

Nombre y apellido..... Número de libreta.....

1	2	3	4	Nota

Los ejercicios (1), (2) y (4) deben ser respondidos enteramente sobre las hojas el enunciado, en los lugares destinados a ello. Por favor no entregue las hojas en las que hace sus cálculos auxiliares. El ejercicio (3) debe ser desarrollado en una hoja separada, justificando todos sus razonamientos. Por favor ponga su nombre y apellido en el enunciado y en todas las hojas restantes que entregue. El puntaje figura junto a cada ejercicio. El parcial se aprueba con 60 o más puntos.

1. (25 puntos) Los siguientes datos corresponden al crecimiento de orugas alimentadas con dietas experimentales que difieren en el contenido de tanino. Se espera que a mayores porcentajes de tanino en la dieta el crecimiento de las orugas sea más lento. Para ello se midieron las variables porcentaje de concentración de tanino en el alimento que se les proporcionó a las larvas de oruga (**tanino**) y la longitud de las orugas al llegar a adultas, en milímetros (**crecimiento**) para 25 orugas. Los datos figuran a continuación.

tanino (%)	crecimiento (mm.)						
0	334.08	2	391.47	4	298.36	6	230.63
0	395.24	2	346.64	4	257.08	6	319.58
0	334.57	2	340.78	4	285.32	8	260.88
0	382.25	2	337.73	6	263.81	8	166.47
0	392.06	4	347.75	6	307.64	8	213.47
2	363.21	4	298.82	6	237.59	8	209.16
						8	217.30

Responda este ejercicio enteramente sobre estas hojas de enunciado. Complete sobre las líneas punteadas donde corresponda. **La salida de R de la regresión figura al final del enunciado.**

- a) (2p.) El modelo lineal plantea la siguiente relación $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ donde (complete)
 $Y_i =$
 $x_i =$
 $\varepsilon_i =$
- b) (2p.) ¿Qué supuestos se hacen sobre los ε_i para obtener intervalos de confianza para β_0 y β_1 ?

- c) (1p.) Este ε_i es: (redondear el ítem que indica la respuesta correcta)
 i. observable
 ii. no observable.
- d) (2p.) Sea $\sigma^2 = Var(\varepsilon_i)$. Luego, otra manera de escribir los supuestos del modelo cuando se quieren obtener intervalos de confianza para β_0 y β_1 es (complete)
 $E(Y_i) =$
 $Var(Y_i) =$
 La distribución de los Y_i es y las Y_i son
- e) (2p.) (Indicar la respuesta correcta con un círculo sobre el ítem). El modelo de regresión lineal implica
 i. que las variables Y y X cumplen una relación lineal perfecta.
 ii. que la media de Y para cada valor de $X = x$ cumple una relación lineal perfecta con x .
 iii. que la media de X para cada valor de $Y = y$ cumple una relación lineal perfecta con y .
 iv. ninguna de las anteriores.
- f) (2p.) (Indicar la respuesta correcta con un círculo sobre el ítem). El valor de R^2 , el coeficiente de determinación entre X e Y , es muy cercano a uno cuando
 i. la pendiente estimada es mayor que cero.

- ii. se cumple el supuesto de homoscedasticidad.
 - iii. el gráfico de dispersión de Y versus X muestra a todas las observaciones muy cerca de una recta.
 - iv. ninguna de las anteriores.
- g) (3p.) Para los datos observados indique el valor estimado para cada uno de los 3 parámetros definidos en los ítems *a*) y *d*). Responda en base a la salida del R que figura más abajo.

$$\hat{\beta}_0 = \dots\dots\dots \quad \hat{\beta}_1 = \dots\dots\dots \quad \hat{\sigma}^2 = \dots\dots\dots$$

h) (1p.) ¿Resulta significativamente distinta de cero la pendiente del modelo propuesto? Responda a nivel 0.05, redondeando la respuesta correcta:

- i. Sí, resulta significativamente distinta de cero la pendiente.
- ii. No, no resulta significativamente distinta de cero la pendiente.

i) (1p.) El p-valor en el cual se basa para dar la respuesta al ítem precedente es

- i. $< 2e-16$ (es decir, menor a $2 \cdot 10^{-16}$)
- ii. $1.22e-08$ (es decir, $1,22 \cdot 10^{-8}$)
- iii. no aparece en la salida del R

j) (2p.) Se desea comparar las longitudes esperadas estimadas por el modelo de regresión de dos orugas adultas: la primera de ellas fue alimentada con alimentos que contenían el 2% de tanino y la segunda fue alimentada con alimentos que contenían el 3% de tanino. Indique cuál respuesta es la correcta.

