Homework 1.

Theory

1. Show that β_0 is unbiased

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 x_i + \epsilon_i \\
\mathrm{E}(\beta_0) &= \mathrm{E}(y_i - \beta_1 x_i - \epsilon_i) \\
\mathrm{E}(\beta_0) &= \mathrm{E}(y_i) - \beta_1 \mathrm{E}(x_i) - \mathrm{E}(\epsilon_i) \\
\mathrm{E}(\beta_0) &= \bar{y} - \beta_1 \bar{x} - 0 \\
\mathrm{E}(\beta_0) &= \bar{y} - \beta_1 \bar{x} \\
\mathrm{E}(\beta_0) &= \beta_0 \\
\beta_0 &= \bar{y} - \beta_1 \bar{x}
\end{aligned}
$$

$$
\begin{aligned}
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\
\mathrm{E}(\hat{\beta}_0) &= \mathrm{E}(\bar{y} - \hat{\beta}_1 \bar{x}) \\
\mathrm{E}(\hat{\beta}_0) &= \mathrm{E}(\bar{y}) - \mathrm{E}(\hat{\beta}_1 \bar{x}) \\
\mathrm{E}(\hat{\beta}_0) &= \bar{y} - \bar{x}\,\mathrm{E}(\hat{\beta}_1) \\
\mathrm{E}(\hat{\beta}_1) &= \beta_1 \qquad\qquad \text{(we showed this in class)} \\
\mathrm{E}(\hat{\beta}_0) &= \bar{y} - \beta_1 \bar{x} \\
\mathrm{E}(\hat{\beta}_0) &= \beta_0
\end{aligned}
$$

2. Determine the variance of β_0

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$$

$$\begin{aligned}
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1\bar{x} && \text{(from class)} \\
\mathrm{Var}(\hat{\beta}_0) &= \mathrm{Var}(\bar{y} - \hat{\beta}_1\bar{x}) \\
\mathrm{Var}(\hat{\beta}_0) &= \mathrm{Var}(\bar{y}) + \mathrm{Var}(\hat{\beta}_1\bar{x}) + 2\,\mathrm{Cov}(\bar{y}, -\hat{\beta}_1\bar{x})
\end{aligned}$$

$$\begin{aligned}
\bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i \\
\mathrm{Var}(\bar{y}) &= \mathrm{Var}(\frac{1}{n}\sum_{i=1}^{n} y_i) \\
\mathrm{Var}(\bar{y}) &= (\frac{1}{n})^2\sum_{i=1}^{n} \mathrm{Var}(y_i) \\
\mathrm{Var}(\bar{y}) &= (\frac{1}{n})^2\sum_{i=1}^{n} \mathrm{Var}(\beta_0 + \beta_1 x_i + \epsilon_i) \\
\mathrm{Var}(\bar{y}) &= (\frac{1}{n})^2\sum_{i=1}^{n} \mathrm{Var}(\beta_0) + \mathrm{Var}(\beta_1 x_i) + \mathrm{Var}(\epsilon_i) \\
\mathrm{Var}(\bar{y}) &= (\frac{1}{n})^2\sum_{i=1}^{n} 0 + 0 + \sigma^2 \\
\mathrm{Var}(\bar{y}) &= (\frac{1}{n})^2.(n\sigma^2) \\
\mathrm{Var}(\bar{y}) &= \frac{\sigma^2}{n}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(-\beta_1\bar{x}) &= (-\bar{x})^2\,\mathrm{Var}(\hat{\beta}_1) \\
\mathrm{Var}(-\beta_1\bar{x}) &= (-\bar{x})^2\,\mathrm{Var}(\hat{\beta}_1) \\
\mathrm{Var}(-\beta_1\bar{x}) &= (-\bar{x})^2\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})} && \text{(we showed this in class)}
\end{aligned}$$

$$\mathrm{Cov}(\bar{y}, \hat{\beta}_1\bar{x}) = \bar{x}\,\mathrm{Cov}(\bar{y}, \hat{\beta}_1)$$

$$\mathrm{Cov}(\bar{y}, \hat{\beta}_1\bar{x}) = \bar{x}\,\mathrm{Cov}(\frac{1}{n}\sum_{i=1}^{n} y_i, \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}) \qquad \text{(definition from class notes)}$$

$$\mathrm{Cov}(\bar{y}, \hat{\beta}_1\bar{x}) = \bar{x}\cdot\frac{1}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\,\mathrm{Cov}(\sum_{i=1}^{n} y_i, \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}))$$

$$\mathrm{Cov}(\bar{y}, \hat{\beta}_1\bar{x}) = \bar{x}\cdot\frac{1}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\,\mathrm{Cov}(\sum_{i=1}^{n} y_i, \sum_{i=1}^{n}(x_i-\bar{x})(y_i) - \bar{y}\sum_{i=1}^{n}(x_i-\bar{x}))$$

$$\mathrm{Cov}(\bar{y}, \hat{\beta}_1\bar{x}) = \bar{x}\cdot\frac{1}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\,\mathrm{Cov}(\sum_{i=1}^{n} y_i, \sum_{i=1}^{n}(x_i-\bar{x})(y_i)) \qquad \text{(special zero from class)}$$

$$\mathrm{Cov}(\bar{y}, \hat{\beta}_1\bar{x}) = \bar{x}\cdot\frac{1}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\sum_{i=1}^{n}(x_i-\bar{x})\sum_{i=1}^{n}\mathrm{Cov}(y_i, y_i)$$

