Diseño de Experimentos Parcial 1

Jhon Ramírez

Marzo 2024

1.

Se conduce un experimento para determinar si el uso de un aditivo químico y un fertilizante estándar aceleran el crecimiento de las plantas. En cada una de 10 localidades se estudiaron dos plantas sembradas en condiciones similares. A una planta de cada localidad se le aplicó el fertilizante puro y a la otra el fertilizante más el aditivo. Después de cuatro semanas el crecimiento en centímetros fue el siguiente: (Vale 7 puntos)

	Localidades									
	1	2	3	4	5	6	7	8	9	10
Sin aditivo	20	31	16	22	19	32	25	18	20	19
Con aditivo	23	34	15	21	22	31	29	20	24	23

Cuadro 1: Tabla de observaciones con y sin aditivo

a Si se supone que el muestreo se lleva a cabo sobre dos distribuciones normales. Pruebe la hipótesis de igualdad de medias al nivel significancia del 5 %.

En este caso como nuestras muestras aunque son pequeñas, provienen de una población normal podemos usar si distribución para hacer una prueba de hipótesis en este caso un t test para igualdad de medias independientes, primero vamos a realizar un var test para verificar si tenemos igualdad de varianzas.

Como resultado de este test tenemos que tiene un $p - value = 0{,}9243$ por lo que no rechazamos la hipótesis de igualdad de varianzas, en consecuencia realizamos un t test con varianzas iguales.

Bien con lo anterior realizamos el t test para diferencia de medias de muestras independientes con varianzas iguales, el resultado de esto resulta: $p - value = 0{,}4304$ por lo no rechazos la hipótesis de igualdad de medias con lo que el uso o no del aditivo no genera una diferencia representativa en la altura de las plantas.

```
#Punto 1
sin_aditivo <-c(20,31,16,22,19,32,25,18,20,19)
con_aditivo <-c(23,34,15,21,22,31,29,20,24,23)
#test de varianzas
var.test(sin_aditivo,con_aditivo)
#test de igualdad de medias
t.test(sin_aditivo,con_aditivo,var.equal = TRUE)

```

b Obtenga e interprete un intervalo de confianza 95 % para la diferencia de medias poluciónales de los procesos del ítem a.

Del test anterior llegamos al intervalo de confianza para el 95 % resulta: $-7{,}2087053{,}208705$ este intervalo contiene al 0 por lo que de este modo también llegamos a no rechazar la hipótesis de diferencia de medias vemos que cumpliendo los supuestos el 95 % de las diferencia de altura están entre los $7,208cm$ de altura por abajo del promedio y el $1,208cm$ por arriba del mismo.

c Si se supone que el muestreo se lleva a cabo sobre dos distribuciones no normales. Pruebe la hipótesis de igualdad al nivel significancia del 5 %.

Ya que en este caso tenemos una muestra pequeña y las muestras no vienen de una población que se distribuya de forma normal, realizaremos un test no parámetrico para igualdad de medias independiente, en este caso el test de Wilcoxon, en este caso se presentan empates, con lo que llegamos a el resultado:$p - value = 0{,}2716$ por lo que de igual forma a los anteriores casos no se tiene evidencia estadística para rechazar la hipótesis de igualdad de medias.

```
#Test de igualdad de medias no parametricas
wilcox.test(sin_aditivo, con_aditivo)

```

2.

En una empresa de manufactura se propone un tratamiento para reducir el porcentaje de productos defectuosos. Para validar esta propuesta se diseñó un experimento en el que se producía con o sin la propuesta de mejora. Cada corrida experimental consistió en producir un lote y la variable de respuesta es el porcentaje de producto defectuoso. Se hicieron 25 réplicas para cada tratamiento. Los datos obtenidos se muestran a continuación: (Vale 8puntos)

Tratamiento	Observaciones								
Con tratamiento	5.3	2.2	4.0	1.1	4.0	2.0	4.0	3.0	2.6
	3.1	2.1	2.1	5.1	1.2	4.1	3.3	4.1	2.1
	3.2	4.0	5.1	2.0	2.2	3.0	4.1		
Sin tratamiento	8.0	8.7	13.2	11.3	7.2	4.5	8.2	6.6	9.1
	9.2	6.7	10.2	12.2	10.6	16.3	13.3	9.2	5.2
	6.4	6.2	7.2	8.0	17.2	4.8			

Cuadro 2: Observaciones con y sin tratamiento

a Si se supone que el muestreo se lleva a cabo sobre dos distribuciones normales. Pruebe la hipótesis de igualdad de varianzas al nivel significancia del 5 %.

