Evaluación 4. Diseño con dos factores

Diana Laura Javier Garcia

18/11/2020

Para el diseño de dos factores realizar el ejemplo resuelto de la sección 10.4, es decir, ejemplos 10.9 y 10.10.

Tastee Bakery Company suministra un producto de panadería a muchos supermercados metropolitanos. La empresa desea estudiar los efectos de dos factores, la altura de la exhibición de los estantes y el ancho de la exhibición de los estantes, sobre la demanda mensual (medida en cajas de 10 unidades cada una) de este producto. El factor “altura de la pantalla” se define para tener tres niveles B (abajo), M (medio) y T (arriba). El factor “ancho de visualización” se define para tener dos niveles: R (regular) y W (ancho). Los tratamientos en este experimento son combinaciones de altura de pantalla y ancho de pantalla. Estos tratamientos son

BR

BW

MR

MW

TR

TW

Aquí, por ejemplo, la notación BR denota el tratamiento “altura de la pantalla inferior y ancho de la pantalla normal”. Para cada combinación de altura y ancho de exhibición, la empresa selecciona aleatoriamente una muestra de m = 3 supermercados del área metropolitana (todos los supermercados utilizados en el estudio tendrán el mismo potencial de ventas). Cada supermercado vende el producto durante un mes utilizando su combinación de altura y ancho de pantalla asignada, y se registra la demanda del producto del mes. Las seis muestras obtenidas en este experimento se dan en la tabla 1. sea la demanda mensual obtenida en el k-ésimo supermercado que utilizó la altura de la pantalla i y la anchura de la pantalla j. Por ejemplo, es la demanda mensual obtenida en el segundo supermercado que utilizó una altura de pantalla media y una pantalla ancha.

Además de dar las seis muestras, la Tabla 10.1 proporciona la media del tratamiento de la muestra para cada combinación de altura y ancho de pantalla. Por ejemplo, es la media de la muestra de tres demandas observadas en los supermercados que utilizan una altura de pantalla inferior y un ancho de pantalla regular. La tabla también proporciona la demanda media muestral para cada nivel de altura de pantalla (B, M y T) y para cada nivel de ancho de visualización (R y W).

### Para cada tipo de experimento presentar el Enunciado y formular con sus propias palabras las hipótesis posibles

Tal como el enunciado lo indica, el problema consiste en estudiar la demanda de productos cuando varía la colocación de los productos en los estantes: se varía la altura (baja, media, alta) y el ancho de las parrillas (regular y ancho). Los datos obtenidos de la muestras se consultaron en en el documento de clase, así se construyo el documento de excel con extensión *csv*, el cual se lee a continuación, como se puede ver, los datos fuueron organizados de tal manera que la primera columna describe las tres alturas, la segunda columna el ancho y finalmente, la tercera muestra la demanda de los productos. Posterior a esto, para tener mayor claridad en la lectura, se renombro cada columna con las etiquetas siguientes: B=bajo, M=medio, T=alto, R=regular y W=ancho. Para poder formular las hipotesis es necesario estudiar los datos de manera descriptiva, esto se presenta en la siguiente sección.

datos<- read.csv("bakery.csv")  
names(datos)

## [1] "alto" "ancho" "demanda"

datos$alto <- factor(datos$alto,labels=c('B','M','T'))  
datos$ancho <- factor(datos$ancho,labels=c('R','W'))  
datos

## alto ancho demanda  
## 1 B R 58.2  
## 2 B R 53.7  
## 3 B R 55.8  
## 4 B W 55.7  
## 5 B W 52.5  
## 6 B W 58.9  
## 7 M R 73.0  
## 8 M R 78.1  
## 9 M R 75.4  
## 10 M W 76.2  
## 11 M W 78.4  
## 12 M W 82.1  
## 13 T R 52.4  
## 14 T R 49.7  
## 15 T R 50.9  
## 16 T W 54.0  
## 17 T W 52.1  
## 18 T W 49.9

names(datos)

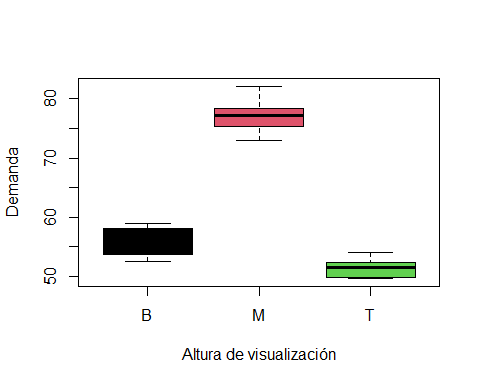
## [1] "alto" "ancho" "demanda"

attach(datos)

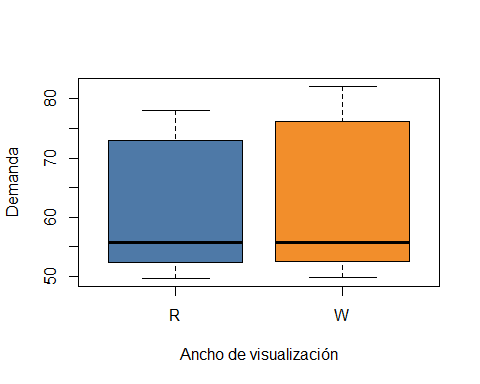
### Estudiar Estadísticamente los datos (resumen estadístico medidas de tendencia y variabilidad, boxplots, histogramas) con el análisis de cada gráfica. Dar una primera idea de las hipótesis más plausibles.

