

Regresión y ANOVA:

ANOVA de 1 factor

Cuervo Fernández, Esther

García Prado, Sergio

Martín Villares, Pablo

19 de octubre de 2017

1. Ejercicio 7

El enunciado del ejercicio es el siguiente: “Se realiza un experimento para estudiar la efectividad de un nuevo tipo de somnífero, para lo cual se administra el mismo a 6 pacientes, mientras que a otros 6 se les da un somnífero estándar y a 6 más un placebo, durante una semana. En la tabla 6.33 se recogen las medias del número de horas de sueño en las 7 noches para los 18 pacientes. (a=placebo, b= estándar, c=nuevo)”

Representamos los datos con un boxplot para observar su distribución de manera gráfica.

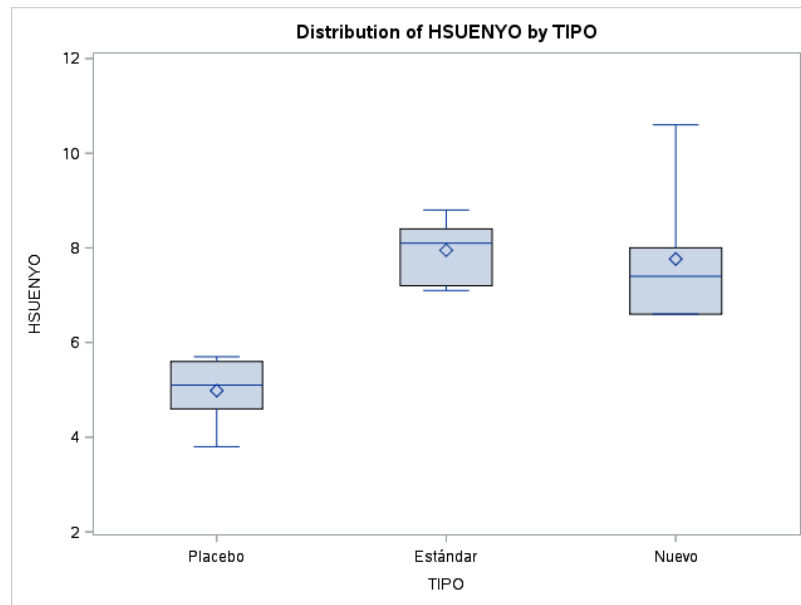


Figura 1: Box Plot de la variable **horas de sueño** condicionada por la variable **tipo de somnífero**

Podemos ver que la distribución de Nuevo y Estandar parece similar, con Placebo teniendo horas de sueño por debajo de las otras dos.

Respecto a un análisis de las medias, obtendremos la comparación de ambos tipos de somnífero contra el control que ofrece el Placebo.

```
proc anova;
  class TIPO;
  model HSUENYO=TIPO;
  means TIPO/dunnett('Placebo');
run;
```

The ANOVA Procedure					
Dependent Variable: HSUENYO					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	33.16333333	16.58166667	15.04	0.0003
Error	15	16.53666667	1.10244444		
Corrected Total	17	49.70000000			

R-Square	Coeff Var	Root MSE	HSUENYO Mean
0.667270	15.21701	1.049974	6.900000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
TIPO	2	33.16333333	16.58166667	15.04	0.0003

proc anova nos proporciona de paso con el resultado del test de la hipótesis nula de que las medias de los tres grupos es igual, con un valor $Pr > F$ 0.0003 podemos descartar esta hipótesis.

El test de dunnett ofrece intervalos de confianza para las diferencias de medias $\mu_{Placebo} - \mu_i$ siendo i Estándar y Nuevo.

The ANOVA Procedure					
Dunnett's t Tests for HSUENYO					
Note: This test controls the Type I experimentwise error for comparisons of all treatments against a control.					
Alpha		0.05			
Error Degrees of Freedom		15			
Error Mean Square		1.102444			
Critical Value of Dunnett's t		2.43926			
Minimum Significant Difference		1.4787			

Comparisons significant at the 0.05 level are indicated by ***.				
TIPO Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Estándar - Placebo	2.9667	1.4880	4.4453	***
Nuevo - Placebo	2.7833	1.3047	4.2620	***

Los resultados de este test nos permiten rechazar la hipótesis de que cualquiera de los somníferos no tienen ningún efecto, ya que hay diferencias significativas entre ambos y el placebo a un nivel de confianza del 95 %.

