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

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

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.

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.

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

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

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 $y_1, \ldots, y_n$?

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

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
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 ε?

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

13. Recall the simple linear model

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

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 2×2 matrix.

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 kth normal equation by differentiating the residual sum of squares.

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

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 $\widehat{b}_j = \boldsymbol{x}_j^T \boldsymbol{y}$.

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

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.

22. Let $\boldsymbol{b} = (b_0, b_1, b_2)$ and $\widehat{\boldsymbol{B}}$ be the least squares estimator for $\boldsymbol{b}$. Further, let $M = \sigma^2 (X^T X)^{-1}$, 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 $\widehat{\boldsymbol{B}} = (X^T X)^{-1} X^T \boldsymbol{Y}$?

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

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?

    (b) The model fit to the height data was

    $$Y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2),$$

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

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

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

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

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

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?

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 $\widehat{\sigma}^2$.

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