Confidence Interval for the Difference Between Two Proportions

$1-\alpha$ Confidence Interval for $p_1 - p_2$:

$$\hat{p}_1 - \hat{p}_2 \pm m = \hat{p}_1 - \hat{p}_2 \pm z^* \sigma_{\hat{p}_1 - \hat{p}_2}$$

$$= \hat{p}_1 - \hat{p}_2 \pm z^* \sqrt{\frac{p_1\,(1-p_1)}{n_1} + \frac{p_2\,(1-p_2)}{n_2}}$$

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

where:

n_1= sample size of group 1

p_1= proportion of group 1

$X_1 \sim B(n_1, p_1)$

$\hat{p}_1 = \frac{X_1}{n_1}$

$n_1\hat{p}_1 \geq 10$ $\quad$ $n_1(1-\hat{p}_1) \geq 10$

n_2= sample size of group 2

p_2= proportion of group 2

$X_2 \sim B(n_2, p_2)$

$\hat{p}_2 = \frac{X_2}{n_2}$

$n_2\hat{p}_2 \geq 10$ $\quad$ $n_2(1-\hat{p}_2) \geq 10$

worldometer 5 years ago

WORLD / COUNTRIES / USA / WASHINGTON

Last updated: August 11, 2020, 15:13 GMT

Washington

Population ≈ 8M

Coronavirus Cases:

65,453

Deaths:

1,701

2.60% of cases

7.17% of closed cases

Recovered:

22,012

33.63% of cases

Closed cases: 23,713

WORLD / COUNTRIES / USA / IDAHO

Last updated: August 11, 2020, 15:13 GMT

Idaho

Population ≈ 2M

Coronavirus Cases:

25,100

Deaths:

239

0.95% of cases

2.49% of closed cases

Recovered:

9,341

37.22% of cases

Closed cases: 9,580

Projections

August 11, 2020

worldometer 4 years ago

WORLD / COUNTRIES / USA / WASHINGTON

Last updated: August 10, 2021, 15:31 GMT

Washington
Population ≈ 8M

Coronavirus Cases:

496,673

Deaths:

6,218

1.25% of cases

2.61% of closed cases

Recovered:

232,455

46.80% of cases

Closed cases: 238,673

WORLD / COUNTRIES / USA / IDAHO

Last updated: August 10, 2021, 15:36 GMT

Idaho
Population ≈ 2M

Coronavirus Cases:

205,033

Deaths:

2,224

1.08% of cases

1.85% of closed cases

Recovered:

117,460

57.29% of cases

Closed cases: 119,684

August 10, 2021

Confidence Interval for the Difference in Two COVID-19 Case Rates

n_1= population of WA= 8,000,000
p_1= Covid case rate in WA
$X_1 \sim B(n_1, p_1)$
$\hat{p}_1$=$\frac{X_1}{n_1}$=496,673/8M = 0.062
$n_1\hat{p}_1 \gg 10$ $\quad n_1(1-\hat{p}_1) \gg 10$

n_2= population of ID= 2,000,000
p_2= Covid case rate in ID
$X_2 \sim B(n_2, p_2)$
$\hat{p}_2$=$\frac{X_2}{n_2}$=205,033/2M = 0.103
$n_2\hat{p}_2 \gg 1\mathbf{0}$ $\quad n_2(1-\hat{p}_2) \gg 10$

95% Confidence Interval for $p_1 - p_2$:

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

Confidence Interval for the Difference in Two COVID-19 Death Rates

n_1= closed cases in WA= 238,673
p_1= death rate in WA
$X_1 \sim B(n_1, p_1)$
$\hat{p}_1$=$\frac{X_1}{n_1}$=0.0261
$n_1\hat{p}_1 \gg 10$ $\quad$ $n_1(1-\hat{p}_1) \gg 10$

n_2= closed cases in ID= 119,684
p_2= death rate in ID
$X_2 \sim B(n_2, p_2)$
$\hat{p}_2$=$\frac{X_2}{n_2}$=0.0185
$n_2\hat{p}_2 \gg 10$ $\quad$ $n_2(1-\hat{p}_2) \gg 10$

95% Confidence Interval for $p_1 - p_2$:

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

Choosing Sample Size for a Given Margin of Error

Margin of Error:

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

Often We Use: $n_1 = n_2 = n$

$\tilde{p}_1 = \tilde{p}_2 =$0.5

$$n = 0.5\left(\frac{z^*}{m}\right)^2$$

Round up to the nearest integer

Hypothesis Test for Comparing Two Proportions

H_0: $p_1 = p_2$

H_a: $p_1 \neq p_2$
(or $p_1 > p_2$
or $p_1 < p_2$)

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

$$\approx \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

where:

n_1= sample size of group 1
p_1= proportion of group 1
$X_1 \sim B(n_1, p_1)$
$\hat{p}_1 = \frac{X_1}{n_1}$
$n_1\hat{p}_1 \geq 10$ $\quad$ $n_1(1-\hat{p}_1) \geq 10$

n_2= sample size of group 2
p_2= proportion of group 2
$X_2 \sim B(n_2, p_2)$
$\hat{p}_2 = \frac{X_2}{n_2}$
$n_2\hat{p}_2 \geq 10$ $\quad$ $n_2(1-\hat{p}_2) \geq 10$

CASE : Evaluation of New Drug

- A researcher believes he has found a drug that will effectively prolong the life of AIDS patients.
- The drug is administered to an experimental group of 50 rats (infected with AIDS virus), and the number of surviving rats (beyond 3 months) are compared with a control group of 50 rats, that were not given any treatment.
- Is there enough evidence to support his theory?

Drug Testing Data

	Control Group	Experimental Group
Sample Size (1)	46	48
Number of Survivors	19	25

(1) Some of the animals died of other reasons

Drug Testing Data

	Control Group	Experimental Group
Sample Size (1)	46	48
Number of Survivors	19	25

$$H_0\colon p_1 = p_2$$

$$H_a\colon p_1 \neq p_2$$

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