Sabiendo esto, realizamos un test de la T para comprobar si existen diferencias significativas entre las medias del somnífero estándar y del nuevo. Haremos esto mediante un contraste $L = 0\mu_{Placebo} + \mu_{Nuevo} - \mu_{Estndar}$. Si este contraste tiene valor 0, implica que las medias son iguales. Obtenemos además el estimador de L , que nos permite conocer, en caso de no ser 0, si L es positivo (el nuevo somnífero es mejor) o negativo (peor). `clparm` nos permite además obtener un intervalo de confianza para el estimador.

```
proc glm;
  class TIPO;
  model HSUENYO=TIPO/clparm;
  contrast 'Eficacia de somnifero nuevo' TIPO -1 1 0;
  estimate 'Eficacia de somnifero nuevo' TIPO -1 1 0;
run;
```

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Eficacia de somnifero nuevo	1	0.10083333	0.10083333	0.09	0.7665

Parameter	Estimate	Standard Error	t Value	Pr > t	95% Confidence Limits	
Eficacia de somnifero nuevo	-0.18333333	0.60620251	-0.30	0.7665	-1.47542339	1.10875673

El resultado del test sobre el contraste da un $Pr > F$ del 0.7665, por lo que no podemos rechazar la hipótesis de que los dos somníferos tienen la misma eficacia. Por tanto, con un nivel de confianza del 95% podemos afirmar que el nuevo somnífero no tiene un efecto significativo respecto a un somnífero estándar.

El estimador nos informa de que el somnífero nuevo produce 0.183 horas de sueño menos que el estándar, con un error de ± 1.2921 horas. Esto coloca al efecto del nuevo somnífero entre 1.47 horas peor y 1.10 horas mejor que el estándar.

2. Ejercicio Acuicultura

El enunciado del ejercicio es el siguiente: “Se quiere estudiar el efecto de distintas dosis de un medicamento para combatir a los parásitos de peces criados en acuicultura. Para ello, se tomaron 60 peces al azar, y se dividieron en 5 grupos de 12 individuos cada uno. El primer grupo no fue medicado, pero a los restantes se les suministró el medicamento en dosis crecientes. Tras una semana de tratamiento, se contabilizaron los parásitos existentes en cada individuo”.

Previamente, se ha realizado un preprocesado de los datos, que fueron suministrados de manera matricial, de tal manera que se tenía 5 columnas cada una de ellas referidas a un tipo de tratamiento, junto con 12 filas, cuyos valores en la celda se presuponen independientes entre sí, referidas a las distintas observaciones de cada tratamiento. Por tanto, lo que se ha llevado a cabo es una transformación de dicha representación a una formada por 2 columnas (grupo y niveles) y 60 filas, donde cada una de ellas indica el tratamiento suministrado, junto con el nivel de parásitos obtenido.

2.1. Representa gráficamente los datos. ¿Te parece que el medicamento es efectivo contra los parásitos? Realiza un contraste de hipótesis para verificarlo.

```

proc boxplot data=peces_f;
  plot NIVELES*GRUPO;
run;

proc glm;
  class GRUPO;
  model NIVELES=GRUPO/clparm;
  contrast 'Eficacia medicamento' GRUPO 0.25 0.25 0.25 0.25 -1/e;
  estimate 'Eficacia medicamento' GRUPO 0.25 0.25 0.25 0.25 -1;
run;

```

Lo primero que se ha hecho es realizar una representación gráfica para conocer mejor la relación entre las variables. Por tanto, se ha elegido representar los datos mediante un *Box Plot* que de la variable `niveles` particionada en los subgrupos determinados por el tipo de `grupo`.