- i. La longitud esperada de la primera estimada por el modelo será 379.7 mm. mayor que la segunda.
- ii. La longitud esperada de la primera estimada por el modelo será 379.7 mm. menor que la segunda.
- iii. La longitud esperada de la primera estimada por el modelo será 19.6 mm. mayor que la segunda.
- iv. La longitud esperada de la primera estimada por el modelo será 19.6 mm. menor que la segunda.

k) (3p.) El residuo de la observación *i*ésima se define por (complete con una fórmula). De acuerdo al ítem g) de este ejercicio, el valor ajustado y el residuo de la observación sexta, es decir del par (2, 363,21), son, respectivamente

$$\text{Valor ajustado} = \dots\dots\dots \quad y \quad \text{Residuo} = \dots\dots\dots$$

l) (2p.) El coeficiente de determinación R^2 para los datos observados en este caso es:
¿Cómo se interpreta este valor? (Indicar la respuesta correcta con un círculo sobre el ítem):

- i. un 76% de la variabilidad de la Y queda explicada por el modelo de regresión
- ii. un 32% de la variabilidad de la Y queda explicada por el modelo de regresión
- iii. un 76% de los pares (x_i, Y_i) estarán sobre la recta estimada
- iv. conociendo el porcentaje de tanino en la alimentación de la larva, podemos predecir la longitud de la oruga adulta con un 76% de confianza

m) (2p.) Se realizará una nueva medición de la longitud de una oruga adulta cuyo alimento tendrá 6% de tanino. Para saber si está bien medida, se comparará la longitud resultante con un intervalo que se armará en base a la muestra analizada en los ítems anteriores. Dicho intervalo resulta ser (complete con valores numéricos)

$$[\dots\dots\dots, \dots\dots\dots]$$

```
sal<-lm(crecimiento ~tanino) Call: lm(formula = crecimiento ~tanino)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-56.309 -15.956  -2.456  15.468  57.553
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  379.772      11.183   33.961 < 2e-16 ***
tanino       -19.624       2.283   -8.597 1.22e-08 ***
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 32.28 on 23 degrees of freedom Multiple
R-squared: 0.7627, Adjusted R-squared: 0.7523 F-statistic: 73.91
on 1 and 23 DF, p-value: 1.221e-08
```

```
> predict(sal,interval="confidence",level=0.95)
```

```
      fit      lwr      upr
1  379.7724 356.6392 402.9056
2  379.7724 356.6392 402.9056
.
.
.
19 262.0272 245.6695 278.3849
20 262.0272 245.6695 278.3849
21 222.7788 199.6456 245.9120
.
.
.
```

```
> predict(sal,interval="prediction",level=0.95)
```

```
      fit      lwr      upr
1  379.7724 309.0992 450.4456
2  379.7724 309.0992 450.4456
.
.
.
19 262.0272 193.2730 330.7814
20 262.0272 193.2730 330.7814
21 222.7788 152.1056 293.4520
.
.
.
```

2. (25 puntos) Una revista de actualidad publica un artículo en el cuál se desea estudiar qué modelos de smartphones son preferidos por los hombres y las mujeres de una determinada ciudad. Concretamente, si p es la proporción de hombres en la ciudad que prefieren un celular con sistema operativo Android por sobre un Blackberry y q es la respectiva proporción en mujeres, se desea estudiar si p y q coinciden. El periodista a cargo de realizar el artículo entrevista a 205 hombres y 198 mujeres de la ciudad y les pregunta sobre sus preferencias al respecto, obteniendo que 113 hombres y 87 mujeres prefieren un celular con Android por sobre un Blackberry.

a) (2p.) La fórmula teórica para un estimador \hat{p} de la proporción p es
 donde las variables involucradas son

.....

El valor \hat{p} observado en la muestra es

b) (2p.) La fórmula teórica para un estimador \hat{q} de la proporción q es
 donde las variables involucradas son

.....

El valor \hat{q} observado en la muestra es

c) (3p.) Una expresión pivote para un intervalo de confianza para $p - q$ es

.....

d) (4p.) La expresión pivote hallada en (c) tiene distribución asintótica

e) (2p.) La fórmula teórica de un intervalo de confianza de nivel asintótico 97% para $p - q$ es

.....

f) (4p.) El intervalo de confianza de nivel asintótico 97% para $p - q$ calculado a partir de los datos es

.....

donde el percentil utilizado para calcularlo es (completar con su valor numérico).

g) (8p.) Proponga un test de hipótesis de nivel asintótico 97% para estudiar la igualdad de proporciones.

i. (2p.) Las hipótesis del test son H_0 : y H_1 :

ii. (6p.) ¿Se rechaza la hipótesis nula al efectuar el test? Justifique su respuesta en los dos renglones dados a continuación.

.....

.....

3. (25 puntos) El tiempo de duración (medidos en minutos) en conversación de un cierto tipo de baterías para celulares tiene distribución normal con media 300 minutos y desvío estándar 30 minutos. El departamento de control de calidad de la fábrica de baterías sospecha que hubo defectos en la producción del último lote de baterías que podrían haber disminuido el tiempo medio de duración de las mismas. Para confirmar estas sospechas decide llevar a cabo un test de hipótesis de nivel 0.05. Si se comprueba que la duración media de ellas es menor que lo deseado se retirará el lote del mercado. Entonces se elige una muestra aleatoria de 16 baterías cuyo tiempo promedio de duración resulta ser de 290 min.
- (12p.) Plantear las hipótesis a testear, definiendo las variables aleatorias involucradas, el estadístico del test y su distribución bajo la hipótesis nula. Dar la región de rechazo y la conclusión a la que se llega con la muestra observada. ¿Se retira finalmente el lote del mercado?
 - (7p.) Hallar la potencia del test planteado en el ítem anterior cuando el verdadero tiempo de duración es de 280 minutos.
 - (6p.) ¿Qué tamaño de muestra debería tomarse si se quiere lograr una potencia de por lo menos 0.95 a un nivel 0.05 cuando el verdadero tiempo de duración es de 280 minutos?

