

# Notes 12

Maryclare Griffin

4/11/2023

These notes are based on Chapters 2 and 6 of KNNL.

We will continue to assume the **normal error linear regression model** for a dependent variable or response  $Y$  and independent variables, predictors, or covariates  $X_1, \dots, X_{p-1}$  is defined as:

$$Y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where:

- The elements of  $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$  are parameters
- The elements of the  $n \times p$  matrix  $\mathbf{X} = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \dots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{n,p-1} \end{pmatrix}$  are known constants
- $\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$  is a random error term elements that are  $\epsilon_i$  that are independent and normally distributed with mean  $E\{\epsilon_i\} = 0$  and variance  $\sigma^2\{\epsilon_i\} = \sigma^2$ .

In practice, we may sometimes want to test the null hypothesis that there is no linear association between the response and the predictors  $H_0: \beta_1 = \dots = \beta_{p-1} = 0$ . When we have a single predictor ( $p = 2$ ), this is the same as tests we have discussed previously, because it is a test of a single parameter  $\beta_1$ . Accordingly, we can compute the studentized statistic  $b_1/s\{b_1\}$  and compare it to quantiles of a  $t$  distribution with  $n - 2$  degrees of freedom. However when we have more than one predictor ( $p > 2$ ), this is a type of test that we have not discussed yet because it is a test of multiple parameters  $\beta_1, \dots, \beta_{p-1}$  simultaneously.

To address this, we will utilize sums of squares:

- The **total sum of squares**  $SSTO = \sum_{i=1}^n (Y_i - \bar{Y})^2$ , which describes the total variability of the response
- The **error sum of squares**  $SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ , which describes the remaining/residual variability of the response having accounted for the predictors

**Note:** The difference between the total sum of squares and the error sum of squares is defined as the **regression sum of squares**,  $SSR = SSTO - SSE$ . The regression sum of squares  $SSR$  describes the total variability of the fitted values  $SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$ .

The sums of squares  $SSTO$ ,  $SSE$ , and  $SSR$  are random variables with  $n - 1$ ,  $n - p$ , and  $p - 1$  degrees of freedom. We define the corresponding **mean squares** to be the sums of squares divided by their degrees of freedom, specifically  $MSTO = \frac{SSTO}{n-1}$ ,  $MSE = \frac{SSE}{n-p}$ , and  $MSR = \frac{SSR}{p-1}$ .

**Note:** The difference between the total mean squares and the mean squared error is not the regression mean squares,  $MSR \neq MSTO - MSE$ .

A test of the null hypothesis that there is no linear association between the response and the predictors  $H_0: \beta_1 = \dots = \beta_{p-1} = 0$  can be obtained from the  $F$ -statistic

$$\frac{MSR}{MSE}$$

First, let's examine the numerator and the denominator to get some intuition. What are their expectations under the normal error regression model?

We have already found  $E\{MSE\}$ ! Note that  $MSE = s^2$ , our unbiased estimator of the noise variance  $\sigma^2$ . We have already shown that  $E\{MSE\} = \sigma^2$ .

What about  $E\{MSR\}$ ? That's going to be trickier, and we're going to need to do a lot of algebra to figure it out.