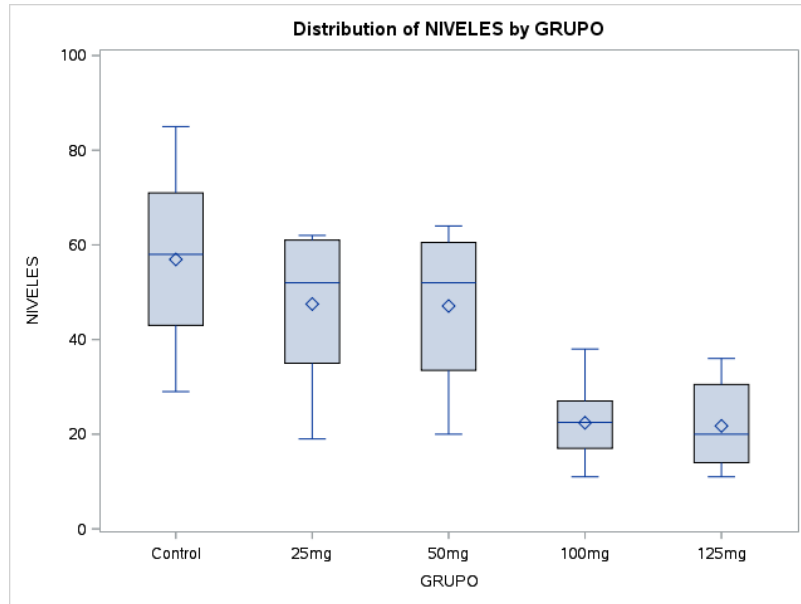


Figura 2: Box Plot de la variable `niveles` condicionada por la variable `grupo`

A partir de dicho gráfico se pueden apreciar dos grupos de medias diferentes, el primero de ellos formado por los tratamientos de `Control`, `25mg` y `50mg`, y el siguiente constituido por los tratamientos de `100mg` y `125mg`.

Para poder asegurar que el tratamiento es efectivo contra los parásitos se ha realizado el siguiente contraste de hipótesis:

$$\begin{aligned}
 H_0 : \mu_{control} &= \frac{\mu_{25mg} + \mu_{50mg} + \mu_{100mg} + \mu_{125mg}}{4} \\
 H_1 : \mu_{control} &\neq \frac{\mu_{25mg} + \mu_{50mg} + \mu_{100mg} + \mu_{125mg}}{4}
 \end{aligned}$$

Si se acepta entonces podemos asegurar que el tratamiento no tiene efecto. Se ha escogido este contraste de hipótesis puesto que lo que se pretende comprobar es si el efecto del medicamento

(en promedio) es diferente de los resultados que se obtienen sin el uso del mismo. Por tanto, se ha planteado como la igualdad de la media del valor de control y el promedio de las medias de cada cantidad de dosis del medicamento.

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Eficacia medicamento	1	4743.704167	4743.704167	26.95	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t	95% Confidence Limits	
Eficacia medicamento	-22.2291667	4.28174419	-5.19	<.0001	-30.8099738	-13.6483596

Dicho hipótesis se ha realizado con un nivel de confianza del 95 %, para la cual se ha obtenido un p-valor ≈ 0 . Por esta razón **tenemos que rechazar la hipótesis nula, lo que significa que el medicamento es efectivo**. Puesto que el estimador obtenido tiene signo negativo, es decir, que los distintos tratamientos disminuyen el nivel de parásitos con respecto del valor de control.

2.2. Obtén estimadores puntuales e intervalos de confianza para las medias de los tratamientos, utilizando dos métodos distintos, y explica las ventajas e inconvenientes de cada uno de ellos.

```
proc means;
  class GRUPO;
  var NIVELES;
run;

proc means clm;
  class GRUPO;
  var NIVELES;
run;

proc univariate cibasic;
  class GRUPO;
  var NIVELES;
run;
```

Para la obtención de los estimadores puntuales así como los intervalos de confianza se han utilizado los *procs* `means` y `univariate` obteniendo los siguientes resultados en cada uno de ellos:

The MEANS Procedure

Analysis Variable : NIVELES						
GRUPO	N Obs	N	Mean	Std Dev	Minimum	Maximum
100mg	12	12	22.4166667	7.7864490	11.0000000	38.0000000
125mg	12	12	21.7500000	8.7399293	11.0000000	36.0000000
25mg	12	12	47.5000000	14.9939382	19.0000000	62.0000000
50mg	12	12	47.0833333	15.4476143	20.0000000	64.0000000
Control	12	12	56.9166667	16.7193863	29.0000000	85.0000000

The MEANS Procedure

Analysis Variable : NIVELES			
GRUPO	N Obs	Lower 95% CL for Mean	Upper 95% CL for Mean
100mg	12	17.4693930	27.3639403
125mg	12	16.1969139	27.3030861
25mg	12	37.9733062	57.0266938
50mg	12	37.2683875	56.8982792
Control	12	46.2936754	67.5396579

The UNIVARIATE Procedure
Variable: NIVELES
GRUPO = Control

Moments			
N	12	Sum Weights	12
Mean	56.9166667	Sum Observations	683
Std Deviation	16.7193863	Variance	279.537879
Skewness	-0.0749322	Kurtosis	-0.8211357
Uncorrected SS	41949	Corrected SS	3074.91667
Coeff Variation	29.3752029	Std Error Mean	4.8264711

