Key Formulas for Statistical Methods

Chapter 3 Descriptive Statistics

Mean $\bar{Y} = \frac{\Sigma Y_i}{n}$ Standard deviation $s = \sqrt{\frac{\Sigma(Y_i - \bar{Y})^2}{n-1}}$

Chapter 4 Probability Distributions

z-score $z = \frac{Y - \mu}{\sigma}$ Standard error $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$

Chapter 5 Statistical Inference: Estimation

Confidence interval for mean $\bar{Y} \pm z\hat{\sigma}_{\bar{Y}}$ with $\hat{\sigma}_{\bar{Y}} = \frac{s}{\sqrt{n}}$

Confidence interval for proportion $\hat{\pi} \pm z\hat{\sigma}_{\hat{\pi}}$ with $\hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$

Chapter 6 Statistical Inference: Significance Tests

$H_0: \mu = \mu_0$ test statistic $\frac{\bar{Y} - \mu_0}{\hat{\sigma}_{\bar{Y}}}$ with $\hat{\sigma}_{\bar{Y}} = \frac{s}{\sqrt{n}}$

This is z test statistic, with approximate normal distribution, for $n \geq 30$

This is t test statistic, t distribution, $df = n - 1$ (assume normal population for small n)

$H_0: \pi = \pi_0$ test statistic $z = \frac{\hat{\pi} - \pi_0}{\sigma_{\hat{\pi}}}$ with $\hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$

Chapter 7 Comparison of Two Groups

Compare means: $(\bar{Y}_2 - \bar{Y}_1) \pm z\hat{\sigma}_{\bar{Y}_2 - \bar{Y}_1}$ with $\hat{\sigma}_{\bar{Y}_2 - \bar{Y}_1} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Test $H_0: \mu_1 = \mu_2$ using $z = \frac{\bar{Y}_2 - \bar{Y}_1}{\hat{\sigma}_{\bar{Y}_2 - \bar{Y}_1}}$

Compare proportions: $(\hat{\pi}_2 - \hat{\pi}_1) \pm z\hat{\sigma}_{\hat{\pi}_2 - \hat{\pi}_1}$ with $\hat{\sigma}_{\hat{\pi}_2 - \hat{\pi}_1} = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$

Chapter 8 Analyzing Association Between Categorical Variables

Chi-squared test of H_0: Independence, $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$, $df = (r-1)(c-1)$

Ordinal measure $\hat{\gamma} = \frac{C - D}{C + D}$, $-1 \leq \hat{\gamma} \leq 1$, $z = \frac{\hat{\gamma}}{\hat{\sigma}_{\hat{\gamma}}}$, $\hat{\gamma} \pm z\hat{\sigma}_{\hat{\gamma}}$

Chapter 9 Linear Regression and Correlation

Linear regression model $E(Y) = \alpha + \beta X$, prediction equation $\hat{Y} = a + bX$

Pearson correlation $r = \left(\frac{s_X}{s_Y}\right) b, \; -1 \le r \le 1$

PRE $r^2 = \frac{\text{TSS} - \text{SSE}}{\text{TSS}}$, $\text{TSS} = \sum (Y - \bar{Y})^2$, $\text{SSE} = \sum (Y - \hat{Y})^2$, $0 \le r^2 \le 1$

Test of independence $H_0 : \beta = 0, \quad t = \frac{b}{\hat{\sigma}_b}, \quad df = n - 2$

Chapter 11 Multiple Regression and Correlation

Multiple regression model $E(Y) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$

Global test $H_0 : \beta_1 = \cdots = \beta_k = 0$

Test statistic $F = \frac{\text{Model mean square}}{\text{Error mean square}} = \frac{R^2/k}{(1 - R^2) / n - (k+1)}$

$df_1 = k, df_2 = n - (k+1)$

Partial test $H_0 : \beta_i = 0$, test statistic $t = \frac{b_i}{\hat{\sigma}_{b_i}}, \quad df = n - (k+1)$

Chapter 12 Comparing Groups: Analysis of Variance Methods

$H_0 : \mu_1 = \cdots = \mu_g$, One-way ANOVA test statistic

$$F = \frac{\text{Between sum of squares}/(g-1)}{\text{Within sum of squares}/(N-g)}, \; df_1 = g - 1, df_2 = N - g$$

Chapter 13 Combining Regression and ANOVA: Analysis of Covariance

$E(Y) = \alpha + \beta X + \beta_1 Z_1 + \cdots + \beta_{g-1} Z_{g-1}$, $Z_i = 1$ or 0 is dummy variable for group i

Chapter 14 Model Building with Multiple Regression

Quadratic regression $E(Y) = \alpha + \beta_1 X + \beta_2 X^2$

Exponential regression $E(Y) = \alpha \beta^X$ (log of mean is linear in X)

Chapter 15 Logistic Regression: Modeling Categorical Responses

Logistic regression $\text{logit}(\pi) = \log(\text{odds}) = \log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta X$

$\pi = \frac{e^{\alpha + \beta X}}{1 + e^{\alpha + \beta X}} = \frac{\text{odds}}{1 + \text{odds}}$