Para tener una primer idea del comportamiento de los datos, se realizó el análisis descriptivo. A continuación se muestran los diagramas de caja, en función del ancho y alto de los almacenes.

boxplot(demanda ~ alto, xlab = "Altura de visualización", ylab = "Demanda", col = palette("Tableau 10"))



boxplot(demanda ~ ancho, xlab = "Ancho de visualización", ylab = "Demanda", col = palette("Classic Tableau"))

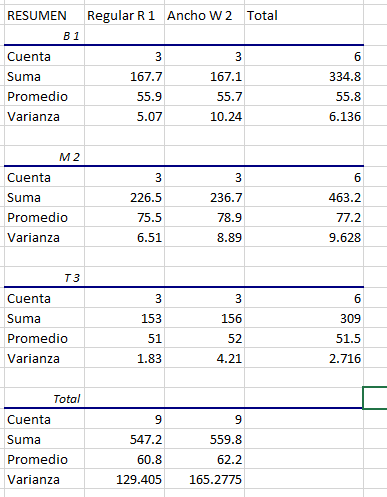


De los diagramas de caja presentados se observa que la demanda es mayor cuando la altura de la pantalla es media (con un valor de aproximadamente 76), a diferencia de esto, cuando las pantallas son colocadas en un nivel bajo B o alto T, la demanda es mucho menor, de casi 20 unidades menos, lo que representa una diferencia importante entre la colocación. Por otra parte, se observa que no existe un diferencia importante entre usar parrillas con un ancho regural R o más amplio W. Así, se resumen las siguientes hipotesis:

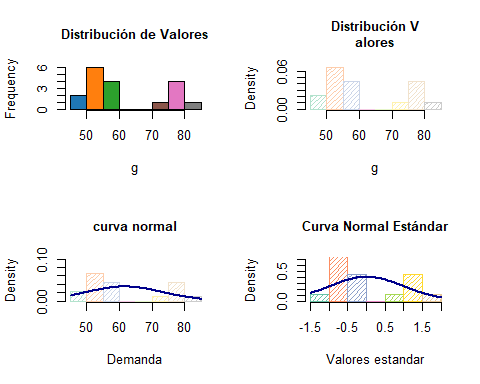
1. La demanda es mayor cuando se emplean parrillas a una altura media.
2. La colocación de la parrilas a una altura baja o alta, da como resultado baja demanda, por lo que no se recomienda esta colocación.
3. La demanda es menor cuando se colocan las parrilas muy alto.
4. EL ancho de las parrillas, no representa una diferencia significativa y por ello se pueden usar cualquiera de las dos.

Por otra parte, el resumen estadístico realizado en excel proporciona valores cualitativos de las observaciones anteriores, asi se ve que los promedios y las varianzas de las muestras con altura B y T (bajo y alto) son similares.

knitr::include\_graphics("cap7.png" )

 Continuando con el análisis descriptivo, a continuación, se presenta el histograma de los datos con el ajuste a una curva normal.

par (mfrow=c(2,2))  
g = datos$demanda  
m<-mean(g)  
std<-sqrt(var(g))  
  
hist(g, breaks = "Sturges", col=palette("Pastel 2"), main= "Distribución de Valores"  
, cex.main=1.0)  
  
hist(g, density=30, breaks = "Sturges", prob=TRUE, main= "Distribución V  
alores", cex.main=1.0, col=palette("Pastel 2"))  
  
hist(g, density=30, prob=TRUE, col=palette("Set 2"),  
xlab="Demanda", ylim=c(0, 0.10),  
main="curva normal", cex.main=1.0)  
  
curve(dnorm(x, mean=m, sd=std),  
col="darkblue", lwd=2, add=TRUE, yaxt="n")  
zg<-(g-m)/std  
  
hist(zg, density=30, prob=TRUE,col=palette("Set 2"),  
xlab="Valores estandar", ylim=c(0, 0.70),  
main="Curva Normal Estándar", cex.main=1.0)  
  
curve(dnorm(x, mean=0, sd=1),  
col="darkblue", lwd=2, add=TRUE, yaxt="n")



De esta manera, se observa que los datos presentan frecuencias muy diferentes, esto puede deberse a que el tamaño de la muestra es de 3 unidades, lo cual da poca variabilidad. Con respecto a la distribución de la demanda es posible observar que la curva normal no se ajusta a los datos, por lo tanto no se puede concluir si la distrbución es normal. La afirmación de esta hipotesis se analizó en la siguiente sección.

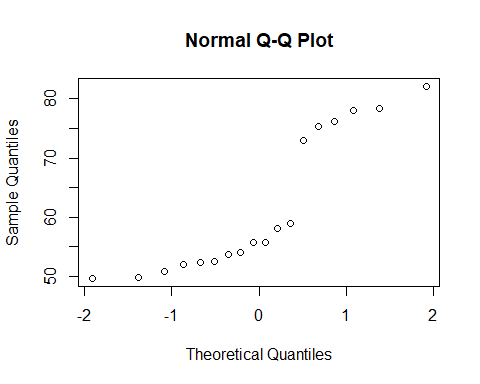
## Verificar supuestos de varianza y normalidad

Varianza y normalidad

Para estudiar normalidad de los datos se realizó se obtuvo la grafica cuantil-cualtil que permite observar cuan cerca está la distribución de un conjutno de datos a alguna distribución. Además se realizó la prueba de Kolmorogov y de Shapiro Wilk. Ambas pruebas plantean la hipótesis nula que una muestra proviene de una distribución normal, y por ende plantea una hipótesis alternativa que sostiene que la distribución no es normal.

