**Exercises for Linear Statistical Models**

1. Suppose we have data $x_1, \ldots, x_n$ and $y_i, \ldots, y_n$, which we model as

$$Y_i = b_0 + b_1 x_i + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2).$$

   (a) Write down the formula for the sum of squares criterion.

   (b) Take the partial derivatives of the sum of squares criterion to derive the normal equations.

   (c) Solve the normal equations to show that the least squares estimates are

$$\widehat{b}_1 = \frac{\sum_{i=1}^n (y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2}$$
$$\widehat{b}_0 = \overline{y} - \widehat{b}_1 \overline{x}$$

2. In terms of the estimates $\widehat{b}_0$ and $\widehat{b}_1$, write down the formulas for the fitted values, the residuals, and $\widehat{\sigma}^2$ (the variance estimate).

3. Explain the difference between an estimate and an estimator.

   **Answer**:

   An estimate is a function of the data, for example $\widehat{b} = \overline{y}$, whereas an estimator is a random variable, a function of the random variables used to model the data, for example $\widehat{B} = \overline{Y}$.

4. Write down how you would explain all of the assumptions of the simple linear model for $y_1, \ldots, y_n$ given $x_1, \ldots, x_n$ to your grandmother (who was a chemical engineer).

5. Suppose we have data $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$. The quantity $sxy$ is given to the left of the equals sign and is always equal to **exactly one** of the three expressions to the right of the equals sign. Circle the correct expression and show why $sxy$ is equal to it.

$$\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) \quad = \quad \sum_{i=1}^n (x_i - \overline{x}) y_i \qquad \sum_{i=1}^n (x_i - \overline{x})\overline{y} \qquad \sum_{i=1}^n \overline{x}(y_i - x_i)$$

   **Answer**:

   The first one is the correct answer. To see why, we subtract zero in a clever way

and regroup the terms:

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\overline{x})y_i &= \sum_{i=1}^{n}(x_i-\overline{x})y_i - \overline{y}\sum_{i=1}^{n}(x_i-\overline{x}) \\ &= \sum_{i=1}^{n}(x_i-\overline{x})y_i - \sum_{i=1}^{n}(x_i-\overline{x})\overline{y} \\ &= \sum_{i=1}^{n}\left[(x_i-\overline{x})y_i - (x_i-\overline{x})\overline{y}\right] \\ &= \sum_{i=1}^{n}\left[(x_i-\overline{x})(y_i-\overline{y})\right]\end{aligned}$$

6. Derive a formula for $\mathrm{Cov}(\widehat{B}_0, \widehat{B}_1)$

   **Answer**:

$$\begin{aligned}\mathrm{Cov}(\widehat{B}_0, \widehat{B}_1) &= \mathrm{Cov}(\overline{Y} - \widehat{B}_1\overline{x}, \widehat{B}_1) \\ &= \mathrm{Cov}(\overline{Y}, \widehat{B}_1) - \overline{x}\mathrm{Cov}(\widehat{B}_1, \widehat{B}_1) \\ &= \mathrm{Cov}(\overline{Y}, \widehat{B}_1) - \overline{x}\mathrm{Var}(\widehat{B}_1) \\ &= \mathrm{Cov}(\overline{Y}, \widehat{B}_1) - \overline{x}\sigma^2/sxx\end{aligned}$$

   The last step uses the formula for $\mathrm{Var}(\widehat{B}_1)$. Now we must consider the first term.

$$\begin{aligned}\mathrm{Cov}(\overline{Y}, \widehat{B}_1) &= \mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^{n}Y_i, \frac{1}{sxx}\sum_{i=1}^{n}(x_i-\overline{x})Y_i\right) \\ &= \frac{1}{n}\frac{1}{sxx}\sum_{i=1}^{n}(x_i-\overline{x})\mathrm{Var}(Y_i) \\ &= \frac{1}{n}\frac{1}{sxx}\sum_{i=1}^{n}(x_i-\overline{x})\sigma^2 \\ &= 0\end{aligned}$$

   The second-to-last step uses the fact that the $Y_i$'s are indepdent. So the answer is $-\overline{x}\sigma^2/sxx$.

7. Use the formulas for the variances and covariances of $\widehat{B}_0$ and $\widehat{B}_1$ to derive a simplified expression for the prediction variance $\mathrm{Var}(\widehat{B}_0 + \widehat{B}_1 x_i)$.

8. Suppose we have data $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$. Which of the following equations

qualify as statistical models for $y_1, \ldots, y_n$?

$$
\begin{aligned}
&Y_i = b_0 + b_1 x_i + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2) \\
&Y_i = b_0 + b_1 x_i \\
&Y_i = e^{\cos(x_i) + \varepsilon_i}, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2) \\
&0 \leq Y_i \leq 10 \\
&Y_i \overset{ind}{\sim} \text{Uniform}(0, 10)
\end{aligned}
$$

**Answer**:

Yes to the first one. It is the simple linear model that fully expresses $Y_i$ as a random variable and explains the assumptions underlying the random term $\varepsilon_i$.

No to the second because because the right side of the equation is not random.

Yes to the third one. Even though this is a nonlinear model, it still specifies $Y_i$ as a random variable and delineates all of the assumptions. This happens to be a linear model for $\log(Y_i)$, a transformation of the response.

No to the fourth one because it merely says that $Y_i$ is between 0 and 10, without specifying the distribution on that interval.

Yes to the fifth one, because it specifies the distribution–uniform–on the interval 0 to 10.

