Appendix 6-1: Hypothesis Test Guide

This appendix provides a guide to the hypothesis tests covered in the chapter.

Hypothesis Tests for Means in the C+E Model: One Population

Case 1: Known Variances

Assumptions

1. The data are $Y_1, Y_2, \ldots, Y_n$ where $Y_j = \mu + \epsilon_j$.
2. Either
   (a) n is large, or
   (b) The ϵ_j are from a $N(0, \sigma^2)$ population.
3. σ^2 is known.

Formulas

Standardized Test Statistic[1]: $Z = \frac{\overline{Y} - \mu_0}{\sigma(\overline{Y})}$.

Hypotheses:	$H_0: \mu = \mu_0$ $H_{a_-}: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_{a^+}: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_{a\pm}: \mu \neq \mu_0$
p-value[2]:	$p_- = P(N(0,1) \leq z^*)$	$p^+ = P(N(0,1) \geq z^*)$	$p\pm = 2\min(p_-, p^+)$

[1] $\sigma(\overline{Y}) = \sqrt{\frac{\sigma^2}{n}}$.

[2] z^* is the observed value of Z, $P(N(0,1) \leq z^*)$ is the proportion of a $N(0,1)$ population less than z^*, and $P(N(0,1) \geq z^*)$ is the proportion of a $N(0,1)$ population greater than z^*.

Case 2: Unknown Variance, n Large

Assumptions 1 and 2a from Case 1 are assumed to hold. This case is treated exactly as Case 1 except that $\sigma(\overline{Y})$ is replaced by $\hat{\sigma}(\overline{Y}) = \sqrt{\frac{S^2}{n}}$ in the computation of the standardized test statistic Z, where S^2 is the sample variance computed from the data.

Case 3: Unknown Variance, n Small

Assumptions

1. The data are $Y_1, Y_2, \ldots, Y_n$ where $Y_j = \mu + \epsilon_j$.
2. The ϵ_j are from a $N(0, \sigma^2)$ population.
3. σ^2 is unknown.

Formulas

Standardized Test Statistic[1]: $t = \frac{\overline{Y} - \mu_0}{\hat{\sigma}(\overline{Y})}$.

Hypotheses:	$H_0: \mu = \mu_0$ $H_{a_-}: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_{a^+}: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_{a\pm}: \mu \neq \mu_0$
p-value[2]:	$p_- = P(t_{n-1} \leq t^*)$	$p^+ = P(t_{n-1} \geq t^*)$	$p\pm = 2\min(p_-, p^+)$

[1] $\hat{\sigma}(\overline{Y}) = \sqrt{\frac{S^2}{n}}$.

[2] t^* is the observed value of t, $P(t_{n-1} \leq t^*)$ is the proportion of a t_{n-1} population less than t^*, and $P(t_{n-1} \geq t^*)$ is the proportion of a t_{n-1} population greater than t^*.

Hypothesis Tests for the Proportion in the Binomial Model: One Population

Case 1: An Exact Test

Assumption

The datum is Y from a $b(n, p)$ population.

Formulas

Test Statistic: Y.

Hypotheses:	$H_0: p = p_0$ $H_{a-}: p < p_0$	$H_0: p = p_0$ $H_{a+}: p > p_0$	$H_0: p = p_0$ $H_{a\pm}: p \neq p_0$
p-value[1]:	$p_- = P(b(n,p_0) \leq y^*)$	$p^+ = P(b(n,p_0) \geq y^*)$	$p\pm$ is given by Equation (6.4)

[1] y^* is the observed value of Y, $P(b(n,p_0) \leq y^*)$ is the proportion of a $b(n,p_0)$ population less than or equal to y^*, and $P(b(n,p_0) \geq y^*)$ is the proportion of a $b(n,p_0)$ population greater than or equal to y^*.

Case 2: An Approximate Test for n Large

Assumptions

1. The datum is Y from a $b(n,p)$ population.
2. n is large: $Y \geq 10$ and $n - Y \geq 10$ is a good rule of thumb overall. If $0.3 \leq \hat{p} = Y/n \leq 0.7$, then $Y \geq 5$ and $n - Y \geq 5$ is a good rule of thumb.

