Universidad Nacional Autónoma de México

Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas

Especialización en Estadística Aplicada

Diseño de Experimentos

Tarea 3

Adriana Haydeé Contreras Peruyero (haydeeperuyero@im.unam.mx)

Alejandro Jiménez Palestino (ajpalestino@gmail.com)

Jesus Alberto Urrutia Camacho (urcajeal@gmail.com)

Ciudad de México

8 de junio de 2021

Índice

1. Problema 1

En una fábrica que produce lotes de tinte están preocupados por la consistencia de la fuerza del tinte. Para medir la fuerza del tinte se pintan cuadrados de tela bajo condiciones controladas. Estos cuadrados de tela se comparan con un estándar, si hay coincidencia se califica con 100. Se seleccionaron al azar 6 lotes de tinte, para cada uno se seleccionaron al azar 6 cuadrados de tela entintada y se calificaron. Los datos son:

lote1	lote2	lote3	lote4	lote5	lote6
94.5	89.0	88.5	100.0	91.5	98.5
93.0	90.0	93.5	99.0	93.0	100.0
91.0	92.5	93.5	100.0	90.0	98.0
89.0	88.5	88.0	98.0	92.5	100.0
96.5	91.5	92.5	95.0	89.0	96.5
88.0	91.5	91.5	97.5	91.0	98.0

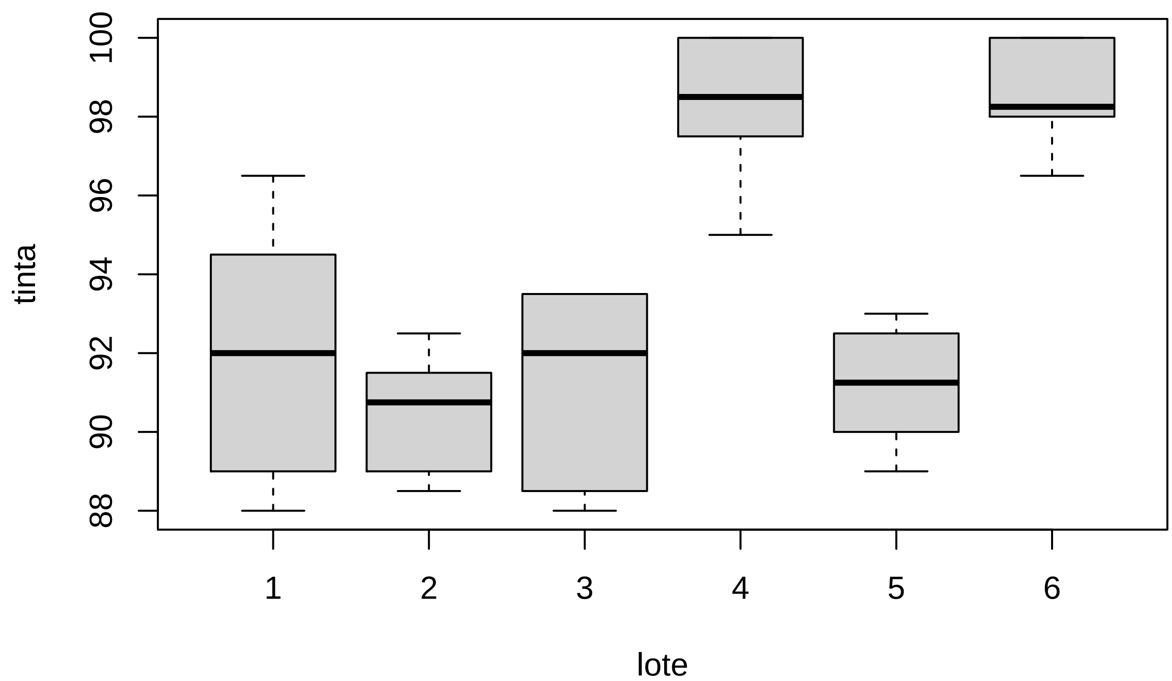
Escriba un modelo lineal para este experimento, explique los términos y realice el analisis de varianza para los datos. Concluya.