$$\sum_{i=1}^{n}(x_i-\bar{x}) = 0$$

$$\mathrm{Cov}(\bar{y}, \hat{\beta}_1\bar{x}) = 0$$

$$\mathrm{Var}(\hat{\beta}_0) = \mathrm{Var}(\bar{y}) + \mathrm{Var}(\hat{\beta}_1\bar{x}) + \mathrm{Cov}(\bar{y}, \hat{\beta}_1\bar{x})$$

$$\mathrm{Var}(\hat{\beta}_0) = \mathrm{Var}(\bar{y}) + \mathrm{Var}(\hat{\beta}_1\bar{x}) + 0$$

$$\boxed{\mathrm{Var}(\hat{\beta}_0) = \frac{\sigma^2}{n} + \bar{x}^2\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})}}$$

3(a). Show that $\sum_{i=1}^{n} \hat{y}_i \cdot e_i = 0$

$$\sum_{i=1}^{n} \hat{y}_i \cdot e_i = 0$$

$$\sum_{i=1}^{n} (\hat{\beta}_0 + \hat{\beta}_1 x_i)(e_i) = 0$$

$$\sum_{i=1}^{n} \hat{\beta}_0(e_i) + \hat{\beta}_1(x_i)(e_i) = 0$$

$$\sum_{i=1}^{n} \hat{\beta}_0(e_i) + \sum_{i=1}^{n} \hat{\beta}_1(x_i)(e_i) = 0$$

$$\hat{\beta}_0 \sum_{i=1}^{n} (e_i) + \hat{\beta}_1 \sum_{i=1}^{n} (x_i)(e_i) = 0$$

$$\sum_{i=1}^{n} (e_i) = 0 \qquad \text{(from property 1)}$$

$$\sum_{i=1}^{n} (x_i)(e_i) = 0 \qquad \text{(from property 4)}$$

$$\hat{\beta}_0 \cdot 0 + \hat{\beta}_1 \cdot 0 = 0$$

$$0 = 0$$

3(b). Show that regression line passes through $(\overline{x}, \overline{y})$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$$\text{let } x_i = \overline{x}$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \overline{x}$$

$$\hat{y}_i = \overline{y} - \beta_1 \overline{x} + \hat{\beta}_1 \overline{x} \qquad \text{(from normal equations)}$$

$$\hat{y}_i = \overline{y}$$

$$\text{this shows that the regression line passes through } (\overline{x}, \overline{y})$$

4. Derive the MLE of σ^2

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 x_i + \varepsilon_i, && i = 1, \ldots, n, \\
\varepsilon_i &= y_i - \beta_0 - \beta_1 x_i, \\
\varepsilon_i &\sim \mathcal{N}(0, \sigma^2), \\
y_i - \beta_0 - \beta_1 x_i &\sim \mathcal{N}(0, \sigma^2) \\
L(\beta_0, \beta_1, \sigma^2 \mid \{(x_i, y_i)\}_{i=1}^n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right) \\
L &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2\right) \\
\ln L &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \\
\frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n \left(y_i - \beta_0 - \beta_1 x_i\right)^2 \\
\frac{\partial \ln L}{\partial \sigma^2} &= 0 \\
0 &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n \left(y_i - \beta_0 - \beta_1 x_i\right)^2 \\
\frac{n}{2\sigma^2} &= \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n \left(y_i - \beta_0 - \beta_1 x_i\right)^2 \\
n &= \frac{1}{\sigma^2} \sum_{i=1}^n \left(y_i - \beta_0 - \beta_1 x_i\right)^2
\end{aligned}
$$

$$
\boxed{\sigma^2_{MLE} = \frac{\sum_{i=1}^n \left(y_i - \beta_0 - \beta_1 x_i\right)^2}{n}}
$$

Methods

Inference and result

- **Linear model fit:**

$$\widehat{\mathrm{RTEN}} = -0.593\ +\ 1.097 \cdot \mathrm{PREP}$$

- **Estimates:**

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.90625 -0.20008  0.06453  0.21065  0.60091

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.59257    0.42265  -1.402    0.168
X            1.09742    0.05615  19.543   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.347 on 41 degrees of freedom
Multiple R-squared:  0.9031,	Adjusted R-squared:  0.9007
F-statistic: 381.9 on 1 and 41 DF,  p-value: < 2.2e-16
```

- **Interpretation:** A 1 unit increase in the preparedness rating of the judge increases their retention score by 1.10 on average.
- **Inference:**

To see whether trial preparation (PREP) is a statistically significant predictor of retention ratings (RTEN), we test the hypotheses below:

$$H_0 : \beta_1 = 0 \qquad \text{vs.} \qquad H_a : \beta_1 \neq 0.$$

The linear model we fit gives the resutls below:

$$\hat{\beta}_1 = 1.097, \quad SE(\hat{\beta}_1) = 0.056, \quad t = 19.54, \quad p < 2 \times 10^{-16}.$$

Since the p-value is significantly less than 0.05, we reject H_0 and conclude that trial preparation is a statistically significant predictor of retention ratings at the 5% confidence level.

A 95% confidence interval for the slope is given by

$$1.097 \pm 2.02 \times 0.056 = (0.985,\ 1.209),$$

so, for a one-unit increase in trial preparation rating, we expect that the retention rating would increase by between 0.99 and 1.21 points.