Formulas

Standardized Test Statistic: $Z = \frac{Y - np_0}{\sigma(Y)}$ where $\sigma(Y) = \sqrt{np_0(1-p_0)}$.

Hypotheses:	$H_0: p = p_0$ $H_{a-}: p < p_0$	$H_0: p = p_0$ $H_{a+}: p > p_0$	$H_0: p = p_0$ $H_{a\pm}: p \neq p_0$
p-value[1]:	$p_- = P(N(0,1) \leq z_l^*)$	$p^+ = P(N(0,1) \geq z_u^*)$	$p\pm = 2\min(p_-, p^+)$

[1] $z_l^* = \frac{y^* - np_0(1-p_0) + 0.5}{\sigma(Y)}$, $z_u^* = \frac{y^* - np_0(1-p_0) - 0.5}{\sigma(Y)}$, where y^* is the observed value of Y. $P(N(0,1) \leq z_l^*)$ is the proportion of a $N(0,1)$ population less than z_l^*, and $P(N(0,1) \geq z_u^*)$ is the proportion of a $N(0,1)$ population greater than z_u^*.

Hypothesis Tests for Differences in Means in the C+E Model: Two Independent Populations

Case 1: Known Variances

Assumptions

1. The data are

$$Y_{1,1}, Y_{1,2}, \ldots, Y_{1,n_1}, \text{ where } Y_{1,j} = \mu_1 + \epsilon_{1,j}, \text{ (population 1)}$$
$$Y_{2,1}, Y_{2,2}, \ldots, Y_{2,n_2}, \text{ where } Y_{2,j} = \mu_2 + \epsilon_{2,j}, \text{ (population 2).}$$

2. The two populations are independent.
3. Either
   (a) n_1 and n_2 are large, or
   (b) The $\epsilon_{1,j}$ are from a $N(0, \sigma_1^2)$ population, and the $\epsilon_{2,j}$ are from a $N(0, \sigma_2^2)$ population.
4. σ_1^2 and σ_2^2 are known.

Formulas

Standardized Test Statistic[1]: $Z = \frac{\overline{Y}_1 - \overline{Y}_2 - \delta_0}{\sigma(\overline{Y}_1 - \overline{Y}_2)}$.

Hypotheses:	$H_0: \mu_1 - \mu_2 = \delta_0$ $H_{a-}: \mu_1 - \mu_2 < \delta_0$	$H_0: \mu_1 - \mu_2 = \delta_0$ $H_{a+}: \mu_1 - \mu_2 > \delta_0$	$H_0: \mu_1 - \mu_2 = \delta_0$ $H_{a\pm}: \mu_1 - \mu_2 \neq \delta_0$
p-value[2]:	$p_- = P(N(0,1) \leq z^*)$	$p^+ = P(N(0,1) \geq z^*)$	$p\pm = 2\min(p_-, p^+)$

[1] $\sigma(\overline{Y}_1 - \overline{Y}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

[2] z^* is the observed value of Z, $P(N(0,1) \leq z^*)$ is the proportion of a $N(0,1)$ population less than z^*, and $P(N(0,1) \geq z^*)$ is the proportion of a $N(0,1)$ population greater than z^*.

Case 2: Unknown Variances, n_1 and n_2 Large

Assumptions 1, 2 and 3a are assumed to hold. This case is treated exactly as Case 1 except that $\sigma(\overline{Y}_1 - \overline{Y}_2)$ is replaced by $\hat{\sigma}(\overline{Y}_1 - \overline{Y}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$ in the computation of the standardized test statistic Z, where S_1^2 and S_2^2 are the sample variances computed from the data from populations 1 and 2 respectively.