9. Under the simple linear model, there is a formula for the least squares estimators of the regression coefficients. Which property of the estimators allows us to conclude that the estimators are normally distributed?

10. Select all that are correct: A confidence interval for the slope $b_1$ . . .

    (a) contains the slope values that we rejected using a hypothesis test

    (b) contains the slope values that we failed to reject using a hypothesis test

    (c) is a set of plausible values of the slope given the data

    (d) has endpoints that are random variables

    **Answer**:

    (a) is wrong. It contains values that we fail to reject.

    (b) is right

    (c) is right, this is good interpretation of the definition of a confidence interval

    (d) is wrong. We *model* the endpoints as relizations of random variables. The endpoints themselves are numbers that we can write down and therefore are non-random.

11. Name two desirable properties of estimators, and explain why they are desirable.

12. Let

$$\boldsymbol{Y} = X\boldsymbol{b} + \boldsymbol{\varepsilon}, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2 I)$$

where $\boldsymbol{Y}$ is $n \times 1$ and $X$ is $n \times p + 1$.

What are the dimensions of $\boldsymbol{b}$ and $\boldsymbol{\varepsilon}$?

What is the expectation of $\boldsymbol{Y}$?

What is the expectation of $X^T\boldsymbol{Y}$?

What is the expectation of $\widehat{\boldsymbol{B}} = (X^TX)^{-1}X^T\boldsymbol{Y}$?

What is the covariance matrix for $\varepsilon$?

What is the covariance matrix for $\widehat{\boldsymbol{B}}$?

**Answer**:

$\boldsymbol{b}$ must be $p + 1 \times 1$ in order that $X\boldsymbol{b}$ is $n \times 1$

$\boldsymbol{\varepsilon}$ must be $n \times 1$ to match the dimensions of $X\boldsymbol{b}$

$$E(\boldsymbol{Y}) = E(X\boldsymbol{b} + \boldsymbol{\varepsilon}) = \boldsymbol{X}\boldsymbol{b} + E(\boldsymbol{\varepsilon}) = \boldsymbol{X}\boldsymbol{b}$$

$$E(X^T\boldsymbol{Y}) = X^TE(\boldsymbol{Y}) = X^TX\boldsymbol{b}$$

$$E((X^TX)^{-1}X^T\boldsymbol{Y}) = (X^TX)^{-1}E(X^T\boldsymbol{Y}) = (X^TX)^{-1}X^TX\boldsymbol{b} = \boldsymbol{b}$$

$\varepsilon_1, \ldots, \varepsilon_n$ independent implies that $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ if $i \neq j$. When $i = j$, $\text{Cov}(\varepsilon_i, \varepsilon_i) = \text{Var}(\varepsilon_i)$, which means that $\text{Cov}(\boldsymbol{\varepsilon}_i) = \sigma^2 I_n$

$$\begin{aligned}
\text{Cov}(\widehat{\boldsymbol{B}}) &= \text{Cov}((X^TX)^{-1}X^T\boldsymbol{Y}) \\
&= (X^TX)^{-1}X^T\text{Cov}(\boldsymbol{Y})X(X^TX)^{-1} \\
&= (X^TX)^{-1}X^T(\sigma^2 I)X(X^TX)^{-1} \\
&= \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1} = \sigma^2(X^TX)^{-1}
\end{aligned}$$

13. Recall the simple linear model

$$Y_i = b_0 + b_1x_i + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2).$$

Write down the design matrix $X$.

Calculate the entries of $X^TX$

Calculate the entries of $(X^TX)^{-1}$

Calculate $X^T\boldsymbol{y}$.

Show that $\widehat{\boldsymbol{b}} = (X^TX)^{-1}X^T\boldsymbol{y}$ produces the same least squares estimates that we originally derived. You'll have to use the formula for the inverse of a $2 \times 2$ matrix.

**Answer:**

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n 1 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} = \begin{bmatrix} n & n\overline{x} \\ n\overline{x} & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

$$(X^T X)^{-1} = \begin{bmatrix} n & n\overline{x} \\ n\overline{x} & \sum_{i=1}^n x_i^2 \end{bmatrix}^{-1} = \frac{1}{n\sum_{i=1}^n x_i^2 - n^2(\overline{x})^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -n\overline{x} \\ -n\overline{x} & n \end{bmatrix}$$

$$X^T \boldsymbol{y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix} = \begin{bmatrix} n\overline{y} \\ \sum_{i=1}^n x_i y_i \end{bmatrix}$$

$$\begin{aligned} (X^T X)^{-1} X^T \boldsymbol{y} &= \frac{1}{n\sum_{i=1}^n x_i^2 - n^2(\overline{x})^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -n\overline{x} \\ -n\overline{x} & n \end{bmatrix} \begin{bmatrix} n\overline{y} \\ \sum_{i=1}^n x_i y_i \end{bmatrix} \\ &= \frac{1}{n\sum_{i=1}^n x_i^2 - n^2(\overline{x})^2} \begin{bmatrix} n\overline{y}\sum_{i=1}^n x_i^2 - n\overline{x}\sum_{i=1}^n x_i y_i \\ -n^2\overline{x}\,\overline{y} + n\sum_{i=1}^n x_i y_i \end{bmatrix} \end{aligned}$$

Let's look at the second entry:

$$\begin{aligned} &= \frac{n\sum_{i=1}^n x_i y_i - n^2\overline{x}\,\overline{y}}{n\sum_{i=1}^n x_i^2 - n^2(\overline{x})^2} \\ &= \frac{\sum_{i=1}^n x_i y_i - n\overline{x}\,\overline{y}}{\sum_{i=1}^n x_i^2 - n(\overline{x})^2} \\ &= \frac{sxy}{sxx} \end{aligned}$$

where the last line applies after applying "the trick" a few times. This is our usual formula for $\widehat{b}_1$

The first entry is a different formula for $\widehat{b}_0$ than you have seen, because we have written $\widehat{b}_0$ in terms of $\widehat{b}_1$. But if you expand $\widehat{b}_0 = \overline{y} - \overline{x}\widehat{b}_1$, you will get the first entry.

