**CONFIDENCE INTERVALS & TESTS OF HYPOTHESIS**

1. **Confidence Interval for Mean – Large Sample (Using Standard Normal Distribution)**

Example: Use the “trees” data set in R

Command: > x <- trees$Height

> xbar <- mean(x) ! Find the mean value of the data set

> std <- sd(x) ! Find the standard deviation of the data set

> serror <- std/length(x) ! Find the standard error of the estimation

> conf <- 0.95 ! Define the confidence coefficient

> Left <- xbar - qnorm(1-(1-conf)/2)\*serror

> Right <- xbar + qnorm(1-(1-conf)/2)\*serror

> "Confidence Interval:"; Left; Right

Note: The operations “<-” and “=” are somewhat different

> y <- c("value 1" =100, "value 2" = 50)

> y

Result: value 1 value 2

100 50

> y <- c("value 1" <-100, "value 2" <- 50) !A vector will be introduced

> y

Result: [1] 100 50

1. **Confidence Interval for Mean – Small Sample (Using Student *t* Distribution)**

Example: Use the “trees” data set in R

Command: > x <- trees$Height

> xbar <- mean(x) ! Find the mean value of the data set

> std <- sd(x) ! Find the standard deviation of the data set

> serror <- std/length(x) ! Find the standard error of the estimation

> conf <- 0.95 ! Define the confidence coefficient

> Left <- xbar - qt(1-(1-conf)/2, df = length(x)-1)\*serror

> Right <- xbar + qt(1-(1-conf)/2, df = length(x)-1)\*serror

> "Confidence Interval:"; Left; Right

1. **Confidence Interval for Variance (Using Chi-square  Distribution)**

Example: Use the “trees” data set in R

Command: > x <- trees$Height

> varx <- (sd(x))^2 ! Find the variance of the data set

> conf <- 0.95 ! Define the confidence coefficient

> Left <- (length(x) - 1)\*varx/qchisq(1-(1-conf)/2, df = length(x)-1)

> Right <- (length(x) - 1)\*varx/qchisq((1-conf)/2, df = length(x)-1)

> "Confidence Interval:"; Left; Right

1. **Hypothesis Testing for Mean – Single Sample – Large Sample**

Example: Use the “trees” data set in R

Command: > x <- trees$Height

> muy0 <- 75 ! Define hypothesized value

> xbar <- mean(x) ! Find the mean value of the data set

> std <- sd(x) ! Find the standard deviation of the data set

> z\_value <- (xbar-muy0)/(std/sqrt(length(x))) ! Find the test statistic

> alpha <- 0.05 ! Define Type I error

> Left <- qnorm(alpha/2) ! Critical value on the left tail

> Right <- qnorm(1-alpha/2) ! Critical value on the right tail

> if (z\_value < Left | z\_value > Right) {

print("REJECT")

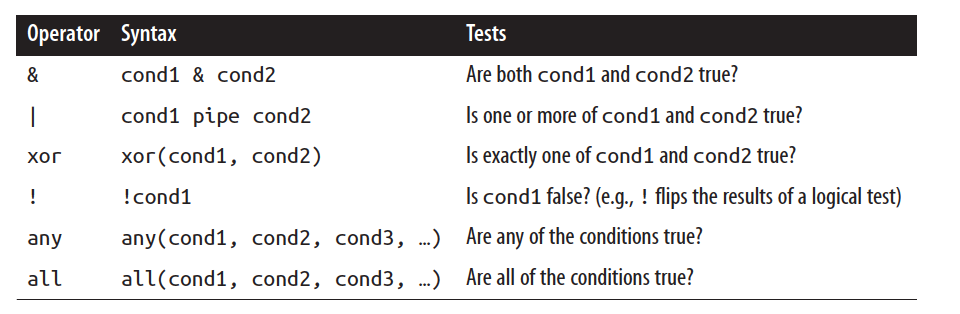
} else {

print("NOT REJECT")

}

Let change the above code for left-tailed test and right-tailed test!

**R’s Boolean Operators**



1. **Hypothesis Testing for Mean – Single Sample – Small Sample**

Let write the code to conduct the test for mean in case of small sample (Hint: use student distribution instead of normal distribution)

Note:

Many software packages only cover the *t*-test and not the *z*-test. In fact, you may not find a *z*-test implementation in some statistical software (including R). This is because the *t*-test results will converge to the z-test results as the sample size gets larger. So, we can use the t.test in R for all cases (i.e, small and large samples). For more details, use ?t.test for help in RStudio

Example: Use t.test for the “trees” data set in R

Command: > t.test(trees$Height, alternative = "two.sided", mu = 75, conf.level = 0.95)

Result:

One Sample t-test

data: trees$Height

t = 0.87381, df = 30, p-value = 0.3892

alternative hypothesis: true mean is not equal to 75

95 percent confidence interval:

73.6628 78.3372

sample estimates:

mean of x

76

Note: alternative = “two-sided” is default, no need to declare! We can use alternative = “less” or “greater”

1. **Hypothesis Testing for Proportion – Single Sample**

Let write the code to conduct the test for proportion (Hint: use standard normal distribution)

Another way: use prop.test in R

Example: Use prop.test for the problem in lecture notes (Slide 24 of Mod 1 – Session 3)

