Institute of Actuaries of India

Subject CS1-Actuarial Statistics (Paper B)

November 2023 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~N(5400,900^2)
> x.mean=5400
> x.sd=900
> #Y~N(3600,1500^2)
> y.mean= 3600
> y.sd=1500
> #P(X>2Y) = P(X-2Y>0)
> # X-2Y ~ N(x.mean - 2y.mean, x.sd^2+4y.sd^2)
>
> f.mean<-x.mean - 2*y.mean
> f.var <- x.sd^2+4*y.sd^2
>
> 1-pnorm(0,mean=f.mean,sd=sqrt(f.var))
[1] 0.2827485
```

[Max 5]

ii)

a)

```
> > dif.mean<-x.mean - y.mean
> dif.mean
[1] 1800
>
> dif.sd <-sqrt(x.sd^2+y.sd^2)
> dif.sd
[1] 1749.286
```

[3]

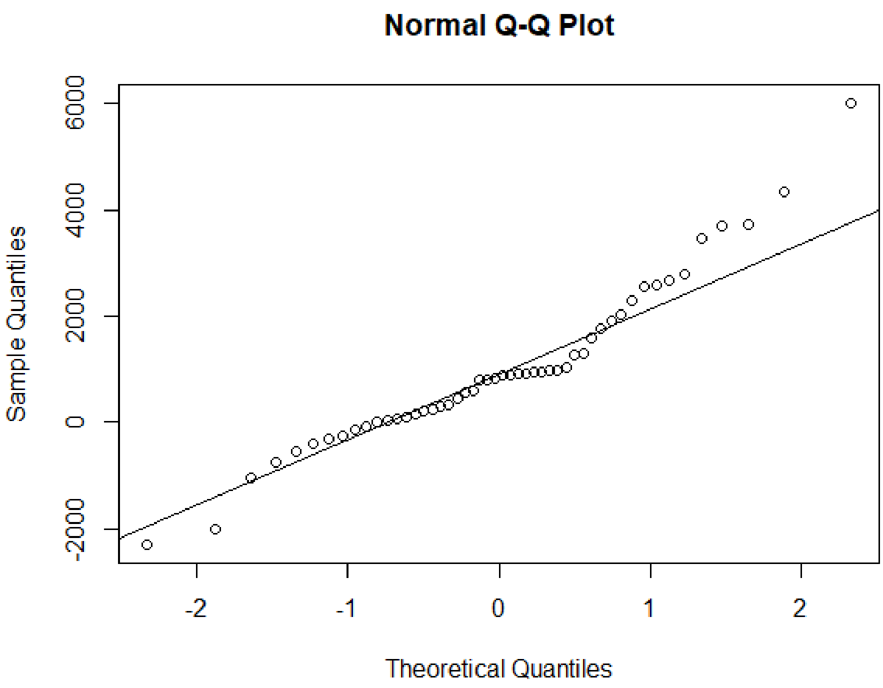
b)

```
> set.seed(1234)
> dif.sample<-rnorm(50,dif.mean,dif.sd)
```

[2]

c)

```
qnorm(dif.sample)
qqline(dif.sample)
```



Above results indicates that despite sample generated from normal distribution it doesn't seem to [4] align with normality probably due to low sample size and high variability.

iii)

a)

```
> sample.mean<-mean(dif.sample)
> sample.mean
[1] 1007.481
> z<- (sample.mean - 1375)/(dif.sd/sqrt(50))
> pnorm(z)
[1] 0.06869143
```

Since p-value is >0.05 we cannot reject the null hypothesis and thus, don't have sufficient evidence to say that mean is less than 1375. [Max 5]

b)