14. The residual sum of squares criterion for the multiple linear model is

$$rss(\boldsymbol{b}^*) = \sum_{i=1}^{n}(y_i - \sum_{j=0}^{p} b_j^* x_{ij})^2$$

Derive the $k$th normal equation by differentiating the residual sum of squares.

15. Show that the set of $p+1$ normal equations can be written as $X^T\boldsymbol{y} = X^T X\widehat{\boldsymbol{b}}$

16. Let

$$X = \begin{bmatrix}\boldsymbol{x}_0 & \cdots & \boldsymbol{x}_p\end{bmatrix}$$

And suppose for this problem that $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_p$ are orthonormal, which means that $\boldsymbol{x}_j^T\boldsymbol{x}_k = 0$ if $j \neq k$ and 1 if $j = k$.

Show that the least squares estimates of $\boldsymbol{b}$ are $\widehat{b}_j = \boldsymbol{x}_j^T\boldsymbol{y}$.

**Answer**:

First, let's write down $X^T X$

$$X^T X = \begin{bmatrix}\boldsymbol{x}_0^T \\ \vdots \\ \boldsymbol{x}_p^T\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_0 & \cdots & \boldsymbol{x}_p\end{bmatrix} = \begin{bmatrix}\boldsymbol{x}_0^T\boldsymbol{x}_0 & \boldsymbol{x}_0^T\boldsymbol{x}_1 & \boldsymbol{x}_0^T\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_0^T\boldsymbol{x}_p \\ \boldsymbol{x}_1^T\boldsymbol{x}_0 & \boldsymbol{x}_1^T\boldsymbol{x}_1 & \boldsymbol{x}_1^T\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_1^T\boldsymbol{x}^p \\ \boldsymbol{x}_2^T\boldsymbol{x}_0 & \boldsymbol{x}_2^T\boldsymbol{x}_1 & \boldsymbol{x}_2^T\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_2^T\boldsymbol{x}^p \\ \vdots & & & & \\ \boldsymbol{x}_p^T\boldsymbol{x}_0 & \boldsymbol{x}_p^T\boldsymbol{x}_1 & \boldsymbol{x}_p^T\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_p^T\boldsymbol{x}_p\end{bmatrix}$$

$$= \begin{bmatrix}1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1\end{bmatrix} = I$$

This means that

$$(X^T X)^{-1}X^T\boldsymbol{y} = X^T\boldsymbol{y} = \begin{bmatrix}\boldsymbol{x}_0^T \\ \vdots \\ \boldsymbol{x}_p^T\end{bmatrix}\boldsymbol{y} = \begin{bmatrix}\boldsymbol{x}_0^T\boldsymbol{y} \\ \vdots \\ \boldsymbol{x}_p^T\boldsymbol{y}\end{bmatrix}$$

In general, the columns of $X$ will not be orthonormal, or even orthogonal. This result applies only in the case when the columns of $X$ are orthonormal.

17. What are the two defining properties of projection matrices?

**Answer:**

Symmetric ($P = P^T$) and idempotent ($PP = P$)

18. Show that $X(X^T X)^{-1}X^T$ is a projection matrix.

19. Suppose that

$$U = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{2} \\ 1/\sqrt{3} & 0 \end{bmatrix} \text{ and } M = UU^T$$

Calculate the entries of $M$ and show that $M$ is a projection matrix.

20. Consider

$$P = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Show that $P$ is a projection matrix.

Describe the space that $P$ projects onto.

21. Show that the fitted values are in the column space of $X$. In other words, show that the fitted values can be written as a linear combination of the columns of $X$, $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_p$.

**Answer:**

The fitted values are $\widehat{\boldsymbol{y}} = X(X^TX)^{-1}X^T\boldsymbol{y}$. We can rewrite this as $\widehat{\boldsymbol{y}} = X\widehat{\boldsymbol{b}}$, which is equal to $\widehat{b}_0\boldsymbol{x}_0 + \cdots + \widehat{b}_p\boldsymbol{x}_p$, which is a linear combination of the columns of $X$. In fact, it's not just any linear combination of the columns of $X$; it's the particular linear combination whose coefficients are the regression coefficients. Since we say that the coefficients of the linear combination are the coordinates of the vector with respect to the basis, the regression coefficient estimates are the coordinates of the fitted values with respect to the basis defined by the columns of $X$.

22. Let $\boldsymbol{b} = (b_0, b_1, b_2)$ and $\widehat{\boldsymbol{B}}$ be the least squares estimator for $\boldsymbol{b}$. Further, let $M = \sigma^2(X^TX)^{-1}$, and $M_{ij}$ be the $(i, j)$ entry of $M$.

What is the expected value of $\begin{bmatrix} 0 & 1 & -1 \end{bmatrix}\widehat{\boldsymbol{B}}$?

What is the variance of $\begin{bmatrix} 0 & 1 & -1 \end{bmatrix}\widehat{\boldsymbol{B}}$?

