Larges samples tests for two sample means

Let $Y_{11}, \ldots, Y_{1n_1}$ be iid $N(\mu_1, \sigma_1^2)$

Let $Y_{21}, \ldots, Y_{2n_2}$ be iid r.v's $N(\mu_2, \sigma_2^2)$

The two samples are independent. $\bar{Y}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} Y_{1j}$, $\bar{Y}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_{2j}$

(1) To test $H_0: \mu_1 - \mu_2 = 0$ (or $\mu_1 - \mu_2 \le 0$) versus $H_1: \mu_1 - \mu_2 > 0$

we reject H_0 if $\bar{Y}_1 - \bar{Y}_2 \ge z_\alpha \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

data

The p-value of the test is $P\left(N(0,1) \ge \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}\right)$ data

(2) To test $H_0: \mu_1 - \mu_2 = 0$ versus $H_1: \mu_1 - \mu_2 < 0$

we reject H_0 if $\bar{Y}_1 - \bar{Y}_2 \le z_\alpha \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

data

The p-value of the test $P\left(N(0,1) \le \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}\right)$

data

(3) To test $H_0: \mu_1 - \mu_2 = 0$ versus $H_1: \mu_1 \ne \mu_2$

we reject H_0 if $|\bar{Y}_1 - \bar{Y}_2| \ge z_{\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

data

The p-value of the test is $P\left(|N(0,1)| \ge \frac{|\bar{Y}_1 - \bar{Y}_2|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}\right)$

data

10.41 Shear strength measurements derived from unconfined compression tests for two types of soils gave the results shown in the following table

Soil type I	Soil type II
$n_1 = 30$	$n_2 = 35$
$\bar{y}_1 = 1.65$	$\bar{y}_2 = 1.43$
$S_1 = 0.26$	$S_2 = 0.22$

Do the soils appear to differ with respect to average shear strength at the 1% significance level?

$H_0: \mu_1 = \mu_2$ $\quad H_1: \mu_1 \neq \mu_2$

Reject H_0 if $\dfrac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \geq z_{\frac{\alpha}{2}}$

$$\frac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{|1.65 - 1.43|}{\sqrt{\frac{(0.26)^2}{30} + \frac{(0.22)^2}{36}}} = 3.65$$

$z_{0.005} = 2.575$

The conclusion is to reject H_0

The p-value is $P(|N(0,1)| \geq 3.65) = 0.0005$

10.13 Charles Dickey discusses studies of the habits of white-tailed deer that indicate that these deer live and feed within very few limited ranges, approximately 150 to 205 acres. To determine whether the range of deer located in two different geographical areas differ, researchers caught, tagged, and fitted 40 deer with small radio transmitters. Several months later, the deer were tracked and identified and the distance y from the release point was recorded. The mean and standard deviation of the distance from the release point were as given in the following table:

	Location 1	Location 2
Sample size	40	40
Sample mean	2980	3205 ft
Sample st. dev.	1140	963 ft
Population mean	μ_1	μ_2

(a) If you have no preconceived reason for believing one one population mean to be larger than the other, what would you choose for you alternative hypotheses? You null hypothesis?

(b) Do the data provide sufficient evidence to indicate that the mean distances differ for the two geographical locations? Test using $\alpha = 0.10$

(a) $H_0: \mu_1 = \mu_2$ $H_a: \mu_1 \neq \mu_2$

The test statistic is $z = \dfrac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \dfrac{290 - 3205}{\sqrt{\frac{(11401)^2}{40} + \frac{(4631)^2}{40}}} = -0.954$

The rejection region, with $\alpha = 0.10$, is $|z| > 1.645$.

Thus, the null hypothesis is not rejected.

The p-value of the test is

$P(|N(0,1)| \geq 0.954) = 2(0.17) = 0.34$

Tests for two proportions

Let Y_1 be Bino (n_1, p_1). Let $\hat{p}_1 = \frac{Y_1}{n_1}$

Let Y_2 be Binom (n_2, p_2) Let $\hat{p}_2 = \frac{Y_2}{n_2}$ Let $\hat{p} = \frac{Y_1 + Y_2}{n_1 + n_2}$

Assume that Y_1 and Y_2 are independent r.v's

(1) To test $H_0: p_1 = p_2$ versus $H_1: p_1 > p_2$

we reject H_0, if $\hat{p}_1 - \hat{p}_2 \geq z_\alpha \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

The p-value of the test is data

$$P\left(N(0,1) \geq \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right)$$

(2) To test $H_0: p_1 = p_2$ versus $H_1: p_1 < p_2$

we reject H_0, if $\hat{p}_1 - \hat{p}_2 \leq z_\alpha \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

The p-value of the test is $P(N(0,1) \leq \quad)$

(3) To test $H_0: p_1 = p_2$ versus $H_1: p_1 \neq p_2$

we reject H_0, if $|\hat{p}_1 - \hat{p}_2| \geq z_{\frac{\alpha}{2}} \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

The p-value of the test is

$$P\left(|N(0,1)| \geq \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right)$$

10.17 A survey examines enrollment in colleges in 1991 and 1992.
The number of responses in 1991 was 1225
The number of responses in 1992 was 1232.
In 1991, 50% of the colleges reported an increase in applications
In 1992, 71% of the colleges reported an increase in applications
Do the data indicate a significant difference in the proportions of colleges reporting in applications for the years 1991 and 1992?
Use $\alpha = 0.01$

p_1 = proportion of colleges reporting an increasing in applications in 1991
in 1992
p_2 = –

$H_0: p_1 = p_2 \quad H_a: p_1 \neq p_2$

Reject H_0, if $\dfrac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}} \geq z_{\frac{\alpha}{2}}$, $\hat{p} = \dfrac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} = 0.606$

$$\frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\hat{p}_1(1-\hat{p}_1)(\frac{1}{n_1}+\frac{1}{n_2})}} = \frac{0.71 - 0.50}{\sqrt{0.606(1-0.606)(\frac{1}{1232}+\frac{1}{1225})}} = 10.65$$

$z_{\frac{\alpha}{2}} = 2.58$. H_0 is rejected

10.23 A political researcher believes that the fraction p_1 of Republicans strongly in favor of the death penalty is greater than the fraction p_2 of Democrats strongly in favor of the death penalty. He acquired independent random samples of 200 Republicans and 200 Democrats and found 46 Republicans and 34 Democrats strongly favoring the death penalty. Does this evidence provide statistical support for the researcher's belief? Use $\alpha = 0.05$.

$H_0: p_1 = p_2 \quad H_1: p_1 > p_2$

We reject H_0 if $\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \geq z_\alpha$

$\hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{46}{200} = 0.23 \qquad \hat{p}_2 = \dfrac{y_2}{n_2} = \dfrac{34}{200} = 0.17$

$\hat{p} = \dfrac{46+34}{200+200} = 0.2$

$\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \dfrac{0.23 - 0.17}{\sqrt{0.2(0.8)\left(\frac{1}{200}+\frac{1}{200}\right)}} = 1.5 \qquad z_\alpha = 1.645$

H_0 is not rejected. the p-value is

$P(N(0,1) \geq 1.5) = 0.0668$