Hypothesis Test Guide

I. Hypothesis Test for a Population Mean

Assumptions

1. Random sample $y_1, \ldots, y_n \sim N(\mu, \sigma^2)$, sample mean $\overline{y}$, sample variance s^2.
2. σ^2 is unknown.

Observed Value of Standardized Test Statistic: $t^* = \frac{\overline{y}-\mu_0}{s/\sqrt{n}}$.

P-values:

H_0	H_a	p-value
$\mu = \mu_0$	$\mu < \mu_0$	$p_- = Pr(t_{n-1} \le t^*)$
$\mu = \mu_0$	$\mu > \mu_0$	$p_+ = Pr(t_{n-1} \ge t^*)$
$\mu = \mu_0$	$\mu \ne \mu_0$	$p\pm = 2\min(p_-, p^+)$

II. Hypothesis Tests for a Population Proportion

Assumptions

1. We are interested in estimating the proportion p having a certain characteristic in the target population.
2. y is the number having the characteristic in a random sample of size n taken from the population.

Case 1: Exact Test

Observed Value of Test Statistic: y^*

P-values:

H_0	H_a	p-value
$p = p_0$	$p < p_0$	$p_- = Pr(b(n, p_0) \le y^*)$
$p = p_0$	$p > p_0$	$p_+ = Pr(b(n, p_0) \ge y^*)$
$p = p_0$	$p \ne p_0$	$p\pm = \sum_{f(y) \le f(y^*)} f(y)$, where $f(y) = \binom{n}{y} p_0^y (1-p_0)^{n-y}$

Case 2: Large Sample Test (with Continuity Correction)

Assumptions $y \ge 10,\ n - y \ge 10$.

Observed Value of Standardized Test Statistic:

$$z_l^* = \frac{y - np_0 + 0.5}{\sqrt{np_0(1-p_0)}},\ z_u^* = \frac{y - np_0 - 0.5}{\sqrt{np_0(1-p_0)}}.$$

P-values:

H_0	H_a	p-value
$p = p_0$	$p < p_0$	$p_- = Pr(N(0,1) \le z_l^*)$
$p = p_0$	$p > p_0$	$p_+ = Pr(N(0,1) \ge z_u^*)$
$p = p_0$	$p \ne p_0$	$p\pm = 2\min(p_-, p^+)$

III. Hypothesis Tests for the Difference of Two Means

In all cases we assume the data are:

$$\text{Population 1: } y_{1,1}, y_{1,2}, \ldots, y_{1,n_1} \sim N(\mu_1, \sigma_1^2), \quad \text{sample mean, } \overline{y}_1, \quad \text{sample variance, } s_1^2$$

$$\text{Population 2: } y_{2,1}, y_{2,2}, \ldots, y_{2,n_2} \sim N(\mu_2, \sigma_2^2), \quad \text{sample mean, } \overline{y}_2, \quad \text{sample variance, } s_2^2$$

We test hypotheses concerning $\mu_1 - \mu_2$.

Case 1: Paired Data

In this case, we take differences $d_i = y_{1,i} - y_{2,i}$, and conduct a one sample hypothesis test for the mean difference μ_D. Since $\mu_D = \mu_1 - \mu_2$, this is also an hypothesis test for $\mu_1 - \mu_2$.

Case 2: Independent Populations, Variances Assumed Equal

Assumption

$\sigma_1^2 = \sigma_2^2$, and their value is unknown.

Observed Value of Standardized Test Statistic:

$$t^{(p)*} = \frac{\overline{y}_1 - \overline{y}_2 - \delta_0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

where

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}.$$

P-values:

H_0	H_a	p-value
$\mu_1 - \mu_2 = \delta_0$	$\mu_1 - \mu_2 < \delta_0$	$p_- = Pr(t_{n_1+n_2-2} \leq t^{(p)*})$
$\mu_1 - \mu_2 = \delta_0$	$\mu_1 - \mu_2 > \delta_0$	$p_+ = Pr(t_{n_1+n_2-2} \geq t^{(p)*})$
$\mu_1 - \mu_2 = \delta_0$	$\mu_1 - \mu_2 \neq \delta_0$	$p\pm = 2\min(p_-, p^+)$

Case 3: Independent Populations, Variances Not Assumed Equal

Observed Value of Standardized Test Statistic:

$$t^{(ap)*} = \frac{\overline{y}_1 - \overline{y}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

P-values:

H_0	H_a	p-value
$\mu_1 - \mu_2 = \delta_0$	$\mu_1 - \mu_2 < \delta_0$	$p_- = Pr(t_\nu \leq t^{(ap)*})$
$\mu_1 - \mu_2 = \delta_0$	$\mu_1 - \mu_2 > \delta_0$	$p_+ = Pr(t_\nu \geq t^{(ap)*})$
$\mu_1 - \mu_2 = \delta_0$	$\mu_1 - \mu_2 \neq \delta_0$	$p\pm = 2\min(p_-, p^+),$

where ν is taken as the largest integer less than or equal to $\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2 \Big/ \left[\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}\right]$.

IV. Hypothesis Tests for the Difference of Two Proportions

Assumptions

1. We are interested in testing $H_0 : p_1 - p_2 = \delta_0$, where p_1 is the proportion having a certain characteristic in population 1 and p_2 is the proportion having a certain characteristic in population 2, and δ_0 is a specified value (usually 0).
2. y_1 is the number having the characteristic in a random sample of size n_1 taken from population 1 and y_2 is the number having the characteristic in a random sample of size n_2 taken from population 2.
3. Sample size is sufficiently large for the CLT to hold. Rule of thumb: y_1, $n_1 - y_1$, y_2, $n_2 - y_2$ all at least 10.

Case 1: $\delta_0 = 0$

Observed Value of Standardized Test Statistic:

$$z_0^* = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

where

$$\hat{p} = \frac{y_1 + y_2}{n_1 + n_2}.$$

P-values:

H_0	H_a	p-value
$p_1 - p_2 = 0$	$p_1 - p_2 < 0$	$p_- = Pr(N(0,1) \leq z_0^*)$
$p_1 - p_2 = 0$	$p_1 - p_2 > 0$	$p_+ = Pr(N(0,1) \geq z_0^*)$
$p_1 - p_2 = 0$	$p_1 - p_2 \neq 0$	$p\pm = 2\min(p_-, p^+)$

Case 2: $\delta_0 \neq 0$

Observed Value of Standardized Test Statistic:

$$z^* = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}},$$

where $\hat{p}_1 = y_1/n_1$ and $\hat{p}_2 = y_2/n_2$.

P-values:

H_0	H_a	p-value
$p_1 - p_2 = \delta_0$	$p_1 - p_2 < \delta_0$	$p_- = Pr(N(0,1) \leq z^*)$
$p_1 - p_2 = \delta_0$	$p_1 - p_2 > \delta_0$	$p_+ = Pr(N(0,1) \geq z^*)$
$p_1 - p_2 = \delta_0$	$p_1 - p_2 \neq \delta_0$	$p\pm = 2\min(p_-, p^+)$