23. Let

$$\boldsymbol{Y} = X\boldsymbol{b} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I)$$

Recall that if $\boldsymbol{Z} \sim N(\boldsymbol{\mu}, \Sigma)$, then $\boldsymbol{a} + M\boldsymbol{Z} \sim N(\boldsymbol{a} + M\boldsymbol{\mu}, M\Sigma M^T)$.

What is the distribution of $\boldsymbol{Y}$?

What is the distribution of $X^T\boldsymbol{Y}$?

What is the distribution of $\widehat{\boldsymbol{B}} = (X^TX)^{-1}X^T\boldsymbol{Y}$?

Does $\boldsymbol{Y}^T\boldsymbol{Y}$ follow a normal distribution? Why or why not?

**Answer:**

Because $\boldsymbol{Y}$ is a linear transformation of $\boldsymbol{\varepsilon}$, and $\boldsymbol{\varepsilon}$ is MVN,

$$\boldsymbol{Y} \sim N(X\boldsymbol{b}, \sigma^2 I)$$

Because $X^T\boldsymbol{Y}$ is a linear transformation of $\boldsymbol{Y}$, and $\boldsymbol{Y}$ is MVN,

$$X^T\boldsymbol{Y} \sim N(X^TX\boldsymbol{b}, \sigma^2 X^TX)$$

Because $\widehat{\boldsymbol{B}}$ is a linear transformatino of $\boldsymbol{Y}$, and $\boldsymbol{Y}$ is MVN,

$$(X^TX)^{-1}X^T\boldsymbol{Y} \sim N((X^TX)^{-1}X^TX\boldsymbol{b}, \sigma^2(X^TX)^{-1}(X^TX)(X^TX)^{-1}) = N(\boldsymbol{b}, \sigma^2(X^TX)^{-1})$$

24. The following is from a longitudinal study measuring heights, weights, etc. of a group of girls. Height2 = height at age 2, Height9 = height at age 9, Height18 = height at age 18, and LegCirc9 = leg circumference at age 9.

Below is output from a regression of Height18 on Height2, Height9, and LegCirc9

```
Call: lm(formula = Height18 ~ Height2 + Height9 + LegCirc9 )

                Estimate Std. Error t value Pr(>|t|)
(Intercept)      35.6880    11.2669   3.167  0.00233
Height2           0.2945     0.1827   1.612  0.11163
Height9           0.9016     0.1195   7.547 1.71e-10
LegCirc9         -0.5986     0.2089  -2.865  0.00559

Residual standard error: 3.404 on 66 degrees of freedom
Multiple R-squared:  0.6997,Adjusted R-squared:  0.6861
F-statistic: 51.27 on 3 and 66 DF,  p-value: < 2.2e-16
```

(a) How many individual girls were in this dataset?

**Answer:**

The degrees of freedom is 66. Since there are 3 covariates in the model + 1 intercept, the degrees of freedom was calculated as $n - 4 = 66$, which means that $n = 70$.

(b) The model fit to the height data was

$$Y_i = b_0 + b_1x_{i1} + b_2x_{i2} + b_3x_{i3} + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2),$$

where the covariates are listed in the same order as in the output above. What is the estimate of $\sigma$?

**Answer:**

In the output, $\widehat{\sigma}$ is "Residual standard error", which is listed as 3.404.

(c) Give an interpretation for the result of the $t$-test for the Height2 regression coefficient. Why does this make sense for height data?

**Answer:**

We do not have sufficient evidence to reject the hypothesis that $b_1$ is zero. In other words, 0 is a plausible value for $b_1$ given the data we have observed. It is not surprising that we don't see evidence for height at age 2 as being important after accounting for height at age 9, since information about height at age 9 likely makes information about height at age 2 irrelevant.

(d) Can we conclude that height at age 2 is not important for predicting height at age 18?

**Answer:**

No, from this model, we can only conclude that we do not have sufficient evidence to say that height at age 2 is important, after accounting for height at age 9 and leg circumference at age 9.

(e) What is the residual sum of squares for this model?

**Answer:**

We know that $\widehat{\sigma}^2 = rss/df$, so $rss = \widehat{\sigma}^2 df = 3.404^2 * 66 = 764.7563$.

(f) What is $syy$ for this dataset, and what is the standard deviation of the response?

**Answer:**

We know that $R^2 = 1 - rss/syy$, so $syy = rss/(1 - R^2) = 764.7563/(1 - .6997) = 2546.641$

25. Show that $X^T\widehat{\boldsymbol{e}} = \mathbf{0}$, that is, the observed residual vector is orthogonal to every column of the design matrix.

26. Let $T = Z/\sqrt{W/m}$. $T$ has a $t$ distribution with $m$ degrees of freedom if what 3 things are true?

**Answer:**

1. $Z \sim N(0,1)$, 2. $W \sim \chi^2_m$, and 3. $Z$ independent of $W$.

27. Suppose we have a hypothesis for the multiple linear model that can be written as

$$H_0 : \boldsymbol{c}^T\boldsymbol{b} = a$$

where $\boldsymbol{c}$ is a $p+1$ by 1 vector, $a$ is a known (hypothesized) scalar, and, as usual, $\boldsymbol{b}$ is the vector of regression coefficients. For example, the hypothesis $H_0 : b_2 - b_1 = 0$ can be written in this form.

(a) Write down the $t$ statistic for this test in terms of $\widehat{\boldsymbol{b}}$, $X$, and $\widehat{\sigma}^2$.

**Answer:**

$$t = (\boldsymbol{c}^T\widehat{\boldsymbol{b}} - a)/\sqrt{\widehat{\sigma}^2\boldsymbol{c}^T(X^TX)^{-1}\boldsymbol{c}}$$

(b) Show that the random version of the $t$ statistic has a $T$ distribution.

