ICMA 151 Statistics for Science I

Academic Year 2023-2024 Trimester 3

Quiz No. 4 (7.5 %)

(Due July 18th by 11.59 PM)

Name Ye Zaw Aung ID 6681078

Show the calculation of all problems. Any answer without calculation details will not be graded.

Problem 1. (20 points) The following data contains scores of students from two national exams when the scores of the two exams are drawn independently.

1 3

```
Exam A:  70 54 90 53 42 81 89 93 82 76 97 64 65 66 44
Exam B: 80 64 63 52 86 88 74 75 76 81 54 68 82 84 77
```

1.1 (2 points) Find the 90% confidence interval for the true population mean of the scores of exam A.

```
a = c(70,54,90,53,42,81,89,93,82,76,97,64,65,66,44)
b = c(80,64,63,52,86,88,74,75,76,81,54,68,82,84,77)
```

1.1.1 Direct Calculation:

```
meon(o) + c(-1,1) + qt(0.05, 14, lower= F) * (sd(o) / sqrt(15))
[1] 63.01830    79.11503
```

1.1.2 Rcode and Result:

```
t.test (a, conf.level = 0.9)
```

1.2 (8 points) Test the null hypothesis $H_0: \mu_A = 80$ against the alternative $H_1: \mu_A \neq 80$ at the significance level 0.01 by direct calculation:

1.2.1 The calculated test statistic (direct calculation) is

(mean(a) - 80) / (sd(a) / sqrt(15))

[1] -1.954977

[1] -1.954977

1.2.2 The p-value (direct calculation) is

2 * pt(abs(-1.954977), 14, lower = F)

[1] 0.07085172

1.2.3 Conclusion is

Since p-value is greater than significance value, we do not reject the null hypothesis

1.2.4 Repeat your analysis using an appropriate R function

t.test(a, mu = 80)

1.3 (2 points) Find the 99% confidence interval for the difference between the true population mean of scores exam A and the true population mean of scores of exam B.

1.3.1 Direct Calculation:

mean(a) - mean(b) + c(-1,1) * qt(0.005, 28, lower = F) * sqrt(var(a)/15 + var(b)/15)

[1] -17.45852 12.39186

1.3.2 Rcode and Result:

t.test(a, b, conf.level = 0.99)

99 percent confidence interval: -17.66195 12.59529

1.4 (8 points) Test the null hypothesis $H_0: \mu_A = \mu_B$ against the alternative $H_1: \mu_A \neq \mu_B$ at the significance level 0.01 by direct calculation:

1.4.1 The calculated test statistic (direct calculation) is

(mean(a) - mean(b)) / sqrt((var(a)/15) + (var(b)/15))

= - 0.4690235

1.4.2 The p-value (direct calculation) is

2 x pt(abs(-0.4690235), 28, lower = F)

[1] 0.6674126

[1] 0.64684.

1.4.3 Conclusion is

Since p-value is greater than significance level, we fail to reject the hypothesis

1.4.4 Repeat your analysis using an appropriate R function

t.test(a, b, conf.level = 0.99)

Problem 2 (10 points) Randomly select 6 individuals who stop using cigarettes and measure their weights before and after quitting smoking. The observations are as follows.

ID	1	2	3	4	5	6
Weights before quitting smoking	152	167	153	118	116	122
Weights after quitting smoking	156	165	151	122	120	125

2.1 (2 points) Find a 99% confidence interval for the mean difference of weight loss.

2.1.1 Direct Calculation:

before = c (152, 167, 153, 118, 116, 122)
after = c (156, 165, 151, 122, 120, 125)
d = before - after.

CI = mean (d) + c(-1,1) * qt (abs (0.005, 5 lower = F) * sd (d) / sqrt(n)

2.1.2 Rcode and Result:

t.test (before, after, paired = T, conf.level = 0.99).

2.2 (8 points) Test the hypothesis that quitting smoking affects weight loss at the significance level of 0.05.

2.2.1 Null hypothesis is H_0 $u_A = u_B$.

2.2.2 Alternative hypothesis is $H_1 : u_A \neq u_B$.

2.2.3 The test statistic (direct calculation) is

mean(d) /(sd(d) / sqrt(6))

[1] -1.49969

2.2.4 The p-value is

2*pt(abs(-1.49969), 5, lower = F).

