Institute of Actuaries of India

Subject CS1-Actuarial Statistics (Paper B)

July 2022 Examination

INDICATIVE SOLUTION

Introduction

The indicative solution has been written by the Examiners with the aim of helping candidates. The solutions given are only indicative. It is realized that there could be other points as valid answers and examiner have given credit for any alternative approach or interpretation which they consider to be reasonable.

Solution 1:

i) X ~ Poisson (15)
ppois(10, 15)
[1] 0.1184644

[4]

ii) X ~ NB(5,0.2)
dnbinom(65,5,0.2)
[1] 0.00013892

[5]

iii) X ~ Binom (100,0.5)
qbinom(0.9, 100, 0.5)
[1] 56

[4]

iv) X ~ Geometric (0.2)
dgeom(x=3, prob=0.2)
[1] 0.1024

[4]

v) X ~ Exp(1/1000)
1 - pexp(2000, 0.001)
[1] 0.1353353

[5]

vi) X ~ N(6,16)
pnorm(3, 6, 4) - pnorm(-1, 6, 4)
[1] 0.1865682

[4]

[26 Marks]

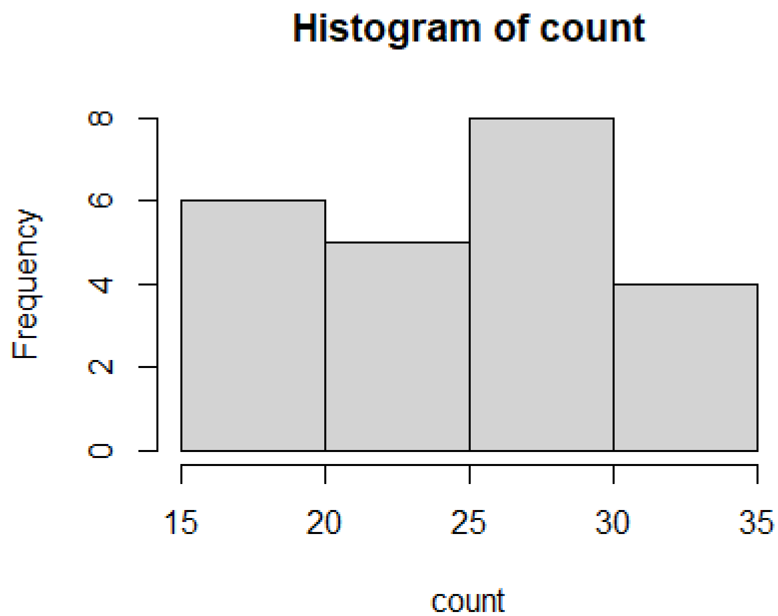
Solution 2:

i) count <- c(31,29,19,18,31,28, 34,27,34,30,16,18, 26,27,27,18,24,22, 28,24,21,17,24)

```
> quantile(count,0.25)

25%

 20

> quantile(count,0.75)

 75%

28.5

> IQR(count)
[1] 8.5
```

[4]

ii) hist(count)



[3]

iii) lambda.hat=mean(x)

print(lambda.hat)
[1] 24.91304

[3]

iv) Ho: The mean fiber count is 25
H1: Mean fiber count is not equal to 25

```
> t.test(count,mu=25)

        One Sample t-test

data:  count
t = -0.076034, df = 22, p-value = 0.9401
alternative hypothesis: true mean is not equal to 25
95 percent confidence interval:
 22.54124 27.28485
sample estimates:
mean of x
 24.91304
```

Based on the p-value the null hypothesis Ho that "the mean fiber count is 25" cannot be rejected. Also 25 lies within the 95% confidence interval.

[5]

v) lambda.hat.sterror=sqrt(lambda.hat/length(x))

print(lambda.hat.sterror)

[1] 1.040757

[2]

vi) lambda.CI.Limits=lambda.hat + c(-1,1)*qnorm(.95)*lambda.hat.sterror

print(lambda.CI.Limits)

[1] 23.20115 26.62494

[3]

vii) > pnorm(30,lambda.hat,sqrt(lambda.hat),lower.tail = FALSE)

[1] 0.1540622

[4]

[24 Marks]

Solution 3:

i)

H0: Annual average rainfall of Belgium and Iran are same

H1: Annual average rainfall of Belgium and Iran are not same

Iran <- c(128,125,133,104,146,132,125,118,129,124)

Belgium <- c(160,128,169,105,151,164,162,177,185,150,182,158,156,123,141,176,162,172)

var.test(Iran, Belgium)

F test to compare two variances

data: Iran and Belgium

F = 0.25802, num df = 9, denom df = 17, p-value = 0.04385

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

95 percent confidence interval:

0.0864436 0.9602591

sample estimates:

ratio of variances

0.258022

Since p-value < 0.05 we fail the variance test thus we reject the null hypothesis that both have equal variance

[6]

ii)

t.test(Iran, Belgium, var.equal = FALSE)

OUTPUT

data: Iran and Belgium

t = -4.9984, df = 25.904, p-value = 3.407e-05

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

95 percent confidence interval:

-42.79403 -17.85041

sample estimates:

mean of x mean of y

126.4000 156.7222

Since, P-value<0.05 we reject the null hypothesis and can conclude that both cities have different amount of rainfall with 95% confidence.

