STAT 151A Homework 2 Solutions, Questions 1-5

Billy Fang

These are rough sketches for the solutions for some of Homework 2. Some computational steps are omitted for brevity.

1

Let $y_i = A + Bx_i + \epsilon_i$ for $i = 1, \dots, n$ be the simple regression model. In simple regression, $\mathbf{b} = \begin{bmatrix} A \\ B \end{bmatrix}$ and

$$\mathbf{X} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}.$$

Then $\mathbf{X}^\top \mathbf{X} = n \begin{bmatrix} 1 & \overline{X} \\ \overline{X} & \frac{1}{n}\sum_i X_i^2 \end{bmatrix}$ so

$$V(\mathbf{b}) = \sigma_\epsilon^2 (\mathbf{X}^\top \mathbf{X})^{-1} = \sigma_\epsilon^2 \frac{1}{n} \frac{1}{\frac{1}{n}\sum_i X_i^2 - \overline{X}^2} \begin{bmatrix} \frac{1}{n}\sum_i X_i^2 & -\overline{X} \\ -\overline{X} & 1 \end{bmatrix} = \sigma_\epsilon^2 \frac{1}{\sum_i (X_i - \overline{X})^2} \begin{bmatrix} \frac{1}{n}\sum_i X_i^2 & -\overline{X} \\ -\overline{X} & 1 \end{bmatrix},$$

where in the last step we used the fact that

$$\frac{1}{n}\sum_i (X_i - \overline{X})^2 = \frac{1}{n}\sum_i (X_i^2 - 2\overline{X}X_i + \overline{X}^2) = \frac{1}{n}\sum_i X_i^2 - 2\overline{X}\frac{1}{n}\sum_i X_i + \overline{X}^2 = \frac{1}{n}\sum_i X_i^2 - \overline{X}^2.$$

Finally, note that $V(A)$ and $V(B)$ are the diagonal elements of $V(\mathbf{b})$, which we computed above.

2

Note that $X^\top X$ is symmetric, and thus $(X^\top X)^{-1}$ is symmetric.

$$H^\top = (X(X^\top X)^{-1}X^\top)^\top = (X^\top)^\top((X^\top X)^{-1})^\top X^\top = X(X^\top X)^{-1}X^\top = H.$$

$$H^2 = X(X^\top X)^{-1}X^\top X(X^\top X)^{-1}X^\top = X(X^\top X)^{-1}X^\top = H.$$

3

(a) The normal equation implies

$$X^\top \widehat{y} = X^\top X\widehat{\beta} = X^\top y.$$

If we assume the first column of X is the all ones vector, then looking at the first entry of both sides of the above equation yields $\sum_{i=1}^n \widehat{y}_i = \sum_{i=1}^n y_i$. Dividing both sides by n yields the desired result $\overline{y} = \overline{\widehat{y}}$.

(b) Again, by the normal equation,

$$\widehat{y}^\top(y - \widehat{y}) = (X\widehat{\beta})^\top(y - X\widehat{\beta}) = \widehat{\beta}^\top(X^\top y - X^\top X\widehat{\beta}) = 0.$$

(c)

$$
\begin{aligned}
\widehat{\mathrm{Cov}}(y, \widehat{y}) &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})(\widehat{y}_i - \overline{\widehat{y}}) \\
&= \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{\widehat{y}})(\widehat{y}_i - \overline{\widehat{y}}) && \text{by part (a)} \\
&= \frac{1}{n}\sum_{i=1}^{n} y_i\widehat{y}_i - \overline{\widehat{y}}\frac{1}{n}\sum_{i=1}^{n}\widehat{y}_i - \overline{\widehat{y}}\frac{1}{n}\sum_{i=1}^{n} y_i + \overline{\widehat{y}}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} \widehat{y}_i^2 - \overline{\widehat{y}}\frac{1}{n}\sum_{i=1}^{n}\widehat{y}_i - \overline{\widehat{y}}\frac{1}{n}\sum_{i=1}^{n}\widehat{y}_i + \overline{\widehat{y}}^2 && \text{by parts (a) and (b)} \\
&= \sum_{i=1}^{n}(\widehat{y}_i - \overline{\widehat{y}})^2 =: \widehat{\mathrm{Var}}(\widehat{y}).
\end{aligned}
$$

4

We are given one of the degrees of freedom in the F-statistic is $n - p - 1 = 246$. There are $p = 5$ variables, so the missing degrees of freedom in the last line is $p = 5$.