**Answer:**

We can write the random version of the statistic as

$$T = \frac{\boldsymbol{c}^T\widehat{\boldsymbol{B}} - a}{\sqrt{\sigma^2\boldsymbol{c}^T(X^TX)^{-1}\boldsymbol{c}}} \frac{1}{\sqrt{\frac{\widehat{\sigma}^2(n-p-1)}{\sigma^2(n-p-1)}}}$$

The first term is $N(0,1)$ because under the null hypothesis, it is a mean-zero normal divided by its standard deviation. The second term is a Chi-squared random variable with $n-p-1$ degrees of freedom divided by $n-p-1$. And the first and second terms are independent because $\widehat{\boldsymbol{B}}$ is independent of $\widehat{\sigma}^2$, due to the fact that $\widehat{\sigma}^2$ is a function of the (random version) of the residuals, and the residuals are independent $\widehat{\boldsymbol{B}}$.

28. Here is a dataset with results from the pinewood derby:

| $i$ | racer | lane | heat | time |
|---|---|---|---|---|
| 1 | Mary | 1 | 1 | 3.123 |
| 2 | Suzy | 2 | 1 | 3.147 |
| 3 | June | 3 | 1 | 3.201 |
| 4 | Mary | 3 | 2 | 3.133 |
| 5 | Suzy | 1 | 2 | 3.168 |
| 6 | June | 2 | 2 | 3.192 |
| 7 | Mary | 2 | 3 | 3.118 |
| 8 | Suzy | 3 | 3 | 3.146 |
| 9 | June | 1 | 3 | 3.225 |

(a) Write down the design matrix for the one-factor model with **racer** as the factor.j

(b) Write down the design matrix for the one-factor model with **lane** as the factor.

(c) Write down the design matrix for the one-factor model with **heat** as the factor.

**Answers:**

Using R's convention to alphatetize the levels, we get

$$\text{racer} \qquad\qquad \text{lane} \qquad\qquad \text{heat}$$

$$X = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

29. Suppose we have the factor model:

$$Y_i = b_0 + b_{j(i)} + \varepsilon_i \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2)$$

where $j(i)$ is the factor level of the $i$th observation. Our dataset has the following design matrix:

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

(a) How many levels of the factor are in our dataset?

**Answer:** 3

(b) What is the rank of the design matrix?

**Answer:** 3 because last 3 columns sum to the first

(c) Compute $X^T X$.

**Answer:**

$$X^T X = \begin{bmatrix} 5 & 2 & 2 & 1 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

(d) State which of the following are estimable functions. For those that are, give a linear combination of observations whose expected value is the estimable function. $b_0 + b_1$, $b_1 + b_2$, $b_2 - b_3$, $b_1 - 0.5(b_2 + b_3)$, $2b_3 - b_1 - b_2$

**Answer:**

$$
\begin{aligned}
b_0 + b_1 &= E(Y_2) \text{ (yes)} \\
b_1 + b_2 &\text{ (no)} \\
b_2 - b_3 &= E(Y_1) - E(Y_4) \text{ (yes)} \\
b_1 - 0.5(b_2 + b_3) &= E(Y_2) - 0.5(E(Y_1) + E(Y_4)) \text{ (yes)} \\
2b_3 - b_1 - b_2 &= 2E(Y_4) - E(Y_2) - E(Y_1) \text{ (yes)}
\end{aligned}
$$

(e) Suppose we pick a solution to the normal equations for which $\widehat{b}_0 = 0$. Then how do we interpret $\widehat{b}_2$?

**Answer:**

$\widehat{b}_2$ is the expected value for factor level 2 (observations 1 and 3)

(f) Suppose we pick a solution to the normal equations for which $\widehat{b}_1 = 0$. Then how do we interpret $\widehat{b}_2$?

**Answer**

$\widehat{b}_2$ is the expected value for factor level 2 minus expected value for factor level 1

30. Suppose that we have six dogs. Dogs 1 and 2 are greyhounds, dogs 3 and 4 are whippets, and dogs 5 and 6 are Italian greyhounds. We model their weights as:

$$
\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_6).
$$

Dog 3 is a whippet that weights 26.0 pounds and is the cutest dog you'll ever meet.

(a) Write these quantities in terms of the notation above:

**Answer:**

$E(Y_1) = b_0 + b_1$

$E(Y_3) = b_0 + b_2$

$E(Y_5) = b_0 + b_3$

(b) Give interpretations in words of the following quantities. If no interpretation exists, write "no interpretation". Hint: use your answers from the previous part.

**Answer:**

$b_0$: no interpretation exists if no constraints are given

$b_0 + b_1$ expected value of factor level 1

$b_3 - b_2$ expected value of factor level 3 minus expected value of factor level 2

(c) Using mathematical notation, write down the hypothesis that all three breeds of the dogs have the same expected weight.

Should we use a $t$ test or an $F$ test? Give the degrees of freedom for the test.

**Answer** $H_0 : b_1 = b_2 = b_3$. We need an F test here because the hypothesis cannot be written as comparing a linear combination of the parameters to a number

(d) Using mathematical notation, write down the hypothesis that we expect a whippet to weigh 15 pounds more than an Italian greyhound.

Should we use a $t$ test or an $F$ test? Give the degrees of freedom for the test.