4. (25 puntos) Varios factores contribuyen al buen funcionamiento de un motor. En particular, es deseable mantener al mínimo el ruido y la vibración del motor. Para estudiar el efecto que tiene la marca del buje sobre la vibración del motor, se examinaron 5 marcas de bujes que se colocaron en 30 motores. Se registró la cantidad de vibración (medida en micras). Los datos figuran a continuación (fueron publicados en la revista *Quality Engineering*, año 1991, pág. 361-369)

Marca 1	13.1	13.8	14.0	14.4	14.0	11.6
Marca 2	16.3	15.7	17.2	14.9	14.4	16.2
Marca 3	13.7	13.9	12.4	13.8	14.9	13.3
Marca 4	15.7	13.7	14.4	16.0	13.9	14.7
Marca 5	13.5	13.4	13.2	12.7	13.4	12.3

Responda este ejercicio enteramente sobre estas hojas de enunciado. Complete sobre las líneas punteadas donde corresponda. **Las salidas de R (ANOVA y medidas resúmenes) para estos datos figuran al final del ejercicio.**

a) (5p.) El modelo de Análisis de la Varianza (ANOVA) para estos datos es

.....

donde

$Y_{ij} =$

$\mu_i =$

$\varepsilon_{ij} =$

e i se mueve entre e $I =$, mientras que j está entre y $J =$(completar con los valores numéricos de este ejercicio).

b) (2p.) Los supuestos de este modelo son

.....

.....

.....

c) (4p.) La fórmula teórica para el estimador de μ_i es

$\hat{\mu}_i =$

El valor observado a partir de los datos para $i = 2$ resulta ser $\hat{\mu}_i =$

d) (2p.) Suponga que se verifican los supuestos del modelo de ANOVA. Se quiere aplicar un test para decidir si existen diferencias en las vibraciones medias del motor según la marca a nivel 0.05. Las hipótesis a testear son (redondee el ítem que corresponda a la opción correcta):

i. $H_0 : \mu_1 = \mu_2 = \dots = \mu_6$ versus $H_1 : \mu_1 \neq \mu_2 \neq \dots \neq \mu_6$.

ii. $H_0 : \mu_1 = \mu_2 = \dots = \mu_5$ versus $H_1 : \mu_1 \neq \mu_2 \neq \dots \neq \mu_5$.

iii. $H_0 : \bar{Y}_1 = \bar{Y}_2 = \dots = \bar{Y}_I$ versus $H_1 : \text{hay al menos dos } k, l \text{ tales que } \bar{Y}_k \neq \bar{Y}_l$.

iv. $H_0 : \mu_1 = \mu_2 = \dots = \mu_I$ versus $H_1 : \text{hay al menos dos subíndices } k, l \text{ tales que } \mu_k \neq \mu_l$.

e) (2p.) El valor del estadístico adecuado para testear dicha hipótesis es, en este caso, (complete con el valor numérico observado) y la conclusión a nivel 0.05 resulta ser (redondee el ítem que corresponda a la opción correcta)

i. Hay diferencias significativas entre las medias poblacionales comparadas

ii. No hay diferencias significativas entre las medias poblacionales comparadas.

f) (6p.) El valor estimado de la varianza de los errores (en cada caso complete con la fórmula y luego con el valor numérico observado)

i. bajo el modelo de ANOVA es

$$\hat{\sigma}^2 = \dots\dots\dots$$

que en este caso da el valor observado $\hat{\sigma}_{\text{obs}}^2 = \dots\dots\dots$

ii. bajo el modelo de Gauss sin sesgo (es decir, las Y_i son observaciones i.i.d.) es

$$\hat{\sigma}^2 = \dots\dots\dots$$

que en este caso da el valor observado $\hat{\sigma}_{\text{obs}}^2 = \dots\dots\dots$

g) (2p.) Se quieren hacer comparaciones múltiples de a pares entre las medias de las vibraciones del motor según la marca del buje, de nivel simultáneo 0.01. Para eso se aplica el método de Tukey, cuya salida figura al final de este enunciado. Una manera de representar el resultado de estas diferencias de Tukey es mediante un gráfico con barras, donde una barra que continua de un renglón al siguiente indica que no se observaron diferencias significativas entre las medias de las marcas respectivas. Indique cuál gráfico corresponde a los datos dados

i. TUKEY (HSD) COMPARISON OF MEANS OF VIBRACION BY MARCA

MARCA	MEAN	HOMOGENEOUS GROUPS
2	15.783	I
4	14.733	I I
3	13.667	.. I I
1	13.483	.. I I
5	13.083 I

ii. TUKEY (HSD) COMPARISON OF MEANS OF VIBRACION BY MARCA

MARCA	MEAN	HOMOGENEOUS GROUPS
2	15.783	I
4	14.733	I I
3	13.667	.. I
1	13.483	.. I
5	13.083	.. I

h) (2p.) Diga Verdadero o Falso (V o F) a nivel 0.01, para cada afirmación que sigue

- i. la media de la marca 2 se diferencia de todas las otras medias.....
- ii. la media de la marca 4 se diferencia de todas las otras medias.....
- iii. la media de la marca 4 no se diferencia de ninguna otra.....
- iv. las medias de las marcas 1 y 2 difieren.....