En este problema estamos suponiendo que dado que las telas son entintadas bajo condiciones controladas, es decir no existe variación debido al proceso de entintado o a las telas, partimos del supuesto que cualquier variación en el puntaje del desempeño del tinte proviene de los lotes.

Bajo ese supuesto podemos decir que se tiene un diseño completamente al azar, unifactorial, de efectos fijos con 6 tratamientos y 6 repeticiones para cada uno.

Donde el modelo lineal es: $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ y_{ij} es la observacion de la j esima u.e. del i esimo tratamiento, μ es la media general, comun a todas las u.e. τ_i es el efecto del tratamiento i esimo ε_{ij} es el error experimental de la unidad ij. Hay en total 6 tratamientos y 6 repeticiones en cada uno por lo que decimos que es un modelo balanceado.

```
boxplot(tinta~lote)
```




```
analisis <- aov(tinta~lote)
summary(analisis)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## lote         5  415.5   83.09   18.59 2.16e-08 ***
## Residuals   30  134.1    4.47                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(analisis)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
## 
## Fit: aov(formula = tinta ~ lote)
## 
## $lote
##            diff        lwr       upr     p adj
## 2-1 -1.50000000  -5.212508  2.212508 0.8193236
## 3-1 -0.75000000  -4.462508  2.962508 0.9891152
## 4-1  6.25000000   2.537492  9.962508 0.0002222
## 5-1 -0.83333333  -4.545842  2.879175 0.9825450
## 6-1  6.50000000   2.787492 10.212508 0.0001259
## 3-2  0.75000000  -2.962508  4.462508 0.9891152
## 4-2  7.75000000   4.037492 11.462508 0.0000074
## 5-2  0.66666667  -3.045842  4.379175 0.9936545
## 6-2  8.00000000   4.287492 11.712508 0.0000042
```

```
## 4-3  7.00000000   3.287492 10.712508 0.0000403
## 5-3 -0.08333333  -3.795842  3.629175 0.9999998
## 6-3  7.25000000   3.537492 10.962508 0.0000229
## 5-4 -7.08333333 -10.795842 -3.370825 0.0000334
## 6-4  0.25000000  -3.462508  3.962508 0.9999446
## 6-5  7.33333333   3.620825 11.045842 0.0000189
```

```r
factor.lote <- c("lote2","lote5" ,"lote3" ,"lote1" ,"lote6" ,"lote4" )
m.lote <- c( mean(lote2), mean(lote5), mean(lote3), mean(lote1), mean(lote6), mean(lote4))
m.lote
```

```
## [1] 90.50000 91.16667 91.25000 92.00000 98.50000 98.25000
```

```r
agrupado <- c("a_", "a_", "a_", "a_", "_b", "_b")

grupos <- data.frame(factor.lote, m.lote, agrupado)

kbl(grupos, booktabs = T, align = "c", caption = "Comparacion de pares por grupo", col.names = c("Lote"
kable_styling(position = "center", latex_options = "hold_position")
```

Cuadro 1: Comparacion de pares por grupo

Lote	Promedio desempeño tinte	Grupo
lote2	90.50000	a_
lote5	91.16667	a_
lote3	91.25000	a_
lote1	92.00000	a_
lote6	98.50000	_b
lote4	98.25000	_b

Utilizando el Análisis de Tukey podemos decir que de los 6 lotes que se muestrearon, los lotes 2, 5, 3 y 1 se desempeñan por debajo de los lotes 6 y 4 dado que estos se agrupan unicamente en 2 grupos. La Recomendacion despues de este análisis es: Identificar las diferencias en la fabricación entre estas dos agrupaciónes dado que tenemos evidencia para decir que los tratamientos son diferentes para estos dos grupos.