Basic Statistical Measures			
Location		Variability	
Mean	56.91667	Std Deviation	16.71939
Median	58.00000	Variance	279.53788
Mode	72.00000	Range	56.00000
		Interquartile Range	28.00000

Basic Confidence Limits Assuming Normality			
Parameter	Estimate	95% Confidence Limits	
Mean	56.91667	46.29368	67.53966
Std Deviation	16.71939	11.84393	28.38748
Variance	279.53788	140.27873	805.84894

The UNIVARIATE Procedure
Variable: NIVELES
GRUPO = 25mg

Moments			
N	12	Sum Weights	12
Mean	47.5	Sum Observations	570
Std Deviation	14.9939382	Variance	224.818182
Skewness	-0.7833979	Kurtosis	-0.7077057
Uncorrected SS	29548	Corrected SS	2473
Coeff Variation	31.5661856	Std Error Mean	4.32837712

Basic Statistical Measures			
Location		Variability	
Mean	47.50000	Std Deviation	14.99394
Median	52.00000	Variance	224.81818
Mode	62.00000	Range	43.00000
		Interquartile Range	26.00000

Basic Confidence Limits Assuming Normality			
Parameter	Estimate	95% Confidence Limits	
Mean	47.50000	37.97331	57.02669
Std Deviation	14.99394	10.62163	25.45788
Variance	224.81818	112.81909	648.10355

The UNIVARIATE Procedure
Variable: NIVELES
GRUPO = 50mg

Moments			
N	12	Sum Weights	12
Mean	47.0833333	Sum Observations	565
Std Deviation	15.4476143	Variance	238.628788
Skewness	-0.5519063	Kurtosis	-1.2651389
Uncorrected SS	29227	Corrected SS	2624.91667
Coeff Variation	32.8090923	Std Error Mean	4.45934214

Basic Statistical Measures			
Location		Variability	
Mean	47.08333	Std Deviation	15.44761
Median	52.00000	Variance	238.62879
Mode	61.00000	Range	44.00000
		Interquartile Range	27.00000

Basic Confidence Limits Assuming Normality			
Parameter	Estimate	95% Confidence Limits	
Mean	47.08333	37.26839	56.89828
Std Deviation	15.44761	10.94302	26.22816
Variance	238.62879	119.74958	687.91663

The UNIVARIATE Procedure
Variable: NIVELES
GRUPO = 100mg

Moments			
N	12	Sum Weights	12
Mean	22.416667	Sum Observations	269
Std Deviation	7.78644899	Variance	60.6287879
Skewness	0.37677956	Kurtosis	0.08885803
Uncorrected SS	6697	Corrected SS	666.916667
Coeff Variation	34.7350884	Std Error Mean	2.24775421

Basic Statistical Measures			
Location		Variability	
Mean	22.41667	Std Deviation	7.78645
Median	22.50000	Variance	60.62879
Mode	27.00000	Range	27.00000
		Interquartile Range	10.00000

Basic Confidence Limits Assuming Normality			
Parameter	Estimate	95% Confidence Limits	
Mean	22.41667	17.46939	27.36394
Std Deviation	7.78645	5.51588	13.22044
Variance	60.62879	30.42496	174.78005

The UNIVARIATE Procedure
Variable: NIVELES
GRUPO = 125mg

Moments			
N	12	Sum Weights	12
Mean	21.75	Sum Observations	261
Std Deviation	8.73992927	Variance	76.3863636
Skewness	0.28193455	Kurtosis	-1.5365018
Uncorrected SS	6517	Corrected SS	840.25
Coeff Variation	40.1835828	Std Error Mean	2.52300026

Basic Statistical Measures			
Location		Variability	
Mean	21.75000	Std Deviation	8.73993
Median	20.00000	Variance	76.38636
Mode	31.00000	Range	25.00000
		Interquartile Range	16.50000

Basic Confidence Limits Assuming Normality			
Parameter	Estimate	95% Confidence Limits	
Mean	21.75000	16.19691	27.30309
Std Deviation	8.73993	6.19132	14.83933
Variance	76.38636	38.33249	220.20583

