STAT 151A Additional Problems: Solutions

Billy Fang

These are rough sketches for the solutions. Some computational steps are omitted for brevity.

1

Intercept p-value: `2 * pt(-2.127, 99)` yields 0.0359. This t-table would tell you the p-value is between 0.02 and 0.05.

Education estimate:

$$t = \frac{\widehat{\beta}_j}{\mathrm{s.e.}(\widehat{\beta}_j)} \implies \widehat{\beta}_j = t \cdot \mathrm{s.e.}(\widehat{\beta}_j) = 11.858 \cdot 0.3489120 \approx 4.1374.$$

Residual standard error degrees of freedom is 99 (same as second degree of freedom in F-statistic)

2

`qt(0.975, 99)` (or this t-table) shows that ≈ 1.984 is the 97.5% quantile of the t-distribution with 99 degrees of freedom. For each j, the confidence interval is

$$\widehat{\beta}_j \pm 1.984 \cdot \mathrm{s.e.}(\widehat{\beta}_j),$$

where $\widehat{\beta}_j$ and $\mathrm{s.e.}(\widehat{\beta}_j)$ are the entries in the first two columns of the table.

3

(a) True.

We state an auxiliary result that will simplify our work. Recall the definition of estimability: $\Lambda^\top \beta$ (for some matrix Λ) is called estimable if $\Lambda^\top \beta = P^\top X \beta$ for some matrix P. After some reformulation, we can write this as

$\Lambda^\top \beta$ is estimable if and only if every column of Λ lies in the column space of $X^\top$

[Check that this is true.]

In simple regression, we have

$$X^\top = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix}.$$

Note that by assumption $n > p = 1$. There are two cases for $C(X^\top)$.

1. If $x_1 = \cdots = x_n$, then $C(X^\top) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ x_1 \end{bmatrix} \right\}$ is a one-dimensional subspace.
2. Otherwise, $C(X^\top) = \mathbb{R}^2$ (i.e. the column space contains all two-dimensional vectors).

Note

$$\beta_1 = \begin{bmatrix} 0 & 1 \end{bmatrix} \beta, \qquad \text{and} \qquad \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \beta.$$

If $\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix}$ is not estimable, then at least one of $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$ does not lie in $C(X^\top)$. This implies Case 2 above cannot happen (otherwise $C(X^\top)$ would contain any two-dimensional vector, which would contradict the previous sentence). Therefore we must have $x_1 = \cdots = x_n$ and $C(X^\top) = \text{span}\left\{\begin{bmatrix}1\\x_1\end{bmatrix}\right\}$. But then $C(X^\top)$ cannot contain $\begin{bmatrix}0\\1\end{bmatrix}$, which implies β_1 is not estimable.

[Note: in the second lab section, I proved the contrapositive of the statement instead. Above, I proved the statement directly, so the argument is "backward" in some sense.]

(b) False. Note

$$\mathbb{E}\widehat{\beta} = \mathbb{E}[(X^\top X)^{-1}X^\top(X\beta+\epsilon)] = \beta + (X^\top X)^{-1}X^\top\mathbb{E}[\epsilon].$$

Thus $\widehat{\beta}$ is unbiased as long as $\mathbb{E}[\epsilon]=0$, even if the components of ϵ are correlated.

4

- Approach 1: see lecture notes on one-way ANOVA
- Approach 2: the function we want to minimize is

$$S(\mu_1,\ldots,\mu_J) = \sum_{j=1}^J \sum_{i\in\text{group } j}(y_i-\mu_j)^2.$$

  Setting the partial derivatives to zero yields

$$\sum_{i\in\text{group } j} y_i = n_j\widehat{\mu}_j.$$

- Approach 3: If X is the $n\times J$ design matrix, where each row is the indicator vector for each observation's group, then

$$X^\top X = \begin{bmatrix} n_1 & & \\ & \ddots & \\ & & n_J\end{bmatrix}.$$

  Moreover,

$$X^\top y = \begin{bmatrix}\sum_{i\in\text{group } 1} y_i\\ \vdots \\ \sum_{i\in\text{group } J} y_i\end{bmatrix}$$

  Thus the normal equation

$$X^\top X\begin{bmatrix}\widehat{\mu}_1\\ \vdots\\ \widehat{\mu}_J\end{bmatrix} = X^\top y$$

  yields the answer.

5 Textbook problems

5.3

Taking the derivative of S with respect to A' and setting it equal to zero yields

$$\sum_{i=1}^n Y_i = nA'.$$

5.4

Recall that the least squares coefficients in simple regression are

$$\begin{aligned} B &= \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_i (X_i - \overline{X})^2} \\ A &= \overline{Y} - B\overline{X} \\ S_E &= \sqrt{\frac{\sum_i (Y_i - \widehat{Y}_i)^2}{n-2}} \\ r &= \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_i (X_i - \overline{X})^2 \sum_i (Y_i - \overline{Y})^2}} \end{aligned}$$

(a) (i) If $X' = X - 10$, then $X_i' = X_i - 10$ and $\overline{X}' = \overline{X} - 10$. Thus $B' = B$ and $A' = A + 10B$ and $r' = r$. Then $\widehat{Y}_i' = \widehat{Y}_i$, so $S_E' = S_E$.