2. Problema 2

Se estudia el acabado de la superficie de metal de ciertas partes hechas por cuatro máquinas. Se realizó un experimento en el cual cada máquina es operada por tres operadores diferentes y se seleccionan y prueban dos especímenes de cada operador. Dada la localización de las máquinas, se utilizaron diferentes operadores en cada máquina los cuales fueron seleccionados al azar. Los datos se presentan a continuación.

Escriba el modelo lineal para este experimento, explique los términos e interprete el análisis de varianza para los datos. Concluya.

A continuación, se describirán las características del diseño total del experimento.

Dado que es un experimentos con diseño de efectos anidados balanceados, la ecuación que describe este modelo es la siguiente:

$$y_{kmo} = \mu + \tau_m + \beta_{m(o)} + \varepsilon_{k(mo)}$$

Cuadro 2: Resultados por operador y máquina

maq	oper	y
Machine1	1	79
		62
	2	94
		74
	3	46
		57
Machine2	1	92
		99
	2	85
		79
	3	76
		68
Machine3	1	88
		75
	2	53
		56
	3	46
		57
Machine4	1	36
		53
	2	40
		56
	3	62
		47

$$ConK = 1, \cdots, 24$$

$$m = 1, 2, 3, 4.$$

$$o = 1, 2, 3.$$

Existen dos factores:

1. Máquina y,
2. Operador.

Con cuatro niveles del factor Máquina y tres niveles el factor operador. Es decir, 4x3, respectivamente.

La relación entre ambos factores es que están **anidados**. Ya que una característica de este diseño es que hay dos factores, donde el factor *Máquina* es fijo, mientras que los *Operadores* son aleatorios, pues los tres "operadores en cada máquina (...) fueron seleccionados al azar". Otra característica es que los niveles j del factor *Operadores* son similares pero no idénticos para las diferentes máquinas, ya que los niveles de *Operadores* están marcados por etiquetas. Esto lleva rechazar que sean niveles cruzados. A continuación, se muestra una tabla que resume la anterior información.

Cuadro 3: Resumen de los términos del modelo

Factores	Factor	Anidación	Tamaño	Hipótesis_Nula
Máquina	Fijo	Principal	1,...,4	Ho : Mui = Muj
Operador	Aleatorio	Anidado	1,...3	sigma = 0

Para elaborar el análisis de ANOVA es necesario tener en consideración que el efecto del factor *Operador* es aleatorio, por lo que su hipótesis nula corresponde a una prueba de variación, la cuál es : $H_o : \sigma^2 = 0$. Mientras que la prueba del factor *Máquina* a ser fijo, su hipótesis nula corresponde a una prueba de igualdad de medias, la cuál es : $\mu_i = \mu_j$.

Entonces, para analizar la ANOVA, se corre el siguiente código. El cuál permite conocer los valores correctos del estadístico F.

```
roper <- as.random(oper)
fmaq <- as.fixed(maq)
modelo <- lm(y ~ fmaq + roper%in%fmaq)
gad(modelo)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
fmaq	3	3617.667	1205.8889	3.423794	0.0727968
fmaq:roper	8	2817.667	352.2083	4.168146	0.0134083
Residual	12	1014.000	84.5000	NA	NA

Con un P valor de 0.0728 no existe efecto significativo del tratamiento de maquinaria sobre las superficies de metal, por lo que no se rechaza la hipótesis de $\mu_i = \mu_j$ con una confianza del 95 %. Por otra parte, hay variabilidad de las superficies de metal en cada operador de cada máquina, por lo que se rechaza la hipótesis nula de $H_o : \sigma^2 = 0$ con un P valor de 0.013 con una confianza del 95 %. Por lo anterior, se debería de tratar de reducir esta variabilidad al entrenar mejor a los operadores.