Para este trabajo se encontró que los datos no se ajustan completamente a una línea recta, por lo que no es suficiente para afirmar que los datos siguen una distribución normmal. En este sentido, el p-valor de la prueba de Shapiro Wilk es menor a 0.05, lo que significa que la hipotesis nula no se cumple, es decir, los datos no se distribuyen normalmente. Sin embargo, el p-valor de la prueba de Kolmorogov es mayor a 0.05, lo que significa que los datos se distribuyen normalmente. Hasta este punto, se han encontrado dos resultados diferentes para estudiar la normalidad. Se recomienda que para afirmar la normalidad se realicen más pruebas.

qqnorm(datos$demanda)



### Prueba de Shapiro Wilk

sw<-shapiro.test(datos$demanda)  
print(sw)

##   
## Shapiro-Wilk normality test  
##   
## data: datos$demanda  
## W = 0.80616, p-value = 0.001858

### Prueba de Kolmogorov

ks<-ks.test(datos$demanda, "pnorm", mean=mean(datos$demanda),sd=sd(datos$demanda))  
print(ks)

##   
## One-sample Kolmogorov-Smirnov test  
##   
## data: datos$demanda  
## D = 0.25388, p-value = 0.1647  
## alternative hypothesis: two-sided

### Prueba de Bartlett

Para verificar la igualdad de varianzas se aplicó la Prueba de Bartlett, la cual permite contrastar la igualdad de varianza en 2 o más poblaciones sin necesidad de que el tamaño de los grupos sea el mismo. Como se puede observar, el p-valor es mayor a 0.05, por lo tanto se cumple la hipotesis de que las varianzas no presentan diferencias significativas.

bartlett.test(demanda ~ alto, data=datos)

##   
## Bartlett test of homogeneity of variances  
##   
## data: demanda by alto  
## Bartlett's K-squared = 1.7275, df = 2, p-value = 0.4216

bartlett.test(demanda ~ ancho, data=datos)

##   
## Bartlett test of homogeneity of variances  
##   
## data: demanda by ancho  
## Bartlett's K-squared = 0.11241, df = 1, p-value = 0.7374

### En la secuencia lógica de la solución Indicar claramente todas las formulas usadas (latex). Es decir no hacer un anexo con las formulas, sino integrarlas a medida que avanza su presentación.

Esta indicación fue realizada en la siguiente sección, que es la que corresponde a los cálculos.

### Presentar las hipótesis (en términos de promedio y tratamiento), presentar el modelo y la ANOVA a partir de las sumas de cuadrados (generados con R) y los grados de libertad obtener los cuadrados medios.

En esta sección se presentan los resultados del ANOVA. Las hipotesis son las siguientes: La hipótesis nula plantea que la media de la variable estudiada es la misma en los diferentes grupos, en contra posición a la hipótesis alternativa de que al menos dos medias son diferentes. La instrucción para realizar el ANOVA es: *aov*. En este trabajo se guarda la instrucción como *modelo*. Y posteriormente, los resultados se ven con el comando *summary*. En términos matemáticos:

modelo <- aov(demanda ~ ancho + alto+ ancho\*alto, data = datos)  
modelo

## Call:  
## aov(formula = demanda ~ ancho + alto + ancho \* alto, data = datos)  
##   
## Terms:  
## ancho alto ancho:alto Residuals  
## Sum of Squares 8.82 2273.88 10.08 73.50  
## Deg. of Freedom 1 2 2 12  
##   
## Residual standard error: 2.474874  
## Estimated effects may be unbalanced

summary(modelo)

## Df Sum Sq Mean Sq F value Pr(>F)   
## ancho 1 8.8 8.8 1.440 0.253   
## alto 2 2273.9 1136.9 185.623 9.42e-10 \*\*\*  
## ancho:alto 2 10.1 5.0 0.823 0.462   
## Residuals 12 73.5 6.1   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

De los resultados obtenidos de la ANOVA se observa que el valor P del factor *ancho* es mayor a 0.05, mientras que el p-valor del factor \* alto\* es menor a 0.05. Retomando la tería sobre la hipotesis nula y alterna de la prueba de ANOVA, se dice que existe una influencia de la altura. Otras caracteristicas que se pueden identificar la tabla ANOVA son los grados de libertad.

1. El número de grados de libertad del factor altura es 2.
2. El número de grados de libertad del factor ancho es 1.
3. El valor p de influencia entre ancho y la antura es mayor a 0.05, esto sugiere que existe una relación entre estos factores.

Es importante identificarla información anterior, ya que es la que permitirá realizar los cálculos a partir de las ecuaciones, como se verá en la siguiente sección.

### Explicar como se calculan los grados de libertad para cada diseño. En Excel generar la tabla ANOVA a partir de las sumas de cuadrados (generadas por R) y los grados de libertad obtener los cuadrados medios, el MSE, los F críticos y los p valores.