En este caso usaremos el var test para verificar la igualdad de varianzas usando el supuesto, para este caso tenemos como resultado: $p - value = 2{,}934e^{-06}$ por lo que rechazamos la hipótesis de igualdad de varianzas.

```
#Punto,2
con_tratamiento<-c(5.3,2.2,4.0,1.1,4.0,2.0,4.0,3.0,2.6,3.1,2.1,2.1,5.1,1.2,4.1,3.3
,4.1,2.1,3.2,4.0,5.1,2.0,2.2,3.0,4.1)
sin_tratamiento<-c(8.0,8.7,13.2,11.3,7.2,4.5,8.2,6.6,9.1,9.2,6.7,10.2,12.2,10.6,
```

```
16.3,13.3,9.2,5.2,6.4,6.2,7.2,8.0,17.2,4.8)
#test de varianzas
var.test(con_tratamiento,sin_tratamiento)

```

b Si se supone que el muestreo se lleva a cabo sobre dos distribuciones normales. Pruebe la hipótesis de igualdad de medias al nivel significancia del 5 %.

Para esto usaremos el t test de igualdad de medias para dos muestras independientes con varianzas diferentes por el resultado anterior, con esto llegamos a que: $p-value = 5{,}68e-09$ de lo que concluimos que rechazamos la hipótesis de igualdad de medias para un nivel de significaría del 5 %.

```
# test de igualdad de medias
t.test(con_tratamiento,sin_tratamiento)

```

c Obtenga e interprete un intervalo de confianza 95 % para la diferencia de medias poblacionales de los procesos del ítem b).

De el mismo t test nos da el intervalo de confianza tal que $-7{,}480315 - 4{,}491351$ con un 95 % de confianza este intervalo contiene a el valor real de la media bajo el supuesto de normalidad.

d Si se supone que el muestreo se lleva a cabo sobre dos distribuciones no normales. Pruebe la hipótesis de igualdad al nivel significancia del 5 %.

Para este caso como no tenemos el supuesto de una muestra normal y tenemos pocos datos realizamos un test no parámetrico en este caso el test para diferencia de medias de Wilcolxon este nos da el siguiente resultado: $p - value = 4{,}807e^{-09}$ con lo que llegamos a la misma conclusión que con el supuesto de normalidad donde rechazamos la hipótesis de igualdad de medias.

```
#test de igualdad de medias no parametrico
wilcox.test(con_tratamiento,sin_tratamiento)

```

3.

Un fisiólogo de animales estudió la función pituitaria de las gallinas, bajo el régimen estándar de muda de pluma forzada que usan los productores de huevo para mantenerlas en producción. Se usaron 25 gallinas en el estudio. Cinco se utilizaron para la medición, una previa al régimen de muda forzada y una al final de cada una de las cuatro etapas del régimen. Las cinco etapas del régimen fueron: 1) premuda (control), 2) ayuno de 8 días, 3) 60 gramos de salvado al día durante 10 días, 4) 80 gramos de salvado al día por 10 días y 5) mezcla de malta durante 42 días. El objetivo era dar seguimiento a las respuestas fisiológicas asociadas con la función pituitaria de las gallinas durante el régimen para explicar por qué vuelven a producir después de una muda forzada. Uno de los compuestos medidos fue la concentración de suero T3. Los datos de la tabla son las medidas de suero T3 en las cinco gallinas sacrificadas al final de cada etapa del régimen. (Vale 10 puntos)

a Proponga el modelo para llevar a cabo el análisis de esta información, haga los supuestos apropiados. Encuentre los parámetros estimados asociados al modelo propuesto.

Tratamiento	Suero T3 (ng/dl) $\times 10^{-1}$				
Premuda	94.09	90.45	99.38	73.56	74.39
Ayuno	98.81	103.55	115.23	129.06	117.61
60 g de salvado	197.18	207.31	177.50	226.05	222.74
80 g de salvado	102.93	117.51	119.92	112.01	101.10
Mezcla de malta	83.14	89.59	87.76	96.43	82.94

Cuadro 3: Niveles de Suero T3 para diferentes tratamientos

El modelo que encuentro adecuado es el siguiente, en este caso $T3$ hace referencia a la medida de suero en la gallina examinada de cada replica, μ a la media, P_i a si la gallina de cada replica recibió premuda, A_i a si la gallina de cada replica recibió ayuno, $60g_i$ a si la gallina de cada replica recibió 60g de salvado, $80g_i$ a si la gallina de cada replica recibió 80g de salvado, Mm_i a si la gallina de cada replica recibió mezcal de malta y ϵ_{ij} el error no estimable dado por la replica $i = 1, 2, 3, 4, 5$ y cada gallina $j = 1, ..., 25$.