```
compon <- lmer(y ~ maq + (1|maq:oper))
summary(compon)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: y ~ maq + (1 | maq:oper)
##
## REML criterion at convergence: 164.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.04636 -0.69740 -0.05224  0.72879  1.48805
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  maq:oper (Intercept) 133.9    11.570
##  Residual              84.5     9.192
## Number of obs: 24, groups:  maq:oper, 12
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   68.667      7.662   8.962
## maqMachine2   14.500     10.835   1.338
## maqMachine3   -6.167     10.835  -0.569
```

```
## maqMachine4  -19.667     10.835  -1.815
##
## Correlation of Fixed Effects:
##             (Intr) mqMch2 mqMch3
## maqMachine2 -0.707
## maqMachine3 -0.707  0.500
## maqMachine4 -0.707  0.500  0.500
# estima los componentes de varianza
```

Respecto a los **componentes de varianza**, es posible afirmar que el factor anidado (operador en Maquinaria) produce el 61.30952 % de varianza. Lo cuál enfatiza la decisión de rechazar la hipótesis nula de $H_0 : \sigma^2 = 0$, ya que este tratamiento aporta variabilidad dentro del resultado de las superficies de metal.

Cuadro 4: Resumen de los términos del modelo

Fuente_varianza	Varianza	Proporción
Operador en Máquina	133.9	0.6130950
Error	84.5	0.3869048
Total	128.4	1.0000000

Finalmente, si ignoraramos que hubiese diferentes operadores, y por lo tanto no hubiese efectos aleatorios, se podría simular el siguiente escenario. Si se deseara comprobar el efecto del efecto de *Máquina*, se aplicaría un factorial de una vía balanceado del efecti fijo *Maquinaria*.

```
sin <- aov(y ~ maq)
summary(sin)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## maq          3   3618  1205.9   6.294 0.0035 **
## Residuals   20   3832   191.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Se podría concluir que existe un efecto en el tratamiento de las máquinas sobre las superficies de metal, al rechazar la hipótesis de $\mu_i = \mu_j$ con un P valor de 0.0035.

3. Problema 3

Se diseña un experimento para estudiar la dispersión de los pigmentos de una pintura. Se estudian cuatro mezclas diferentes de un pigmento particular. El procedimiento consiste en preparar una mezcla particular y en aplicarla después a un panel utilizando tres métodos (con brocha, por rocío y con rodillo). La respuesta medida es el porcentaje de reflectancia (coeficiente de reflexión) del pigmento. Se repite el experimento completo en 3 días. Los datos obtenidos se presentan en la tabla 5:

Suponga que las mezclas y los métodos de aplicación son factores fijos. Escriba un modelo lineal para este experimento, explique los términos y realice el análisis de varianza para los datos. Concluya.

Este experimento se trata de un modelo de parcelas divididas en bloques al azar de las parcelas grandes. En este caso tenemos que la parcela grande es el recipiente donde se preparan las diferentes mezclas y el tratamiento es el tipo de receta de mezcla que tiene 4 factores y el bloque es el día en que se prepara la

Cuadro 5: Resultados del coeficiente de reflexión del pigmento de una pintura

mezcla	dia	metodo	y3
1	1	1	69.0
		2	73.1
		3	75.2
	2	1	69.8
		2	74.0
		3	76.2
	3	1	70.8
		2	73.8
		3	75.8
2	1	1	70.9
		2	74.4
		3	78.2
	2	1	69.6
		2	75.2
		3	77.9
	3	1	71.2
		2	73.8
		3	79.4
3	1	1	79.3
		2	79.0
		3	83.5
	2	1	79.0
		2	79.7
		3	84.6
	3	1	77.4
		2	80.7
		3	85.7
4	1	1	71.2
		2	74.9
		3	77.4
	2	1	69.3
		2	73.1
		3	76.5
	3	1	72.4
		2	73.4
		3	77.5

mezcla y se repite el procdimiento. Las parcelas pequeñas son las muestras de las mezclas y sus tratamiento son los métodos de aplicación que tiene 3 factores, brocha, por rocío y con rodillo.

Un experimento de parcelas divididas en bloques al azar tiene por modelo:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \gamma_k + (\tau\gamma)_{ik} + \epsilon_{ijk}$$

con $i = 1, 2, 3, 4,\ j = 1, 2, 3$ y $k = 1, 2, 3$.