$$\begin{aligned}
E\{MSR\} &= E\left\{\frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{p-1}\right\} \\
&= \frac{1}{p-1} E\left\{\mathbf{Y}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}\mathbf{1}'\right) \left(\mathbf{H} - \frac{1}{n} \mathbf{1}\mathbf{1}'\right) \mathbf{Y}\right\} \\
&= \frac{1}{p-1} E\left\{\mathbf{Y}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{Y}\right\} \\
&= \frac{1}{p-1} E\left\{(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})\right\} \\
&= \frac{1}{p-1} E\left\{(\boldsymbol{\beta}' \mathbf{X}' + \boldsymbol{\epsilon}') \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})\right\} \\
&= \frac{1}{p-1} E\left\{\boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \boldsymbol{\epsilon} + \boldsymbol{\epsilon}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \boldsymbol{\epsilon}\right\} \\
&= \frac{1}{p-1} \left(\boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{X}\boldsymbol{\beta} + 2E\left\{\boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \boldsymbol{\epsilon}\right\} + E\left\{\boldsymbol{\epsilon}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \boldsymbol{\epsilon}\right\}\right) \\
&= \frac{1}{p-1} \left(\boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{X}\boldsymbol{\beta} + 2\boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{E}\{\boldsymbol{\epsilon}\} + E\left\{\text{tr}\left(\boldsymbol{\epsilon}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \boldsymbol{\epsilon}\right)\right\}\right) \\
&= \frac{1}{p-1} \left(\boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{X}\boldsymbol{\beta} + E\left\{\text{tr}\left(\boldsymbol{\epsilon}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \boldsymbol{\epsilon}\right)\right\}\right) \\
&= \frac{1}{p-1} \left(\boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{X}\boldsymbol{\beta} + E\left\{\text{tr}\left(\left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \boldsymbol{\epsilon}\boldsymbol{\epsilon}'\right)\right\}\right) \\
&= \frac{1}{p-1} \left(\boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{X}\boldsymbol{\beta} + \text{tr}\left(\left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) E\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\}\right)\right) \\
&= \frac{1}{p-1} \left(\boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{X}\boldsymbol{\beta} + \text{tr}\left(\left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \sigma^2 \mathbf{I}_n\right)\right) \\
&= \frac{1}{p-1} \left(\boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{X}\boldsymbol{\beta} + \sigma^2 \text{tr}\left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right)\right) \\
&= \frac{1}{p-1} \left(\boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{X}\boldsymbol{\beta} + \sigma^2 \left(\text{tr}(\mathbf{H}) - \text{tr}\left(\frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right)\right)\right) \\
&= \frac{1}{p-1} \left(\boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{X}\boldsymbol{\beta} + \sigma^2 \left(\text{tr}\left(\mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'\right) - \frac{1}{n} \text{tr}(\mathbf{1}'_n \mathbf{1}_n)\right)\right) \\
&= \frac{1}{p-1} \left(\boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{X}\boldsymbol{\beta} + \sigma^2 \left(\text{tr}\left((\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X}\right) - 1\right)\right) \\
&= \frac{1}{p-1} \left(\boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{X}\boldsymbol{\beta} + \sigma^2 (p-1)\right) \\
&= \frac{1}{p-1} \left(\boldsymbol{\beta}' \mathbf{X}' \left(\mathbf{H} - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n\right) \mathbf{X}\boldsymbol{\beta}\right) + \sigma^2 \\
&= \frac{1}{p-1} \left(\sum_{k=0}^{p-1} \sum_{l=0}^{p-1} \beta_k \beta_l A_{ij}\right) + \sigma^2 \\
&= \sigma^2 + \frac{1}{p-1} \sum_{k=1}^{p-1} \sum_{l=1}^{p-1} \beta_k \beta_l \left(\sum_{i=1}^n (X_{ik} - \bar{X}_k) (X_{i,l} - \bar{X}_l)\right).
\end{aligned}$$

If we want to work out why the last step holds, we need to examine  $\mathbf{A}$  more carefully:

$$\begin{aligned}
\mathbf{A} &= \left( \mathbf{X}'\mathbf{H}\mathbf{X} - \frac{1}{n}\mathbf{X}\mathbf{1}_n\mathbf{1}'_n\mathbf{X} \right) \\
&= \left( \mathbf{X}'\mathbf{X} - \frac{1}{n}\mathbf{X}\mathbf{1}_n\mathbf{1}'_n\mathbf{X} \right) \\
&= \left( \begin{pmatrix} n & \sum_{i=1}^n X_{i1} & \cdots & \sum_{i=1}^n X_{i,p-1} \\ \sum_{i=1}^n X_{i1} & \sum_{i=1}^n X_{i1}^2 & \cdots & \sum_{i=1}^n X_{i1}X_{i,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{i,p-1} & \sum_{i=1}^n X_{i1}X_{i,p-1} & \cdots & \sum_{i=1}^n X_{i,p-1}^2 \end{pmatrix} - \right. \\
&\quad \left. \frac{1}{n} \begin{pmatrix} n & \sum_{i=1}^n X_{i1} & \cdots & \sum_{i=1}^n X_{i,p-1} \\ \sum_{i=1}^n X_{i1} & \sum_{i=1}^n X_{i1}^2 & \cdots & \sum_{i=1}^n X_{i1}X_{i,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{i,p-1} & \sum_{i=1}^n X_{i1}X_{i,p-1} & \cdots & \sum_{i=1}^n X_{i,p-1}^2 \end{pmatrix} \right) \\
&= \left( \begin{pmatrix} n & \sum_{i=1}^n X_{i1} & \cdots & \sum_{i=1}^n X_{i,p-1} \\ \sum_{i=1}^n X_{i1} & \sum_{i=1}^n X_{i1}^2 & \cdots & \sum_{i=1}^n X_{i1}X_{i,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{i,p-1} & \sum_{i=1}^n X_{i1}X_{i,p-1} & \cdots & \sum_{i=1}^n X_{i,p-1}^2 \end{pmatrix} - \right. \\
&\quad \left. \begin{pmatrix} n & \sum_{i=1}^n X_{i1} & \cdots & \sum_{i=1}^n X_{i,p-1} \\ \sum_{i=1}^n X_{i1} & \frac{1}{n}(\sum_{i=1}^n X_{i1})^2 & \cdots & \frac{1}{n}(\sum_{i=1}^n X_{i1})(\sum_{i=1}^n X_{i,p-1}) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{i,p-1} & \frac{1}{n}(\sum_{i=1}^n X_{i1})(\sum_{i=1}^n X_{i,p-1}) & \cdots & \frac{1}{n}(\sum_{i=1}^n X_{i,p-1})^2 \end{pmatrix} \right) \\
&= \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & (\sum_{i=1}^n X_{i1}^2) - \bar{X}_1^2 & \cdots & (\sum_{i=1}^n X_{i1}X_{i,p-1}) - \bar{X}_1\bar{X}_{p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & (\sum_{i=1}^n X_{i1}X_{i,p-1}) - \bar{X}_1\bar{X}_{p-1} & \cdots & (\sum_{i=1}^n X_{i,p-1}^2) - \bar{X}_{p-1}^2 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \sum_{i=1}^n (X_{i1} - \bar{X}_1)^2 & \cdots & \sum_{i=1}^n (X_{i1} - \bar{X}_1)(X_{i,p-1} - \bar{X}_{p-1}) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \sum_{i=1}^n (X_{i1} - \bar{X}_1)(X_{i,p-1} - \bar{X}_{p-1}) & \cdots & \sum_{i=1}^n (X_{i,p-1} - \bar{X}_{p-1})^2 \end{pmatrix}
\end{aligned}$$

Now let's examine  $E\{MSR\} = \sigma^2 + \frac{1}{p-1} \sum_{k=1}^{p-1} \sum_{l=1}^{p-1} \beta_k \beta_l (\sum_{i=1}^n (X_{ik} - \bar{X}_k)(X_{i,l} - \bar{X}_l))$ . It is:

- Equal to  $E\{MSE\}$  under  $H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0!$
- Greater than  $E\{MSE\}$  under  $H_a : \beta_k \neq 0$  for at least one value of  $k > 0!$  We're not going to delve into exactly why this is true especially if more than one  $\beta_1, \dots, \beta_{p-1}$  is greater than 0, but trust me that it is true!

This suggests that the ratio  $\frac{MSR}{MSE}$  is a good test statistic for testing  $H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$ ; we expect it to be close to 1 when  $H_0$  is true and greater than 1 otherwise.

To actually obtain a formal test of  $H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$  that uses  $\frac{MSR}{MSE}$ , we need to determine it's distribution.

Under the null  $H_0 : \beta_1 = \cdots = \beta_{p-1} = 0$ , the  $F$ -statistic has an  $F$  distribution with  $p-1$  and  $n-p$  degrees of freedom, denoted by the  $F(p-1, n-p)$  distribution.

**Note:** Let  $v_1$  and  $v_2$  be independent  $\chi^2(\nu_1)$  and  $\chi^2(\nu_2)$  random variables. We define an  $F$  random variable as  $\frac{\frac{v_1}{\nu_1}}{\frac{v_2}{\nu_2}}$ .