Estos dos métodos ofrecen los mismos resultados, sin embargo lo que les diferencia es la cantidad de información que aportan cada uno de ellos. En el caso de **means** la información es mucho más compacta y resumida, lo cual facilita su comprensión, sin embargo este método no da los intervalos

de confianza sin la opción CLM. Mientras que en el caso de `univariate` se obtiene información mucho más extendida y ampliada, como los valores de la varianza e intervalos de confianza para los estimadores de la esperanza y la varianza. Para casos en que se requiere de un resumen sencillo del conjunto de datos como es en este caso, es más apropiado el uso del `proc means`

2.3. Analiza las diferencias significativas entre pares de medias de los 5 grupos utilizando diferentes métodos, comentando y comparando los resultados. ¿Cambian las conclusiones si utilizas $\alpha = 0,1$?

```
proc anova;
  class GRUPO;
  model NIVELES=GRUPO;
  means GRUPO/lsd duncan snk tukey bon scheffe lines alpha=0.1;
run;
```

Para la realización de los distintos tests se han utilizando los siguientes métodos: *Método de la T*, *Duncan*, *NewmanKeuls*, *Tukey*, *Bonferroni* y *Scheffé*. Estos se dividen en dos grupos, los que aseguran los resultados para experimentos múltiples con una confianza global de $1-\alpha$ (*NewmanKeuls*, *Tukey*, *Bonferroni* y *Scheffé*) y los que no aseguran dicho nivel de confianza (*Método de la T*, *Duncan*).

The ANOVA Procedure			
t Tests (LSD) for NIVELES			
Note: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.			
Alpha	0.05		
Error Degrees of Freedom	55		
Error Mean Square	176		
Critical Value of t	2.00404		
Least Significant Difference	10.854		

Means with the same letter are not significantly different.			
t Grouping	Mean	N	GRUPO
A	56.917	12	Control
A			
A	47.500	12	25mg
A			
A	47.083	12	50mg
B	22.417	12	100mg
B			
B	21.750	12	125mg

The ANOVA Procedure
Duncan's Multiple Range Test for NIVELES

Note: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	55
Error Mean Square	176

Number of Means	2	3	4	5
Critical Range	10.85	11.42	11.79	12.06

Means with the same letter are not significantly different.			
Duncan Grouping	Mean	N	GRUPO
A	56.917	12	Control
A			
A	47.500	12	25mg
A			
A	47.083	12	50mg
B	22.417	12	100mg
B			
B	21.750	12	125mg

The ANOVA Procedure
Student-Newman-Keuls Test for NIVELES

Note: This test controls the Type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha	0.05
Error Degrees of Freedom	55
Error Mean Square	176

Number of Means	2	3	4	5
Critical Range	10.853515	13.045579	14.348969	15.274968

Means with the same letter are not significantly different.			
SNK Grouping	Mean	N	GRUPO
A	56.917	12	Control
A			
A	47.500	12	25mg
A			
A	47.083	12	50mg
B	22.417	12	100mg
B			
B	21.750	12	125mg

The ANOVA Procedure
Tukey's Studentized Range (HSD) Test for NIVELES

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	55
Error Mean Square	176
Critical Value of Studentized Range	3.98855
Minimum Significant Difference	15.275

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	GRUPO
A	56.917	12	Control
A			
A	47.500	12	25mg
A			
A	47.083	12	50mg
B	22.417	12	100mg
B			
B	21.750	12	125mg

The ANOVA Procedure
Bonferroni (Dunn) t Tests for NIVELES

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	55
Error Mean Square	176
Critical Value of t	2.92470
Minimum Significant Difference	15.84

Means with the same letter are not significantly different.

Bon Grouping	Mean	N	GRUPO
A	56.917	12	Control
A			
A	47.500	12	25mg
A			
A	47.083	12	50mg
B	22.417	12	100mg
B			
B	21.750	12	125mg

**The ANOVA Procedure
Scheffe's Test for NIVELES**

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	55
Error Mean Square	176
Critical Value of F	2.53969
Minimum Significant Difference	17.262

Means with the same letter are not significantly different.			
Scheffe Grouping	Mean	N	GRUPO
A	56.917	12	Control
A			
A	47.500	12	25mg
A			
A	47.083	12	50mg
B	22.417	12	100mg
B			
B	21.750	12	125mg

Se ha probado con un valor $\alpha = 0,05$, para el cual se han obtenido los mismos resultados en todos los tests: Existen dos grupos de medias, el formado por **Control**, **25mg** y **50mg**, y el siguiente constituido por los tratamientos de **100mg** y **125mg** tal y como se podía intuir en el gráfico de la figura .