El procedimiento ANOVA para un experimento factorial de dos factores divide la suma total de cuadrados (SSTO) en cuatro componentes: la suma de cuadrados del factor 1-SS(1), la suma de cuadrados del factor 2 la suma de la interacción de los cuadrados- , y la suma del error de los cuadrados-(SSE). La fórmula para esta partición es la siguiente:

Los pasos para calcular estas sumas de cuadrados, así como lo que se mide por las sumas de cuadrados, se pueden resumir de la siguiente manera: Paso 1: Calcule SSTO, que mide la cantidad total de variabilidad:

Paso 1: Calcule SSTO, que mide la cantidad total de variabilidad:

2: Calcule que mide la cantidad de variabilidad debido a los diferentes niveles del factor de paso 1

Paso 3: Calcule $ 5 S (2) $, que mide la cantidad de variabilidad debido a los diferentes niveles del factor 2

Paso 4: Calcular SS(interacción), que mide la cantidad de variabilidad debido a la interacción entre los factores 1 y 2

Paso 5: Calcular SSE que mide la cantidad de variabilidad debido a la enorme:

Para calcular las entradas en la tabla ANOVA, se debe identificar los niveles de cada factor, asi, el factor 1 de ancho tiene dos niveles, el factor 2 de alto tiene tres niveles:

a=3

b=2

m=3

Realizando los cálculos tenemos:

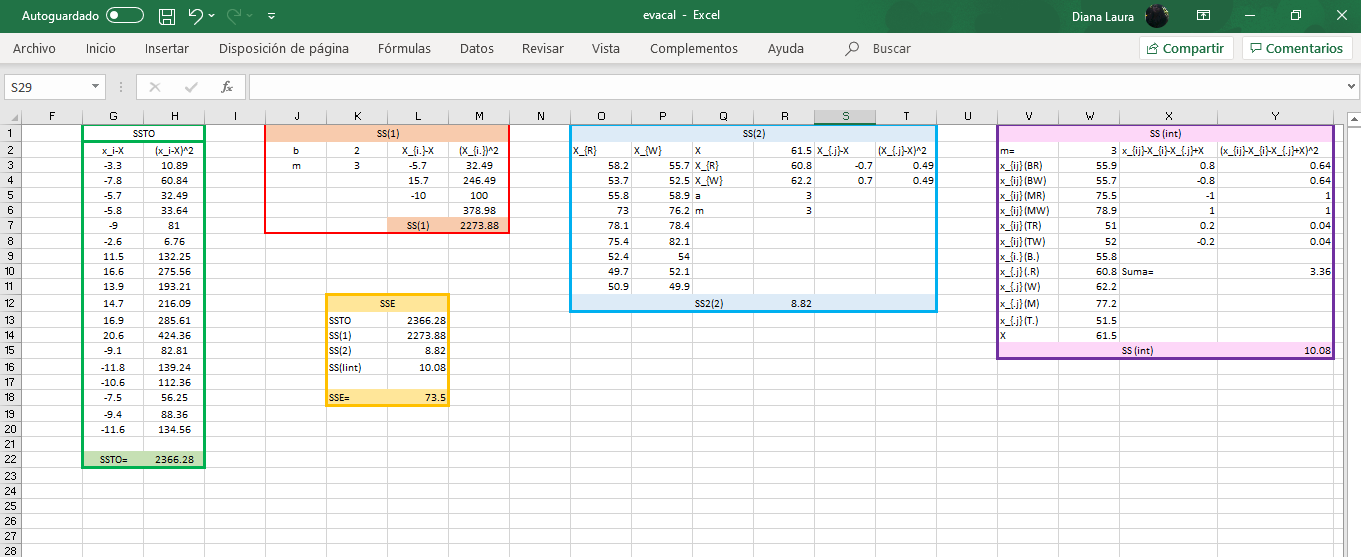
Para SSTO:

Para SS(1):

Para SS(2):

Finalmente SSE:

knitr::include\_graphics("cap5.png" )



Recodando que la hipótesis nula H0 establece que no existe interacción entre los factores 1 y 2 versus la hipótesis alternativa Ha de que la interacción existe. Para un nivel de significancia si:

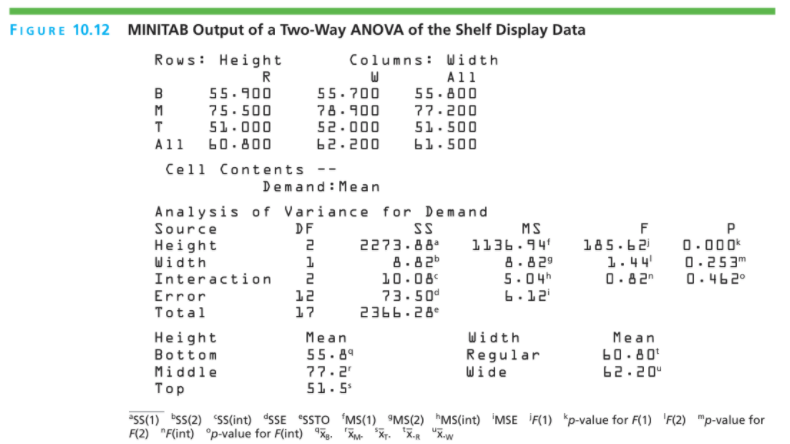
$ F(int)=$

es mayor que el punto basado en (a-1) (b-1) numerador y ab (m-1) denominador de grados delibertad. Para este problema, se estableció basado en (a-1) (b-1) =2 numerador y ab (m-1)=12 grados de denominador de libertad.

Así, se observa que es menor que 3.89, esto significa que no existe poca o ninguna interacción entre la altura de exhibición del estante y el ancho. Esta conclusión coincidde con la observación realizada en la sección de estadistica descriptiva realizada anteriormente, corroborando así que la demanda depende unicamente de la altura.

