Homework 3 Solution

Stephanie DeGraaf

October 4, 2018

1. (a) The X matrix for the simple linear regression model is

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

and $X^T X$ is given by

$$X^T X = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}.$$

$X^T X$ is invertible as long as $det(X^T X) \neq 0$. The determinant of $X^T X$ is

$$det(X^T X) = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2.$$

Hence $det(X^T X) = 0$ only when all x_i are constant. Given that $x_1, ..., x_n$ are not all constant, $X^T X$ is invertible. This implies that every $\lambda^T \beta$ is estimable; hence, β_0 and β_1 are estimable.

(b) The regression line is given by $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, where $\hat{\beta}_0 = \bar{y} - \hat{\beta}\bar{x}$. At the value $\bar{x}$, the regression line is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} = \bar{y}.$$

Thus the point $(\bar{x}, \bar{y})$ falls on the regression line.

(c) The covariance of β is given by $Cov(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$.

$$X^T X = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{bmatrix} = \begin{bmatrix} n & 0 \\ 0 & \sum_{i=1}^n x_i^2 \end{bmatrix}.$$

Then the covariance is

$$Cov(\hat{\beta}) = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & \sigma^2/\sum_{i=1}^n x_i^2 \end{bmatrix}.$$

Thus $Cov(\hat{\beta}_0, \hat{\beta}_1) = 0$, which implies that $\hat{\beta}_0$ and $\hat{\beta}_1$ are uncorrelated.

(d) Two jointly normal distributed random variables are independent as long as they are uncorrelated. Since $e_1, ..., e_n$ are jointly normal, $Y \sim N(X\beta, \sigma^2 I_n)$ and so $\hat{\beta} \sim N(\beta, \sigma^2(X^TX)^{-1})$. Hence, $\hat{\beta}$ has a multivariate normal distribution, so $\hat{\beta}_0$ and $\hat{\beta}_1$ are independent.

2. (a) The X matrix for this model is

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}.$$

Let $\lambda^T = (1, \bar{x}_1, \bar{x}_2, ..., \bar{x}_p)$, so $\beta_0 + \beta_1\bar{x}_1 + \dots \beta_p\bar{x}_p = \lambda^T\beta$. Then λ is a linear combination of the rows of X, by taking the sum of all the rows divided by n. Specifically, for $a = \frac{1}{n}(1, 1, ..., 1)$, $\lambda = aX$. Therefore $\beta_0 + \beta_1\bar{x}_1 + \dots \beta_p\bar{x}_p$ is estimable.

(b) Writing out the normal equations for this model gives us

$$\begin{bmatrix} n & \sum_{i=1}^n x_{i1} & \dots & \sum_{i=1}^n x_{ip} \\ - & - & \dots & - \\ \vdots & \vdots & & \vdots \\ - & - & \dots & - \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1}y_i \\ \vdots \\ \sum_{i=1}^n x_{ip}y_i \end{bmatrix}.$$

Thus the least squares solution $\hat{\beta}$ must satisfy the equation

$$n\hat{\beta}_0 + n\bar{x}_1\hat{\beta}_1 + \dots + n\bar{x}_p\hat{\beta}_p = \sum_{i=1}^n y_i,$$

which implies that $\hat{\beta}$ satisfies

$$\hat{\beta}_0 + \bar{x}_1\hat{\beta}_1 + \dots + \bar{x}_p\hat{\beta}_p = \bar{y}.$$

Hence $\bar{y}$ is the least squares estimate.

(c) Assuming the e_i are independent,

$$Var(\bar{y}) = \frac{1}{n^2}Var\left(\sum_{i=1}^n y_i\right) = \frac{1}{n^2}\sum_{i=1}^n Var(y_i) = \frac{1}{n^2}n\sigma^2 = \sigma^2/n.$$

For $\hat{\sigma}^2 = RSS/(n-p-1)$, we know $\mathbb{E}(\hat{\sigma}^2) = \sigma^2$; thus, a good estimate for $Var(\bar{y}) = \sigma^2/n$ would be

$$\frac{\hat{\sigma}^2}{n} = \frac{RSS}{n(n-p-1)}.$$

3. (a) The variable $I(Weight + 3 * Height)$ is a linear combination of the $Weight$ and $Height$ variables that are already included in the model, so the coefficient for this variable is not estimable.

