Test 1 Formula Sheet

STAT 021

Swarthmore College

Summary statistics

- Sample mean: $\frac{1}{n}\sum_{i=1}^{n} x_i$
- Sample variance: $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- Population mean: For a random variable X, the expectation of X is

$$E(X) = \sum(\text{possibilities} \times \text{probabilities})$$

- Population variance: For a random variable X, the variance of X is

$$Var(X) = \sum((\text{possibilities} - E(X))^2 \times \text{probabilities})$$

Population proportions

- To find a $(1-\alpha)100\%$ CI for p: $\hat{p} \pm z^*_{\alpha/2} \times SE(\hat{p})$
- If p_0 is the true value of p, then $\frac{\hat{p}-p_0}{SE(\hat{p})} \sim N(0,1)$ for large enough n.
- To find a $(1-\alpha)100\%$ CI for $p_1 - p_2$: $\hat{p}_1 - \hat{p}_2 \pm z^*_{\alpha/2} \times SE(\hat{p}_1 - \hat{p}_2)$
- If $p_1 - p_2$ is the true difference in two independent population proportions, then $\frac{\hat{p}_1-\hat{p}_2}{SE(\hat{p}_1-\hat{p}_2)} \sim N(0,1)$ for large enough n_1 and n_2.

Population means

- To find a $(1-\alpha)100\%$ CI for μ: $\hat{p} \pm t^*_{(n-1),\alpha/2} \times SE(\bar{x})$
- If μ_0 is the true value of μ, then $\frac{\bar{x}-\mu_0}{SE(\bar{x})} \sim t_{(n-1)}$ for any $n > 2$.
- To find a $(1-\alpha)100\%$ CI for $\mu_1 - \mu_2$ for two independent populations: $\bar{x}_1 - \bar{x}_2 \pm t^*_{\nu,\alpha/2} \times SE(\bar{x}_1 - \bar{x}_2)$, but the formula for ν is complicated and you don't need to know it.
- If $\mu_1 - \mu_2 = 0$ is the true difference in two independent population means, then $\frac{\bar{x}-\mu_0}{SE(\bar{x})} \sim t_\nu$ for any $n > 2$.

Linear Regression Formulas and Definitions

- Linear model: $Y = \beta_0 + \beta_1 x + \epsilon$

In the model above, if we assume that the mean of ϵ is 0 and the variance of ϵ is some unknown number, σ^2, then the mean of the random variable Y is $\beta_0 + \beta_1 x$ and the variance of Y is σ^2.

- Fitted/estimated model: $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_1 x_i$
- Residuals: $e_i = \hat{y}_i - y_i$

In the fitted model, we solve for the least squares estimates of the parameters using these equations:

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

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

- To find a $(1-\alpha)100\%$ CI for β_1: $\hat{\beta}_1 \pm t^*_{(n-2),\alpha/2} \times SE(\hat{\beta}_1)$
- If $\beta_1 = 0$ is the true regression slope, then $\frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} \sim t_{(n-2)}$ for any $n > 2$.

Information in an ANOVA table includes:

- Regression model sums of squares: $SS_{Mod} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$
- Error sums of squares: $SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
- Total sums of squares: $SS_{Tot} = \sum_{i=1}^{n}(y_i - \bar{y})^2$
- Relationship among the sums of squares terms: $SS_{Tot} = SS_{Mod} + SSE$

The sums of squares terms are used to calculate the following statistics:

- $\hat{\sigma} = \sqrt{\frac{SSE}{n-2}}$
- $R^2 = 1 - \frac{SSE}{SS_{Tot}} = \frac{SS_{Mod}}{SS_{Tot}}$