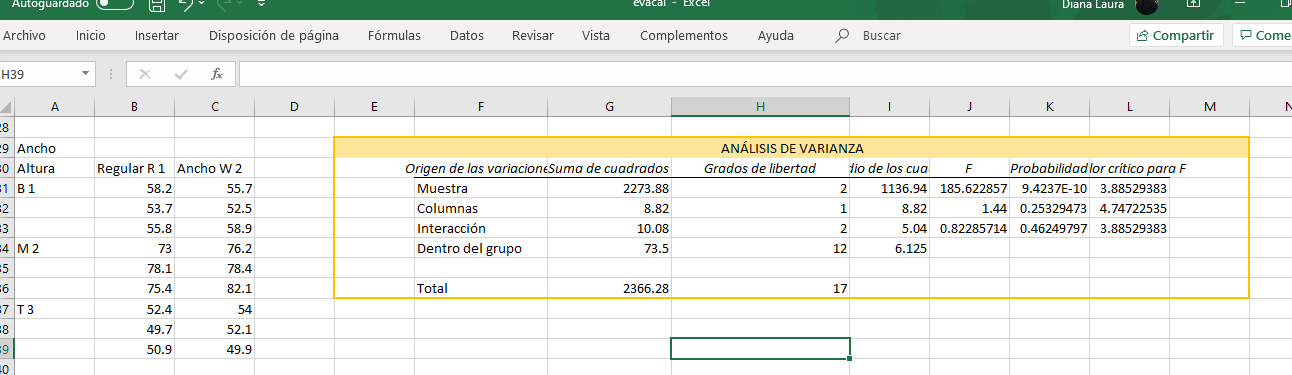
Si ahora se compara con los resultados de ANOVA mostrados en el libro proporcionado, vemos que los resultados coinciden:

knitr::include\_graphics("cap6.png" )



Finalmente, se presenta la tabla de ANOVA generada en excel

knitr::include\_graphics("cap8.png" )



Como se puede observar en la Tabla enterior, el analisis realizado por excel brinda la misma información que se ha obtenido por comandos de R y por fórmulas matématicas.

### Obtener los intervalos simultáneos de Tukey e interpretar (ver precisiones). Los cálculos se usaran usando las formulas y los comandos en R.

### Para cada experimento, verificar los resultados de comparaciones múltiples mediante la prueba de Tukey . Obtener la estadística q mediante la tabla anexa. Los cálculos se usaran usando las formulas y los comandos en R

### Interpretación de las diferencias entre promedios usando el criterio de Tukey. No olvidar presentar las gráficas de medias y/o si necesario las gráficas de interacción.

Es importante mencionar que estos incisos se realizaron a la par. En primer lugar se realizó la prueba de Tukey, que El intervalo de confianza se obtuvo de la siguiente manera:

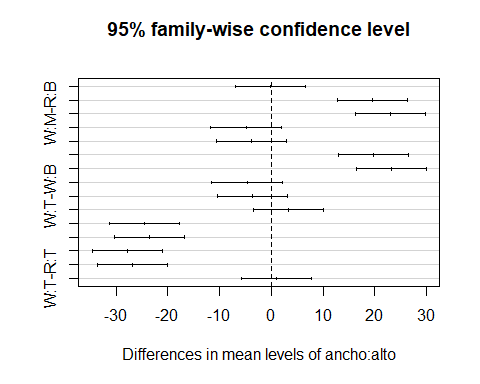
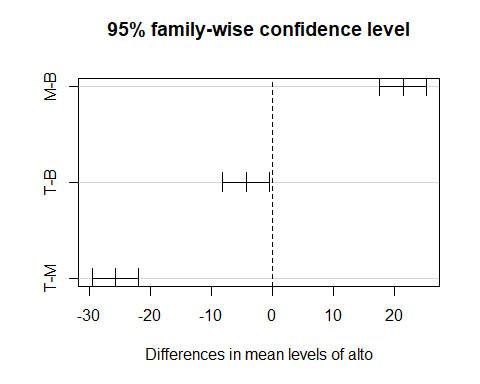
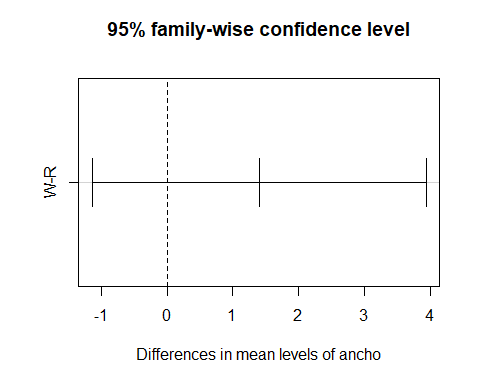
intervals = TukeyHSD(modelo)  
intervals

