Statistical Inference Theory - End Semester Lab

CB.SC.I5DAS20032

Problem 1

Calculate and analyze the correlation coefficient between the number of study hours and the number of sleeping hours of different students. Number of Study Hours:2,4,6,8,10 Number of Sleeping Hours:10,9,8,7,6

Solution

```
study = c(2,4,6,8,10)
sleep = c(10,9,8,7,6)

cor(study, sleep)

## [1] -1
```

Problem 2

A random sample of 16 length values from a normal population showed a mean of 41.5 inches, and the sum of squares of deviations from this mean is equal to 135 inches. Check whether the assumption of a mean of 43.5 inches for the unknown population mean is reasonable from the perspective of 99% confidence levels.

Solution

h0: m = m0 reject if abs(t0) > t(alpha/2, n-1) h1: m!= m0

```
n = 16
xBar = 41.5

dev = 135

m0 = 43.5
a = 0.01

sd = dev/(n - 1)

t0 = (xBar - m0)/(sd/sqrt(n))

t0

## [1] -0.8888889

qt(1 - a/2, df = 15)

## [1] 2.946713
```

As test statistic abs(t0) = 0.8889 falls outside the rejection area, we **do not reject** h0.

Problem 3

Suppose we want to estimate the proportion of families in a town, which have two or more children. A random sample of 144 families shows that 48 families have two or more children. Setup a 95 per cent confidence interval estimate of the population proportion of families having two or more children.

Solution

```
z = qnorm(1 - 0.025)

z

## [1] 1.959964

buff = z*sqrt((48/144)*(96/144)/144)

buff

## [1] 0.07699466

48/144 + c(-buff, buff)

## [1] 0.2563387 0.4103280
```

Problem 4

A random sample of 20 bullets produced by a machine shows an average diameter of 3.5 mm and a SD of 0.2 mm. Assuming that the diameter measurement follows N(μ,σ). Obtain a 95% interval estimate for the true variance.

Solution

```
sd = 0.2
n = 20

c(19*0.004/pchisq(0.025, 19, lower.tail = FALSE), 19*0.04/pchisq(1 - 0.025,
19, lower.tail = FALSE))

## [1] 0.076 0.760
```

Problem 5

In order to promote fairness in grading, each application was graded twice by different graders. Based on the grades, can we see if there is a difference between the two graders? The data is Grader1:3,0,5,2,5,5,5,4,4,5 Grader2:2,1,4,1,4,3,3,2,3,5

Solution

```
g1 = c(3,0,5,2,5,5,5,4,4,5)
g2 = c(2,1,4,1,4,3,3,2,3,5)
```

```r
t.test(g1, g2, paired = TRUE)

##
##  Paired t-test
##
## data:  g1 and g2
## t = 3.3541, df = 9, p-value = 0.008468
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  0.325555 1.674445
## sample estimates:
## mean difference
##               1

qt(1 - 0.05, 8)

## [1] 1.859548
```

As t0 = 3.3541 is greater than 1.859548 and p value is lesser than critical value, we **reject** the null hypothesis that there is no difference between the graders.

Problem 6

A group of 5 patients treated with medicine. A is of weight :42,39,38,60,41 kgs Second group of 7 patients from the same hospital treated with medicine B is of weight :38,42,56,64,68,69,62 kgs Find whether there is any difference between medicines?

Solution

h0: σ1^2 = σ2^2 h1: σ1^2 != σ2^2

```r
a = c(42,39,38,60,41)
b = c(38,42,56,64,68,69,62)

sd(a)/var(b)

## [1] 0.05885281

qf(0.05, 4, 6, lower.tail = TRUE)

## [1] 0.1622552
```

We **reject** the null hypothesis.

Problem 7

Researchers are interested in the mean age of a certain population. A random sample of 10 individuals drawn from the population of interest has a mean of 27. Assuming that the population is approximately normally distributed with variance 20,can we conclude that the mean is different from 30 years ? (α=0.05) .

Solution

h0: m = 30 reject if abs(t0) > t(a/2, n - 1) h1: m != 30

Student's T statistic is taken even though population variance is known due to small sample size.

```
n = 10
xBar = 27
var = 20

t0 = (xBar - 30)/(sqrt(var/n))

t0

## [1] -2.12132

qt(1 - 0.05/2, n-1)

## [1] 2.262157
```

Null hypothesis is **not rejected**.

Problem 8

Joon believes that 50% of first-time brides in the United States are younger than their grooms. She performs a hypothesis test to determine if the percentage is the same or different from 50%. Joon samples 100 first-time brides and 53 reply that they are younger than their grooms. For the hypothesis test, she uses a 1% level of significance.

Solution

h0: m = 0.50 reject of abs(z0) > z(a/2) h1: m != 0.50

```
z = (53 - 50)/sqrt(100*0.53*0.47)

z

## [1] 0.6010829

qnorm(1 - 0.01/2)

## [1] 2.575829
```

We **do not reject** the null hypothesis.

Problem 9

Test at 5% level of significance with an alternative hypothesis that one of the two population variances is greater than the other based on their sample variances equal to 16 and 25 and respective degrees of freedom are 11 and 9.

Solution

```r
25/16 < qf(0.05, 9, 11)

## [1] FALSE
```

We **do reject** the null hypothesis.

Problem 10

Four coins were tossed 160 times and the following results were obtained. Number of heads : 0 1 2 3 4 Observed Frequency :17 52 54 31 6 Under the assumption that the coins are balanced, find expected frequencies of getting 0,1,2,3, or 4 heads and test the goodness of fit.

Solution

```r
obs = c(17, 52, 54, 31, 6)
exp = dbinom(0:4, 4, 0.5)

exp*160

## [1] 10 40 60 40 10

chisq.test(obs, p = exp)

##
##  Chi-squared test for given probabilities
##
## data:  obs
## X-squared = 12.725, df = 4, p-value = 0.0127
```

p value is lesser than critical value. Null hypothesis is **rejected**.