```
t.test(dif.sample, mu=1375, alternative = "less")
        One Sample t-test

data:  dif.sample
t = -1.6786, df = 49, p-value = 0.0498
alternative hypothesis: true mean is less than 1375
95 percent confidence interval:
    -Inf 1374.558
sample estimates:
mean of x
 1007.481
```

Since p-value is < 0.05 we can reject the null hypothesis and can imply that mean is less than 1375. [3]

iv)

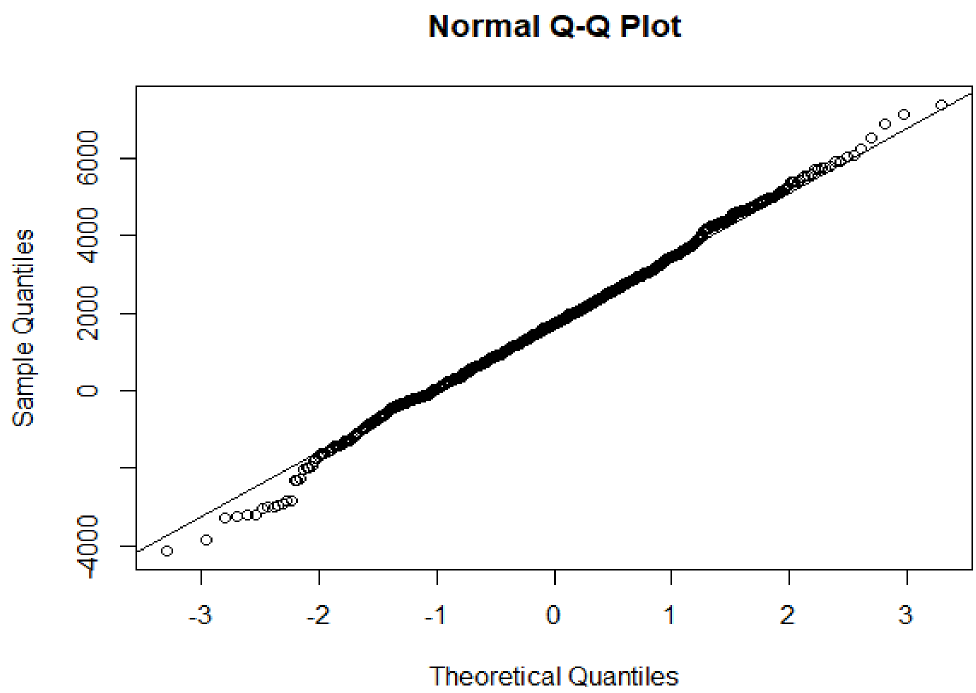
a)

```
dif.sample2<-rnorm(1000,dif.mean,dif.sd)
```

[1]

b)