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = demanda ~ ancho + alto + ancho \* alto, data = datos)  
##   
## $ancho  
## diff lwr upr p adj  
## W-R 1.4 -1.141948 3.941948 0.2532947  
##   
## $alto  
## diff lwr upr p adj  
## M-B 21.4 17.587972 25.2120277 0.0000000  
## T-B -4.3 -8.112028 -0.4879723 0.0272608  
## T-M -25.7 -29.512028 -21.8879723 0.0000000  
##   
## $`ancho:alto`  
## diff lwr upr p adj  
## W:B-R:B -0.2 -6.987459 6.587459 0.9999982  
## R:M-R:B 19.6 12.812541 26.387459 0.0000058  
## W:M-R:B 23.0 16.212541 29.787459 0.0000010  
## R:T-R:B -4.9 -11.687459 1.887459 0.2218018  
## W:T-R:B -3.9 -10.687459 2.887459 0.4307117  
## R:M-W:B 19.8 13.012541 26.587459 0.0000052  
## W:M-W:B 23.2 16.412541 29.987459 0.0000009  
## R:T-W:B -4.7 -11.487459 2.087459 0.2558536  
## W:T-W:B -3.7 -10.487459 3.087459 0.4831328  
## W:M-R:M 3.4 -3.387459 10.187459 0.5660838  
## R:T-R:M -24.5 -31.287459 -17.712541 0.0000005  
## W:T-R:M -23.5 -30.287459 -16.712541 0.0000008  
## R:T-W:M -27.9 -34.687459 -21.112541 0.0000001  
## W:T-W:M -26.9 -33.687459 -20.112541 0.0000002  
## W:T-R:T 1.0 -5.787459 7.787459 0.9954333

print(intervals)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = demanda ~ ancho + alto + ancho \* alto, data = datos)  
##   
## $ancho  
## diff lwr upr p adj  
## W-R 1.4 -1.141948 3.941948 0.2532947  
##   
## $alto  
## diff lwr upr p adj  
## M-B 21.4 17.587972 25.2120277 0.0000000  
## T-B -4.3 -8.112028 -0.4879723 0.0272608  
## T-M -25.7 -29.512028 -21.8879723 0.0000000  
##   
## $`ancho:alto`  
## diff lwr upr p adj  
## W:B-R:B -0.2 -6.987459 6.587459 0.9999982  
## R:M-R:B 19.6 12.812541 26.387459 0.0000058  
## W:M-R:B 23.0 16.212541 29.787459 0.0000010  
## R:T-R:B -4.9 -11.687459 1.887459 0.2218018  
## W:T-R:B -3.9 -10.687459 2.887459 0.4307117  
## R:M-W:B 19.8 13.012541 26.587459 0.0000052  
## W:M-W:B 23.2 16.412541 29.987459 0.0000009  
## R:T-W:B -4.7 -11.487459 2.087459 0.2558536  
## W:T-W:B -3.7 -10.487459 3.087459 0.4831328  
## W:M-R:M 3.4 -3.387459 10.187459 0.5660838  
## R:T-R:M -24.5 -31.287459 -17.712541 0.0000005  
## W:T-R:M -23.5 -30.287459 -16.712541 0.0000008  
## R:T-W:M -27.9 -34.687459 -21.112541 0.0000001  
## W:T-W:M -26.9 -33.687459 -20.112541 0.0000002  
## W:T-R:T 1.0 -5.787459 7.787459 0.9954333

plot(intervals)



Analisando los resultados del análisis de Tukey se observa lo siguiente:

1. Comparando entre la anchura de los estantes, sea regular o ancho, no existe una diferencia importante, ya que para el nivel de confianza propuesto de 95%, el valor de p es mayor a 0.05.
2. Con respecto a la altura de los estantes se observa que el valor-p es menor a 0.05 de cualquier par de comparación entre las alturas, lo que significa que existe una diferencia significativa en este factor.

Ahora bien, si se analiza cada combinación algunos de los intervalos con diferencia no estadisticamente significativa son los siguientes:

W:B-R:B

R:T-R:B

W:T-R:B

R:T-W:B

W:T-W:B

W:M-R:M

W:T-R:T

Esto se puede verificar de manera grafica, ya que el intervalo de confianza incluye al cero, lo que sgnifica que existe una probabilidad de que sean iguales.

Con el fin de corroborar lo anterior, se calcularon los intervalos de confianza. Para ello tomaron en cuenta las siguientes definiciones:

1. Considere la diferencia entre los efectos de niveles i y i’ del factor 1 sobre el valor medio de la variable de respuesta.
2. Una estimación puntual de esta diferencia es
3. Para un intervalo de confianza la diferencia es

donde el punto se basa en ab (m -1) grados de libertad y MSE es el error cuadrado medio encontrado en el ANOVA bidireccional

1. Un intervalo de confianza 100 (1-A) por ciento simultáneo de Tukey para esta diferencia (en el conjunto de todas las posibles diferencias emparejadas entre los efectos de los diferentes niveles del factor 1) es:

$ $

Considerando intervalos de confianza simultáneos del 95 por ciento de Tukey, con MSE=6.12:

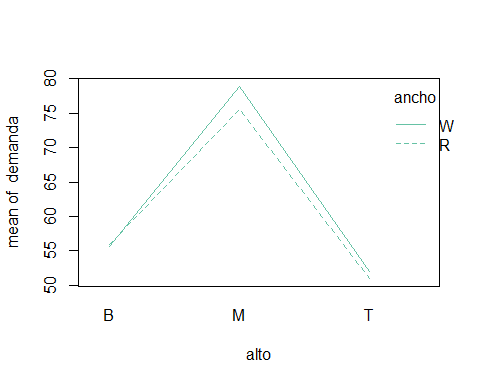
Este intervalo dice que tenemos un 95 por ciento de confianza en que cambiar de una altura de pantalla inferior a una media La altura de la exhibición aumentará la demanda media del producto de panadería entre 17.5925 y 5.2075 cajas por mes.