[6]

iii) Confidence interval can be read from part b

95 percent confidence interval:
-42.79403 -17.85041

[2]

iv) The confidence interval (-42.8,-17.8) does not contain 0, therefore the assumption of equal means is not true. This result is in line with the conclusion in part (b).

[3]

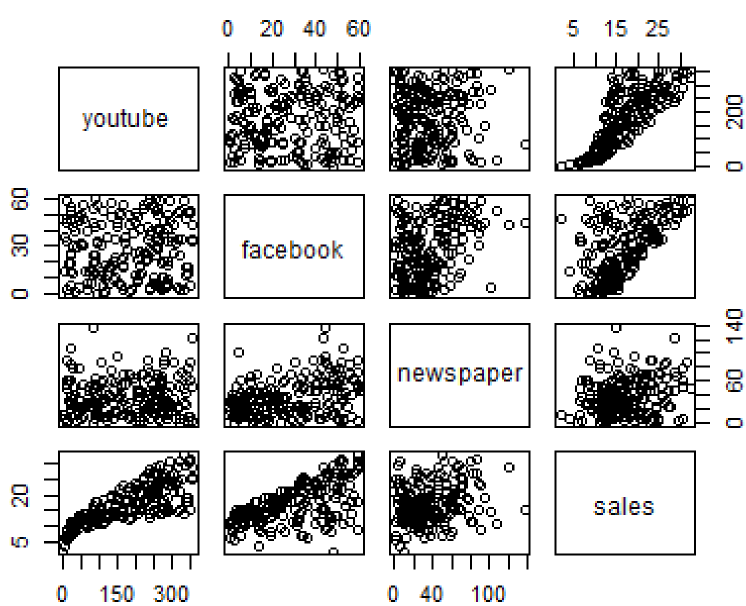
[17 Marks]

Solution 4:

marketing = read.csv("data.csv")

data_size = dim(marketing)

i) plot(marketing)



The last row of the plot indicates how various advertising channel budgets impact the sales. We can clearly see that youtube and facebook sales increase linearly with increase in the advertising budget. The newspaper (3rd plot) sales however shows no particular trend.

[5]

ii)

```
> Model <- lm(sales ~ youtube + facebook + newspaper, data = marketing)
> summary (Model)

Call:
lm(formula = sales ~ youtube + facebook + newspaper, data = marketing)

Residuals:
    Min      1Q  Median      3Q     Max
-10.5932  -1.0690   0.2902   1.4272   3.3951

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.526667   0.374290   9.422   <2e-16 ***
youtube      0.045765   0.001395  32.809   <2e-16 ***
facebook     0.188530   0.008611  21.893   <2e-16 ***
newspaper   -0.001037   0.005871  -0.177     0.86
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.023 on 196 degrees of freedom
Multiple R-squared:  0.8972,      Adjusted R-squared:  0.8956
F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16
```

[6]

iii) It can be seen that from the **estimates column and from p values** that, changes in the youtube and facebook advertising budgets are significantly associated to changes in sales while changes in the newspaper budget is not. [3]

iv)

```
> cor(marketing$youtube,marketing$sales)
[1] 0.7822244
> cor(marketing$facebook,marketing$sales)
[1] 0.5762226
> cor(marketing$newspaper,marketing$sales)
[1] 0.228299
```

The pairwise plot and the above correlation indicated the same conclusion on newspaper having a very low / no particular trend with respect to sales.

[3]

v)

```
> Model1 <- lm(sales ~ youtube + facebook , data = marketing)
> summary(Model1)
```

```
Call:
lm(formula = sales ~ youtube + facebook, data = marketing)

Residuals:
    Min      1Q   Median      3Q     Max
-10.5572  -1.0502   0.2906   1.4049   3.3994

Coefficients:
          Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.50532    0.35339   9.919   <2e-16 ***
youtube      0.04575    0.00139  32.909   <2e-16 ***
facebook     0.18799    0.00804  23.382   <2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.018 on 197 degrees of freedom
Multiple R-squared:  0.8972,      Adjusted R-squared:  0.8962
F-statistic: 859.6 on 2 and 197 DF,  p-value: < 2.2e-16
```

Sales = 3.5 + 0.045*youtube + 0.187*facebook

[4]

vi) Adjusted R squared for Model in part (b) and that of part (e) is 0.89, hence there is no particular improvement after removing newspaper parameter. However, a model with less parameters is considered better, hence we can consider Model 1 calculated in part (f) to be a good fit. [3]

vii)
```
> marketing[which.max(marketing$sales),]

youtube facebook newspaper sales

176  332.28    58.68     50.16  32.4
```

Maximum sales generated is 32.4 thousand dollars. [2]

viii)
```
> PredTest = predict(Model1)
> PredTest[176]

29.74023
```

[4]

ix) (Observed ILI - Estimated ILI)/Observed ILI
```
> (32.4-29.74023)/32.4
[1] 0.08209167
```

[3]

[33 Marks]

****