```
qqnorm(dif.sample2)
qqline(dif.sample2)
```



With larger sample size, it indicates normality. [3]

[26 Marks]

Solution 2:

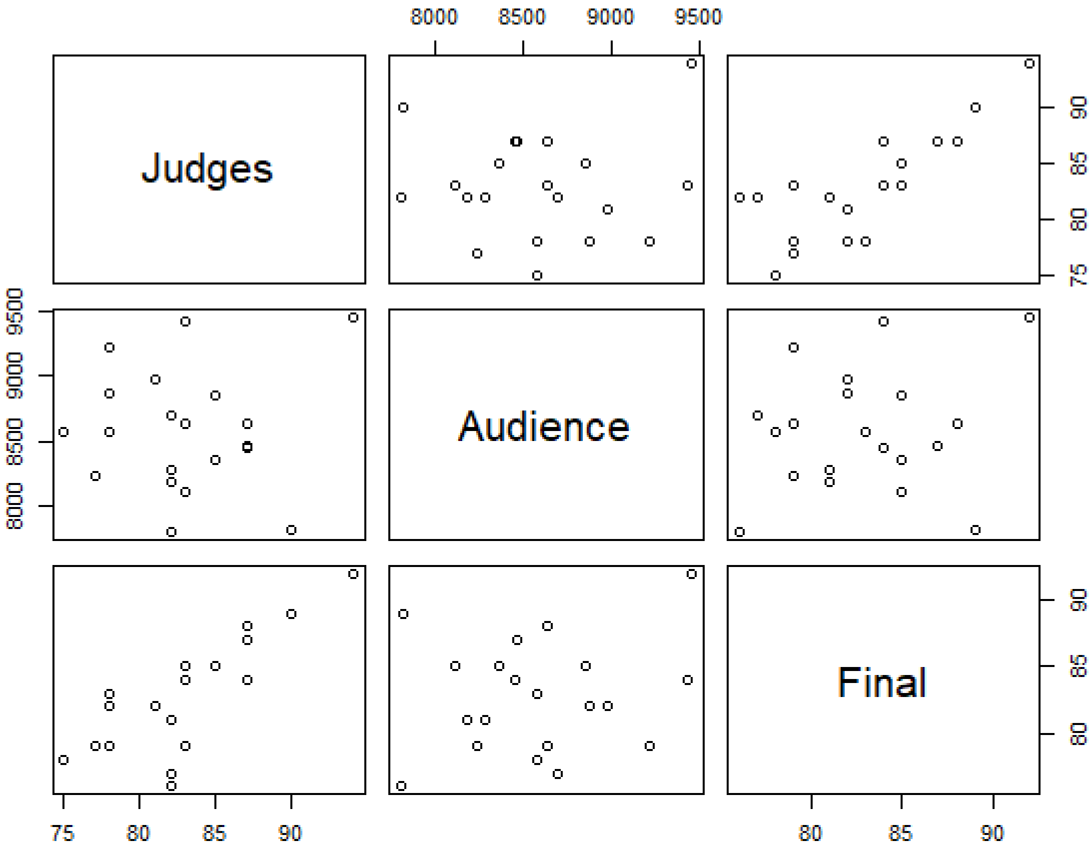
i)

```
dance<-read.csv("dance.csv")
head(dance)
```

[2]

ii)

```
plot(dance)
```



[4]

iii) Judges score and Final score seems to have a linear relationship
Audience score is quite scattered and doesn't show any strong linear relationship with either Judges or Final score [3]

iv)

a)

```
m1<-lm(Final~Judges+Audience,data=dance)
```

[2]

b)

```
> summary(m1)

Call:
lm(formula = Final ~ Judges + Audience, data = dance)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.2783 -0.7971  0.1841  1.6334  3.7680 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.617827  15.129865   0.702    0.492    
Judges       0.720273   0.128870   5.589 3.26e-05 ***
Audience     0.001449   0.001294   1.120    0.278    
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.604 on 17 degrees of freedom
Multiple R-squared:  0.6601,  Adjusted R-squared:  0.6201
F-statistic: 16.51 on 2 and 17 DF,  p-value: 0.0001038

Final = 10.617827 + 0.720273 * Judges + 0.001449 *Audience
```
[4]

c)
```
> confint(m1)
                  2.5 %     97.5 %
(Intercept) -21.3033977 42.5390512
Judges        0.4483811  0.9921649
Audience     -0.0012812  0.0041793
>
```
[3]

d) > #Judges score seems significant since confidence interval doesn't contain 0.
Sum of audience score doesn’t seem significant sine it contains 0.
Alternate:
> #P-value is <.01 for judges score showing significance.

[3]

v)
```
audience.count<-c(110,100,90,120,100,100,100,100,110,110,100,
                  100,110,90,100,110,120,120,100,100)
sum(audience.count)
[1] 2090
```
[2]

vi)
```
dance$Audience2<-dance$Audience/audience.count
```
[2]

vii)
```
> cor.test(dance$Final,dance$Audience2)

        Pearson's product-moment correlation

data:  dance$Final and dance$Audience2
t = 4.9045, df = 18, p-value = 0.0001142
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4716109 0.8982071
sample estimates:
      cor
0.7562948

> #correlation between audience score and final score is quite high
```
[3]

viii)
```
m2<-lm(Final~Judges+Audience2,data=dance)
> summary(m2)