   (b) This is actually the estimate for $\beta_2 + \beta_4$.

   (c) This is the same model, according to R, as the model including the $Weight + 3 * Height$ variable. Thus, the RSS can be computed from the previous model output, where

$$RSS = (RSE)^2(n - p - 1) = (5.142)^2(247) = 6530.721.$$

   (d) The RSS for the sub-model m can be computed by

$$RSS(m) = (5.696)^2(249) = 8078.66.$$

   The F-statistic for comparing model M to model m is

$$F = \frac{(RSS(m) - RSS(M))/p}{RSS(M)/(n - (k + p + 1))} = \frac{(8078.66 - 6530.721)/2}{6530.721/247} = 29.2725.$$

   The p-value for this F-statistic is given by $P(X \geq 29.2725)$ where $X \sim F_{2,247}$.

4. (a) Each t statistic for coefficient j has the form

$$t_j = \frac{\hat{\beta}_j}{RSE * \sqrt{V[j,j]}}$$

   where $V = (X^T X)^{-1}$. Thus, the Residual Standard Error (RSE) can be computed from any of the coefficients. Using the values for β_0, we compute

$$-0.104 = \frac{-1.07425}{RSE * \sqrt{3.740212022}}.$$

   Solving for RSE, we find

$$\mathbf{RSE} = \frac{-1.07425}{-0.104\sqrt{3.740212022}} \approx 5.34.$$

   The standard error for each coefficient is computed as $RSE * \sqrt{V[j,j]}$; thus,

$$\mathbf{se_{weight}} = RSE * \sqrt{V[3,3]} = 5.34 * \sqrt{2.632523e - 05} \approx 0.0274.$$

   The t-statistic for $Weight$ is then easily computed as

$$\mathbf{t_{weight}} = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{0.12373}{0.02740375} = 4.515075.$$

The coefficient for $Thigh$ is

$$\hat{\boldsymbol{\beta}}_{\mathbf{4}} = t_{thigh} * se(thigh) = 2.444 * 0.14952 = 0.3654269.$$

Now that we know $se(thigh)$ and RSE, the final entry of $(X^TX)^{-1}$ can be computed by

$$\mathbf{V[5,5]} = \left(\frac{se(thigh)}{RSE}\right)^2 = \left(\frac{0.14952}{5.341016}\right)^2 = (0.02799467)^2 = 0.0007837.$$

To compute the F-statistic, we first compute the RSS as

$$RSS = (RSE)^2(247) = 5.341016^2 * (247) = 7046.03.$$

Next, we compute the TSS from the value of R^2, which we know is given by $R^2 = 1 - \frac{RSS}{TSS}$, or

$$0.5349 = 1 - \frac{7046.035}{TSS} \Rightarrow TSS = 15149.51.$$

Finally, the F-statistic can be computed as

$$\mathbf{F} = \frac{(TSS - RSS)/p}{RSS/(n-p-1)} = \frac{8103.47/4}{5.341016}247 \approx 71.017.$$

(b) The fit value is the same for both intervals: it is the fitted value from the regression for the given X values:

$$\begin{aligned}\mathbf{fit_{conf}} = \mathbf{fit_{pred}} &= \hat{y}(x_0)\\ &= -1.07425 + 0.18901(30) + 0.12373(180) - 0.46074(72) + 0.3654269(60)\\ &= 15.61978.\end{aligned}$$

Since the confidence interval is symmetric, the upper bound can easily be found by

$$\mathbf{upr_{conf}} = 15.61978 + (15.61978 - 14.56726) = 16.6723.$$

This also implies that $t * se_{conf} = 15.61978 - 14.56726 = 1.05252$. From part (a), $\hat{\sigma}$ is 5.341016. The t-value for a 95% confidence interval is the 0.975 quantile of a t-distribution with 247 degrees of freedom, which is 1.969615. We can then solve for $x_0^T(X^TX)^{-1}x_0$ by

$$1.05252 = 1.969615 * 5.341016 * \sqrt{x_0^T(X^TX)^{-1}x_0},$$

which indicates that $x_0^T(X^TX)^{-1}x_0 = 0.01001077$. Thus, the prediction interval is calculated by

$$15.61978 \pm 1.969615(5.341016)\sqrt{1 + 0.01001077},$$

so the upper and lower bounds are

$$\begin{aligned}\mathbf{lwr_{pred}} &= 5.04772\\ \mathbf{upr_{pred}} &= 26.19184.\end{aligned}$$