**The ANOVA Procedure
t Tests (LSD) for NIVELES**

Note: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.1
Error Degrees of Freedom	55
Error Mean Square	176
Critical Value of t	1.67303
Least Significant Difference	9.0612

Means with the same letter are not significantly different.			
t Grouping	Mean	N	GRUPO
A	56.917	12	Control
B	47.500	12	25mg
B			
B	47.083	12	50mg
C	22.417	12	100mg
C			
C	21.750	12	125mg

**The ANOVA Procedure
Duncan's Multiple Range Test for NIVELES**

Note: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.1
Error Degrees of Freedom	55
Error Mean Square	176

Number of Means	2	3	4	5
Critical Range	9.06	9.57	9.91	10.14

Means with the same letter are not significantly different.			
Duncan Grouping	Mean	N	GRUPO
A	56.917	12	Control
B	47.500	12	25mg
B			
B	47.083	12	50mg
C	22.417	12	100mg
C			
C	21.750	12	125mg

The ANOVA Procedure
Student-Newman-Keuls Test for NIVELES

Note: This test controls the Type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha	0.1
Error Degrees of Freedom	55
Error Mean Square	176

Number of Means	2	3	4	5
Critical Range	9.0611953	11.351701	12.708923	13.670537

Means with the same letter are not significantly different.			
SNK Grouping	Mean	N	GRUPO
A	56.917	12	Control
A			
A	47.500	12	25mg
A			
A	47.083	12	50mg
B	22.417	12	100mg
B			
B	21.750	12	125mg

The ANOVA Procedure
Tukey's Studentized Range (HSD) Test for NIVELES

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.1
Error Degrees of Freedom	55
Error Mean Square	176
Critical Value of Studentized Range	3.56960
Minimum Significant Difference	13.671

Means with the same letter are not significantly different.			
Tukey Grouping	Mean	N	GRUPO
A	56.917	12	Control
A			
A	47.500	12	25mg
A			
A	47.083	12	50mg
B	22.417	12	100mg
B			
B	21.750	12	125mg

**The ANOVA Procedure
Bonferroni (Dunn) t Tests for NIVELES**

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.1
Error Degrees of Freedom	55
Error Mean Square	176
Critical Value of t	2.66822
Minimum Significant Difference	14.451

Means with the same letter are not significantly different.			
Bon Grouping	Mean	N	GRUPO
A	56.917	12	Control
A			
A	47.500	12	25mg
A			
A	47.083	12	50mg
B	22.417	12	100mg
B			
B	21.750	12	125mg

**The ANOVA Procedure
Scheffe's Test for NIVELES**

Note: This test controls the Type I experimentwise error rate.

Alpha	0.1
Error Degrees of Freedom	55
Error Mean Square	176
Critical Value of F	2.04997
Minimum Significant Difference	15.509

Means with the same letter are not significantly different.			
Scheffe Grouping	Mean	N	GRUPO
A	56.917	12	Control
A			
A	47.500	12	25mg
A			
A	47.083	12	50mg
B	22.417	12	100mg
B			
B	21.750	12	125mg

Después se ha realizado el mismo conjunto de tests con $\alpha = 0,1$ obteniendo los mismos resultados que para el caso anterior en los tests que aseguran un nivel de confianza global $1 - \alpha$, mientras que en el resto (*Método de la T, Duncan*), estos indican que la media del grupo **Control** es diferente del grupo formado por los tratamientos 25mg y 50mg, por lo que afirman la existencia de 3 grupos de medias diferentes.

2.4. Realiza el test de Dunnett para comparar los efectos de las distintas dosis con el grupo control y comenta el resultado.

```
proc anova;
  class GRUPO;
  model NIVELES=GRUPO;
  means GRUPO/dunnett('Control');
run;
```

The ANOVA Procedure Dunnett's t Tests for NIVELES				
Note: This test controls the Type I experimentwise error for comparisons of all treatments against a control.				
Alpha	0.05			
Error Degrees of Freedom	55			
Error Mean Square	176			
Critical Value of Dunnett's t	2.51460			
Minimum Significant Difference	13.619			

Comparisons significant at the 0.05 level are indicated by ***.				
GRUPO Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
25mg - Control	-9.417	-23.036	4.202	
50mg - Control	-9.833	-23.452	3.786	
100mg - Control	-34.500	-48.119	-20.881	***
125mg - Control	-35.167	-48.786	-21.548	***

Tras aplicar el test de *Dunnett* usando el tratamiento **control** para compararlo con el resto de tratamientos con un nivel de confianza de $\alpha = 0,05$ para todo el experimento, se ha concluido que las medias de los tratamientos 100mg y 125mg son diferentes de las de **control**. Sin embargo, las de 25mg y 50mg son tomadas como iguales.

