Hypothesis tests in linear regression

- Slope:
  - H_0: $\beta_1 = \beta_{1,0}$
  - H_1: $\beta_1 \neq \beta_{1,0}$
  - Reject H_0: $\beta_1 = \beta_{1,0}$ if $|t_0| > t_{\alpha/2,n-2}$

$$T_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2 / S_{xx}}} = \frac{\hat{\beta}_1 - \beta_{1,0}}{se(\hat{\beta}_1)}$$

- Intercept:
  - H_0: $\beta_0 = \beta_{0,0}$
  - H_1: $\beta_0 \neq \beta_{0,0}$
  - Reject H_0: $\beta_1 = \beta_{1,0}$ if $|t_0| > t_{\alpha/2,n-2}$

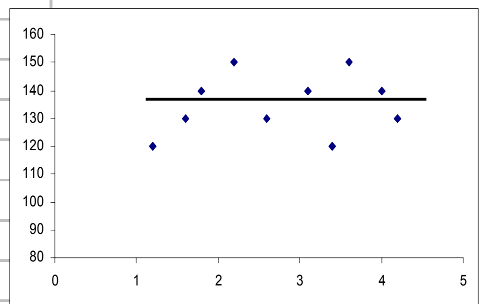
$$T_0 = \frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]}} = \frac{\hat{\beta}_0 - \beta_{0,0}}{se(\hat{\beta}_0)}$$

Hypothesis tests in linear regression

- Special case:
  - H_0: $\beta_1 = 0$
  - H_1: $\beta_1 \neq 0$
- Fail to reject H_0: $\beta_1 = 0$ is equivalent to concluding that there is no linear relationship between x and Y.

Example: Sales vs. Advertising

Advertising (mil.$)	Sales (mil.$)
1.2	120
1.6	130
1.8	140
2.2	150
2.6	130
3.1	140
3.4	120
3.6	150
4.0	140
4.2	130



Analysis of variance

- Test for significance of regression.
  - Analysis of variance identity

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

  - Symbolical equation $SS_T = SS_R + SS_E$

SS_T : *Total correct sum of squares*

SS_R : *Re gression sum of squares* SS_E : *Error sum of squares*

Analysis of variance

- Test statistics

$$F_0 = \frac{SS_R / 1}{SS_E /(n-2)} = \frac{MS_R}{MS_E}$$

- Reject H_0: $\beta_1 = 0$, if $f_0 > f_{\alpha,1,n-2}$

Confidence intervals

- Confidence interval on the slope

$$\hat{\beta}_1 - t_{\alpha/2,n-2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \le \beta_1 \le \hat{\beta}_1 + t_{\alpha/2,n-2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$$

- Confidence interval on the intercept

$$\hat{\beta}_0 - t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]} \le \beta_0 \le \hat{\beta}_0 + t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]}$$

Confidence interval on the mean response

$$\hat{\mu}_{Y|x_0} - t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]}$$

$$\leq \mu_{Y|x_0} \leq \hat{\mu}_{Y|x_0} + t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]}$$

$$where\ \hat{\mu}_{Y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$$