20INMCAL204-Assignment

Jayaraj J Pillai[a,1,1,*], Siju K.S[b,2], Jobin Jose[b,2]

[a]*Department of Computer Applications RB402*
[b]*Department of Mathematics AB304*

Abstract

Contents

*Corresponding author
Email addresses: jayaraj.inmca2025@saintgits.org (Zameel Kb), siju.swamy@saintgits.org (Siju K.S), jobin.jose@saintgits.org (Jobin Jose)
[1]Student, 20INMCAL204.
[2]Course Faculty, 20INMCAL204.

1. Question-1

Case Calculate the coefficient of correlation for the ages of husbands and wives:

Ageof Husband(Y ears)	23	27	28	29	30	31	33	35	36	39
Ageof Wife(Y ears)	18	22	23	24	25	26	28	29	30	32

1.1. R-code

```
# loading data
age_husband=c(23, 27, 28, 29, 30, 31, 33, 35, 36, 39)
age_wife=c(18, 22, 23, 24, 25, 26, 28, 29, 30, 32)
```

1.2. Calculating correlation coefficient

```
resp2=cor.test(age_husband,age_wife,method="pearson")
resp2
```

```
##
##  Pearson's product-moment correlation
##
## data:  age_husband and age_wife
## t = 29.737, df = 8, p-value = 1.773e-09
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9803815 0.9989771
## sample estimates:
##       cor
## 0.9955072
```

1.3. Result

The coefficient of correlation for the ages of husbands and wives is 0.9955072. Since it's greater than 0.7 ,then ages are highly positively correlated.

2. Question-2

Case The rank of the same 15 students in two subjects Statistics and Accountancy are given below, the two numbers within the brackets denoting the ranks of the same students in Statistics and Accountancy respectively. (0,10),(2,7),(3,2),(4,6),(5,4),(6,8),(7,3),(8,1),(9,11),(10,15),(11,9),(12,5),(13,14),(14,12),(15,13). Use Spearman's formula to find the rank correlation coefficient

2.1. R-code

```
# loading data
statistics=c(0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15)
accountancy=c(10, 7, 2, 6, 4, 8, 3, 1, 11, 15, 9, 5, 14, 12, 13)
```

2.2. Calculating correlation coefficient

```
res3=cor.test(statistics,accountancy,method="spearman")
res3
```

```
##
##  Spearman's rank correlation rho
##
## data:  statistics and accountancy
## S = 272, p-value = 0.05243
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.5142857
```

2.3. Result

The rank correlation coefficient between the two subjects Statistics and Accountancy is 0.5142857.Since it's less than 0.7 ,then ages are less positively correlated.

3. Question-3

Case From the following data, obtain the two regression equations

Sales :	91	97	108	121	67	124	51	73	111	57
Purchases :	71	75	69	97	70	91	39	61	80	47

3.1. R-code

```
#loading data
Sales=c(91,97,108,121,67,124,51,73,111,57)
Purchases=c(71,75,69,97,70,91,39,61,80,47)
```

```
marketing=data.frame(sales=c(91,97,108,121,67,124,51,73,111,57),purchases=c(71,75,69,97,70,91,39,61,8
```

```
model=lm(marketing$sales~marketing$purchases )
model
```

```
##
## Call:
## lm(formula = marketing$sales ~ marketing$purchases)
##
## Coefficients:
##         (Intercept)  marketing$purchases
##              -5.188                1.360
```

```
model=lm(marketing$purchases~marketing$sales )
model
```

```
##
## Call:
## lm(formula = marketing$purchases ~ marketing$sales)
##
## Coefficients:
##     (Intercept)  marketing$sales
##         14.8113           0.6132
```

3.2. Results

purchases=14.8113 + 0.6132(sales)

Sales=-5.188+-5.188(purchases)

4. Question-4

CaseThe production manager of a company maintains that the flow time in days (y), depends on the number of operations (x) to be performed. The following data give the necessary information:

x	2	2	3	4	4	5	6	6	7	7
y	8	13	14	11	20	10	22	26	22	25

Calculate the value of the Karl Pearson's Correlation Coefficient.