De manera similar, un intervalo de confianza simultáneo del 95 por ciento de Tukey para la diferencia entre los efectos de una altura de media M y arriba T, sobre la demanda es:

Este intervalo indica que para 95 por ciento de confianza en que cambiar de una altura de pantalla superior a una media la altura de la pantalla aumentará la demanda media del producto de panadería entre 21,8925 y 29,5075 casos por mes.

Finalmente, de estos tres puntos, se presenta el diagrama de interacción. Como se puede observar, el diagrama de interacción es consistente con las conclusiones anteriores, sin embargo, queda claro que proporciona información más fácil de interpretar, y no está demás confirmar que la demanda es mayor cuando las parrillas son colocadas a la altura media con una ligera preferencia sobre un ancho mayor al regular.

interaction.plot(alto, ancho, demanda)



### Aunque no está incluido, realizar comparaciones múltiples mediante el método LSD (fórmulas en el Montgomery). Los cálculos se usaran usando las formulas y los comandos en R

Para determinar los factores que influyen en la respuesta se obtiene la diferecia minima significativa (LSD)

Observación: se instalo la paquetería *agricolae*:

install.packages('agricolae', repos="http://cran.rstudio.com/")

## Installing package into 'C:/Users/diana/OneDrive/Documentos/R/win-library/4.0'  
## (as 'lib' is unspecified)

## package 'agricolae' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\diana\AppData\Local\Temp\RtmpKSqqsS\downloaded\_packages

library(agricolae)

## Warning: package 'agricolae' was built under R version 4.0.3

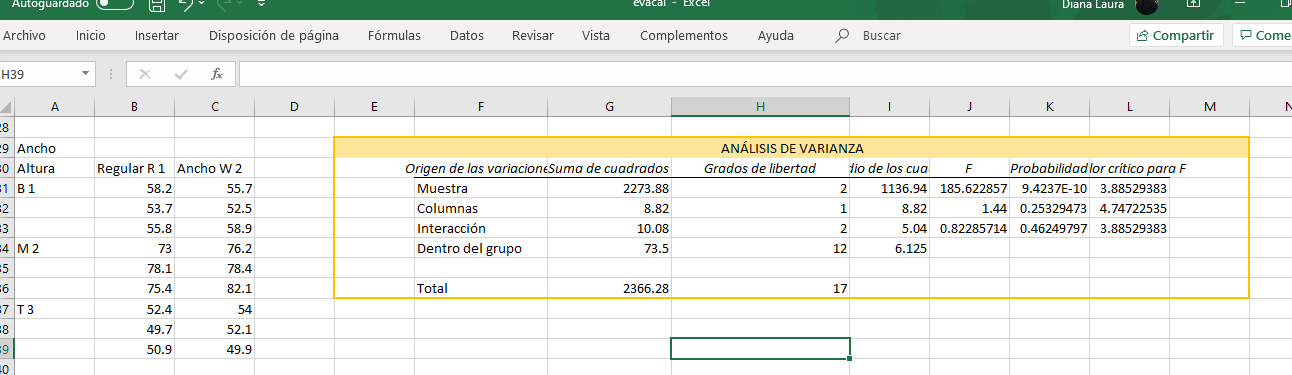
LSD<-LSD.test(modelo,"demanda")  
LSD

## $statistics  
## MSerror Df Mean CV t.value LSD  
## 6.125 12 61.5 4.024185 2.178813 7.625845  
##   
## $parameters  
## test p.ajusted name.t ntr alpha  
## Fisher-LSD none demanda 18 0.05  
##   
## $means  
## demanda std r LCL UCL Min Max Q25 Q50 Q75  
## 49.7 49.7 NA 1 44.30771 55.09229 49.7 49.7 49.7 49.7 49.7  
## 49.9 49.9 NA 1 44.50771 55.29229 49.9 49.9 49.9 49.9 49.9  
## 50.9 50.9 NA 1 45.50771 56.29229 50.9 50.9 50.9 50.9 50.9  
## 52.1 52.1 NA 1 46.70771 57.49229 52.1 52.1 52.1 52.1 52.1  
## 52.4 52.4 NA 1 47.00771 57.79229 52.4 52.4 52.4 52.4 52.4  
## 52.5 52.5 NA 1 47.10771 57.89229 52.5 52.5 52.5 52.5 52.5  
## 53.7 53.7 NA 1 48.30771 59.09229 53.7 53.7 53.7 53.7 53.7  
## 54 54.0 NA 1 48.60771 59.39229 54.0 54.0 54.0 54.0 54.0  
## 55.7 55.7 NA 1 50.30771 61.09229 55.7 55.7 55.7 55.7 55.7  
## 55.8 55.8 NA 1 50.40771 61.19229 55.8 55.8 55.8 55.8 55.8  
## 58.2 58.2 NA 1 52.80771 63.59229 58.2 58.2 58.2 58.2 58.2  
## 58.9 58.9 NA 1 53.50771 64.29229 58.9 58.9 58.9 58.9 58.9  
## 73 73.0 NA 1 67.60771 78.39229 73.0 73.0 73.0 73.0 73.0  
## 75.4 75.4 NA 1 70.00771 80.79229 75.4 75.4 75.4 75.4 75.4  
## 76.2 76.2 NA 1 70.80771 81.59229 76.2 76.2 76.2 76.2 76.2  
## 78.1 78.1 NA 1 72.70771 83.49229 78.1 78.1 78.1 78.1 78.1  
## 78.4 78.4 NA 1 73.00771 83.79229 78.4 78.4 78.4 78.4 78.4  
## 82.1 82.1 NA 1 76.70771 87.49229 82.1 82.1 82.1 82.1 82.1  
##   
## $comparison  
## NULL  
##   
## $groups  
## demanda groups  
## 82.1 82.1 a  
## 78.4 78.4 ab  
## 78.1 78.1 ab  
## 76.2 76.2 ab  
## 75.4 75.4 ab  
## 73 73.0 b  
## 58.9 58.9 c  
## 58.2 58.2 cd  
## 55.8 55.8 cde  
## 55.7 55.7 cde  
## 54 54.0 cde  
## 53.7 53.7 cde  
## 52.5 52.5 cde  
## 52.4 52.4 cde  
## 52.1 52.1 cde  
## 50.9 50.9 de  
## 49.9 49.9 e  
## 49.7 49.7 e  
##   
## attr(,"class")  
## [1] "group"

