1	2	3	4

Calif.

Estadística (Química)

Segundo Parcial - 21/11/2013

Apellido y Nombre: Libreta Nro.:

Lor ejercicios 1., 3. y 4. deben ser respondidos en las hojas de enunciado, mientras que para el ejercicio 2., de ser necesario puede anexar alguna/s hoja/s.

Criterio de aprobación: Para aprobar este examen es necesario sumar 60 puntos.

* Colocar nombre, apellido y L.U. en cada hoja entregada.

* Antes de retirarse debe firmar la hoja de asistencia.

---

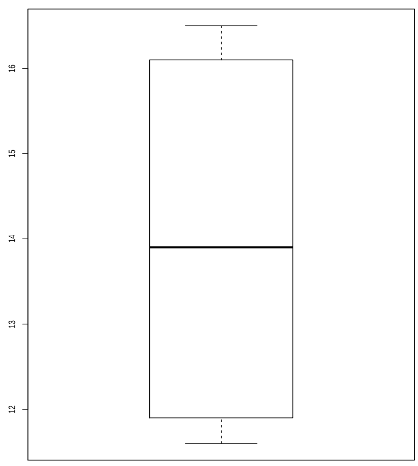
1. *(20p)* Para probar la eficacia de un nuevo servidor web, un inquieto cibernético decidió enviarse a sí mismo, por correo electrónico, el mismo mensaje varias veces y registrar el tiempo, en segundos, transcurrido entre el envío y la recepción. Luego de realizarlo 18 veces obtuvo los siguientes valores, ordenados de menor a mayor:

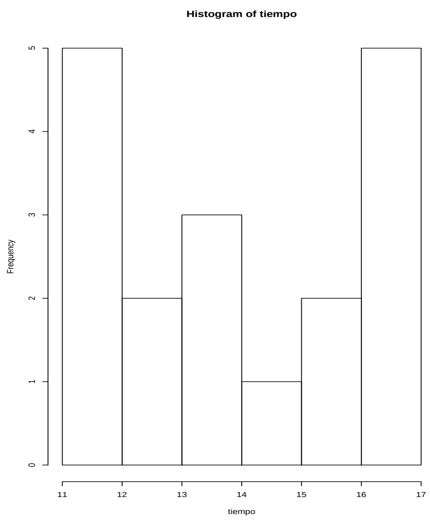
   11.6 11.6 11.7 11.7 11.9 12.4 12.7 13.6 13.8 14.0 14.4 15.2 15.6 16.1 16.3 16.3 16.4 16.5

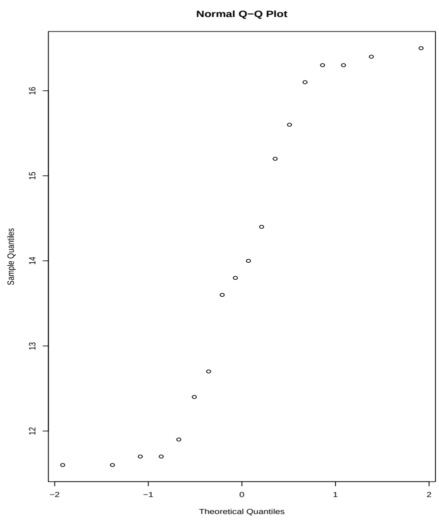
   Él considera que el servidor es eficiente si el tiempo medio que tarda un mensaje en recibirse luego de ser enviado es menor a 15 segundos. ¿Tiene evidencia significativa para decir que el nuevo servidor es eficiente? Para responder a esta pregunta responda los siguientes ítems basándose en la siguiente información:

```
>tiempo<-c(11.6,11.6,11.7,11.7,11.9,12.4,12.7,13.6,13.8,14.0,14.4,15.2,15.6,16.1,16.3,16.3,16.4,16.5)
```

```
> boxplot(tiempo)                   > hist(tiempo)                    > qqnorm(tiempo)
```








```
> shapiro.test(tiempo)

        Shapiro-Wilk normality test

data:  tiempo
W = 0.8729, p-value = 0.01989
```

a) *(4p)* Si se reemplaza el valor 14.4 por 20, cómo se ve afectada la mediana de la muestra:
| aumenta | disminuye | no cambia | (Tache lo que no corresponda.)