$$T3_{ij} = \mu + P_i + A_i + 60g_i + 80g_i + Mm_i + \epsilon_{ij}$$

En este caso supondremos independencia entre ϵ_{ij} y cada una de las etapas del tratamiento. Veamos este modelo de forma matricial:

$$\mathbf{T}3 = \mathbf{X}\theta + \epsilon$$

donde $\mathbf{X}$ es de la forma:

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Esto ya que las primeras 5 gallinas solo fueron medidas como control, mientras que las ultimas recibieron todas las etapas del proceso. Mientras que el vector θ es de la forma:

$$\begin{bmatrix} \mu \\ P_i \\ A_i \\ 60g_i \\ 80g_i \\ Mm_i \end{bmatrix}$$

El mejor estimador insesgado para este vector de parámetros resulta:

$$\theta^0 = (\mathbf{X}\mathbf{X}^t)^{-}\mathbf{X}^t\mathbf{T}3$$

usando una inversa generalizada para $(\mathbf{X}\mathbf{X}^t)$

con lo anterior el código que realiza esto en R resulta:

```
## Punto 3
T3=c(94.09,90.45,99.38,73.56,74.39,98.81,103.55,115.23,129.06,117.61,197.18
,207.31,177.50,226.05,222.74,102.93,117.51,119.92,112.01,101.10,83.14,89.59
,87.76,96.43,82.94)

```

```r
mu=rep(1,25)
P=rep(1,25)
A=c(rep(0,5),rep(1,20))
g60=c(rep(0,10),rep(1,15))
g80=c(rep(0,15),rep(1,10))
Mm=c(rep(0,20),rep(1,5))

X=cbind(mu,P,A,g60,g80,Mm)
X

Xt=t(X)

XtX=Xt%*%X
W=XtX[-1, -1]
W_i=solve(W)
igXtX=rbind(c(0, rep(0, ncol(W_i))), cbind(0, W_i))
igXtX
theta=igXtX%*%Xt%*%T3
theta

```

Dando como resultado el vector estimado como:

$$\begin{bmatrix} \mu & 0{,}000 \\ P & 86{,}374 \\ A & 26{,}478 \\ g60 & 93{,}304 \\ g80 & -95{,}462 \\ Mm & -22{,}722 \end{bmatrix}$$

b Encuentre los 25 valores ajustados ($\hat{y}$) y sus correspondientes $\hat{e}$.

Con lo anterior usado el siguiente código en R llegamos a las siguientes estimaciones para la variable respuesta $T3$

```r
T3_est=X%*%theta
T3_est

```

Dándonos como resultado las siguientes estimaciones:

$$\hat{\mathbf{T}}\mathbf{3} = \begin{bmatrix} 86{,}374 \\ 86{,}374 \\ 86{,}374 \\ 86{,}374 \\ 86{,}374 \\ 112{,}852 \\ 112{,}852 \\ 112{,}852 \\ 112{,}852 \\ 112{,}852 \\ 206{,}156 \\ 206{,}156 \\ 206{,}156 \\ 206{,}156 \\ 206{,}156 \\ 110{,}694 \\ 110{,}694 \\ 110{,}694 \\ 110{,}694 \\ 110{,}694 \\ 87{,}972 \\ 87{,}972 \\ 87{,}972 \\ 87{,}972 \\ 87{,}972 \end{bmatrix}$$

Y con esto obtenemos el vector de residuales e:

$$\hat{\mathbf{e}} = \begin{bmatrix} 7{,}716 \\ 4{,}076 \\ 13{,}006 \\ -12{,}814 \\ -11{,}984 \\ -14{,}042 \\ -9{,}302 \\ 2{,}378 \\ 16{,}208 \\ 4{,}758 \\ -8{,}976 \\ 1{,}154 \\ -28{,}656 \\ 19{,}894 \\ 16{,}584 \\ -7{,}764 \\ 6{,}816 \\ 9{,}226 \\ 1{,}316 \\ -9{,}594 \\ -4{,}832 \\ 1{,}618 \\ -0{,}212 \\ 8{,}458 \\ -5{,}032 \end{bmatrix}$$

c Construya el análisis de varianza con: Causa de variación, grados de libertad, sumas de cuadrados, cuadrados medios, valor F. Concluya sobre la hipótesis de interés. Calcule e interprete el coeficiente de determinación apropiado.