We can compute the F-statistic using only R^2; see page 115 of the textbook.

$$F = \frac{n-p-1}{p}\frac{R^2}{1-R^2} = \frac{246}{5}\frac{0.7228}{1-0.7228} \approx 128.29.$$

Finally, the p-value can be computed in R using `1 - pf(128.29, 5, 246)` which returns zero, so the p-value is extremely small.

5

(a) Using linearity of expectation and the fact that the least squares estimate is unbiased (recall: $\mathbb{E}[b] = \mathbb{E}[(X^\top X)^{-1}X^\top y] = (X^\top X)^{-1}X^\top \mathbb{E}[X\beta + \epsilon] = \beta$), we have

$$
\begin{aligned}
\mathbb{E}[\delta] &= \mathbb{E}[\widehat{Y}_0] - \mathbb{E}[Y_0] \\
&= \mathbb{E}[x_0^\top b] - \mathbb{E}[x_0^\top \beta + \epsilon] \\
&= x_0^\top \mathbb{E}[b] - x_0^\top \beta \\
&= x_0^\top \beta - x_0^\top \beta = 0.
\end{aligned}
$$

Next, recall that the covariance matrix of the least squares estimate is $V(b) = \sigma_\epsilon^2(X^\top X)^{-1}$, so

$$
\begin{aligned}
V(\delta) &= \mathbb{E}[\delta^2] \\
&= \mathbb{E}[(\widehat{Y}_0 - \mathbb{E}[Y_0])^2] \\
&= \mathbb{E}[(x_0^\top b - x_0^\top \beta)^2] \\
&= \mathbb{E}[x_0^\top (b-\beta)(b-\beta)^\top x_0] \\
&= x_0^\top \mathbb{E}[(b-\beta)(b-\beta)^\top] x_0 \\
&= x_0^\top V(b) x_0 \\
&= \sigma_\epsilon^2 x_0^\top (X^\top X)^{-1} x_0.
\end{aligned}
$$

(b) Using $\mathbb{E}[b] = \beta$ again, we have

$$\mathbb{E}[D] = \mathbb{E}[x_0^\top(b-\beta) - \epsilon_0] = x_0^\top \mathbb{E}[b-\beta] = 0.$$

Below, we will use the fact that ϵ_0 is independent of b (recall b depends only on $\epsilon_1, \ldots, \epsilon_n$) to have $\mathbb{E}[\epsilon_0 x_0^\top (b-\beta)] = \mathbb{E}[\epsilon_0]\mathbb{E}[x_0^\top (b-\beta)] = 0$.

$$\begin{aligned}
V(D) &= \mathbb{E}[(x_0^\top (b-\beta) - \epsilon_0)^2] \\
&= \mathbb{E}[x_0^\top (b-\beta)(b-\beta)^\top x_0] - 2\mathbb{E}[\epsilon_0 x_0^\top (b-\beta)] + \mathbb{E}[\epsilon_0^2] \\
&= x_0^\top \mathbb{E}[(b-\beta)(b-\beta)^\top] x_0 + \sigma_\epsilon^2 \\
&= \sigma_\epsilon^2 x_0^\top (X^\top X)^{-1} x_0 + \sigma_\epsilon^2. && \text{recall } \operatorname{Cov}(b) = \sigma^2 (X^\top X)^{-1}
\end{aligned}$$

The variance of D is larger because it accounts for the extra variance in the error ϵ_0.

A quicker alternate solution: using the fact that Y_0 (which depends only on ϵ_0) and $\widehat{Y}_0$ (which depends only on $\epsilon_1, \ldots, \epsilon_n$) are independent, we have

$$V(D) = \operatorname{Var}(\widehat{Y}_0) + \operatorname{Var}(Y_0) = \sigma_\epsilon^2 x_0^\top (X^\top X)^{-1} x_0 + \sigma_\epsilon^2.$$

(c) Note that $k = 3$ (education, income, percentage of women), and $n = 102$ (see top of page 98 of the textbook).