DESCRIPTIVE STATISTICS

VARIABLE	N	MEAN	SD	MINIMUM	MAXIMUM
VIBRACION	30	14.150	1.2886	11.600	17.200

MARCA	MEAN	SAMPLE SIZE	GROUP STD DEV
1	13.483	6	1.0167
2	15.783	6	1.0147
3	13.667	6	0.8165
4	14.733	6	0.9395
5	13.083	6	0.4792
TOTAL	14.150	30	0.8766

```
> marca.f<-as.factor(marca)
> salida<-aov(vibracion ~marca.f)
> summary(salida)
          Df Sum Sq Mean Sq F value    Pr(>F)
marca.f    4  28.943   7.2358   9.4159 8.742e-05 ***
```

```
Residuals  25  19.212   0.7685
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> shapiro.test(salida$res)
```

Shapiro-Wilk normality test

```
data:  salida$res W = 0.9738, p-value = 0.6469
```

```
> levene.test(vi,marca.f)
```

modified robust Brown-Forsythe Levene-type test based on the absolute deviations from the median

```
data:  vi Test Statistic = 0.4981, p-value = 0.7373
```

```
> marca.tuk<-TukeyHSD(salida,"marca.f",ordered=FALSE,conf.level=0.99)
```

```
> marca.tuk
```

Tukey multiple comparisons of means
99% family-wise confidence level

```
Fit: aov(formula = vibracion ~marca.f)
```

```
$marca.f
```

	diff	lwr	upr	p adj
2-1	2.3000000	0.4591004	4.1408996	0.0010545
3-1	0.1833333	-1.6575663	2.0242330	0.9960738
4-1	1.2500000	-0.5908996	3.0908996	0.1300460
5-1	-0.4000000	-2.2408996	1.4408996	0.9310288
3-2	-2.1166667	-3.9575663	-0.2757670	0.0026227
4-2	-1.0500000	-2.8908996	0.7908996	0.2619868

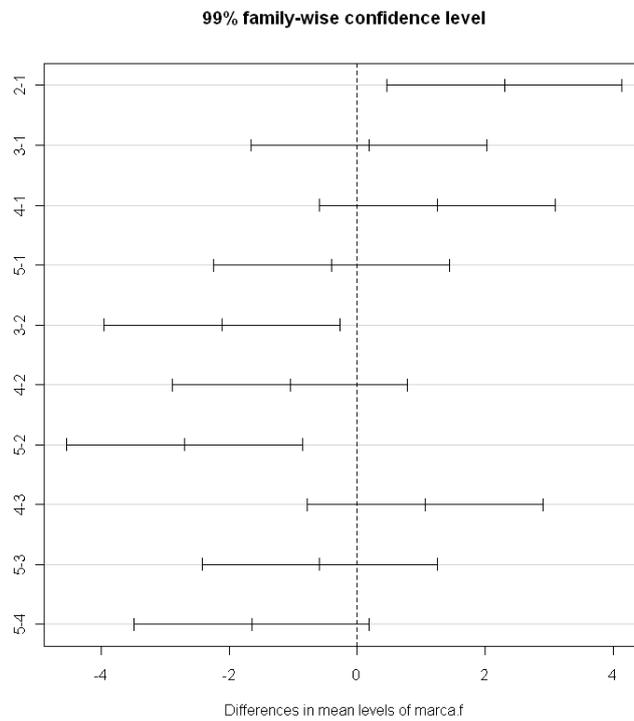
5-2 -2.7000000 -4.5408996 -0.8591004 0.0001419

4-3 1.0666667 -0.7742330 2.9075663 0.2482551

5-3 -0.5833333 -2.4242330 1.2575663 0.7773363

5-4 -1.6500000 -3.4908996 0.1908996 0.0242883

> plot(marca.tuk)



1	2	3	Nota

Apellido y Nombre:
e-mail:

LU N°:

Estadística (Químicos)
Recuperatorio del Segundo Parcial – Segundo Cuatrimestre 2007

Puntaje y criterio de aprobación: El puntaje figura en cada ejercicio. Para aprobar es necesario obtener al menos 60 puntos.

Observación: En todos los ejercicios, defina claramente las variables aleatorias y los parámetros involucrados.