(ii) If $X' = 10X$, then $X_i' = 10X_i$ and $\overline{X}' = 10\overline{X}$. Thus $B' = B/10$ and $A' = A$ and $r' = r$. Then $\widehat{Y}_i' = \widehat{Y}_i$, so $S_E' = S_E$.

(iii) If $X' = 10X - 10$, then applying (ii) then (i) yields $B' = B/10$ and $A' = A + B$ and $r' = r$. Again, $\widehat{Y}_i' = \widehat{Y}_i$, so $S_E' = S_E$.

(b) (i) If $Y'' = Y + 10$, then $Y_i'' = Y_i + 10$ and $\overline{Y}'' = \overline{Y} + 10$. Thus $B'' = B$ and $A'' = A + 10$ and $r' = r$. Then $\widehat{Y}_i'' = \widehat{Y}_i + 10$ so $S_E'' = S_E$.

(ii) If $Y'' = 5Y$, then $Y_i'' = 5Y_i$ and $\overline{Y}'' = 5\overline{Y}$. Thus $B'' = 5B$ and $A'' = 5A$ and $r' = r$. Then $\widehat{Y}_i'' = 5\widehat{Y}_i$ so $S_E'' = 5S_E$.

(iii) If $Y'' = 5Y + 10$, then applying (ii) and then (i) yields $B'' = 5B$ and $A'' = 5A + 10$ and $r' = r$. Then $\widehat{Y}_i'' = 5\widehat{Y} + 10$ so $S_E'' = 5S_E$.

(c) If $X' = c_1X + c_2$ and $Y' = c_3Y + c_4$ with $c_1 \neq 0$, then $B' = \frac{c_3}{c_1}B$ and $A' = c_3(A - \frac{c_2}{c_1}B) + c_4$ and $r' = \operatorname{sign}(c_1)\operatorname{sign}(c_3) \cdot r$. Then $\widehat{Y}_i' = c_3\widehat{Y}_i + c_4$ so $S_E' = |c_3|S_3$.

6.6

Recall that in simple regression, the standard error of the slope coefficient is

$$\mathrm{SE}(B) = \sqrt{\frac{S_E^2}{\sum_i (X_i - \overline{X})^2}}.$$

(a) Since $B' = B/10$ and $S_E' = S_E$, we have $\mathrm{SE}(B') = \mathrm{SE}(B)/10$ and $t_0' = t_0$.

(b) Since $B'' = 5B$ and $S_E'' = 5S_E$, we have $\mathrm{SE}(B'') = 5\,\mathrm{SE}(B)$ and $t_0'' = t_0$.

(c) Hypothesis tests for the slope do not change because the t statistic stays the same. If $X' = c_1X + c_2$ and $Y' = c_3Y + c_4$ with $c_1 \neq 0$, then the new confidence interval is of the form $\frac{c_3}{c_1}B \pm q|\frac{c_3}{c_1}|S_E$ where q is the appropriate quantile of the t-distribution. So the center of the interval changes according to $B' = \frac{c_3}{c_1}B$, and the width of the interval scales by $|c_3/c_1|$.

9.8

[This is also proved in the lecture notes on normal regression theory.]

Our goal is to show $\mathrm{Cov}(e, b) = 0$. [Then, since (e, b) is jointly Gaussian, this implies e and b are independent, and consequently B_j and S_E are independent.]

As noted in the hint,

$$b - \beta = (X^\top X)^{-1}X^\top(X\beta + \epsilon) - \beta = (X^\top X)^{-1}X^\top\epsilon,$$

so

$$\begin{aligned}\operatorname{Cov}(e,b) &= \mathbb{E}[e(b-\beta)^\top] \\ &= \mathbb{E}[e\epsilon^\top X(X^\top X)^{-1}] \\ &= \mathbb{E}[e\epsilon^\top]X(X^\top X)^{-1}.\end{aligned}$$

Recall $e = (I-H)\epsilon$ (e.g., see Lab 3 notes), where $H = X(X^\top X)^{-1}X^\top$. Thus,

$$\begin{aligned}\operatorname{Cov}(e,b) &= \mathbb{E}[(I-H)\epsilon\epsilon^\top]X(X^\top X)^{-1} \\ &= (I-H)(\sigma^2 I)X(X^\top X)^{-1} \\ &= \sigma^2\big[I - X(X^\top X)^{-1}X^\top\big]X(X^\top X)^{-1} \\ &= 0.\end{aligned}$$