Call:
lm(formula = Final ~ Judges + Audience2, data = dance)

Residuals:
    Min      1Q  Median      3Q     Max
-1.9408 -1.0269  0.1129  1.0466  1.6075

Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.20323    5.84491   0.206    0.839
Judges       0.56604    0.06545   8.648 1.24e-07 ***
Audience2    0.42002    0.05366   7.827 4.91e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.258 on 17 degrees of freedom
Multiple R-squared:  0.9207,  Adjusted R-squared:  0.9114
F-statistic: 98.73 on 2 and 17 DF,  p-value: 4.39e-10
```

[3]

ix) #Adjusted R-square =0.9114 for model2 vs 0.6201 for model1.
#This indicates model 2 is better

Alternate:
R square can be used [3]

[34 Marks]

Solution 3:

i) log(y) = 0.5306
poisson distribution is used to model response variable. [2]

ii)
```
> Claim.mean<-round(exp(.5306),1)
> Claim.mean
[1] 1.7
```
[2]

iii) log(y) = 1.1394 * GenderF + 1.1394 * GenderM -1.4271 *HealthNonDiabetic
where GenderF =1 if Gender = F else 0
GenderM =1 if Gender = M else 0
HealthNonDiabetic =1 if Health= NonDiabetic else 0

[Max 4]

iv) For model 2, "-1" is used in Glm R formula to not take the intercepts while fitting the model. Thus, no intercept exits. [2]

v) AIC (=59.4) of model 2 which is lower than AIC (71.1) of model 1. This indicates that model 2 is better. [2]

vi)
```
#AIC = - 2LogL(Model) + 2*Parameters
> #LogL(Model) = Parameters - AIC/2
> aic<-59.403
> par<-3
>
> L<- par - aic/2
> L
[1] -26.7015
```

[Max 3]

vii)
```
#Total claims = Mean claims * Total policies
> x<-Claim.mean*20
> poisson.test(x=X,T=20,r=1.5,conf.level = 0.99)

            Exact Poisson test
```

```
data:  X time base: 20
number of events = 34, time base = 20, p-value = 0.4639
alternative hypothesis: true event rate is not equal to 1.5
99 percent confidence interval:
 1.042837 2.605372
sample estimates:
event rate
       1.7

>
> # We cannot reject the null hypothesis that parameter is equal to 1.5.
```

[Max 5]

[20 Marks]

Solution 4:

i)

```
q4<-matrix(c(455,251,309,400,
             458,322,246,426,
             587,292,217,470,
             531,340,120,547),
           ncol=4,nrow=4)

n<-ncol(q4)
m<-mean(rowMeans(q4))
s<-mean(apply(q4,1,var))
v<-var(rowMeans(q4)) - mean(apply(q4,1,var))/n
Z<- n/(n+s/v)

n
[1] 4
> m
[1] 373.1875
> s
[1] 3967.854
> v
[1] 16843.22
> Z
[1] 0.9443816
```

[1+2+2+3+2]

ii)

```
Z*rowMeans(q4[3:4,])+(1-Z)*m
[1] 231.3532 455.8799
```

[3]

iii) Risk Volumes are required to apply EBCT2 [2]

iv)

```
obs<-c(61,71,15,3)
>
> #Combine 2 and 3 since value of 3 is less than 5
> obs.comb<-c(61,71,15 + 3)
>
> p<-0.2
> exp<-dbinom(c(0:1),3,p)
> exp[3]<-1-pbinom(1,3,p)
> sum(exp)
[1] 1
>
```

```
> chisq.test(x=obs.comb,p=exp)

        Chi-squared test for given probabilities

data:  obs.comb
X-squared = 6.7371, df = 2, p-value = 0.03444
```

Since p-value <0.5, we have sufficient evidence to reject the null hypothesis that cancellation follows bin(3,0.2)

[Max 5]

[20 Marks]

****