We can get some intuition for why  $\frac{MSR}{MSE} \sim F_{p-1, n-p}$  by recalling the definition of a  $\chi^2(\nu)$  random variable and rewriting

$$\begin{aligned} \frac{MRS}{MSE} &= \frac{\frac{MSR}{\sigma^2}}{\frac{MSE}{\sigma^2}} \\ &= \frac{\sum_{i=1}^n \left( \frac{\hat{Y}_i - \bar{Y}}{\sigma} \right)^2}{p-1} \\ &= \frac{\sum_{i=1}^n \left( \frac{Y_i - \hat{Y}_i}{\sigma} \right)^2}{n-p} \end{aligned}$$

Although we're not going to delve into exactly why, the quantities  $\sum_{i=1}^n \left( \frac{\hat{Y}_i - \bar{Y}}{\sigma} \right)^2$  and  $\sum_{i=1}^n \left( \frac{Y_i - \hat{Y}_i}{\sigma} \right)^2$  are independent, the quantity  $\sum_{i=1}^n \left( \frac{\hat{Y}_i - \bar{Y}}{\sigma} \right)^2$  is equivalent to a sum of  $p - 1$  squared standard normal random variables, and the quantity  $\sum_{i=1}^n \left( \frac{Y_i - \hat{Y}_i}{\sigma} \right)^2$  is equivalent to a sum of  $n - p$  squared standard normal random variables. Thus, the ratio  $\frac{MRS}{MSE}$  can be expressed as the ratio of two  $\chi^2$  random variables divided by their degrees of freedom.

Now we can construct a decision rule for a test, keeping in mind that it makes sense to conclude  $H_a$  when the ratio  $\frac{MSR}{MSE}$  is larger than we would expect to observe under the null.

**Note:** Let  $F(\nu_1, \nu_2)$  be a random variable distributed according to a  $F$  distribution with  $\nu_1$  and  $\nu_2$  degrees of freedom. The  $\alpha$  quantile of an  $F$  distribution with  $\nu_1$  and  $\nu_2$  degrees of freedom is denoted by  $F(\alpha; \nu_1, \nu_2)$ , and defined as satisfying:

$$P(F(\nu_1, \nu_2) \leq F(\alpha; \nu_1, \nu_2)) = \alpha.$$

Under the normal errors linear regression model, the decision rule based on a the test statistic  $\frac{MSR}{MSE}$  for a level  $1 - \alpha$  test of the null hypothesis  $H_0: \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$  versus the alternative hypothesis  $H_a: \beta_k \neq 0$  for some  $k > 0$  is:

- If  $\frac{MSR}{MSE} \leq F(1 - \alpha; p - 1, n - p)$ , conclude the null  $H_0$
- If  $\frac{MSR}{MSE} > F(1 - \alpha; p - 1, n - p)$ , conclude the alternative  $H_a$

**Note:** We can think of a test of the null hypothesis  $H_0: \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$  versus the alternative hypothesis  $H_a: \beta_k \neq 0$  for at least one  $k > 0$  as a test of the null hypothesis that there is a linear statistical association between the response  $\mathbf{Y}$  and all of the predictors  $\mathbf{X}_1, \dots, \mathbf{X}_{p-1}$  versus the alternative hypothesis that there is a linear association between the response  $\mathbf{Y}$  and at least one predictor  $\mathbf{X}_k$  with  $k > 0$ .