Case 3: Variances Unknown, but Assumed to Be Equal, n_1 and n_2 Not Both Large

Assumptions

1. The data are
$$Y_{1,1}, Y_{1,2}, \ldots, Y_{1,n_1}, \text{ where } Y_{1,j} = \mu_1 + \epsilon_{1,j}, \text{ (population 1)}$$
$$Y_{2,1}, Y_{2,2}, \ldots, Y_{2,n_2}, \text{ where } Y_{2,j} = \mu_2 + \epsilon_{2,j}, \text{ (population 2).}$$
2. The two populations are independent.
3. The $\epsilon_{1,j}$ are from a $N(0, \sigma_1^2)$ population, and the $\epsilon_{2,j}$ are from a $N(0, \sigma_2^2)$ population.
4. σ_1^2 and σ_2^2 are unknown, but are assumed to be equal.

Formulas

Standardized Test Statistic[1]: $t^{(p)} = \frac{\overline{Y}_1 - \overline{Y}_2 - \delta_0}{\hat{\sigma}_p(\overline{Y}_1 - \overline{Y}_2)}$.

Hypotheses:	$H_0: \mu_1 - \mu_2 = \delta_0$ $H_{a_-}: \mu_1 - \mu_2 < \delta_0$	$H_0: \mu_1 - \mu_2 = \delta_0$ $H_{a^+}: \mu_1 - \mu_2 > \delta_0$	$H_0: \mu_1 - \mu_2 = \delta_0$ $H_{a\pm}: \mu_1 - \mu_2 \neq \delta_0$
p-value[2]:	$p_- = P(t_{n_1+n_2-2} \leq t^{(p)*})$	$p^+ = P(t_{n_1+n_2-2} \geq t^{(p)*})$	$p\pm = 2\min(p_-, p^+)$

[1] $\hat{\sigma}_p(\overline{Y}_1 - \overline{Y}_2) = \sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$, where $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$, and S_1^2 and S_2^2 are the sample variances computed from the data from populations 1 and 2 respectively.

[2] $t^{(p)*}$ is the observed value of $t^{(p)}$, $P(t_{n_1+n_2-2} \leq t^{(p)*})$ is the proportion of a $t_{n_1+n_2-2}$ population less than $t^{(p)*}$, and $P(t_{n_1+n_2-2} \geq t^{(p)*})$ is the proportion of a $t_{n_1+n_2-2}$ population greater than $t^{(p)*}$.

Case 4: Variances Unknown, and Not Assumed to Be Equal, n_1 and n_2 Not Both Large

Assumptions

1. The data are
$$Y_{1,1}, Y_{1,2}, \ldots, Y_{1,n_1}, \text{ where } Y_{1,j} = \mu_1 + \epsilon_{1,j}, \text{ (population 1)}$$
$$Y_{2,1}, Y_{2,2}, \ldots, Y_{2,n_2}, \text{ where } Y_{2,j} = \mu_2 + \epsilon_{2,j}, \text{ (population 2).}$$
2. The two populations are independent.
3. The $\epsilon_{1,j}$ are from a $N(0, \sigma_1^2)$ population, and the $\epsilon_{2,j}$ are from a $N(0, \sigma_2^2)$ population.
4. σ_1^2 and σ_2^2 are unknown, and are not assumed to be equal.

Formulas

Standardized Test Statistic[1]: $t^{(ap)} = \frac{\overline{Y}_1 - \overline{Y}_2 - \delta_0}{\hat{\sigma}(\overline{Y}_1 - \overline{Y}_2)}$.

Hypotheses:	$H_0: \mu_1 - \mu_2 = \delta_0$ $H_{a_-}: \mu_1 - \mu_2 < \delta_0$	$H_0: \mu_1 - \mu_2 = \delta_0$ $H_{a^+}: \mu_1 - \mu_2 > \delta_0$	$H_0: \mu_1 - \mu_2 = \delta_0$ $H_{a\pm}: \mu_1 - \mu_2 \neq \delta_0$
p-value[2]:	$p_- = P(t_\nu \leq t^{(ap)*})$	$p^+ = P(t_\nu \geq t^{(ap)*})$	$p\pm = 2\min(p_-, p^+)$