(c) The Res.Df is simply the difference $n - p - 1$. Thus,

$$\mathbf{Res.Df_1} = 248$$
$$\mathbf{Res.Df_2} = 247.$$

The RSS for model 2 was computed in part (a) as

$$\mathbf{RSS_2} = 7046.035.$$

From the F-statistic, we can compute the RSS for model 1, since

$$F = 1.6203 = \frac{RSS_1 - RSS_2}{RSS_2/247}.$$

Hence,

$$\mathbf{RSS_1} = 1.6203\left(\frac{7046.035}{247}\right) + 7046.035 = 7092.256.$$

The Df is simply the difference in degrees of freedom between the two models; hence

$$\mathbf{Df} = 248 - 247 = 1.$$

Finally, the Sum of Sq value is the difference between RSS for each model; hence

$$\textbf{Sum of Sq} = RSS_1 - RSS_2 = 46.22142.$$

5. (a) The Res.Df is the difference $n - p - 1$. Thus,

$$\mathbf{Res.Df_1} = 138$$
$$\mathbf{Res.Df_2} = 137.$$

The F-statistic for testing one parameter is identical to the square of the t-statistic for that parameter: $F = T^2$. Hence,

$$\mathbf{F} = t^2_{skipped} = (03.197)^2 = 10.22081.$$

The p-value for this F-statistic comes from an F-distribution with 1 and 137 degrees of freedom; hence

$$\mathbf{Pr}(> \mathbf{F}) = 0.00172.$$

The RSS for model 2 can be computed from the RSE and Df as

$$\mathbf{RSS_2} = RSE^2(137) = (0.3295)^2(137) = 14.87412.$$

The RSS for model 1 can be computed from the F-statistic and RSS_1 by

$$10.22081 = \frac{RSS_1 - 14.87412}{14.87412/137},$$

which implies

$$\mathbf{RSS_1} = 15.9838.$$

Finally, the Sum of Squares is the difference between the RSSs:

$$\textbf{Sum of Sq} = 15.9838 - 14.87412 = 1.109676.$$

(b) The fit value is the same for both intervals: it is the fitted value from the regression for the given X values:

$$\mathbf{fit_{conf}} = \mathbf{fit_{pred}} = \hat{y} = 1.38955 + 0.41183(3.4) + 0.01472(25) - 0.08311(2) = 2.991518.$$

Since the confidence interval is symmetric, the lower bound can easily be found by

$$\mathbf{lwr_{conf}} = 2.991518 - (3.064408 - 2.991518) = 2.918628.$$

The t-value for a 95% confidence interval is given as 1.977431. From part (a), $\hat{\sigma} = RSE = 0.3295$. From the formula for the confidence interval, we can solve for the quantity $\sqrt{x_0^T (X^T X)^{-1} x_0}$, since

$$t * se_{conf} = 3.064408 - 2.991518 = 0.07289 = 1.977431 * 0.3295 * \sqrt{x_0^T (X^T X)^{-1} x_0}.$$

Hence,

$$x_0^T (X^T X)^{-1} x_0 = 0.01251476.$$

Now from the formula for a prediction interval, we compute

$$t * se_{pred} = 1.977431 * 0.3295 * \sqrt{1 + 0.01251476} = 0.6556279.$$

Finally, the upper and lower bounds of the prediction interval are

$$\mathbf{lwr_{pred}} = 2.991518 - 0.6556279 = 2.33589;$$
$$\mathbf{upr_{pred}} = 2.991518 + 0.6556279 = 3.647146.$$

6. (a) F: there are infinitely many solutions.

(b) T: this is an underdetermined system of linear equations.

(c) T: high leverage implies that this point has a combination of extreme X values, which will influence the regression line to be close to this point.