**Example 1:** Again, consider data from a company that manufactures refrigeration equipment, called the Toluca company. They produce refrigerator parts in lots of different sizes, and the amount of time it takes to produce a lot of refrigerator parts depends on the number of parts in the lot and several other variable factors. Let  $X$  be the number of refrigerator plots in a lot, and let  $Y$  refer to the amount of time it takes to produce a size of lot  $X$ . Suppose a cost analyst in the Toluca Company is interested in testing whether or not there is a linear association between work hours and lot size, i.e. the null hypothesis  $H_0: \beta_1 = 0$  at level  $\alpha = 0.05$  using the ratio  $\frac{MSR}{MSE}$ .

```
load("~/Dropbox/Teaching/STAT525/Spring2023/bookdata/toluca.RData")
n <- nrow(data) # Extract number of observations
Y <- data$Y # Extract response
X <- data$X # Extract predictor
linmod <- lm(Y~X) # Fit linear model
Y.hat <- linmod$fitted.values # Obtain fitted values
# Compute sums of squares
```

```

SSR <- sum((Y.hat - mean(Y))^2)
SSE <- sum((Y - Y.hat)^2)
# Compute mean squares
MSR <- SSR/1
MSE <- SSE/(n - 2)

F.statistic <- MSR/MSE

alpha <- 0.05
fquantile <- qf(1 - alpha, 1, n - 2)

```

We obtain a test statistic is  $\frac{MSR}{MSE} = 105.876$ . We compare this to the 0.95 quantile of an  $F$  distribution with 1 and 23 degrees of freedom,  $F(0.95; 1, 23) = 4.279$ . Because the test statistic exceeds  $F(0.95; 1, 23)$ , we conclude  $H_a: \beta_1 \neq 0$ , i.e. we conclude that there is evidence of a linear association between work hours and lot size at level  $\alpha = 0.05$ .

When we are performing a test, it can also be helpful to compute the corresponding  $p$ -value, which is the probability of observing a test statistic that is more extreme than the observed value if the null  $H_0$  is true. When we are performing a level  $1 - \alpha$  test of the null hypothesis  $H_0: \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$ , the  $p$ -value is

$$P\left(F(p-1, n-p) > \frac{MSR}{MSE}\right).$$

**Example 2:** Consider the same data. What is the  $p$ -value of the test of whether or not there is a linear association between work hours and lot size, i.e. the  $p$ -value of the test of the null hypothesis  $H_0: \beta_1 = 0$  based on the test statistic  $\frac{MSR}{MSE}$ ?

```
pvalue <- 1 - pf(F.statistic, 1, n - 2)
```

We obtain a  $p$ -value of  $4.448828 \times 10^{-10}$ .

To conclude, we'll work through one more example.

**Example 3:** Consider data from portrait studios in 21 cities run by Dwaine Studios, Inc. The studios specialize in portraits of children. Let  $X_1$  be the number of persons aged 16 or younger in a city, let  $X_2$  refer to per capita disposable income in a city, and let  $Y$  be the sales of portraits of children in that city from one of the 21 studios. The portrait studio is interested in testing whether or not there is a linear association between the number of persons aged 16 or younger and per capita disposable income and the sales of portraits of children having accounted for per capita disposable income, i.e. the null hypothesis  $H_0: \beta_1 = \beta_2 = 0$  at level  $\alpha = 0.05$ .

```

load("~/Dropbox/Teaching/STAT525/Spring2023/bookdata/dwaine.RData")
n <- nrow(data)
X1 <- data$X1 # Extract the first predictor
X2 <- data$X2 # Extract the second predictor
Y <- data$Y # Extract the response
linmod <- lm(Y~X1+X2) # Obtain the linear regression coefficients
summary(linmod)

```

```

##
## Call:
## lm(formula = Y ~ X1 + X2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.4239  -6.2161   0.7449   9.4356  20.2151
##

```

```

## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -68.8571    60.0170  -1.147  0.2663
## X1           1.4546     0.2118   6.868  2e-06 ***
## X2           9.3655     4.0640   2.305  0.0333 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.01 on 18 degrees of freedom
## Multiple R-squared:  0.9167, Adjusted R-squared:  0.9075
## F-statistic: 99.1 on 2 and 18 DF,  p-value: 1.921e-10

```

From the printed regression results, we can see that we observe a  $p$ -value for a test of the null hypothesis  $H_0: \beta_1 = \beta_2 = 0$  that is less than  $\alpha = 0.05$ . Accordingly, we conclude  $H_a: \beta_1 \neq 0$  or  $\beta_2 \neq 0$ , i.e. we conclude that there is evidence of a linear association between the number of persons aged 16 or younger and per capita disposable income and the sales of portraits of children having accounted for per capita disposable income at level  $\alpha = 0.05$ .