**Answer:** $H_0 : b_2 - b_1 = 15$. We can use either a t-test or an F test here.

31. The SAT data contains average SAT scores from each state. Each state was grouped into 1 of 9 regions: ENC, ESC, MA, MTN, NE, PAC, SA, WNC, WSC. Consider the following model for the average SAT math score from the $i$th state:

$$Y_i = b_0 + b_{j(i)} + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2),$$

where $j(i)$ is the region number (1-9) of the $i$th state. Below is R output from fitting the model:

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  551.000      8.093  68.088  < 2e-16 ***
regionESC      1.750     12.139   0.144 0.886059
regionMA     -52.333     13.215  -3.960 0.000284 ***
regionMTN    -11.500     10.316  -1.115 0.271285
regionNE     -49.167     10.957  -4.487 5.52e-05 ***
regionPAC    -36.200     11.445  -3.163 0.002899 **
regionSA     -61.333     10.093  -6.077 3.08e-07 ***
regionWNC     29.857     10.596   2.818 0.007339 **
regionWSC    -11.750     12.139  -0.968 0.338599

Residual standard error: 18.1 on 42 degrees of freedom
Multiple R-squared:  0.7733,Adjusted R-squared:  0.7302
F-statistic: 17.91 on 8 and 42 DF,  p-value: 2.832e-11
```

(a) Which region's coefficient did R set to zero (the reference region)?

**Answer:** ENC. We know because it doesn't appear in the table.

(b) What is the estimated expected SAT score in the reference region?

**Answer:** The intercept: 551.0

(c) What is the estimated expected SAT score in the NE region?

**Answer:** $Intercept + regionNE = 551.0 + (-49.167) = 501.833$

(d) Which region has the highest estimated expected SAT score, and what is the estimate?

**Answer:** regionWNC is the largest positive coefficient, so it must have the highest average score. If all of the region coefficients were negative, then the highest would be the reference region.

(e) What is the $t$ statistic for testing whether MA has the same expected SAT score as the reference region?

**Answer:** -3.960

(f) How many degrees of freedom in the $t$ statistic from the previous part?

**Answer:** 42

(g) In terms of $b_j$'s, write down the hypothesis tested by the $F$ test in the output.

**Answer:** $H_0 : b_1 = b_2 = \cdots = b_8$

(h) In the $F$ test, what is the $rss$ for the full model?

**Answer:** We can get it from the Residual standard error ($\widehat{\sigma}$)

$$\widehat{\sigma} = \sqrt{rss_1/df_1} \quad \Longrightarrow \quad rss_1 = \widehat{\sigma}^2 * df_1 = 18.1^2 * 42 = 13759.6$$

(i) What is the $rss$ for the reduced model? You can use knowledge of the $F$ statistic or the $R^2$ statistic to solve this. Try it both ways and make sure you get the same answer.

**Answer:**

$$F = \frac{(rss_0 - rss_1)/(J-1)}{rss_1/df_1} \quad \Longrightarrow \quad rss_0 = F * rss_1/df_1 * (J-1) + rss_1$$
$$= 17.91 * 18.1^2 * 8 + 13759.6 = 60699.6$$

(j) Think carefully: how many "states" are in this dataset?

**Answer:** There are 42 degrees of freedom, and $J = 9$ regions (levels), so there must be 51 "states" in the dataset. Turns out that District of Columbia is included as a state in the dataset

32. Consider the following model for 3pm temperatures:

$$Y_i = b_0 + b_{j(i)} + c_1 x_i + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2),$$

where $j(i)$ is the month associated with the $i$th observation, and $x_i$ is the 7am temp (atobstemp). Below is output from fitting the model to data from months 1 through 3.

```
                  Estimate Std. Error t value Pr(>|t|)
b0 - (Intercept) 15.06816    1.62505   9.272   <2e-16
b2 - mon02        2.90206    1.64191   1.767   0.0789
b3 - mon03        3.36943    1.56601   2.152   0.0328
c1 - atobstemp    1.00450    0.04983  14.471   <2e-16