Command: > prop.test(33,200, p =0.08, alternative = "two.sided", conf.level = 0.99)

Result:

1-sample proportions test with continuity correction

data: 33 out of 200, null probability 0.08

X-squared = 18.495, df = 1, p-value = 1.703e-05

alternative hypothesis: true p is not equal to 0.08

99 percent confidence interval:

0.1063495 0.2459364

sample estimates:

p

0.165

Command: > prop.test(33,200, p =0.08, alternative = "two.sided", conf.level = 0.99, correct = FALSE) ! This is the procedure used in lecture

Result:

1-sample proportions test without continuity correction

data: 33 out of 200, null probability 0.08

X-squared = 19.633, df = 1, p-value = 9.383e-06

alternative hypothesis: true p is not equal to 0.08

99 percent confidence interval:

0.1083804 0.2431329

sample estimates:

p

0.165

Note:

* alternative = “two-sided” is default, no need to declare! We can use alternative = “less” or “greater”.
* The default value of conf.level is 0.95 if not declare.
* correct = TRUE is the default value and it must be used when either the expected number of successes or failures is less than < 5

1. **Hypothesis Testing for Mean – Two Samples (all sample sizes!)**

Example 1:

Command: > x <- c(16.03, 16.04, 16.05, 16.05, 16.02, 16.01, 15.96, 15.98, 16.02, 15.99)

> y <- c(16.02, 15.97, 15.96, 16.01, 15.99, 16.03, 16.04, 16.02, 16.01, 16.00)

> t.test(x, y)

Note: Some parameters take the default value in the above command

* alternative = "two.sided"
* mu = 0
* paired = FALSE
* var.equal = FALSE
* conf.level = 0.95

Result:

Welch Two Sample t-test

data: x and y

t = 0.79894, df = 17.493, p-value = 0.435

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.01635123 0.03635123

sample estimates:

mean of x mean of y

16.015 16.005

Example 2:

Command: > x <- c(16.03, 16.04, 16.05, 16.05, 16.02, 16.01, 15.96, 15.98, 16.02, 15.99)

> y <- c(16.02, 15.97, 15.96, 16.01, 15.99, 16.03, 16.04, 16.02, 16.01, 16.00)

> t.test(x, y, var.equal = TRUE)

Result:

Two Sample t-test

data: x and y

t = 0.79894, df = 18, p-value = 0.4347

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.01629652 0.03629652

sample estimates:

mean of x mean of y

16.015 16.005

Example 3:

Command: > x <- c(16.03, 16.04, 16.05, 16.05, 16.02, 16.01, 15.96, 15.98, 16.02, 15.99)

> y <- c(16.02, 15.97, 15.96, 16.01, 15.99, 16.03, 16.04, 16.02, 16.01, 16.00)

> t.test(x, y, paired = TRUE)

Result:

Paired t-test

data: x and y

t = 0.62284, df = 9, p-value = 0.5488

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.02631997 0.04631997

sample estimates:

mean of the differences

0.01

1. **Hypothesis Testing for Proportion – Two Samples**

Example 1: We have two groups of individuals:

* Group A with lung cancer: *n* = 500
* Group B, healthy individuals: *n* = 500

The number of smokers in each group is as follow:

* Group A with lung cancer: *n* = 500, 490 smokers, *p*A= 490/500 = 0.98
* Group B, healthy individuals: *n* = 500, 400 smokers, *p*B= 400/500 = 0.80

We want to know, whether the proportions of smokers are the same in the two groups of individuals?

Command: > prop.test(x = c(490,400), n = c(500,500), correct = FALSE)

Result:

2-sample test for equality of proportions without continuity

correction

data: c(490, 400) out of c(500, 500)

X-squared = 82.737, df = 1, p-value < 2.2e-16

alternative hypothesis: two.sided

95 percent confidence interval:

0.1428536 0.2171464

sample estimates:

prop 1 prop 2

0.98 0.80

Example 2: Problem in lecture (Slide 46 – Module 1\_Session 3)

Command: > prop.test(x = c(430,308), n = c(755,616), correct = FALSE)

Result:

2-sample test for equality of proportions without continuity

correction

data: c(430, 308) out of c(755, 616)

X-squared = 6.5998, df = 1, p-value = 0.0102

alternative hypothesis: two.sided

95 percent confidence interval:

0.0165606 0.1225122

sample estimates:

prop 1 prop 2

0.5695364 0.5000000

Note: correct = TRUE is the default value and it must be used when either the expected successes or failures are less than < 5

1. **Hypothesis Testing for Variance – Two Samples (The F Test)**

Example:

Command: > x <- c(16.03, 16.04, 16.05, 16.05, 16.02, 16.01, 15.96, 15.98, 16.02, 15.99)

> y <- c(16.02, 15.97, 15.96, 16.01, 15.99, 16.03, 16.04, 16.02, 16.01, 16.00)

> var.test(x, y)

Note: Some parameters take the default value in the above command

* alternative = "two.sided"
* ratio = 1
* conf.level = 0.95

Result:

F test to compare two variances

data: x and y

F = 1.4103, num df = 9, denom df = 9, p-value = 0.6168

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.3502877 5.6776841

sample estimates:

ratio of variances

1.410256