2.5. Se desea comparar el efecto de las dosis bajas del medicamento, 25-50 mg con el de las dosis altas, 100-125 mg. Construye un intervalo de confianza para la diferencia entre ambos efectos e interpreta el resultado obtenido


```

proc glm;
class GRUPO;
model NIVELES=GRUPO/clparm;* se obtiene I.C. para el contraste pedido;
contrast 'dosis bajas contra dosis altas' GRUPO .5 .5 -.5 -.5 0/e;
estimate 'dosis bajas contra dosis altas' GRUPO .5 .5 -.5 -.5 0;
run;

```

En este caso se ha planteado el contraste de hipótesis como la igualdad del promedio de medias de dosis altas con las dosis bajas, por tanto puede escribirse como:

$$H_0 : \frac{\mu_{25mg} + \mu_{50mg}}{2} = \frac{\mu_{100mg} + \mu_{125mg}}{2}$$

$$H_1 : \frac{\mu_{25mg} + \mu_{50mg}}{2} \neq \frac{\mu_{100mg} + \mu_{125mg}}{2}$$

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
dosis bajas contra dosis altas	1	7625.520833	7625.520833	43.33	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t	95% Confidence Limits	
dosis bajas contra dosis altas	-25.2083333	3.82970843	-6.58	<.0001	-32.8832405	-17.5334261

Se ha realizado dicho contraste con un nivel de confianza $1 - \alpha$ con $\alpha = 0,05$, para el cual se obtiene un p-valor ≈ 0 por lo que se debe rechazar la hipótesis nula, es decir, no existe igualdad de medias entre los resultados obtenidos con dosis bajas y dosis altas.

El intervalo de confianza para el estimador de la diferencia de medias obtenido es $[-32,88, -17,53]$. Esto puede interpretarse como que las dosis altas disminuyen el nivel de parásitos entre 17,53 y 32,88 respecto de las dosis bajas.

2.6. Verifica las hipótesis del modelo con ayuda de gráficos de residuos

Para verificar la hipótesis del modelo se han realizado distintas pruebas, entre ellas se ha estudiado la hipótesis de igualdad de varianzas entre los distintos tratamientos. También se ha analizado la distribución de los residuos obtenidos, para comprobar que sigue una distribución normal.