Residual standard error: 8.704 on 173 degrees of freedom
Multiple R-squared:  0.5712,Adjusted R-squared:  0.5639
F-statistic: 78.14 on 3 and 176 DF,  p-value = 0.0459
```

(a) Is this an additive model or an interaction model?

**Answer:** Additive model since the effect of 7am temp does not depend on month, and vice versa.

(b) Give interpretations for all 4 regression coefficients.

**Answer:** $b_0$ is the expected 3pm temperature in month 1 when the 7am temperature is 0; in other words, the intercept for month 1. $b_2$ is the intercept for month 2 minus the intercept for month 1. $b_3$ is the intercept for month 3 minus the intercept for month 1. $c_1$ is the expected increase in 3pm temperature when 7am temperature increase by 1 degree; in other words, the slope with respect to 7am temperature.

(c) Draw a plot by hand with the three fitted regression lines.

(d) It goes without saying that the month 3 temperatures (March) are higher on average than the month 1 temperatures (January). This model says something slightly different. Explain what this model says and why it is also plausible.

**Answer:** This model says that for a given 7am temperature, the expected 3pm temperature is 3.36 degrees higher in month 3 than it would be for the same 7am temperature in month 1. Since the slope with respect to 7am temperature is very close to 1, the model basically says that the 3pm temperature in January is roughly 15 degrees higher than the 7am temperature, and in March it's likely to to be about 18.4 degrees higher. This is consistent with my experience that it's quite possible to have a day in March that is quite cold in the morning, but warms up considerably by the afternoon.

(e) I altered some numbers in the R output. See how many inconsistencies you can find, and explain why.

**Answer:** The t-statistic is supposed to be the estimate divided by its standard error, so one of the values for $c_1$ is not right. An F-statistic of 78.18 will be produce a miniscule p-value, so either the F-statistic or the p-value of 0.0459 must be wrong (probably the p-value is wrong). Can you find any others?

(f) Write down the reduced model for the F-test that appears in the R output.

**Answer:** $Y_i = b_0 + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2)$

33. Below is an incomplete table of estimated expected values for a two factor additive model for levels $j = 1, 2, 3$ and $k = 1, 2, 3$. Complete the table with numbers and write out what estimates R would produce for $b_0$, $b_1$, $b_2$, $b_3$, $c_1$, $c_2$, and $c_3$.

| | | $k$ | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| | 1 | | 1 | |
| $j$ | 2 | 0 | 7 | 8 |
| | 3 | | 5 | |

**Answer:** Using the additive property, we get

| | | $k$ | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| | 1 | -6 | 1 | 2 |
| $j$ | 2 | 0 | 7 | 8 |
| | 3 | -2 | 5 | 6 |

$\widehat{b}_0 = -6, \quad \widehat{b}_1 = 0, \quad \widehat{b}_2 = 6, \quad \widehat{b}_3 = -2, \quad \widehat{c}_1 = 0, \quad \widehat{c}_2 = 7, \quad \widehat{c}_3 = 1$

You can double check your answers by ensuring that $\widehat{b}_0 + \widehat{b}_j + \widehat{c}_k$ matches the values in the table

34. In our housing example, we have a dataset with sale prices for houses in the St. Louis area. Let $y_i$ be the sale price of house $i$, let $x_i$ be its size in square feet, and let $j(i)$ be a factor variable for the zip code. Each zip code is given a label 1 through $J$.

    We model the data as:

$$Y_i = b_0 + b_{j(i)} + (c_0 + c_{j(i)})x_i + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2)$$

    If we apply R's convention that $b_1 = 0$ and $c_1 = 0$,

    (a) Explain in words what this model assumes about the relationship between expected sale price and zip code and size.

    **Answer:** For each zip code, the expected sale prices are a linear function of square footage. Each zip code has its own linear relationship, with different intercepts and slopes for each zip code.

    (b) What is the interpretation of $b_0 + b_j$?

    **Answer:** Intercept for zip code $j$

    (c) What is the interpretation of $c_0 + c_j$?

    **Answer:** Slope for zip code $j$

    (d) What is the interpretation of $c_3 - c_2$?

    **Answer:** Slope for zip code 3 minus slope for zip code 2.

(e) Under R's default constraint that $b_1 = 0$ and $c_1 = 0$, give interpretations for the following parameters:

**Answers:**

$b_0$: intercept for zip code 1

$c_0$: slope for zip code 1

$b_2$: intercept for zip code 2 minus intercept for zip code 1

$c_3$: slope for zip code 3 minus slope for zip code 1

(f) In terms of the parameters, what is the expected sale price for a 2000 square foot house in the fourth zip code?

**Answers:** $b_0 + b_4 + (c_0 + c_4)2000$

35. In a memory experiment, subjects of different ages were given different strategies for remembering words from a list. The number of words memorized by subject $i$ is $y_i$. We fit the following model to the data:

$$Y_i = b_0 + b_{j(i)} + c_{k(i)} + (bc)_{j(i),k(i)} + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2),$$

where $j(i)$ is the age group of observation $i$ (Older, Younger), and $k(i)$ is the memory process of observation $i$ (Adjective, Counting, Imagery, Intentional, Rhyming). Below is output from fitting the model.

```
                                        Estimate Std. Error t value Pr(>|t|)
b0 - (Intercept)                         11.0000     0.8959  12.279  < 2e-16
b2 - AgeYounger                           3.8000     1.2669   2.999  0.00350
c2 - ProcessCounting                     -4.0000     1.2669  -3.157  0.00217
c3 - ProcessImagery                       2.4000     1.2669   1.894  0.06139
c4 - ProcessIntentional                   1.0000     1.2669   0.789  0.43201
c5 - ProcessRhyming                      -4.1000     1.2669  -3.236  0.00170
(bc)22 - AgeYounger:ProcessCounting      -4.3000     1.7917  -2.400  0.01846
(bc)23 - AgeYounger:ProcessImagery        0.4000     1.7917   0.223  0.82385
(bc)24 - AgeYounger:ProcessIntentional    3.5000     1.7917   1.953  0.05387
(bc)25 - AgeYounger:ProcessRhyming       -3.1000     1.7917  -1.730  0.08702
---------------------------------------------------------------------------
Residual standard error: 2.833 on 90 degrees of freedom
Multiple R-squared:  0.7293,Adjusted R-squared:  0.7022
F-statistic: 26.93 on 9 and 90 DF,  p-value: < 2.2e-16
```

(a) In 20 words or fewer, give an interpertation of the estimate 11.000 in the first row.

**Answer:** Estimate of the expected number of words memorized by Older subjects using the Adjective strategy.

(b) Find the p-value 0.00217. Explain in words what ages and processes are being compared in that test.

**Answer:** Comparing Older subjects using the Counting strategy to older subjects using the Adjective strategy.

(c) What is the estimated expected number of words memorized by Older subjects using the Rhyming process?

**Answer:** This is $\widehat{b}_0 + \widehat{c}_5 = 11 - 4.1 = 6.9$

(d) What is the estimated expected number of words memorized by Younger subjects using the Imagery process?

**Answer:** $\widehat{b}_0 + \widehat{b}_2 + \widehat{c}_3 + \widehat{(bc)}_{23} = 11 + 3.8 + 2.5 + 0.4 = 17.7$

(e) The $(bc)_{22}$ coefficient is significant. What is the interpretation of this coefficient, and what do we learn about memory strategies from this significant result?

**Answer:** This is comparing the difference between younger and older subjects using the counting strategy to the difference between younger and older subjects using the adjective strategy. Equivalently, it's comparing the difference between the counting and adjective strategies for younger subjects to the difference betweeen the counting and adjective strategies for older subjects. We learn that the counting strategy is relatively worse than the adjective strategy for older subjects, compared to the same comparison for younger subjects.

(f) Using information supplied here, write a formula for the missing F statistic. Use only numbers in your formula.