Ejercicio 1 (30 puntos) Se realizó un experimento para determinar en qué forma cuatro temperaturas específicas de cocido afectan la densidad de cierto tipo de ladrillos. La experiencia dio como resultado los siguientes datos de densidad a las cuatro temperaturas consideradas:

100°	125°	150°	175°
21.8	21.4	21.9	21.9
21.9	21.7	21.8	21.7
21.7	21.5	21.8	21.8
21.6	21.5	21.6	21.7
21.7	21.6	21.5	21.6
21.5	21.4	21.8	21.8
21.8	21.3		

- Plantee la hipótesis que se desea testear definiendo claramente los parámetros involucrados en este problema.
- Testee las hipótesis planteadas en a) con nivel de significación 0.05. ¿Qué test/salida emplea para tomar esta decisión?
- Indique los supuestos necesarios bajo los cuales el test F del ANOVA es válido. En base a las salidas del Statistix verifique si se cumplen los supuestos. Indique en cada caso en qué test y/o gráfico basa su conclusión.
- Usando algún método para calcular todos los intervalos de confianza de nivel simultáneo 0.95 para la diferencia de medias entre las temperaturas, construya un intervalo para la diferencia entre las temperaturas 100° y 150°.
- Si efectivamente existen diferencias ¿Qué pares de temperaturas difieren significativamente entre sí con un nivel de significación simultáneo del 5%? Explique en qué método/s y en qué salida/s basa su conclusión.

ONE-WAY AOV FOR: F100 F125 F150 F175					
SOURCE	DF	SS	MS	F	P
BETWEEN	3	0.31337	0.10446	5.96	0.0039
WITHIN	22	0.38548	0.01752		
TOTAL	25	0.69885			
		CHI-SQ	DF	P	
BARTLETT'S TEST OF		-----	-----	-----	
EQUAL VARIANCES		0.61	3	0.8947	
COCHRAN'S Q			0.3245		
LARGEST VAR / SMALLEST VAR			2.0606		

COMPONENT OF VARIANCE FOR BETWEEN GROUPS 0.01340
 EFFECTIVE CELL SIZE 6.5

SAMPLE VARIABLE	GROUP MEAN	SIZE	STD DEV
F100	21.714	7	0.1345
F125	21.486	7	0.1345
F150	21.733	6	0.1506
F175	21.750	6	0.1049
TOTAL	21.665	26	0.1324

CASES INCLUDED 26 MISSING CASES 6

KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV

VARIABLE	MEAN RANK	SAMPLE SIZE
F100	15.4	7
F125	5.8	7
F150	16.7	6
F175	17.1	6
TOTAL	13.5	26

KRUSKAL-WALLIS STATISTIC 10.2659
 P-VALUE, USING CHI-SQUARED APPROXIMATION 0.0164

PARAMETRIC AOV APPLIED to RANKS

SOURCE	DF	SS	MS	F	P
BETWEEN	3	579.815	193.272	5.11	0.0078
WITHIN	22	832.185	37.8266		
TOTAL	25	1412.00			

TOTAL NUMBER OF VALUES THAT WERE TIED 25
 MAX. DIFF. ALLOWED BETWEEN TIES 0.00001

CASES INCLUDED 26 MISSING CASES 6

BONFERRONI COMPARISON OF MEANS

VARIABLE	MEAN	HOMOGENEOUS GROUPS
F175	21.750	I
F150	21.733	I
F100	21.714	I
F125	21.486	.. I

THERE ARE 2 GROUPS IN WHICH THE MEANS ARE NOT SIGNIFICANTLY DIFFERENT FROM ONE ANOTHER.

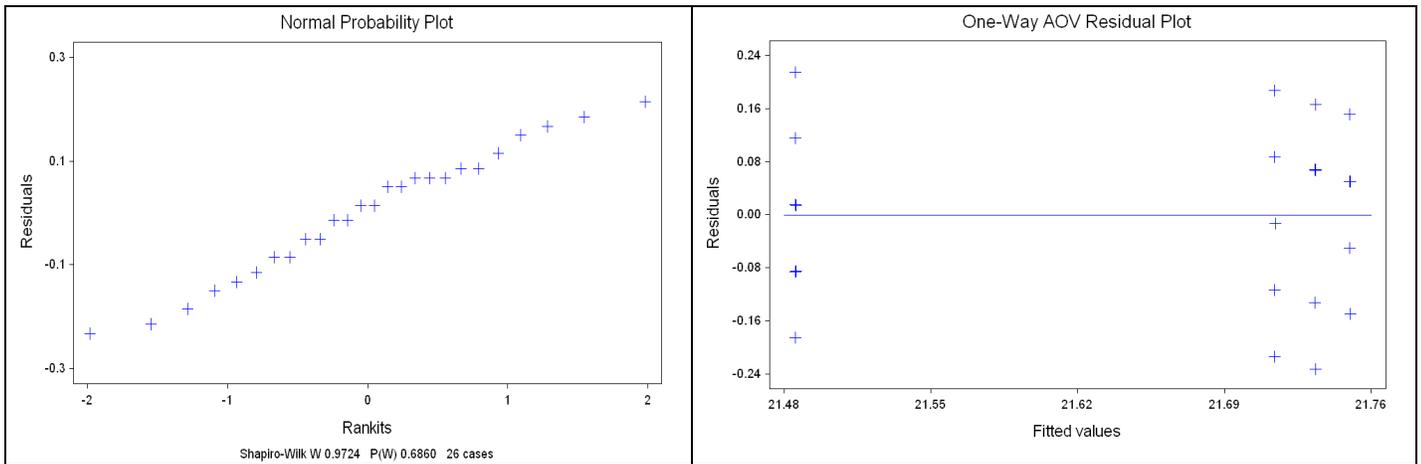
CRITICAL T VALUE 2.899 REJECTION LEVEL 0.050
 STANDARD ERRORS AND CRITICAL VALUES OF DIFFERENCES VARY BETWEEN COMPARISONS BECAUSE OF UNEQUAL SAMPLE SIZES.