4.1. R-code

```
# loading data
x=c(2, 2, 3, 4, 4, 5, 6, 6, 7, 7)
y=c(8, 13, 14, 11, 20, 10, 22, 26, 22, 25)
```

4.2. Calculating correlation coefficient

```
res=cor.test(x,y,method="pearson")
res
```

```
##
##  Pearson's product-moment correlation
##
## data:  x and y
## t = 3.543, df = 8, p-value = 0.007586
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2990255 0.9457651
## sample estimates:
##       cor
## 0.781513
```

4.3. Result

The value of the Karl Pearson's correlation coefficient is 0.781513.Since it's greater than 0.7 ,than ages are highly positively correlated.

5. Question-5

CaseCalculate Spearman's rank correlation coefficient between advertisement cost and sales from the following data

Advertisement Cost ('000 Rs)	39	65	62	90	82	75	25	98	36	78
Sales (Lakhs Rs.)	47	53	58	86	62	68	60	91	51	84

5.1. R-code

```r
# loading data
Advertisement=c(9, 65, 62, 90, 82, 75, 25, 98, 36, 78)
Sales=c(47, 53, 58, 86, 62, 68, 60, 91, 51, 84)
```

5.2. Calculating correlation coefficient

```r
ress3=cor.test(Advertisement,Sales,method="spearman")
ress3
```

```
##
##  Spearman's rank correlation rho
##
## data:  Advertisement and Sales
## S = 20, p-value = 0.001977
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.8787879
```

5.3. Result

The Spearman's rank correlation coefficient between advertisement cost and sales is 0.8787879.

6. Question-6

Case From the data given below, find the two regression equations

Marks in Economics	25	28	35	32	31	36	29	38	34	32
Marks in Statistics	43	46	49	41	36	32	31	30	33	39

6.1. R-code

```r
#loading data
Economics=c(25, 28, 35, 32, 31, 36, 29, 38, 34, 32)
Statistics=c(43, 46, 49, 41, 36, 32, 31, 30, 33, 39)
```

```r
marks=data.frame(Economics=c(25, 28, 35, 32, 31, 36, 29, 38, 34, 32),Statistics=c(43, 46, 49, 41, 36, 32, 31, 30, 33, 3
```

```r
model=lm(marks$Economics~marks$Statistics)
model
```

```
##
## Call:
## lm(formula = marks$Economics ~ marks$Statistics)
##
## Coefficients:
##       (Intercept)  marks$Statistics
##           40.8794           -0.2337
```

```r
model=lm(marks$Statistics~marks$Economics)
model
```

```
 ##
## Call:
## lm(formula = marks$Statistics ~ marks$Economics)
 ##
## Coefficients:
##      (Intercept)  marks$Economics
##          59.2571           -0.6643
```

6.2. Result

p-value of the F-statistic is 0.2599.

7. Question-7

Case A die was thrown 9000 times and of these 3220 yielded a 3 or 4.Can the die be regarded as unbiased?

7.1. R-code

```r
prop.test(x=3220, n=9000, p=1/3, alternative = "two.sided")
```

```
##
 ## 1-sample proportions test with continuity correction
 ##
 ##data:  3220 out of 9000, null probability 1/3
 ##X-squared = 24.09, df = 1, p-value = 9.193e-07
 ##alternative hypothesis: true p is not equal to 0.3333333
 ##95 percent confidence interval:
##  0.3478820 0.3677956
## sample estimates:
 ##         p
 ##0.3577778
```

7.2. Result and discussion

Since the p-value is less than 0.05, the null hypothesis is rejected at 5 percent significance level. So it is reasonable to conclude that the coin is biased.

8. Question-8

Case In a random sample of 400 persons from a large population, 120 are females. Can it be said that males and females are in the ratio 5:3 in the population? Use 1% significance?

8.1. R-code

```
prop.test(x=120,n=420,p=3/8,alternative = "two.sided")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  120 out of 420, null probability 3/8
## X-squared = 13.907, df = 1, p-value = 0.0001921
## alternative hypothesis: true p is not equal to 0.375
## 95 percent confidence interval:
##  0.2434784 0.3319409
## sample estimates:
##         p
## 0.2857143
```

8.2. Result

Since the p-value is less than 0.05, the null hypothesis is rejected at 5 percent significance level. So it is reasonable to conclude that the ratio of male to female is 5:3.