```
Analysis of Variance Table
-----------------------------------------------
Model 1: Words ~ Age + Process
Model 2: Words ~ Age * Process
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1     94 912.6
2     90 722.3  4     190.3 ?????? 0.0002793 ***
```

$$F = \frac{(912.6 - 722.3)/4}{722.3/90}$$

36. There is a weather station in Ithaca that measures hourly temperatures. Suppose you would like to build a model for predicting the 3pm temperature (`temp3p`) from the 7am temperature (`temp7a`) and the day of year (`doy`). We expect a sinusoidal pattern in `doy` for a full year of data, but we'll analyze data from day 121 to 269 of the year, which we can approximate with a quadratic. Here is the full quadratic model in `temp7a` and `doy`:

```
Call:
lm(formula = temp3p ~ temp7a + doy + I(doy^2) )

Residuals:
    Min      1Q  Median      3Q     Max
-10.953  -2.557   0.263   2.821  12.587

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.240e+00  3.020e+00  -2.729  0.00646 **
temp7a       4.771e-01  2.952e-02  16.160  < 2e-16 ***
doy          2.322e-01  3.405e-02   6.818 1.57e-11 ***
I(doy^2)    -5.415e-04  8.635e-05  -6.271 5.27e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.889 on 1028 degrees of freedom
Multiple R-squared:  0.396,Adjusted R-squared:  0.3942
F-statistic: 224.6 on 3 and 1028 DF,  p-value: < 2.2e-16
```

(a) Write out the statistical model that was assumed, defining all of your notation.

**Answer:**

$$
\begin{aligned}
Y_i &= b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i2}^2 + \varepsilon_i \\
x_{i1} &= \text{ 7am temperature} \\
x_{i2} &= \text{ day of year} \\
\varepsilon_i &\overset{ind}{\sim} N(0, \sigma^2)
\end{aligned}
$$

(b) The estimated quadratic coefficient on day of year is negative. Why does this make sense?

**Answer:** For a given 7am temperature, the 3pm temperature can be warmer in the middle of the year (the summer) than earlier or later in the year.

(c) How do you interpret the intercept? Should we trust our estimate of the intercept?

**Answer:** Expected 3pm temperature when 7am temperature is zero on day of the year 0. We should not trust this value to be predictive of that situation because we only have data between days 121 and 269, and we don't expect the quadratic trend to continue beyond the range of the data.

(d) What is our estimate of the day of the year with the maximum 3pm temperature?

**Answer:** This is $-0.2322/(-2 * 0.0005415) = 214.4$, about August 2nd.

(e) Here is a model that adds an interaction between 7am temperature and day of year. Write out the statistical model that was assumed, defining all of your

notation.

```
Call:
lm(formula = temp3p ~ temp7a * doy + I(doy^2) )

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.159e+01  3.119e+00  -3.715 0.000214 ***
temp7a       9.064e-01  1.137e-01   7.969 4.24e-15 ***
doy          2.402e-01  3.388e-02   7.089 2.51e-12 ***
I(doy^2)    -4.971e-04  8.651e-05  -5.746 1.20e-08 ***
temp7a:doy  -2.171e-03  5.557e-04  -3.906 9.98e-05 ***

Residual standard error: 3.863 on 1027 degrees of freedom
Multiple R-squared:  0.4048,Adjusted R-squared:  0.4025
F-statistic: 174.6 on 4 and 1027 DF,  p-value: < 2.2e-16
```

**Answer:**

$$
\begin{aligned}
Y_i &= b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i2}^2 + b_4 x_{i1} x_{i2} + \varepsilon_i \\
x_{i1} &= \text{ 7am temperature} \\
x_{i2} &= \text{ day of year} \\
\varepsilon_i &\overset{ind}{\sim} N(0, \sigma^2)
\end{aligned}
$$

(f) What are the slopes with respect to 7am temperature on days 121 and 269? Are these numbers practically different from the slope in the first model?

**Answer:** The slopes are $0.9064 - 0.002171 * 121 = 0.6437$ and $0.9064 - 0.002171 * 269 = 0.3224$. Yes, they are quite different. The first is larger than the slope in the first model, and the second is smaller.

(g) What are the intercepts with respect to 7am temperature on days 121 and 169?

**Answer:** To find these, we set 7am temp ($x_{i1}$) to zero, and plug in the two values of day of year:

$$
\begin{aligned}
\widehat{b_0} + \widehat{b_2} 121 + \widehat{b_3} 121^2 &= 10.20 \\
\widehat{b_0} + \widehat{b_2} 269 + \widehat{b_3} 269^2 &= 17.05
\end{aligned}
$$

(h) Without doing any calculations, can you find the p-value for the F test that compares these two models?

**Answer:** The only difference between the two models is the interaction term. When two models differ by one term, the F test for comparing the two models is equivalent to the t-test for that term, so the p-value is 9.98e-05.