

Statistics for Engineers (MAT2001)- Lab

Experiment-VII: Two Sample Z-test

1 Two sample Sample Z-test

```
prop.test(c(x1,x2), c(n1,n2), correct=, alternate = ).
```

- x1 and x2 are the number of successes in sample 1 and 2 respectively.
- n1 and n2 are the sample sizes or number of trials.
- correct = TRUE (use a continuity correction factor) or FALSE (do not).
- alternate = "two.sided" (default), "less", or "greater".

The Trial Urban District Assessment (TUDA) is a study sponsored by the government of student achievement in large urban school district. In 2009, 1311 of a random sample of 1900 eighth-graders from Houston performed at or above the basic level in mathematics . In 2011, 1440 of a random sample of 2000 eighth-graders from Houston performed at or above the basic level . (The study reports the proportions).

(A)Is there an increase in the proportion of eighth-graders who performed at or above the basic level in mathematics from 2009 to 2011 at the 5% significance level?

(B) Compute the 95% confidence interval for the difference in proportion of eighth- graders who performed at or above the basic level in mathematics from 2009 to 2011.

(A) Let p1 and p2 be the proportions of eighth-graders that performed at or above the basic level in mathematics in 2011 and 2009,respectively .We wish the test $H_0 : p_1 = p_2$ against $H_1 = p_1 > p_2$

```
prop.test(c(1440, 1311), c(2000, 1900), alternative = "greater",
          correct = FALSE)

##
## 2-sample test for equality of proportions without continuity correction
##
## data:  c(1440, 1311) out of c(2000, 1900)
## X-squared = 4.2197, df = 1, p-value = 0.01998
```

```
## alternative hypothesis: greater
## 95 percent confidence interval:
## 0.005972807 1.000000000
## sample estimates:
## prop 1 prop 2
## 0.72 0.69
```

The p-value = 0.02 < 0.05 so we reject H_0 . Thus, there is evidence that there is an increase from 2009 to 2011 in the proportion of eighth-graders who performed at or above the basic level at the 5% significance level.

(B) $H_0 : p_1 = p_2$ $H_1 = p_1 \neq p_2$

```
prop.test(c(1440, 1311), c(2000, 1900),
          correct = FALSE)

##
## 2-sample test for equality of proportions without continuity correction
##
## data: c(1440, 1311) out of c(2000, 1900)
## X-squared = 4.2197, df = 1, p-value = 0.03996
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## 0.001369833 0.058630167
## sample estimates:
## prop 1 prop 2
## 0.72 0.69
```

Thus, we are 95% confident that the percent of eighth-graders who performed at or above the basic level in mathematics in 2011 is between 0:14% and 5:86% higher than in 2009.

The use of helmet among recreational alpine skiers and snowboarders are generally low. A study from Norway wanted to examine if helmet use reduces the risk of head injury. In the study, they compared the helmet use among skiers and snowboarders that was injured with a control group. The control group consisted of skiers and snowboarders that was uninjured. 96 of 578 people with head injuries used a helmet and 656 of 2992 people in the uninjured group used a helmet. Is helmet use lower among skiers and snowboarders who had head injuries?

$H_0 : p_1 = p_2$ $H_1 = p_1 < p_2$

```
prop.test(c(96, 656), c(578, 2992), alternative = "less",
          correct = FALSE)

##
## 2-sample test for equality of proportions without continuity correction
```

```
##
## data:  c(96, 656) out of c(578, 2992)
## X-squared = 8.2336, df = 1, p-value = 0.002056
## alternative hypothesis: less
## 95 percent confidence interval:
##  -1.00000000 -0.02482216
## sample estimates:
##      prop 1      prop 2
## 0.1660900 0.2192513
```

The $p\text{-value} = 0.0021 < 0.01$ so we have strong evidence that helmet use is lower among skiers and snowboarders who had head injuries compared to uninjured skiers and snowboarders.

In the built-in data set named quine, children from an Australian town is classified by ethnic background, gender, age, learning status and the number of days absent from school.

```
# library(MASS) # load the MASS package
# quine
# table(quine$Eth, quine$Sex)
# table
```

Assuming that the data in quine follows the normal distribution, find the 95% confidence interval estimate of the difference between the female proportion of Aboriginal students and the female proportion of Non-Aboriginal students, each within their own ethnic group.

```
lib <- library(MASS) # load the MASS package
df <- quine
table(df$Eth, df$Sex)

##
##      F  M
##  A 38 31
##  N 42 35

prop.test(table(df$Eth, df$Sex), correct = FALSE)

##
## 2-sample test for equality of proportions without continuity correction
##
## data:  table(df$Eth, df$Sex)
## X-squared = 0.0040803, df = 1, p-value = 0.9491
## alternative hypothesis: two.sided
```

```
## 95 percent confidence interval:  
## -0.1564218  0.1669620  
## sample estimates:  
##      prop 1      prop 2  
## 0.5507246 0.5454545
```

The 95% confidence interval estimate of the difference between the female proportion of Aboriginal students and the female proportion of Non-Aboriginal students is between -15.6% and 16.7%.