Para hacerlo usamos el siguiente código en R:

```
# Calcular la suma total de los cuadrados
SST <- sum((T3 - mean(T3))^2)

# Calcular la suma de los cuadrados del modelo
SSM <- sum((T3_est - mean(T3))^2)

# Calcular la suma de los cuadrados del error
SSE <- sum(e^2)

# Calcular los grados de libertad
df_modelo <- length(theta) - 1
df_error <- length(T3) - length(theta)
df_total <- length(T3) - 1

```

```
# Calcular las medias de los cuadrados
MS_modelo <- SSM / df_modelo
MS_error <- SSE / df_error

# Calcular el valor F
F_value <- MS_modelo / MS_error

# Calcular el coeficiente de determinacion
R_cuadrado_ajustado <- 1 - (SSE / df_error) / (SST / df_total)

```

Con esto llegamos a la siguiente tabla:

Causa de variación	Grados de libertad	Sumas de cuadrados	Cuadrados medios	Valor F
Modelo	5	48568.88	9713.775	59.341
Error	19	3110.189	163.6942	
Total	24	51679.07		

Cuadro 4: Análisis de Varianza

Coeficiente de Determinación Ajustado
0.9239797

Cuadro 5: Coeficiente de Determinación Ajustado

En este caso la hipótesis a ala que responde este análisis de varianzas es:

$$\begin{cases} H_0 : P = A = 60g = 80g = Mm = 0 \\ H_1 : \text{ Al menos uno diferente de 0.} \end{cases}$$

Con lo anterior es claro que se rechaza la hipótesis nula de por lo que por lo menos una componente del vector θ es diferente de 0. Por otro lado el coeficiente de determinación ajustado tiene un valor de 0,9239 por lo que el modelo explica gran parte de la variación presente en la cantidad de $T3$ en las gallinas.

d Construya el conjunto generador de funciones estimables y diga cuáles de las siguientes funciones paramétricas son estimables:

$$L_1 = 4\tau_1 - \tau_2 - \tau_3 - \tau_4 - \tau_5 \quad L_2 = \tau_2 - \tau_3 \quad L_3 = \tau_1 + \tau_2 - 2\tau_3 + \tau_5$$

Para las funciones estimables, juzgue la hipótesis a un nivel de significancia del 5 %.

Para este caso llamaremos $P = \tau_1, A = \tau_2, 60g = \tau_3, 80g = \tau_4, Mm = \tau_5$, ahora bien una base generadora para las funciones estimables resulta;

$$(X^t X)^- (X^t X)$$

Por lo que la base resulta:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Bien con lo anterior buscamos los λ^t asociados a cada una de las hipótesis requeridas tal que:

$$\lambda_1 = \begin{bmatrix} 0 & 4 & -1 & -1 & -1 & -1 \end{bmatrix} \lambda_2 = \begin{bmatrix} 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix} \lambda_3 = \begin{bmatrix} 0 & 1 & 1 & -2 & 0 & 5 \end{bmatrix}$$

Operando estos con la matriz generadora debería volver a darnos el mismo vector, de ser asi serian estimables, veamos esto con el siguiente codigo en R:

```
> # generadora
> Gen=igXtX%*%XtX
> Gen=round(Gen)
> Gen
    mu  P  A  g60  g80  Mm
     0  0  0    0    0   0
P    1  1  0    0    0   0
A    0  0  1    0    0   0
g60  0  0  0    1    0   0
g80  0  0  0    0    1   0
Mm   0  0  0    0    0   1
> lambda1=c(0,4,-1,-1,-1,-1)
> lambda2=c(0,0,1,-1,0,0)
> lambda3=c(0,1,1,-2,0,5)
> lambda1%*%Gen
     mu P   A g60 g80  Mm
[1,]  4 4 -1  -1  -1  -1
> lambda2%*%Gen
     mu P A g60 g80 Mm
[1,]  0 0 1  -1   0  0
> lambda3%*%Gen
     mu P A g60 g80 Mm
[1,]  1 1 1  -2   0  5

```

Con esto es claro que la única prueba estimable es la segunda, por lo que veremos esa hipótesis usando el estadístico en R:

```
        t(lambda2)%*%theta/(sqrt(t(lambda2)%*%igXtX%*%lambda2*MS_error))

```

El anterior resultado es: $-4{,}768022$ y teniendo en cuenta que el percentil 0,025 de una t con 19 grados de libertadas es $-2{,}063$ se rechaza la hipótesis $L_2 = \tau_2 - \tau_3$.