b) *(3p)* Defina las variables aleatorias involucradas y el parámetro de interés:
....
....

c) *(4p)* Las hipótesis a testear son H_0: .... vs. H_1: ....

d) *(4p)* De la lista de tests obtenidos con el R, que figura al final del ejercicio, qué número elige para testear las hipótesis: ....
¿Por qué?
....
....
....
....

e) *(3p)* ¿Qué decisión toma a nivel 5 %?
☐ Rechazo H_0, porque ....
☐ No rechazo H_0, porque ....

f) *(2p)* Escriba la conclusión en términos del problema: ....
....

<u>Salidas del R</u>:

Nº1:
```
> t.test(tiempo,alternative="two.sided",mu=15)

        One Sample t-test

data:  tiempo
t = -2.2441, df = 17, p-value = 0.03843
alternative hypothesis: true mean is not equal to 15
95 percent confidence interval:
 13.03830 14.93948
sample estimates:
mean of x
 13.98889
```

Nº2:
```
> t.test(tiempo,alternative="less",mu=15)

        One Sample t-test

data:  tiempo
t = -2.2441, df = 17, p-value = 0.01921
alternative hypothesis: true mean is less than 15
95 percent confidence interval:
     -Inf 14.77268
sample estimates:
mean of x
 13.98889
```

Nº3:
```
> wilcox.test (tiempo , alternative ="two.sided", mu=15, exact=FALSE)

        Wilcoxon signed rank test with continuity correction

data:  tiempo
V = 43.5, p-value = 0.07058
alternative hypothesis: true location is not equal to 15
```

Nº4:
```
> wilcox.test (tiempo , alternative ="less", mu=15, exact=FALSE)

        Wilcoxon signed rank test with continuity correction

data:  tiempo
V = 43.5, p-value = 0.03529
alternative hypothesis: true location is less than 15
```

Nº5:
```
> SIGN.test (tiempo, md=15, alternative ="two.sided")

        One-sample Sign-Test

data:  tiempo
s = 7, p-value = 0.4807
alternative hypothesis: true median is not equal to 15
95 percent confidence interval:
 12.04622 15.95378
sample estimates:
median of x
       13.9

                  Conf.Level  L.E.pt  U.E.pt
Lower Achieved CI     0.9037 12.4000 15.6000
Interpolated CI       0.9500 12.0462 15.9538
Upper Achieved CI     0.9691 11.9000 16.1000
```

Nº6:
```
> SIGN.test (tiempo, md=15, alternative ="less")

        One-sample Sign-Test

data:  tiempo
s = 7, p-value = 0.2403
alternative hypothesis: true median is less than 15
95 percent confidence interval:
     -Inf 15.58942
sample estimates:
median of x
       13.9

                  Conf.Level L.E.pt  U.E.pt
Lower Achieved CI     0.8811   -Inf 15.2000
Interpolated CI       0.9500   -Inf 15.5894
Upper Achieved CI     0.9519   -Inf 15.6000
```

Apellido y Nombre: Libreta Nro.:

2. *(30p)* Se quieren comparar dos métodos para la determinación de cromo en pasturas. Para ello se realizan diez mediciones en la misma muestra, cinco con cada método, obteniéndose los siguientes resultados.

Método 1: media muestral 1.48; desviación estándar muestral 0.28
Método 2: media muestral 2.33; desviación estándar muestral 0.31

Resolver los siguientes items suponiendo que las muestras son normales.

a) *(3p)* Definir las variables aleatorias y los parámetros involucrados en el problema.

b) *(6p)* ¿Hay evidencias a nivel 0.05 de que las varianzas de los dos métodos son diferentes? Escribir las hipótesis a testear, el estadístico del test, su distribución bajo la hipótesis nula y la región de rechazo. ¿Cuál es su conclusión?

c) *(7p)* ¿Hay evidencias significativas de que las medias de ambos métodos difieren? Escribir las hipótesis a testear, el estadístico del test y su distribución bajo la hipótesis nula. Acotar el p-valor. ¿Cuál es su conclusión?