The ANOVA Procedure

Levene's Test for Homogeneity of NIVELES Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
GRUPO	4	405946	101486	2.97	0.0273
Error	55	1880948	34199.1		

```

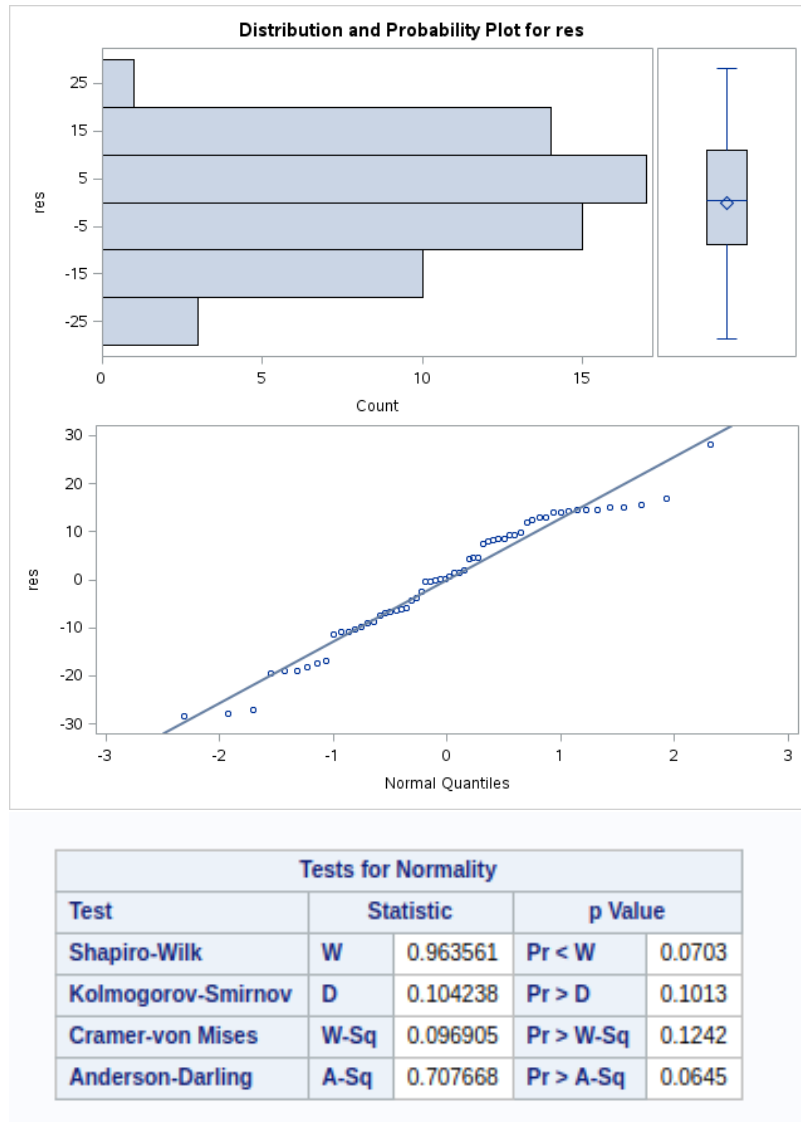
proc anova;
  class grupo;
  model niveles=grupo;
  means grupo/hovtest ;
run;

proc glm;
  class grupo;
  model niveles=grupo;
  output out= resal p=pred r=res student=rst;
  proc print data=resal;
run;

proc univariate plot normal data=resal;
  var res;
run;

```

En el caso de la hipótesis de igualdad de varianzas, esta se llevado a cabo a partir del test de *Levene* para el cual se ha obtenido un p-valor $\approx 0,027$. Por tanto, a un nivel de confianza del 95 % tendríamos que rechazar la hipótesis de que las varianzas son iguales. Sin embargo dicha hipótesis se puede aceptar si se toma un nivel de confianza del 98 %.



En cuanto a la hipótesis de normalidad de la distribución de residuos, se puede admitir que esta es cierta debido a los resultados obtenidos en los distintos tests de normalidad. Esta idea se refleja en el gráfico de normalidad de los residuos.

2.7. Suponiendo que los grupos de datos correspondieran a cinco medicamentos elegidos aleatoriamente entre todos los posibles en el mercado, ¿cómo cambiaría el modelo adecuado? ¿Podría afirmarse que existe variación significativa entre los efectos de los distintos medicamentos? Estima las componentes de la varianza bajo dicho supuesto

En este caso, en lugar de plantear el problema como un análisis de la varianza (ANOVA) de un factor con k tratamientos distintos, lo deberemos plantear como un análisis de la varianza desde el punto de vista de un factor aleatorio. De esta manera, el conjunto de pasos a realizar para llevar a

```
proc glm;
  class grupo;
  model niveles=grupo;
  random grupo/test;
run;
```

cabo el test de hipótesis no varía. Sin embargo la interpretación es distinta, ya que ahora en lugar de estudiar si existen tratamientos diferentes entre si y sus diferencias, el problema se plantea como conocer si todos los niveles extraídos como muestra pertenecen a una misma población:

$$H_0 : \sigma_\tau = 0$$

$$H_1 : \sigma_\tau > 0$$

The GLM Procedure					
Source	Type III Expected Mean Square				
GRUPO	Var(Error) + 12 Var(GRUPO)				

The GLM Procedure					
Tests of Hypotheses for Random Model Analysis of Variance					
Dependent Variable: NIVELES					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
GRUPO	4	12373	3093.233333	17.58	<.0001
Error: MS(Error)	55	9680.000000	176.000000		

Al igual que se concluyó en anteriores apartados, en este caso también hay que rechazar la hipótesis nula, debido a que el p-valor toma un valor extremadamente bajo p-valor ≈ 0 . Puesto que hemos rechazado la hipótesis nula, podemos estimar el valor de la varianza entre los niveles tal y como se indica a continuación:

$$Var(Error) + 12Var(GRUPO) = 3093,23$$

$$Var(Error) = 176$$

$$\Rightarrow Var(GRUPO) = \frac{3093,23 - 176}{12} = 243,1025$$