Formula Sheet – Exam 4

Chapter 9: Hypothesis Testing

Regions of rejection

This is the same for means and proportions, just replace μ with p

Hypothesis for comparing means	$H_0: \mu_1 - \mu_2 = 0$ $H_A: \mu_1 - \mu_2 \neq 0$	$H_0: \mu_1 - \mu_2 \geq 0$ $H_A: \mu_1 - \mu_2 < 0$	$H_0: \mu_1 - \mu_2 \leq 0$ $H_A: \mu_1 - \mu_2 > 0$
Region for Rejection (large samples)	$z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$	$z < -z_{\alpha}$	$z > z_{\alpha}$
Region for Rejection (small samples)	$t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$	$t < -t_{\alpha}$	$t > t_{\alpha}$

- Difference in Population Means

$$z = \frac{(\overline{x}_1 - \overline{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Difference in Population Proportions

$$z = \frac{(\overline{p}_1 - \overline{p}_2)}{\sqrt{\overline{p}(1-\overline{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \text{where} \quad \overline{p} = \frac{n_1\overline{p}_1 + n_2\overline{p}_2}{n_1 + n_2}$$

Differences in Small Samples using the t-statistic

- Difference in population mean

$$t = \frac{(\overline{x}_1 - \overline{x}_2)}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \text{where} \quad s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Chapter 12: Regression Formula

Regression Model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad \text{where} \quad \varepsilon_i \sim N(0, \sigma) \text{ and independent}$$

Sum of Squares

$$SS_{XX} = \sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \qquad SS_{XY} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \qquad SS_{YY} = \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2$$

Ordinary Least Squares Estimators

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} \qquad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$$

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

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$$

Sum of Square Errors, Variance and Standard Error

$$SSE = \sum (e_i)^2 = \sum (y_i - \bar{y})^2 = SS_{YY} - \hat{\beta}_1 SS_{XY}$$

$$s^2 = \frac{SSE}{n-2}$$

$$s = \sqrt{\frac{SSE}{n-2}}$$

Sampling Distribution of Estimated Slope

$$\sigma_{\hat{\beta}_1} = \frac{\sigma}{\sqrt{SS_{XX}}}$$ (population)

$$s_{\hat{\beta}_1} = \frac{s}{\sqrt{SS_{XX}}}$$ (estimated value)

$$t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}$$ (has a *t*-distribution with *n*–2 degrees of freedom)

$$\hat{\beta}_1 \pm t_{\alpha/2} s_{\hat{\beta}_1}$$ $(1-\alpha)$ confidence interval for β_1

Correlation and Fit

$$R = \frac{SS_{XY}}{\sqrt{SS_{XX} SS_{YY}}}$$ (Correlation Coefficient)

$$R^2 = 1 - \frac{SSE}{SS_{YY}}$$ (Coefficient of Determination)

Estimation and prediction confidence intervals

A $(1-\alpha)$ confidence interval for the mean value of *y* at $x = x_t$ is

$$\hat{\beta}_0 + \hat{\beta}_1 x_t \pm t_{\alpha/2} s \sqrt{\frac{1}{n} + \frac{(x_t - \bar{x})^2}{SS_{XX}}}$$

A $(1-\alpha)$ confidence interval for the value of *y* at $x = x_t$ is

$$\hat{\beta}_0 + \hat{\beta}_1 x_t \pm t_{\alpha/2} s \sqrt{1 + \frac{1}{n} + \frac{(x_t - \bar{x})^2}{SS_{XX}}}$$