Cada término significa lo siguiente:

- y_{ijk} es el porcentaje de reflectancia (coeficiente de reflexión) del pigmento del método de aplicación k en el día j con la mezcla i.
- τ_i es el efecto de la mezcla i.
- β_j es el efecto del día j, el cual es aleatorio. Estos serían los bloques.
- γ_k es el efecto del tipo de aplicación k, brocha, por rocío o con rodillo.
- $(\tau\beta)_{ij}$ es el efecto aleatorio de la interacción mezcla-día.
- $(\tau\gamma)_{ik}$ es la interacción mezcla - método. Nivel i del tipo de mezcla con nivel k del tipo de aplicación.
- ϵ_{ijk} es el error experimental del coeficiente de reflexión con el método de aplicación k en el día j con la mezcla i.

Las supocisiones del modelo son las siguientes:

- $\beta_j, (\tau\beta)_{ij}, \epsilon_{ijk}$ son independientes.
- $\beta_j \sim N(0, \sigma^2_\beta)$
- $(\tau\beta)_{ij} \sim N(0, \sigma^2_{\tau\beta})$
- $\epsilon_{ijk} \sim N(0, \sigma^2)$

Vamos a hacer el análisis de varianza.

La tabla de análisis de varianza es la siguiente

F.V.	E(CM)
Mezcla τ_i	$\sigma^2 + 4\sigma^2_{\tau\beta} + 12\theta^2_\tau$
Día β_j	$\sigma^2 + 12\sigma^2_\beta$
Mezcla - Día $\tau\beta_{ij}$	$\sigma^2 + 4\sigma^2_{\tau\beta}$
Método γ_k	$\sigma^2 + 12\theta^2_{\tau\beta}$
Mezcla - Método $(\tau\gamma)_{ij}$	$\sigma^2 + 3\theta^2_{\tau\gamma}$
Error ϵ_{ijk}	σ^2

- Para probar no efecto del tipo de mezcla, el estadístico F se construye como $CM_\tau/CM_{\tau\beta}$.
- Para probar no efecto del método de aplicación, el estadístico F se construye como CM_γ/CM_ϵ.
- Para probar que no hay interacción entre $Mezcla - Aplicación$, el estadístico F se construye como $CM_{\tau\beta}/CM_\epsilon$.

En R, la tabla de análisis de varianza la obtenemos de la siguiente forma.

```
dia<-as.random(dia)
mezcla<-as.fixed(mezcla)
metodo<-as.fixed(metodo)

#anova del modelo
modelo3<-aov(y3~mezcla+dia+mezcla*dia+metodo+mezcla*metodo)
#ANOVA
```

```r
m3<-gad(modelo3)
m3
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
mezcla	3	353.249722	117.7499074	134.032357	0.0000069
dia	2	2.335556	1.1677778	1.558769	0.2407156
metodo	2	254.302222	127.1511111	169.723396	0.0000000
mezcla:dia	6	5.271111	0.8785185	1.172661	0.3684166
mezcla:metodo	6	11.257778	1.8762963	2.504511	0.0666047
Residual	16	11.986667	0.7491667	NA	NA

```r
#pvalor de mezcla-metodo
p3<-m3$`Pr(>F)`[5]

estimates(modelo3)
```

```
## $tm
##               mezcla dia metodo   n
## mezcla             0   3      3 1.8
## dia                4   1      3 1.8
## metodo             4   3      0 1.8
## mezcla:dia         0   1      3 1.8
## mezcla:metodo      0   3      0 1.8
## Res                1   1      1 1.0
##
## $mse
##               Mean square estimates
## mezcla        "Res + mezcla:dia + mezcla"
## dia           "Res + dia"
## metodo        "Res + metodo"
## mezcla:dia    "Res + mezcla:dia"
## mezcla:metodo "Res + mezcla:metodo"
## Residual      "Res"
##
## $f.versus
##               F-ratio versus
## mezcla        "mezcla:dia"
## dia           "Residual"
## metodo        "Residual"
## mezcla:dia    "Residual"
## mezcla:metodo "Residual"
```

Como el $p-$valor de la interacción `mezcla:metodo` es igual 0.0666047 a entonces no hay interacción. Lo que si tenemos en este casi es efecto de método y efecto de mezcla.

Vamos a realizar ahora Tukey para ver los efectos principales de mezcla y de método.

Para mezcla tenemos los siguientes resultados.

