STAT 2131:
Applied Statistical Methods I
HW #2
Due Thursday, September 3

1. 1.22 from KNNL.

2. 2.16 from KNNL.

3. We have looked at using ordinary least squares/maximum likelihood estimation for the simple linear regression model. This problem considers an alternative estimation procedure. For simplicity, we will assume that the variance σ^2 is known. You observe outcomes y_i from $i = 1, \ldots, n$ subjects and assume that the data follow the linear model

$$y_i = \beta_0 + x_i\beta_1 + \epsilon_i$$

where x_i are known deterministic covariates, β_0 and β_1 are unknown deterministic parameters, and ϵ_i are independent and identically distributed mean zero Gaussian random variables with *known* variance 1. Given some $\lambda \geq 0$, you decide to estimate β_1 with $\tilde{b}_1$ that minimizes the penalized sum-of-squares

$$\mathrm{PSS}_\lambda(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 + \lambda\beta_1^2$$

so that

$$\left(\tilde{b}_0, \tilde{b}_1\right) = argmin_{\beta_0, \beta_1 \in \mathbb{R}} \mathrm{PSS}_\lambda(\beta_0, \beta_1).$$

   (a) What are $\tilde{b}_0, \tilde{b}_1$ when $\lambda = 0$? When $\lambda = \infty$?

   (b) A collaborator claims that $\tilde{b}_1$ always has smaller variance than the best linear unbiased estimator of β_1. Prove or disprove this claim.

   (c) Show that $\tilde{b}_1$ is a biased estimate for β_1 when $\lambda > 0$.

4. Suppose $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ for $i = 1, \ldots, n$, where $x_1, \ldots, x_n$ are known constants and

$$E(\epsilon_i) = 0, \quad i = 1, \ldots, n$$
$$\mathrm{Cov}(\epsilon_i, \epsilon_j) = \begin{cases} \sigma^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, \quad i, j = 1, \ldots, n.$$

Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the ordinary least squares estimates for β_0 and β_1 and define

$$\hat{\epsilon}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i, \quad i = 1, \ldots, n.$$

(a) Prove that the vector of $\hat{\epsilon}$'s is orthogonal to the design. That is, prove

$$\sum_{i=1}^{n} \hat{\epsilon}_i = \sum_{i=1}^{n} \hat{\epsilon}_i x_i = 0.$$

(b) Let $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, $\bar{\hat{\epsilon}} = n^{-1} \sum_{i=1}^{n} \hat{\epsilon}_i$ and $\bar{\hat{y}} = n^{-1} \sum_{i=1}^{n} y_i$. Using part (a), show that

$$\sum_{i=1}^{n} \left(\hat{\epsilon}_i - \bar{\hat{\epsilon}}\right) \left(\hat{y}_i - \bar{\hat{y}}\right) = 0.$$

(c) Define $\hat{\sigma}^2 = (n-2)^{-1} \sum_{i=1}^{n} \hat{\epsilon}_i^2$. Show that $E\left(\hat{\sigma}^2\right) = \sigma^2$.

5. You just started working for a company that is interested in looking at how radio advertising affected their sales. The data consist of the amount of sales Y_i in \$100 and the amount of radio advertising time X_i in hours for the ith month for $i = 1, \ldots, 24$.

The following R code was run:

```
#Radio is the number of radio advertisements for each month, in hours
#sales is the profit for each month, in dollars
fit <- lm(sales ~ Radio, data = Sales)
```

The output was collated into the following tables:

Number of Observations 24

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr(>F)
Model	1	1952.65488	1952.65488	712.51	<.0001
Error	22	60.29179	2.74054		
Corrected Total	23	2012.94667			

Root MSE	1.65546	R-Square	0.9700
Dependent Mean	114.88255	Adj R-Sq	0.9687
Coeff Var	1.44100		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr(>|t|)
Intercept	1	101.57570	0.60225	168.66	<.0001
Radio	1	1.15806	0.04338	26.69	<.0001

(a) Suppose conditional on the variable Radio, sales from each month are independent. Write down the model assumed when running the above R code. Clearly state any assumptions and define all notations.

(b) Assume that the model in part (a) is appropriate. Provide a point estimate and a 95% confidence interval for the change in expected monthly amount of sales if monthly radio advertisement increases by 10 hours.

(c) Assume that the model in part (a) is appropriate, and that the residuals in part (a) are normally distributed. The manager plans to have no radio advertisement next month.

   (i) Provide him with a point estimate and a 95% prediction interval for the amount of sales next month.

   (ii) Why is the normality assumption for the prediction interval important, and why is it not critical we make that assumption in part (b)? Please give an intuitive explanation rather than fancy mathematical derivations.

(d) Perform a test to study whether radio advertisement tends to increase the amount of sales. State clearly the hypothesis, the testing statistic, the p-value and your conclusion.