(d) T: Under the assumption of normality, $e/\sigma \sim N(0, 1)$, and we can write

$$RSS/\sigma^2 = \frac{e^T (I - H) e}{\sigma^2} = (e/\sigma)^T (I - H)(e/\sigma).$$

Since $I - H$ is symmetric and idempotent, $RSS/\sigma^2 \sim \chi^2_{n-p-1}$.

(e) F: $\hat{Y} = HY \Rightarrow \sum (\hat{y}_i/\sigma)^2 = (Y/\sigma)^T H (Y/\sigma)$. However, $Y \sim N(X\beta, \sigma^2 I)$, which is not a standard $N(0, 1)$ variable. Thus, the sum of squares of the fitted values are not chi-square.

(f) F: if $\lambda^T \beta$ is estimable, the least squares estimate $\lambda^T \hat{\beta}$ is the unique estimator. By Gauss-Markov, $\hat{\beta}$ is BLUE and thus has the smallest variance of all linear, unbiased estimators. However, if you have a biased estimator, you could get a smaller variance.

(g) F:

$$R^2 = \frac{(k-1)F}{n-k+(k-1)F} \sim Beta\left(\frac{k-1}{2}, \frac{n-k}{2}\right)$$

(h) T: by definition, $\hat{\sigma}^2 = RSS/(n-p-1) = e^T(I-H)e/(n-p-1)$; hence,

$$\begin{aligned} Var(\hat{\sigma}^2) &= Var\left(\frac{e^T(I-H)e}{(n-p-1)}\right) \\ &= \frac{1}{(n-p-1)^2} Var\left(e^T(I-H)e\frac{\sigma^2}{\sigma^2}\right) \\ &= \frac{\sigma^4}{(n-p-1)^2} Var\left(\frac{e^T(I-H)e}{\sigma^2}\right). \end{aligned}$$

We know $e^T(I-H)e/\sigma^2 \sim \chi^2_{n-p-1}$. The variance of this chi-square distribution is $2*(n-p-1)$; hence,

$$Var(\hat{\sigma}^2) = \frac{\sigma^4}{(n-p-1)^2} 2(n-p-1) = \frac{2\sigma^4}{n-p-1}.$$

(i) T: The p-value will depend on the particular random permutation of your data.

(j) F: Consider a simple linear regression where $y_i = \beta_0 + \beta_1 x_i$ for $i = 1, ..., n$. Let $x_1 \neq x_2$ with $x_1 + x_2 = 2\bar{x}$, and for all other x_j, let $x_j = \bar{x}$. The (1,2) entry of the hat matrix will be

$$h_{1,2} = \frac{1}{n} + \frac{(x_1-\bar{x})(x_2-\bar{x})}{\sum(x_k-\bar{x})^2} = \frac{1}{n} - \frac{1}{2} < 0.$$

Thus $h_{1,2}$ is not necessarily between 0 and 1.

(k) F: the slope can be anything

(l) T: in simple linear regression, we can write the slope as $\hat{\beta}_1 = r(s_y/s_x)$. Assuming unit standard deviations, $\hat{\beta}_1 = r \leq 1$.

(m) F: removing variables will always increase the RSS, but removing variables will also increase $n-p-1$. Thus, $RSE = \sqrt{RSS/(n-p-1)}$ may increase or decrease.

(n) T: if (X^TX) is invertible, then $\hat{\beta} = (X^TX)^{-1}X^TY$, and this is a unique solution. Further, any linear function of β given by $\lambda^T\beta$ is given by $\lambda^T(X^TX)X^TY$.

(o) F: each residual $\hat{e}_i$ has variance $\sigma^2(1-h_{ii})$.

(p) F: normality does not affect this relationship

(q) F: normality does not affect this relationship

(r) T: if normality is violated, even if the residuals and fitted values are uncorrelated, we are no longer guaranteed independence.

(s) F: a small p-value only indicates that one of the coefficients in the model is nonzero; it does not indicate that the model is valid or useful.

(t) F: a small p-value only indicates that if $\beta_2 = 0$, it is unlikely that $\hat{\beta}/se(\hat{\beta})$ would be as extreme as it was.

(u) F: $\hat{\beta}_2$ could be small but with even smaller standard error: this would generate a large t-statistic and small p-value.