d) *(4p)* Si se tuviera la información adicional de que la media del método 2 no puede ser menor que la del método 1. ¿Cuáles serían las hipótesis a testear? ¿Hay evidencia significativa para rechazar la hipótesis nula a nivel 5 %?

e) *(4p)* Supongamos ahora que el desvío estándar poblacional es conocido y vale σ =0.3 para ambos métodos. ¿Cuál sería el estadístico del test para las hipótesis del item anterior? ¿Y la región de rechazo?

f) *(6p)* Realizando la misma cantidad de determinaciones por método y suponiendo que la verdadera media del método 2 excede en 0.5 a la verdadera media del método 1, ¿cuántas mediciones deberían hacerse por método, si se quiere que la probabilidad de error de tipo II del test del ítem anterior sea a lo sumo 0.1?

Apellido y Nombre: Libreta Nro.:

3. *(20p)* Un gobierno provincial realiza una evaluación de opción múltiple a estudiantes de quinto año en 5 localidades, calificándolos entre 0 y 100. Desea saber si hay diferencias significativas entre los desempeños de las distintas localidades, con el objeto de reforzar el presupuesto educativo en aquellas con peor desempeño. A pesar de su buena voluntad a futuro, no tiene presupuesto asignado para pagar las correcciones de estos exámenes, así que el ministro elige apenas 20 exámenes de cada localidad al azar entre las respectivas pilas y se pone a corregir.

   a) *(4p)* Describir las variables aleatorias involucradas, los parámetros y los supuestos para realizar un test F de Análisis de la varianza. ¿Cuáles son las hipótesis del test?

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   ....

   A continuación se detallan las salidas obtenidas en R:

```
> anovanotas<-aov(notas~localidad)
> shapiro.test(anovanotas$res)

        Shapiro-Wilk normality test

data:  anovanotas$res
W = 0.9837, p-value = 0.2559

> levene.test(notas,localidad)

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

data:  notas
Test Statistic = 0.4128, p-value = 0.799

> summary(anovanotas)
            Df Sum Sq Mean Sq F value   Pr(>F)
localidad    4   4796  1199.1   5.298 0.000678 ***
Residuals   95  21501   226.3
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

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

Fit: aov(formula = notas ~ localidad)

$localidad
     diff         lwr      upr     p adj
2-1 -6.75 -19.9796199  6.47962 0.6171542
3-1 10.00  -3.2296199 23.22962 0.2278144
4-1 12.50  -0.7296199 25.72962 0.0734032
5-1  4.10  -9.1296199 17.32962 0.9100341
3-2 16.75   3.5203801 29.97962 0.0058434
4-2 19.25   6.0203801 32.47962 0.0009858
5-2 10.85  -2.3796199 24.07962 0.1602573
4-3  2.50 -10.7296199 15.72962 0.9845509
5-3 -5.90 -19.1296199  7.32962 0.7279397
5-4 -8.40 -21.6296199  4.82962 0.3995901
```

A continuación se calculan las medias muestrales de notas por localidad

```
> tapply(notas,localidad,mean)
    1     2     3     4     5
63.95 57.20 73.95 76.45 68.05
```

b) *(4p)* ¿Se cumplen los supuestos del modelo para aplicar el test Anova? Fundamente brevemente:

....

....

....

c) *(4p)* ¿Qué conclusión puede sacar el ministro?

....

....

....

d) *(4p)* ¿Existen pares de localidades para las cuales se puede decir que la media de sus notas es significativamente diferente con nivel simultáneo 5 %? Fundamente brevemente

....

....

....

e) *(4p)* Si se decide subir el presupuesto en el mayor grupo de localidades de peores resultados en el que no se encuentren diferencias significativas a nivel simultáneo 5 %. ¿En qué localidades deberían subir el presupuesto? Fundamente brevemente

....

....

....