TUKEY (HSD) COMPARISON OF MEANS

VARIABLE	MEAN	HOMOGENEOUS GROUPS
F175	21.750	I
F150	21.733	I
F100	21.714	I
F125	21.486	.. I

THERE ARE 2 GROUPS IN WHICH THE MEANS ARE NOT SIGNIFICANTLY DIFFERENT FROM ONE ANOTHER.

CRITICAL Q VALUE 3.928 REJECTION LEVEL 0.050
 STANDARD ERRORS AND CRITICAL VALUES OF DIFFERENCES



Ejercicio 2 (35 puntos) Se consideran dos fórmulas químicas A y B para un nuevo producto que se utilizará para teñir telas. La empresa productora está interesada en telas especialmente resistentes a perder color tras la exposición al sol. Diez piezas de diferentes tejidos se cortan en dos mitades y a cada una se le aplica uno de los dos tintes. Los 20 trozos de tela se exponen al sol durante un periodo de tiempo, al cabo del cual se mide la intensidad del color, obteniéndose:

		Telas									
Fórmulas		1	2	3	4	5	6	7	8	9	10
A		7.2	4.3	5.8	6.5	4.9	6.8	6.3	7.0	6.5	6.2
B		5.1	4.1	5.5	4.1	5.0	5.1	5.3	7.3	4.8	5.8

(bajos resultados indican menos intensidad, es decir, mayor pérdida de color)

Se supone que la diferencia entre la intensidad de color con la fórmula A y la intensidad de color con la fórmula B sigue una distribución $N(\mu_A - \mu_B, \sigma^2)$. Se desea saber si existen diferencias entre las medias de ambas fórmulas.

- Plantee las hipótesis correspondientes y proponga un test de nivel 0.05 para las hipótesis planteadas.
- Construir un intervalo de confianza de nivel 0.95 para la diferencia de intensidad media entre las fórmulas. ¿Tiene razones la empresa productora para sospechar que la fórmula B es mejor que la fórmula A?
- Suponiendo la varianza poblacional $\sigma^2 = 1$, calcular la probabilidad de cometer error de tipo II cuando la diferencia de intensidad de color entre A y B es de 0.97.

\bar{X}_A	\bar{X}_B	\bar{X}_{A-B}	s_A	s_B	s_{A-B}
6.15	5.21	0.94	0.92	0.91	0.97

Ejercicio 3: (35puntos) El valor energético (en kcal. por cada 100g.) de galletitas de agua (Y) se relaciona con la cantidad de grasas totales (en g.) (X) involucradas en su producción. Un experimentador toma una muestra de tamaño 49 (es decir, compra 49 paquetes de galletitas) para verificar la adecuación de un modelo de regresión lineal a esta relación. Utilizando las salidas que se dan a continuación, responda a las siguientes preguntas:

- Expresar el modelo de regresión lineal indicando claramente los parámetros y variables involucradas. Escribir los supuestos necesarios para que sean válidas las conclusiones respecto de los tests y los intervalos de confianza.
- Analice la adecuación del modelo. Indique en que salidas/gráficos se basan sus conclusiones.

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF Y

PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P
CONSTANT	14.9243	39.0974	0.38	0.7044
X	4.40329	0.99116	4.44	0.0001