0.1939809

2.2.5 Conclusion is

Since p-value is greater than significance level, we fail to reject the null hypothesis.

2.2.6 Repeat your analysis using an appropriate R function

t.test(before, after., paired = T., mu =0).

Problem 3 (10 points) To compare the time spent in producing one type of product by 5 machines, randomly select a sample size of 4 from each machine and record the time spent by each machine. The observations are as follows.

```
##    Machine.1 Machine.2 Machine.3 Machine.4 Machine.5
## 1       19.0      16.2      17.6      17.4      16.5
## 2       18.3      15.6      16.8      18.3      16.3
## 3       18.4      15.3      17.2      16.6      16.8
## 4       18.1      16.9      14.1      17.0      16.8
```

Perform an analysis of variance at the significance level of 0.05 and test if the producing times among different machines are different.

3.1 The null hypothesis is

$H_0: u_1 = u_2 = u_3 = u_4$

3.2 The alternative hypothesis is

H_1: At least one of them is different.

3.3 Use R to construct an ANOVA table.

Rcode:

```
values = c(Machine.1, Machine.2, Machine.3, Machine.4, Mahine5
keys = rep(1:5, each = 4)
summary (aov(values ~ as.factor(keys))
```

Source	SS	DF	MS	F	P-value
Group	14.76	4	3.691	4.923	0.00976
Error	11.24	15	0.750		
Total	26.	19			

3.4 Draw a conclusion from the test.

Since p-value is less than the significance level, we reject the null hypothesis

Problem 4. (20 points) In a study on the sugar fructose changing as an effect of temperature variation, the results are as follows:

temperatures (x) (unit : °C)	The changes of Sugar fructose (y) (unit : % of weights)
1.0	10.03
1.1	10.03
1.2	10.22
1.3	10.71
1.4	10.54
1.5	10.80
1.6	10.76
1.7	11.15
1.8	11.10
1.9	12.42
2.0	12.44

4.1 (2 points) Construct a scatterplot of the changes of Sugar fructose (y) versus temperature (x).

x = c(1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0)

Rcode: y = c(10.03, 10.03, 10.22, 10.71, 10.54, 10.80, 10.76, 11.15, 11.1, 12.42, 12.44)

Result: plot(x, y)

4.2 (3 points) Compute the estimates of slope and intercept by using mathematical formula.

$$b_1 = \frac{\Sigma_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\Sigma_{i=1}^{n} (x_i - \bar{x})^2}$$

(slope)

b_0 (intercept) $= \bar{y} - b_1 \bar{x}$

sum((x - mean(x)) * (y - mean(y))) / sum((x-mean(x))^2)

mean(y) - b_1 * mean(x)

$b_1 = 2.304545$

$b_0 = 7.470455$

4.3 Compute the least-squares line for predicting changes of Sugar fructose from temperature.

The equation of the fitted regression line is

Rcode: $\hat{y} = b_0 + b_1 x$

plot(x,y)

abline (7.470455,2.304545)

Result:

4.4 (1 point) Predict the changes of Sugar fructose at a temperature of 2.3 °C.

when $x = 2.3^{\circ}C$,

$$\hat{y} = 7.470455 + 2.304545 (2.3).$$

$$= 12.77091$$

4.5 (2 points) Compute the coefficient of determination R^2 between temperature and the changes of Sugar fructose and interpret result.

summary (lm (y~x))

Multiple R-squared: 0.8408.

4.6 (2 points) Compute the coefficient of correlation and interpret the result.

r = sqrt (0.8408)

= 0.9169515

4.7 (2 points) Test the null hypothesis if the intercept is zero at the significant level 0.05

summary(lm(y~x))

p-value = 1.44e-07

since p-value is lesser than significance level, we reject the null hypothesis.

4.8 (2 points) Test the null hypothesis if the slope is zero at the significant level 0.05

summary(lm(y~x))

p-value = 7.10e-05

Sine p-value is lesser than significance level, we reject the null hypothesis.

4.9 (4 points) Test the null hypothesis if the intercept is 3 at the significant level 0.05

t_cal = (7.470455 - 3) / 0.5123

p-value = 2*pt(abs(t_cal), 9, lower = F)

= 1.098388e-05

Since p-value < 0.05, we reject the null hypothesis.