Inferences about the Regression Coefficients

1. **Statistical model for a linear regression**

$$y = \beta_0 + \beta_1 x + \epsilon$$

2. **Assumptions**

   (a) The relationship is linear and $\mathrm{E}(\epsilon_i) = 0$ for all i.

   (b) The errors have the same variance.

   (c) $\epsilon_i \sim N(0, \sigma^2)$ for all i.

3. **Least squares estimation**

   - The best fitting straight line is given by

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

   - The least-square estimates of β_1 and β_0 are

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad \text{and} \quad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

   where $S_{xx} = \sum_i x_i^2 - n(\overline{x})^2$ and $S_{xy} = \sum_i x_i y_i - n(\overline{x})(\overline{y})$.

   - Residuals: $\hat{\epsilon}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i),\ i = 1, \ldots, n$
   - Residual sum of square (error sum of square):

$$\mathrm{SSE} = \sum_{i=1}^{n} \hat{\epsilon}_i^2 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}, \quad \text{where} \quad S_{yy} = \sum_{i=1}^{n} y_i^2 - n(\overline{y})^2$$

   - Estimate of the error variance σ^2:

$$S^2 = \frac{\mathrm{SSE}}{n-2}$$

4. **Correlation**

   - Coefficient of determination:

$$R^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}}$$

   - Sample correlation coefficient:

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

5. **Inference about the regression parameters**

Estimation

Parameter	Point Estimate	Confidence Interval
β_1	$\hat{\beta}_1$	$\hat{\beta}_1 \pm t_{\alpha/2,n-2} S/\sqrt{S_{xx}}$
β_0	$\hat{\beta}_0$	$\hat{\beta}_0 \pm t_{\alpha/2,n-2} S\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}}$

Testing

	Case (a)	Case (b)	Case (c)
Step 1	$H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$ (2-sided alternative)	$H_0 : \beta_1 \leq 0$ $H_1 : \beta_1 > 0$ (1-sided alternative)	$H_0 : \beta_1 \geq 0$ $H_1 : \beta_1 < 0$ (1-sided alternative)
Step 2	$\alpha = ?$	$\alpha = ?$	$\alpha = ?$
Step 3	$t = \frac{\hat{\beta}_1}{S/\sqrt{S_{xx}}}$ Rejection region $\lvert t \rvert \geq t_{\alpha/2,n-2}$	$t = \frac{\hat{\beta}_1}{S/\sqrt{S_{xx}}}$ Rejection region $t \geq t_{\alpha,n-2}$	$t = \frac{\hat{\beta}_1}{S/\sqrt{S_{xx}}}$ Rejection region $t \leq -t_{\alpha,n-2}$
Step 4	Substitute $\hat{\beta}_1$ $t = ?$ Decision	$t = ?$ Decision	$t = ?$ Decision
Step 5	$p = 2\times$ area	$p =$ area	$p =$ area

Estimation of the mean response for a specified x value

A 100 $(1-\alpha)\%$ confidence interval for the expected response $y^* = \beta_0 + \beta_1 x^*$:

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2,n-2} S \sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}$$

Prediction of a single response for a specified x value

A 100 $(1-\alpha)\%$ prediction interval for a future trial:

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2,n-2} S \sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}$$