```r
lsmeans(modelo3, pairwise~mezcla)
```

```
## NOTE: Results may be misleading due to involvement in interactions

## $lsmeans
##  mezcla lsmean    SE df lower.CL upper.CL
##  1        73.1 0.289 16     72.5     73.7
```

```
## 2        74.5 0.289 16     73.9     75.1
## 3        81.0 0.289 16     80.4     81.6
## 4        74.0 0.289 16     73.4     74.6
##
## Results are averaged over the levels of: dia, metodo
## Confidence level used: 0.95
##
## $contrasts
##  contrast estimate    SE df t.ratio p.value
##  1 - 2      -1.433 0.408 16  -3.513 0.0138
##  1 - 3      -7.911 0.408 16 -19.389 <.0001
##  1 - 4      -0.889 0.408 16  -2.179 0.1715
##  2 - 3      -6.478 0.408 16 -15.876 <.0001
##  2 - 4       0.544 0.408 16   1.334 0.5556
##  3 - 4       7.022 0.408 16  17.210 <.0001
##
## Results are averaged over the levels of: dia, metodo
## P value adjustment: tukey method for comparing a family of 4 estimates
```

En la segunda tabla `$contrasts` podemos observar las comparaciones multiples para las parejas de medias entre diferentes tipos de mezclas. Si analizamos sus p-valores, vemos que la mezcla 1 es diferente a la mezcla 2 y la mezcla 3. De igual forma, la mezcla 2 es difernete a la mezcla 3 y la mezlca 3 es diferente a la mezcla 4.

En la primera tabla de la salida, `$lsmeans` podemos ver que la mezcla 3 es la que tiene el mayor promedio de respuestas ya que su media es la mayor. Esto es el indicativo de que es la mejor mezcla. La función `HSD.test` del paquete `agricole` también nos muestra las medias por grupos.

```r
T_mezcla <- HSD.test(modelo3,"mezcla")
T_mezcla$groups
```

	y3	groups
3	80.98889	a
2	74.51111	b
4	73.96667	bc
1	73.07778	c

Vamos a hacer por último el mismo análisis pero para método de aplicación.

```r
lsmeans(modelo3, pairwise~metodo)
```

```
## NOTE: Results may be misleading due to involvement in interactions

## $lsmeans
##  metodo lsmean   SE df lower.CL upper.CL
##  1        72.5 0.25 16     72.0     73.0
##  2        75.4 0.25 16     74.9     76.0
##  3        79.0 0.25 16     78.5     79.5
##
## Results are averaged over the levels of: mezcla, dia
## Confidence level used: 0.95
##
## $contrasts
##  contrast estimate    SE df t.ratio p.value
##  1 - 2       -2.93 0.353 16  -8.301 <.0001
##  1 - 3       -6.50 0.353 16 -18.395 <.0001
##  2 - 3       -3.57 0.353 16 -10.094 <.0001
```

```
##
## Results are averaged over the levels of: mezcla, dia
## P value adjustment: tukey method for comparing a family of 3 estimates
```

Podemos observar que los contrastes entre los tres tipos de método de aplicación son significativos. Además, al tener una media mayor el método de aplicación 3 vemos que es el mejor método de aplicación.

```
T_metodo <- HSD.test(modelo3,"metodo")
T_metodo$groups
```

	y3	groups
3	78.99167	a
2	75.42500	b
1	72.49167	c