[1] $\hat{\sigma}(\overline{Y}_1 - \overline{Y}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$, where S_1 and S_2 are the sample standard deviations computed from the data from populations 1 and 2 respectively. Under H_0, the distribution of $t^{(ap)}$ is approximately t_ν where the degrees of freedom ν is taken as the largest integer less than or equal to $\left[\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2\right] \Bigg/ \left[\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}\right]$

[2] $t^{(ap)*}$ is the observed value of $t^{(ap)}$, $P(t_\nu \leq t^{(ap)*})$ is the proportion of a t_ν population less than $t^{(ap)*}$, and $P(t_\nu \geq t^{(p)*})$ is the proportion of a t_ν population greater than $t^{(ap)*}$.

Hypothesis Tests for Proportions in the Binomial Model: Two Independent Populations

Case 1: A Test of the Equality of Two Proportions

Assumptions

1. The data are Y_1 from a $b(n_1, p_1)$ population and Y_2 from a $b(n_2, p_2)$ population.
2. The two populations are independent.
3. n_1 and n_2 are large: $Y_i \geq 10$ and $n_i - Y_i \geq 10,\ i = 1, 2$, is a good rule of thumb overall. If $0.3 \leq \hat{p}_i = Y_i/n_i \leq 0.7$, then $Y_i \geq 5$ and $n_i - Y_i \geq 5,\ i = 1, 2$, is a good rule of thumb.

Formulas

Standardized Test Statistic[1]: $Z = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_0(\hat{p}_1 - \hat{p}_2)}$.

Hypotheses:	$H_0: p_1 - p_2 = 0$ $H_{a-}: p_1 - p_2 < 0$	$H_0: p_1 - p_2 = 0$ $H_{a+}: p_1 - p_2 > 0$	$H_0: p_1 - p_2 = 0$ $H_{a\pm}: p_1 - p_2 \neq 0$
p-value[2]:	$p_- = P(N(0,1) \leq z^*)$	$p^+ = P(N(0,1) \geq z^*)$	$p\pm = 2\min(p_-, p^+)$

[1] $\hat{\sigma}_0(\hat{p}_1 - \hat{p}_2) = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$, where $\hat{p} = \frac{Y_1 + Y_2}{n_1 + n_2}$.

[2] z^* is the observed value of Z, $P(N(0,1) \leq z^*)$ is the proportion of a $N(0,1)$ population less than z^*, and $P(N(0,1) \geq z^*)$ is the proportion of a $N(0,1)$ population greater than z^*.

Case 2: A Test of the General Difference of Two Proportions

Assumptions

1. The data are Y_1 from a $b(n_1, p_1)$ population and Y_2 from a $b(n_2, p_2)$ population.
2. The two populations are independent.
3. n_1 and n_2 are large: a guideline is given in Case 1.

Formulas

Test Statistic[1]: $Z = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}(\hat{p}_1 - \hat{p}_2)}$.

Hypotheses[2]:	$H_0: p_1 - p_2 = \delta_0$ $H_{a-}: p_1 - p_2 < \delta_0$	$H_0: p_1 - p_2 = \delta_0$ $H_{a+}: p_1 - p_2 > \delta_0$	$H_0: p_1 - p_2 = \delta_0$ $H_{a\pm}: p_1 - p_2 \neq \delta_0$
p-value[3]:	$p_- = P(N(0,1) \leq z^*)$	$p^+ = P(N(0,1) \geq z^*)$	$p\pm = 2\min(p_-, p^+)$

[1] $\hat{\sigma}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$, where $\hat{p}_1 = Y_1/n_1$ and $\hat{p}_2 = Y_2/n_2$.

[2] $\delta_0 \neq 0$.

[3] z^* is the observed value of Z, $P(N(0,1) \leq z^*)$ is the proportion of a $N(0,1)$ population less than z^*, and $P(N(0,1) \geq z^*)$ is the proportion of a $N(0,1)$ population greater than z^*.