

Homework 1.

Question 1.

1. Show that β_0 is unbiased

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

$$E(\beta_0) = E(y_i - \beta_1 x_i - \epsilon_i)$$

$$E(\beta_0) = E(y_i) - \beta_1 E(x_i) - E(\epsilon_i)$$

$$E(\beta_0) = \bar{y} - \beta_1 \bar{x} - 0$$

$$E(\beta_0) = \bar{y} - \beta_1 \bar{x}$$

$$E(\beta_0) = \beta_0$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

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

$$E(\hat{\beta}_0) = E(\bar{y} - \hat{\beta}_1 \bar{x})$$

$$E(\hat{\beta}_0) = E(\bar{y}) - E(\hat{\beta}_1 \bar{x})$$

$$E(\hat{\beta}_0) = \bar{y} - \bar{x} E(\hat{\beta}_1)$$

$$E(\hat{\beta}_1) = \beta_1 \quad (\text{we showed this in class})$$

$$E(\hat{\beta}_0) = \bar{y} - \beta_1 \bar{x}$$

$$E(\hat{\beta}_0) = \beta_0$$

2. Determine the variance of β_0

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

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (\text{from class})$$

$$\text{Var}(\hat{\beta}_0) = \text{Var}(\bar{y} - \hat{\beta}_1 \bar{x})$$

$$\text{Var}(\hat{\beta}_0) = \text{Var}(\bar{y}) + \text{Var}(\hat{\beta}_1 \bar{x}) + 2 \text{Cov}(\bar{y}, -\hat{\beta}_1 \bar{x})$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\text{Var}(\bar{y}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n y_i\right)$$

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

$$\text{Var}(\bar{y}) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^n \text{Var}(\beta_0 + \beta_1 x_i + \epsilon_i)$$

$$\text{Var}(\bar{y}) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^n \text{Var}(\beta_0) + \text{Var}(\beta_1 x_i) + \text{Var}(\epsilon_i)$$

$$\text{Var}(\bar{y}) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^n 0 + 0 + \sigma^2$$

$$\text{Var}(\bar{y}) = \left(\frac{1}{n}\right)^2 \cdot (n\sigma^2)$$

$$\text{Var}(\bar{y}) = \frac{\sigma^2}{n}$$

$$\text{Var}(-\beta_1 \bar{x}) = (-\bar{x})^2 \text{Var}(\hat{\beta}_1)$$

$$\text{Var}(-\beta_1 \bar{x}) = (-\bar{x})^2 \text{Var}(\hat{\beta}_1)$$

$$\text{Var}(-\beta_1 \bar{x}) = (-\bar{x})^2 \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})} \quad (\text{we showed this in class})$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$0 = 0$$

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

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

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

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

$$\hat{y}_i = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} \quad (\text{from normal equations})$$

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

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

4. Inference and result

- **Linear model fit:**

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

- **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:

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

The linear model we fit gives the results 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.