Apellido y Nombre: Libreta Nro.:

4. *(30p)* Las materias primas que utiliza una fábrica en la producción de una fibra sintética son almacenadas en un lugar donde no se tiene control sobre la humedad. Para analizar si la humedad del almacenamiento influye sobre la humedad de la materia prima almacenada, se tomaron muestras de la materia prima durante 13 días. Cada día se midió el porcentaje de humedad de la muestra y del lugar de almacenamiento. Los datos obtenidos son los siguientes:

Porcentajes de humedad:													
de la materia prima	29	32	35	36	39	42	43	44	48	48	50	55	57
de almacenamiento	5	9	8	9	10	12	10	12	11	11	14	13	15

a) *(3p)* Proponiendo un modelo de regresión lineal para los datos, $Y_i = \alpha + \beta X_i + \epsilon_i$, donde

$X_i =$ ....,

$Y_i =$ ....,

los supuestos que se hacen sobre los ϵ_i son:

....

b) *(6p)* A continuación se encuentra la salida de R del ajuste lineal junto con algunas medidas resumen de X e Y y algunas probabilidades de la distribución t. La salida del R se encuentra incompleta. Completar sobre las líneas punteadas con los valores que deberían verse en la salida.

```
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  29.00   36.00   43.00   42.92   48.00   57.00
> summary(y)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   5.00    9.00   11.00   10.69   12.00   15.00

> summary(lm(y~x))
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-1.7883 -1.0787  0.2489  1.0057  1.5665

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) ........    1.71599  -0.783 .......
x            0.28040    0.03926  ...... 1.89e-05 ***

Residual standard error: 1.169 on 11 degrees of freedom
Multiple R-squared: 0.8226,     Adjusted R-squared: 0.8065
F-statistic: 51.01 on 1 and 11 DF,  p-value: 1.887e-05
```

```
> pt(0.783,11)                  > qt(0.783,12)
[1] 0.774922                   [1] 0.8094746
> qt(0.783,11)                  > pt(1.71599,11)
[1] 0.8120221                  [1] 0.9429221
> pt(0.783,12)                  > pt(1.71599,12)
[1] 0.7755978                  [1] 0.9440786
```

c) *(2p)* Completar el cuadro con las estimaciones de los parámetros que se obtienen con los datos observados

parámetro	estimación
α	
β	
σ	

d) *(3p)* Calcular un intervalo de confianza de nivel 0.95 para la ordenada al origen. Escribir los extremos numéricos de dicho intervalo:

(. . . . . , ...)

e) *(5p)* Decidir si la pendiente es significativa al 5 %. Para ello se necesitan testear las siguientes hipótesis.
$H_0 =$ . . . . vs $H_1 =$ . . . .
El estadístico del test es . . . .
y su distribución bajo H_0 es . . . .
El p-valor para estos datos es . . . .
y por lo tanto la hipótesis nula es | rechazada | no rechazada | (Tachar lo que no corresponda)

f) *(4p)* Marque con una cruz la opción correcta. No justificar.

i.

	Con un nivel de significación de 5 % la humedad del lugar de almacenamiento influye en la humedad de la materia prima.
	Con un nivel de significación de 5 % la humedad del lugar de almacenamiento **no** influye en la humedad de la materia prima.

ii.

	El porcentaje de la humedad media estimada de las telas aumenta 0.28 unidades cuando el porcentaje de la humedad del lugar de almacenamiento aumenta una unidad.
	El porcentaje de la humedad media estimada de las telas disminuye 0.28 unidades cuando el porcentaje de la humedad del lugar de almacenamiento aumenta una unidad.

g) *(2p)* El valor del coeficiente de determinación R^2 es . . . . Esto significa que el porcentaje de variabilidad en la humedad de las telas **no** explicada por el modelo de regresión lineal es . . . .

h) *(2p)* Se desea predecir cuál será la humedad de una muestra almacenada en un lugar con una humedad de 45 %. Dar dicha predicción . . . .

i) *(1p)* El intervalo de predicción de una nueva observación siempre es más . . . . que el intervalo de confianza para su esperanza.

j) *(2p)* A continuación se da un diagrama de dispersión de los datos. Sobre el mismo gráfico dibujar la recta de regresión, no a ojo! Las cuentas realizadas deben figurar al costado del mismo.