The coefficients of the least squares regression have been computed in Section 5.2.2:

$$b = \begin{bmatrix} -6.7943 \\ 4.1866 \\ 0.0013136 \\ -0.0089052 \end{bmatrix},$$

and thus our estimate for an occupation with

$$x_0 = \begin{bmatrix} 1 \\ 13 \\ 12000 \\ 50 \end{bmatrix}$$

is

$$\widehat{Y}_0 = x_0^\top b = 62.94944.$$

The matrix $X^\top X$ can be written as

$$X^\top X = \begin{bmatrix} n & \sum_i x_{i1} & \sum_i x_{i2} & \sum_i x_{i3} \\ \sum_i x_{i1} & \sum_i x_{i1}^2 & \sum_i x_{i1}x_{i2} & \sum_i x_{i1}x_{i3} \\ \sum_i x_{i2} & \sum_i x_{i2}x_{i1} & \sum_i x_{i2}^2 & \sum_i x_{i2}x_{i3} \\ \sum_i x_{i3} & \sum_i x_{i3}x_{i1} & \sum_i x_{i3}x_{i2} & \sum_i x_{i3}^3 \end{bmatrix} = \begin{bmatrix} 102 & 1095 & 693386 & 2956 \\ 1095 & 12513 & 8121410 & 32281 \\ 693386 & 8121410 & 6534383460 & 14093097 \\ 2956 & 32281 & 14093097 & 187312 \end{bmatrix},$$

whose values are read from Table 5.1 in the textbook. Its inverse can be computed on a computer.

We can also get

$$X^\top y = \begin{bmatrix} \sum_i y_i \\ \sum_i y_i x_{i1} \\ \sum_i y_i x_{i2} \\ \sum_i y_i x_{i3} \end{bmatrix} = \begin{bmatrix} 4777 \\ 55326 \\ 37748108 \\ 131909 \end{bmatrix},$$

whose values again are from Table 5.1.

An unbiased estimate of σ_ϵ^2 is

$$S_E^2 = \frac{\text{RSS}}{n-k-1}.$$

To compute RSS,

$$\text{RSS} = \|y - Xb\|^2 = \|y\|^2 - 2b^\top X^\top y + b^\top X^\top X b$$

Note $\|y\|^2 = \sum_i y_i^2 = 253618$, which can be found in Table 5.1. The other quantities we have already computed. Thus,

$$\begin{aligned} \text{RSS} &= 6051.621, \\ S_E^2 &= \frac{6033.57}{102 - 3 - 1} \approx 61.75 \\ S_E &\approx 7.846. \end{aligned}$$

Alternatively, you could use the fact that S_E was computed already on page 98.

Finally, we will need $t_{0.95}$, the 95% quantile of the t distribution with $n - k - 1$ degrees of freedom, which can be found in R using `qt(0.95, 102 - 3 - 1)` or looking at a t table.

$$t_{0.95} = 1.29025.$$

We now have everything we need to compute the confidence intervals.

(i) Using the expression for $V(\delta)$ computed in (a), we know

$$\frac{\widehat{Y}_0 - \mathbb{E}Y_0}{S_E\sqrt{x_0^\top (X^\top X)^{-1} x_0}}$$

follows the t-distribution with $n - k - 1$ degrees of freedom, so a 90% confidence interval for $\mathbb{E}[Y_0]$ is

$$\widehat{Y}_0 \pm t_{0.95} S_E \sqrt{x_0^\top (X^\top X)^{-1} x_0} \approx 62.95 \pm 2.76.$$

(ii) Using the expression for $V(D)$ computed in (a), we know

$$\frac{\widehat{Y}_0 - Y_0}{S_E\sqrt{1 + x_0^\top (X^\top X)^{-1} x_0}}$$

follows the t-distribution with $n - k - 1$ degrees of freedom, so a 90% confidence interval for Y_0 is

$$\widehat{Y}_0 \pm t_{0.95} S_E \sqrt{1 + x_0^\top (X^\top X)^{-1} x_0} \approx 62.95 \pm 13.34.$$

(d) Now, we have

$$x_0 = \begin{bmatrix} 1 \\ 0 \\ 50000 \\ 100 \end{bmatrix}.$$

The estimated variance is

$$S_E^2[1 + x_0^\top (X^\top X)^{-1} x_0] \approx 334.31 \approx 18.21^2,$$

which is larger than part (ii) of (c) above, which had variance $13.34^2 \approx 177.96$. The increase is due to the increase in $x_0^\top (X^\top X)^{-1} x_0$, which captures the fact that the x_0 in part (d) is farther from the original data than the the x_0 in part (c).