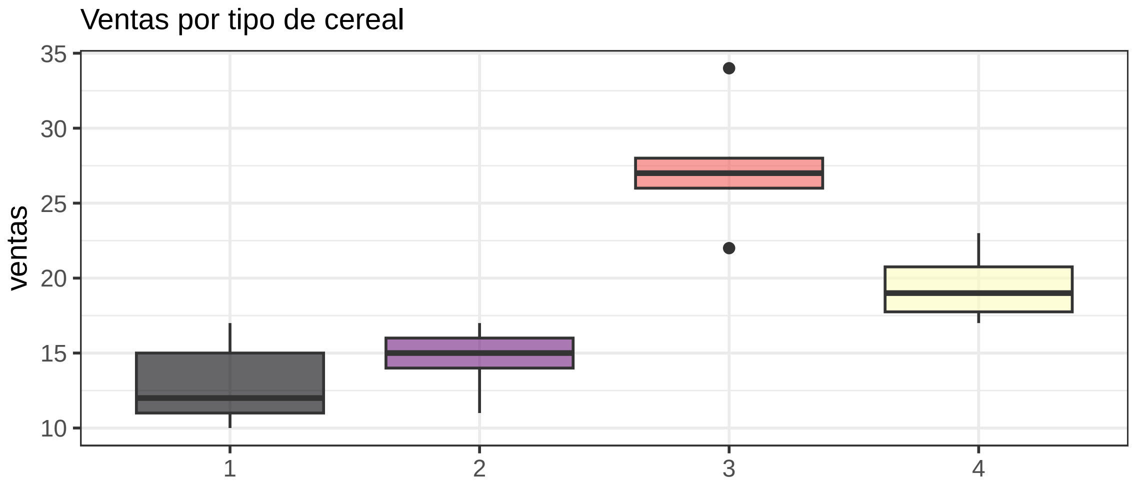
Ejercicio 3

Compararemos 4 distintos diseños de empaque de un nuevo cereal, asignados aleatoriamente a 5 tiendas como unidades muestrales, y las ventas en un periodo de 2 semanas.

En el siguiente Boxplot, podemos observar que el empaque que más se vendió es el 3, aunque con una mayor variabilidad entre las tiendas que las venden y el empaque que menos se vendió es el empaque 1.



Ajustaremos un modelo de regresión lineal múltiple del número de ventas promedio por cada tipo de empaque.

```
##
## Call:
## lm(formula = ventas ~ cereal, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -5.40  -1.75  -0.40   1.70   6.60
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   13.000      1.421   9.150 1.59e-07 ***
## cereal2        1.600      2.009   0.796   0.4383
## cereal3       14.400      2.009   7.167 3.25e-06 ***
## cereal4        6.500      2.131   3.050   0.0081 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.177 on 15 degrees of freedom
## Multiple R-squared:  0.8056, Adjusted R-squared:  0.7667
## F-statistic: 20.71 on 3 and 15 DF,  p-value: 1.367e-05
```

Las expresiones del número de ventas promedio por cada tipo de empaque son.

$E(ventas; cereal1) = \hat{\beta}_0 = 13$

$E(ventas; cereal2) = \hat{\beta}_0 + \hat{\beta}_1 = 13 + 1.6 = 14.6$

$E(ventas; cereal3) = \hat{\beta}_0 + \hat{\beta}_2 = 13 + 14.4 = 27.4$

$E(ventas; cereal4) = \hat{\beta}_0 + \hat{\beta}_3 = 13 + 6.5 = 19.5$

Las hipótesis que se contrastan con la prueba F asociada a la tabla ANOVA son, la hipotesis nula $H_0 : \hat{\beta}_1 = \hat{\beta}_2 = \hat{\beta}_3 = 0$ contra la alternativa de $H_a : \hat{\beta}_i \neq 0$ para al menos un $i = 1, 2, 3$. Esta prueba se presenta en la salida o `summary` anterior, en donde podemos observar un $p - value : 1.367e - 05$ de la prueba `F-statistic` con valor de 20.71 con 3 y 15 grados de libertad. Como el valor del $p - value < 0.05$, i.e., considerando un nivel de significancia estadística $\alpha = 0.05$, podemos concluir que se rechaza la hipótesis nula H_0, por lo que al menos un β_i es distinto de cero en el modelo planteado.

Para ver si el diseño del empaque afecta las ventas promedio, plantearemos algunas pruebas de hipótesis, usando un nivel de confianza del 95%. Nos preguntamos si $E(ventas; cereal1) \neq E(ventas; cereal2)$, $E(ventas; cereal1) \neq E(ventas; cereal3)$, $E(ventas; cereal1) \neq E(ventas; cereal4)$, $E(ventas; cereal2) \neq E(ventas; cereal3)$, $E(ventas; cereal2) \neq E(ventas; cereal4)$, $E(ventas; cereal3) \neq E(ventas; cereal4)$. Entonces, planteamos las siguientes pruebas.