R-SQUARED	0.2957	RESID. MEAN SQUARE (MSE)	7647.90
ADJUSTED R-SQUARED	0.2808	STANDARD DEVIATION	87.4523

SOURCE	DF	SS	MS	F	P
REGRESSION	1	150942	150942	19.74	0.0001
RESIDUAL	47	359451	7647.90		
TOTAL	48	510393			

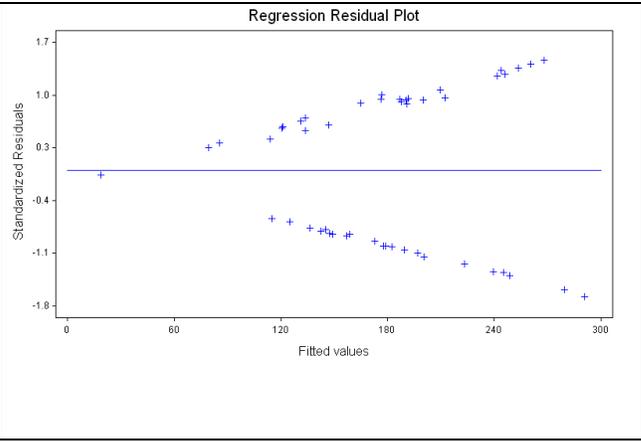
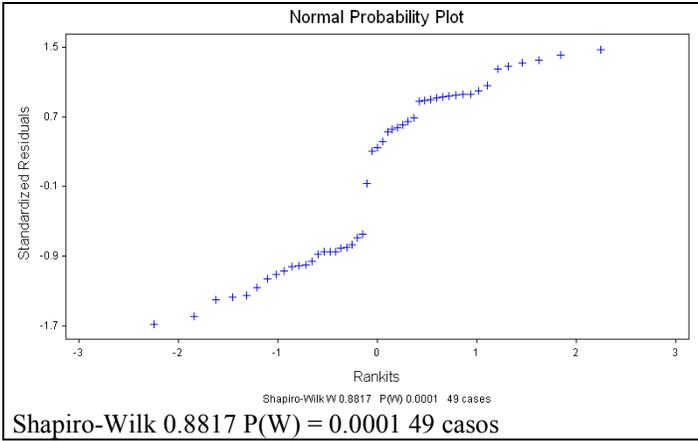
CASES INCLUDED 49 MISSING CASES 0

PREDICTED/FITTED VALUES OF Y

LOWER PREDICTED BOUND	-126.96	LOWER FITTED BOUND	-1.1412
PREDICTED VALUE	58.957	FITTED VALUE	58.957
UPPER PREDICTED BOUND	244.87	UPPER FITTED BOUND	119.06
SE (PREDICTED VALUE)	92.414	SE (FITTED VALUE)	29.874

UNUSUALNESS (LEVERAGE)	0.1167
PERCENT COVERAGE	95.0
CORRESPONDING T	2.01

PREDICTOR VALUES: X = 10.000

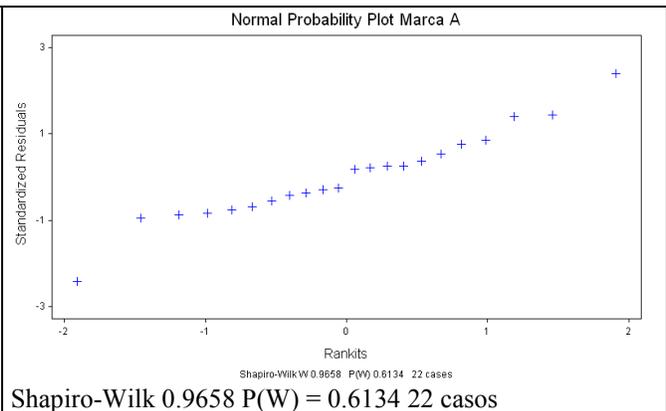
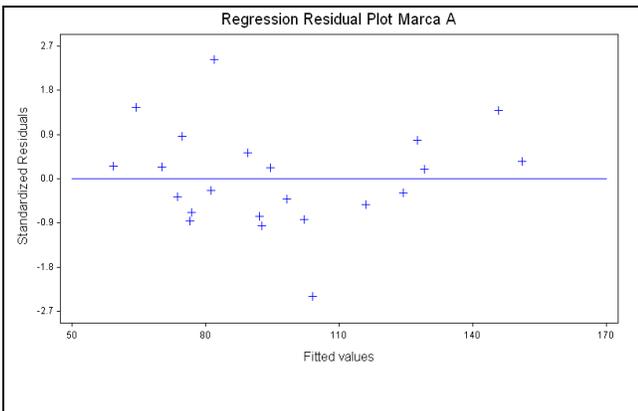


El experimentador observa que los 49 paquetes en verdad son de dos marcas diferentes: 22 paquetes son de la marca A y los 27 restantes son de la marca B. Entonces decide ajustar un modelo de regresión lineal para cada una de las dos marcas. A continuación figuran las salidas de estos dos nuevos ajustes, y los gráficos correspondientes.

- c) Analice la adecuación del modelo para cada uno de los dos ajustes (o sea, para cada una de las dos marcas). Indique en qué salidas/gráficos se basan sus conclusiones.
- d) Al investigador le interesa calcular el valor energético esperado para 100g. de galletitas de agua producidas con $X = 10g.$ de grasas totales. Por suerte, este valor está dentro del rango de las X consideradas en este experimento. Calcúlelo, si con los datos de los que dispone está en condiciones de hacerlo. Sino, no haga nada en este ítem y pase al siguiente.
- e) Ídem d) pero ahora con la información extra de que la marca de dichas galletitas es A.
- f) Dé un intervalo de confianza de 0.95 del valor esperado cuya estimación calculó en el ítem d) o en el ítem e) según lo que haya contestado. (Dé sólo uno, si contestó afirmativamente en ambas preguntas elija uno y justifique su elección). Para ese mismo ajuste, obtenga la estimación de la recta por cuadrados mínimos y diga cuáles parámetros son significativos al nivel 0.05.