9. Question-9

Case The adjoining table shows the number of motor registrations in a certain territory for a term of 5 years and the sale of motor tyres by a firm in that territory for the same period.

Years	Motor registration	No.of tyres sold
1	600	1250
2	630	1100
3	720	1300
4	750	1350
5	800	1500

Find the regression equation to estimate the sale of Tyres when motor registration is known. Estimate sale of Tyres when registration is 850.

9.1. R-code

```
marketing=data.frame(X=c(600,630,720,750,800),Y=c(1250,1100,1300,1350,1500))
```

```
cor.test(marketing$X,marketing$Y)
```

```
##
##  Pearson's product-moment correlation
##
## data:  marketing$X and marketing$Y
## t = 2.8396, df = 3, p-value = 0.06567
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.1156646  0.9901769
## sample estimates:
##       cor
## 0.8537215
```

```r
model <- lm(Y ~ X , data = marketing)
summary(model)$coef
```

```
##               Estimate  Std. Error  t value    Pr(>|t|)
## (Intercept) 255.035971 370.0726639 0.6891511 0.54024551
## X             1.492806   0.5257011 2.8396474 0.06566563
```

9.2. Result

The p-value of the F-statistic is 0.06567.

10. Question-10

Case In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers? Justify your answer statistically.

10.1. R-code

```r
prop.test(x=325,n=600,p=1/2,alternative = "greater")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  325 out of 600, null probability 1/2
## X-squared = 4.0017, df = 1, p-value = 0.02273
## alternative hypothesis: true p is greater than 0.5
## 95 percent confidence interval:
##  0.5072614 1.0000000
## sample estimates:
##         p
## 0.5416667
```

10.2. Result

Since the p-value is less than 0.05, the null hypothesis is rejected at 5 percent significance level. So it is reasonable to conclude that the ratio of smokers to non smokers is greater than 50%.

11. Question-11

Case The mean height of 500 male students who showed above average participation in college athletics was 75.2 inches with a standard deviation of 2.3 inches; while 500 male students who showed no interest in such participation had mean height of 68.5 inches with a standard deviation 2.9 inches. Test the hypothesis that male students who participate in college athletics are taller than other male students.

11.1. R-code

```
library(BSDA)
```

```
## Loading required package: lattice
```

```
##
## Attaching package: "BSDA"
```

```
## The following object is masked from "package:datasets":
##
##     Orange
```

```
zsum.test(mean.x = 75.2, n.x = 500, sigma.x = 75.2,
mean.y = 68.5, n.y = 500, sigma.y = 2.9, alternative = "greater")
```

```
##
##  Two-sample z-Test
##
## data:  Summarized x and y
## z = 1.9908, p-value = 0.02325
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  1.164169       NA
## sample estimates:
## mean of x mean of y
##      75.2      68.5
```

11.2. Result

Since the p-value is less than 0.05, the null hypothesis is rejected at 5 percent significance level. So it is reasonable to conclude that the mean height of male students shows above average participation are significantly taller.

12. Question-12

Case Certain pesticide is packed into bags by a machine. A random sample of 10 bags drawn and their contents are found to weigh (in kg.) as follows:

50, 49, 52, 44, 45, 48, 46, 45, 49, 45.
Test if the average packing can be taken to be 50 kg.

12.1. R-code

```
weight=c(50, 49, 52, 44, 45, 48, 46, 45, 49, 45)
```

```
data=t.test(weight, mu=50)
data
```

```
##
##  One Sample t-test
##
## data:  weight
## t = -3.1993, df = 9, p-value = 0.01084
## alternative hypothesis: true mean is not equal to 50
## 95 percent confidence interval:
##  45.39089 49.20911
## sample estimates:
## mean of x
##      47.3
```

12.2. Results

The p-value of the test is 0.01084

13. Question-13

CaseIn two large populations there are 30% and 25% respectively of fair haired people. Is this difference likely be hidden in samples of 1200 and 900 respectively from the two populations.

