,59	17,75	<0,01
Residual	4745,9	87	54,55		
Total (Corr.)	5714,49	88			

$F - tabla_{1,87}^{\alpha=0,01} = 6{,}9$ Como F-.ratio > F-tabla existe efecto lineal de estatura significativo.

Coeficiente de determinación R^2= $\frac{968{,}59}{5714{,}49} x100 = 16{,}95\%$ El efecto lineal de estatura explica un 16,95% de la variabilidad del peso.

c)$m_{peso/estatura=175}$=65,224+0,4987(175-165)=70,211Kg

$s_{residual}$=$\sqrt{54{,}55}$=7,39

Intervalo del 95% de probabilidad:

[70,211-2x7,39, 70,211+2x7,39]=[55,431, 84,991]

11.- Se ha medido durante 13 días el tiempo medio de respuesta de un sistema informático (en segundos) y la carga media en el mismo (en consultas por minuto). Con los datos se ha realizado un ajuste de regresión lineal.

a) Completa la tabla siguiente y analiza qué parámetros del modelo de regresión son significativos (α=5%).

Multiple Regression Analysis

Dependent variable: Trespuesta

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	0,0747486	0,190814		
CARGA	0,789228	0,084857		
CARGA^2	-0,0385481	0,0075167		

b) Teniendo en cuenta el signo de los parámetros del modelo indica aproximadamente cómo será el gráfico de la evolución del tiempo medio de respuesta en función de la carga media diaria.

c) Completa la siguiente tabla e indica la conclusión de la misma (α=5%). Calcula el coeficiente de determinación e interprétalo.

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	17,1757				
Residual					
Total (Corr.)	17,5169				

a) **Multiple Regression Analysis**

Dependent variable: Trespuesta

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	0,0747486	0,190814	0,3874	>0,05
CARGA	0,789228	0,084857	9,28	<0,05
CARGA^2	-0,0385481	0,0075167	-5,133	<0,05

$t - tabla_{13-3}^{\alpha=0,05} = 2{,}228$ La ordenada no difiere de cero pues el t-statistic es menor en valor absoluto que la t-tabla. Sin embargo hay efectos significativos de la carga y de la carga al cuadrado al ser sus t-statistics mayores en valor absoluto que la t-tabla.

b)El efecto lineal es positivo por lo que a mayor carga mayor tiempo de respuesta. El efecto cuadrático es negativo por lo que cada vez habrá menos efecto lineal positivo.

c) **Analysis of Variance**

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	17,1757	2	8,58785	251,6954	<0,05
Residual	0,3412	10	0,03412		
Total (Corr.)	17,5169	12			

$F - tabla_{2,10}^{\alpha=0,05} = 4{,}1$ Como F-ratio es mayor que F-tabla el modelo incluye efectos significativos, explica variabilidad del tiempo de respuesta.

$Coeficiente\ de\ determinación\ R^2 = \frac{17{,}1757}{17{,}5169} x100 = 98{,}05\%$ La carga explica con sus efectos lineal y cuadrático el 98,05% de la variabilidad observada en el tiempo de respuesta.