Planteamiento de la hipótesis nula:

$\hat{\beta}_0 = \hat{\beta}_0 + \hat{\beta}_1 \rightarrow \hat{\beta}_1 = 0$

$\hat{\beta}_0 = \hat{\beta}_0 + \hat{\beta}_2 \rightarrow \hat{\beta}_2 = 0$

$\hat{\beta}_0 = \hat{\beta}_0 + \hat{\beta}_3 \rightarrow \hat{\beta}_3 = 0$

$\hat{\beta}_0 + \hat{\beta}_1 = \hat{\beta}_0 + \hat{\beta}_2 \rightarrow \hat{\beta}_1 = \hat{\beta}_2$

$\hat{\beta}_0 + \hat{\beta}_1 = \hat{\beta}_0 + \hat{\beta}_3 \rightarrow \hat{\beta}_1 = \hat{\beta}_3$

$\hat{\beta}_0 + \hat{\beta}_2 = \hat{\beta}_0 + \hat{\beta}_3 \rightarrow \hat{\beta}_2 = \hat{\beta}_3$

Tenemos términos redundantes, por lo que nos quedaría la siguiente prueba de hipótesis.

$H_0 : \hat{\beta}_1 = \hat{\beta}_2 = \hat{\beta}_3 = 0$ VS $H_a : \hat{\beta}_i \neq 0$ p.a. $i = 1, 2, 3$.

En la prueba global F asociada a la tabla ANOVA descrita anteriormente, se rechzó H_0. Por lo que podemos concluir que al menos un diseño de empaque afecta las ventas promedio, sin embargo no nos dice explícitamente cuál o cuáles en un análisis simultáneo.

Para esto, podemos plantear una prueba de hipótesis simultánea asociada a la igualdad de las ventas promedio entre todos los posibles pares de diferentes empaques, que puede resolverse con la prueba lineal general simultánea. A continuación, se muestra la salida de la prueba.

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = ventas ~ cereal, data = data)
##
## Linear Hypotheses:
##        Estimate Std. Error t value Pr(>|t|)
## 1 == 0    1.600      2.009   0.796   0.8549
## 2 == 0   14.400      2.009   7.167   <0.001 ***
## 3 == 0    6.500      2.131   3.050   0.0363 *
## 4 == 0  -12.800      2.009  -6.370   <0.001 ***
## 5 == 0   -4.900      2.131  -2.299   0.1420
## 6 == 0    7.900      2.131   3.707   0.0102 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Finalmente realizaremos una prueba de hipótesis para argumentar en favor o en contra de la hipótesis de que el diseño de empaque 3 es el que más aumenta las ventas en comparación con el resto de empaques. Esto es, que $E(ventas; cereal3) > E(ventas; cereal1)$, $E(ventas; cereal3) > E(ventas; cereal2)$, $E(ventas; cereal3) > E(ventas; cereal4)$. Entonces planteamos, las siguientes hipótesis.

$\hat{\beta}_0 + \hat{\beta}_2 > \hat{\beta}_0 \rightarrow \hat{\beta}_2 > 0$

$\hat{\beta}_0 + \hat{\beta}_2 > \hat{\beta}_0 + \hat{\beta}_1 \rightarrow \hat{\beta}_2 > \hat{\beta}_1$

$\hat{\beta}_0 + \hat{\beta}_2 > \hat{\beta}_0 + \hat{\beta}_3 \rightarrow \hat{\beta}_2 > \hat{\beta}_3$

Hipótesis nula: $H_0 : \hat{\beta}_2 \leq 0,\ \hat{\beta}_2 \leq \hat{\beta}_1,\ \hat{\beta}_2 \leq \hat{\beta}_3$

Hipótesis alternativa: $H_a : \hat{\beta}_2 > 0,\ \hat{\beta}_2 > \hat{\beta}_1,\ \hat{\beta}_2 > \hat{\beta}_3$

A continuación se muestra la salida, donde podemos observar que se rechaza H_0, por lo que podemos afirmar que el diseño de empaque 3 es el que más aumenta las ventas en comparación con el resto de empaques.

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = ventas ~ cereal, data = data)
##
## Linear Hypotheses:
##        Estimate Std. Error t value  Pr(>t)
## 1 <= 0   14.400      2.009   7.167 < 0.001 ***
## 2 <= 0   12.800      2.009   6.370 < 0.001 ***
## 3 <= 0    7.900      2.131   3.707 0.00304 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```