13.1. R-code

```
prop.test(x=c(360,225),n=c(1200,900),alternative = "two.sided")
```

```
##
##  2-sample test for equality of proportions with continuity correction##
## data:  c(360, 225) out of c(1200, 900)
## X-squared = 6.1512, df = 1, p-value = 0.01313
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.01065384 0.08934616
## sample estimates:
## prop 1 prop 2
##    0.30    0.25
```

13.2. Result

Since the p-value is less than 0.05, the null hypothesis is rejected at 5 percent significance level. So it is reasonable to conclude that difference in proportion is significant.

14. Question-14

Case It is claimed that a random sample of 100 tyres with a mean life of 15269 kms is drawn from a population of tyres which has a mean life of 15200 kms and a standard deviation of 1248 kms. Test the validity of the claim at (i). 5% and (ii) 1% level of significance.

14.1. R-code

```
zsum.test(mean.x = 15200, n.x = 100, sigma.x = 1248, mu=15269, alternative = "less", conf.level = 0.95)
```

```
##
##  One-sample z-Test
##
## data:  Summarized x
## z = -0.55288, p-value = 0.2902
## alternative hypothesis: true mean is less than 15269
## 95 percent confidence interval:
##        NA 15405.28
## sample estimates:
## mean of x
##     15200
```

```
zsum.test(mean.x = 15200, n.x = 100, sigma.x = 1248, mu=15269, alternative = "less", conf.level = 0.99)
```

```
##
##  One-sample z-Test
##
## data:  Summarized x
## z = -0.55288, p-value = 0.2902
## alternative hypothesis: true mean is less than 15269
## 99 percent confidence interval:
##        NA 15490.33
## sample estimates:
## mean of x
##     15200
```

14.2. Result

Since the p-value is greater than 0.05, the null hypothesis is not rejected at 5 percent and 1 percent significance level. So it is reasonable to conclude that the company's claim is not valid.

15. Question-15

Case Certain pesticide is packed into bags by a machine. A random sample of 10 bags drawn and their contents are found to weigh (in kg.) as follows:

50, 49, 52, 44, 45, 48, 46, 45, 49, 45.
Test if the average packing can be taken to be 50 kg.

15.1. R-code

```
weight=c(50, 49, 52, 44, 45, 48, 46, 45, 49, 45)
```

```
data=t.test(weight, mu=50)
data
```

```
##
##  One Sample t-test
##
## data:  weight
## t = -3.1993, df = 9, p-value = 0.01084
## alternative hypothesis: true mean is not equal to 50
## 95 percent confidence interval:
##  45.39089 49.20911
## sample estimates:
## mean of x
##      47.3
```

15.2. Results

The p-value of the test is 0.01084.

16. Question-16

Case A sample of 400 students of MCA and 400 students of M.Tech classes was taken to know their opinion about autonomous colleges. 290 of the MCA and 310 of the M.Tech students favored the autonomous status. Present these facts in the form of a table and test at 5% level, that the opinion regarding autonomous status of colleges is independent of the level of classes of students.

16.1. R-code

```
Fav=c(290,310)
Agnst=c(110,90)
tab1=data.frame(Fav,Agnst)
```

```
chisq.test(tab1)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tab1
## X-squared = 2.4067, df = 1, p-value = 0.1208
```

16.2. Result

Since the p-value is greater than 0.05, the null hypothesis is not rejected at 5 percent significance level. So it is reasonable to conclude that the opinion is not independent of the class of students.

17. Question-17

Case An IQ test was administered to 5 persons before and after they were trained. The results are given below:

Candidates	1	2	3	4	5
IQbef oretraining	110	120	123	132	125
IQaf tertraining	120	118	125	136	121

Test whether there is any change in IQ after the training programme.

```r
before_training=c(110, 120, 123, 132, 125)
after_training=c(120, 118, 125, 136, 121)
```

```r
data=t.test(x=before_training, y=after_training, paired=TRUE)
data
```

```
##
##  Paired t-test
##
## data:  before_training and after_training
## t = -0.8165, df = 4, p-value = 0.4601
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -8.800874  4.800874
## sample estimates:
## mean difference
##              -2
```

17.1. Results

p-value is less than the significance level 0.05 , therefore we will reject the null hypothesis.