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF Y (marca A)					
PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P	
CONSTANT	6.78528	1.13718	5.97	0.0000	
X	2.30802	0.02809	82.17	0.0000	
R-SQUARED	0.9970	RESID. MEAN SQUARE (MSE)	2.08728		
ADJUSTED R-SQUARED	0.9969	STANDARD DEVIATION	1.44474		
SOURCE	DF	SS	MS	F	P
REGRESSION	1	14093.6	14093.6	6752.14	0.0000
RESIDUAL	20	41.7457	2.08728		
TOTAL	21	14135.4			
CASES INCLUDED 22 MISSING CASES 0					

PREDICTED/FITTED VALUES OF Y			
LOWER PREDICTED BOUND	26.347	LOWER FITTED BOUND	28.050
PREDICTED VALUE	29.866	FITTED VALUE	29.866
UPPER PREDICTED BOUND	33.384	UPPER FITTED BOUND	31.681
SE (PREDICTED VALUE)	1.6865	SE (FITTED VALUE)	0.8701
UNUSUALNESS (LEVERAGE)	0.3627		
PERCENT COVERAGE	95.0		
CORRESPONDING T	2.09		
PREDICTOR VALUES: X = 10.000			



UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF Y (MarcaB)

PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P
CONSTANT	7.43921	2.81579	2.64	0.0140
X	6.63879	0.07299	90.95	0.0000

R-SQUARED	0.9970	RESID. MEAN SQUARE (MSE)	26.8384
ADJUSTED R-SQUARED	0.9969	STANDARD DEVIATION	5.18058

SOURCE	DF	SS	MS	F	P
REGRESSION	1	222028	222028	8272.77	0.0000
RESIDUAL	25	670.960	26.8384		
TOTAL	26	222699			

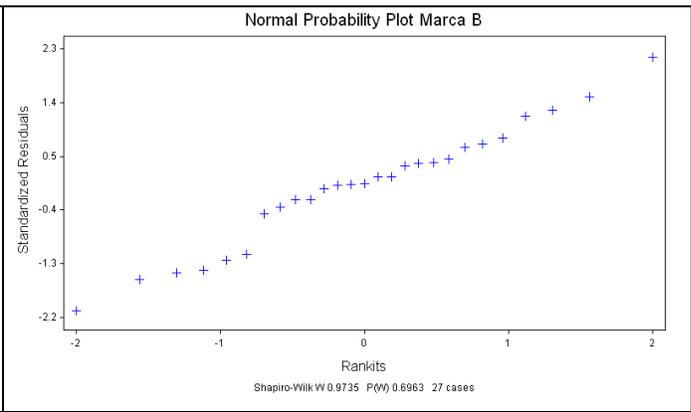
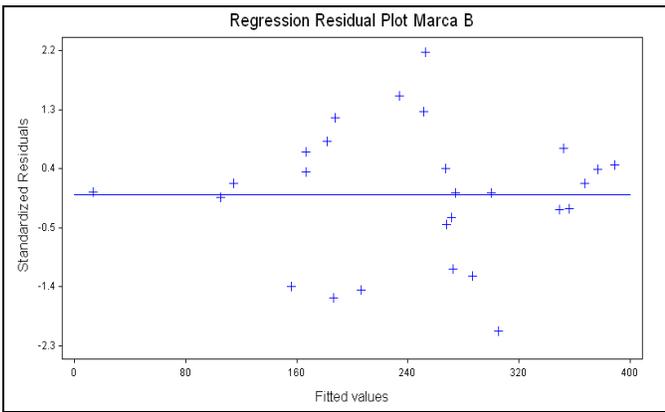
CASES INCLUDED 27 MISSING CASES 0

PREDICTED/FITTED VALUES OF Y

LOWER PREDICTED BOUND	62.276	LOWER FITTED BOUND	69.402
PREDICTED VALUE	73.827	FITTED VALUE	73.827
UPPER PREDICTED BOUND	85.378	UPPER FITTED BOUND	78.253
SE (PREDICTED VALUE)	5.6085	SE (FITTED VALUE)	2.1488

UNUSUALNESS (LEVERAGE)	0.1720
PERCENT COVERAGE	95.0
CORRESPONDING T	2.06

PREDICTOR VALUES: X = 10.000



```

> temper<-c(tempe$F100,tempe$F125,tempe$F150[-7],tempe$F175[-7])
> clase<-c(rep(100,7),rep(125,7),rep(150,6),rep(175,6))
> clase<-as.factor(clase)
> salida<-aov(temper~clase)
> summary(salida)

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
clase	3	0.3134	0.10446	5.962	0.00391 **
Residuals	22	0.3855	0.01752		

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

> TukeyHSD(salida)
  Tukey multiple comparisons of means
  95% family-wise confidence level

```

```

Fit: aov(formula = temper ~ clase)

```

\$clase	diff	lwr	upr	p adj
125-100	-0.22857143	-0.42504476	-0.0320981	0.0186686
150-100	0.01904762	-0.18544830	0.2235435	0.9937559
175-100	0.03571429	-0.16878164	0.2402102	0.9616341
150-125	0.24761905	0.04312312	0.4521150	0.0138606
175-125	0.26428571	0.05978979	0.4687816	0.0082423
175-150	0.01666667	-0.19554878	0.2288821	0.9962263