### Comprobar los resultados en Excel/yo/ en MegaStat

Para comprobar los resultados mostrados, se obtuvo la Tabla de ANOVA en excel, es importante mencionar que el análisis de datos proporciona también un resumen estadistico de la media y la desviación estandar, por lo que son estos resultados los primeros que dan una idea del comportamiento de los datos. Retomando la tabla mostrada en la sección de análisis de ANOVA, en esta sección se vuelve a presentar para analisarla con mayor detalle.

knitr::include\_graphics("cap8.png" )



De la Tabla, se observa que el análisis de varianza da el número de grados de libertad, la suma de cuadrados, el valor F y los valores criticos de F para cada factor. Si se compara con los resultados obtenidos en R, y con los cálculados mediante formulas puede notarse ligeras varicaciones, las cuales están asociadas principalemente al redondeo de números. Haciendo enfasis en la relación entre los factores alura y ancho, se corrobara lo ya mencionado: para un intervalo de onfianza del 95%, solo la altura tiene un valor F (probablidad) menor a 0.05. La tabla se calculó en el documento *evacal.exc* en la hoja *ANOVAE3*.

### Presentar claramente sus conclusiones para cada tipo de experimento, de acuerdo con los resultados obtenidos, explicando las diferencias entre los tres tipos de experimentos.

En este experimento en partícular, que consistió en estudiar la influencia de la altura y el ancho de las parrillas de aparadores en la demanda o venta, se indentificó en primer lugar los factores, estos fueron, como se mencionó, la altura y el ancho, así el experimento fue la combinación de tres alturas (bajo B, media M y alto T) y dos tipos de ancho (regular R y ancho W), las muestras fueron de tamaño tres, lo que dió en total 18 datos. Una vez, que se identificó el problema, se procedió a realizar un analisis estadistico descriptivo que consistió en obtener el diagrama de caja y el histograma. Con estos elementos fue posible identificar la tendencia de los datos tanto el promedio como la dispersión. Para este caso en partícular, se observó que la demanda es mayor cuando se colocan las parrillas a una altura media, y la demanda es menor cuando estan abajo o arriba, teniendo valores aún menores cuando son colocadas a una altura alta. Con respecto al diagrama de caja del factor ancho se vio que no existe una diferencia entre usar un tamaño regular y uno más amplio, esto formó las hipotesis. Si bien es cierto, que el diagrama de caja es un método muy rápido para identificar los valores medios es importante tomar en cuenta que se deben obtener los valores cuantitativos. Por otra parte, dado que una de las hipotesis es la normalidad de los datos, fue necesario comprobar de manera gráfica, lo anterior se realizó superponiedo la curva normal sobre el histograma, lo que dió como primer resultado que los datos no se distribuyen con normalidad.

Una vez que se identificaron las hipotesis, el siguiente paso fue comprobarlas de manera cuantitativa, para ello se realizarón pruebas normalidad y varianza. En las primeras se encontró una contradicción, ya que mientras Shapiro mostrada normalidad de los datos, Kolmorogov dio como resultado lo contrario. Esto puede deberse a que el tamaño de las muestras fueron muy pequeñas, lo cual dió poca variabilidad.

Posteriormente, se obtuvo al análisis ANOVA mediante comandos en R, por formulas matématicas y por Excel. En los tres casos se obtuvieron los mismos resultados, salvo por diferencias de decimas que pueden deberse a redondeo o por los mismos programas. El ANOVA en cualquiera de los métodos dio el número de grados de libertad, la suma de cuadrado de los errores y el valor-p. Con respecto al valor p, que es el que mayor información ofrece, se observó que para un intervalo de confianza del 95%, el factor *ancho* (p-valor < 0.05) no influye en la demanda o las ventas. Por el contrario a esto, el factor altura (\*p-valor<0.05) influye en la ventas. De forma más precisa, se encontró que cuando se colocan las parrillas a una altura media, las ventas son mayores. Estos resultados se pudieron estudiar en la sección de los intervalor de Tukey, en donde se puede analizar una combinación en particular. Finalmente, se obtuvo el diagrama de interacción, el cual permitió corroborar de manera grafica lo anterior. A partir de este trabajo se sugiere colocar las parrillas a una altura media, y dado que existe poca diferencia entre usar parrillas con ancho regular o mayor, se puede decir que se pueden emplear cualquiera de las dos, pero con una preferencia a las más anchas.