18. Question-18

Case The number of scooter accidents per month in a certain town were as follows:

12, 8, 20, 2, 14, 10, 15, 6, 9, 4

Are these frequencies in agreement with the belief that accident conditions were the same during the ten month period?

18.1. R-code

```r
observed <- c(12, 8, 20, 2, 14, 10, 15, 6, 9, 4)
expected <- c(.1, .1, .1, .1, .1, .1, .1, .1, .1, .1)
```

```r
#perform Chi-Square Goodness of Fit Test
reschgf=chisq.test(x=observed, p=expected)
reschgf
```

```
##
##  Chi-squared test for given probabilities
##
## data:  observed
## X-squared = 26.6, df = 9, p-value = 0.001628
```

18.2. Results & discussions

Since the p-value =0.001628 is less than 0.05, we reject the null hypothesis.

19. Question-19

Case In an experiment on the immunization of humans from Covid 19, the following results were obtained. Find the expected frequency from this cross table.

Status	Died	Survived
Vaccinated	2	10
Not vaccinated	6	6

19.1. R-code

```
vaccinated=c(2,10)
not_vaccinated=c(6,6)
```

```
data=t.test(x=vaccinated,y=not_vaccinated,paired = TRUE)
data
```

```
##
##  Paired t-test
##
## data:  vaccinated and not_vaccinated
## t = 0, df = 1, p-value = 1
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -50.82482  50.82482
## sample estimates:
## mean difference
##               0
```

19.2. Results & discussions

Since p-test is greater than significance level 0.05, we fail to reject the null hypothesis.

20. Question-20

Case Novartis Pvt.Ltd has the head office in Kolkata and a branch at Mumbai. The personal director wanted to know if the workers at the two places would like the introduction of work at home plan and a survey was conducted for this purpose. Out of a sample of 500 workers at Kolkata, 62% favours the new plan. At Mumbai, out of a sample of 400 workers, 41% were against the new plan. Is there any significant difference between the two groups in their attitude towards the new plan at 5% level?

20.1. R-code

```
n1=500
n2=400
x1=310
x2=236
```

```
Fav=c(x1,x2)
Agnst=c(n1-x1,n2-x2)
tab2=data.frame(Fav,Agnst)
tab2
```

```
##  Fav Agnst
## 1 310  190
## 2 236  164
```

```
prop.test(x=c(x1,x2),n=c(n1,n2),p=NULL,alternative="two.sided",conf.level=0.95,correct=    TRUE)
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(x1, x2) out of c(n1, n2)
## X-squared = 0.71714, df = 1, p-value = 0.3971
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.03654011  0.09654011
## sample estimates:
## prop 1 prop 2
##   0.62   0.59
```

20.2. Result

Since the p-value(0.3971) is greater than 0.05 ,null hypothesis is not rejected at 5% level.

21. Question-22

Case The following table show the distribution of digits in number chosen at random from a telephone directory.

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	1026	1107	997	966	1075	933	1107	972	964	853

Test whether the digits may be taken to occur equally frequently in the directory at 5% significance level.

21.1. R-code

```
O=c(1026, 1107, 997, 966, 1075, 933, 1107, 972, 964, 853)
E=c(0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)
chisq.test(O, p=E)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  O
## X-squared = 58.542, df = 9, p-value = 2.558e-09
```

21.2. Result

Since the p-value is less than 0.05, the null hypothesis is rejected at 5 percent significance level. So it is reasonable to conclude that the number of occurrence is independent

22. Question-24

Case The mean return of two stocks in NSE-50 and their risk (variability) are as given below.Examine whether the difference in the variability in risk is significant at 5% level of significance.

Specification	*Returnfor40weeks*	*Returnfor60weeks*
Mean	1258	1243
Risk(SD)	34	28

22.1. R-code

```
x <- rnorm(n=40,mean=1258, sd = 34)y
<- rnorm(n=60,mean=1243, sd = 28) #
var test in R
var.test(x, y, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  x and y
## F = 2.4522, num df = 39, denom df = 59, p-value = 0.001803
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.397974 4.452717
## sample estimates:
## ratio of variances
##           2.452213
```

22.2. Result

Since the p-value is greater than 0.05, the null hypothesis is not rejected at 5 percent significance level. So it is reasonable to conclude that the variability in the risk is significant.