Math Coursework 1

Prathamesh Kulkarni (Student ID: 36004026)

30/10/2021

Q1)

Given n = 10 and $\lambda = 5$

a) Sample observation after applying $\lambda = 5$ and n = 10 to a rpois function we get these values 6, 6, 3, 6, 8, 8, 4, 1, 5, 4

b) The index of dispersion is calculated by dividing variance by mean of the sample generated. The index of dispersion for the above data is 0.9346405 which is closer to 1 as the perfect poisson has an index of dispersion of 1.

c) The expression for the method of moment estimator of λ will be 1.1

$\frac{1}{n}\sum_{i=1}^{n} X_i = \lambda$

because it follows a poission distribution.

d) the method of moment estimate of λ is 1.2

$\frac{1}{n}\sum_{i=1}^{n} x_i = \lambda$

e) The difference between estimate and actual value is 0.002777341.

```
estimated_probability <- dpois(6, 5.1)
actual_probability <- dpois(6, 5)
estimated_probability - actual_probability
```

```
## [1] 0.002777341
```

f) The first five bootstrap means for N = 100 obtained from the sample are 5.4, 4.3, 5.1, 5.8, 4.8

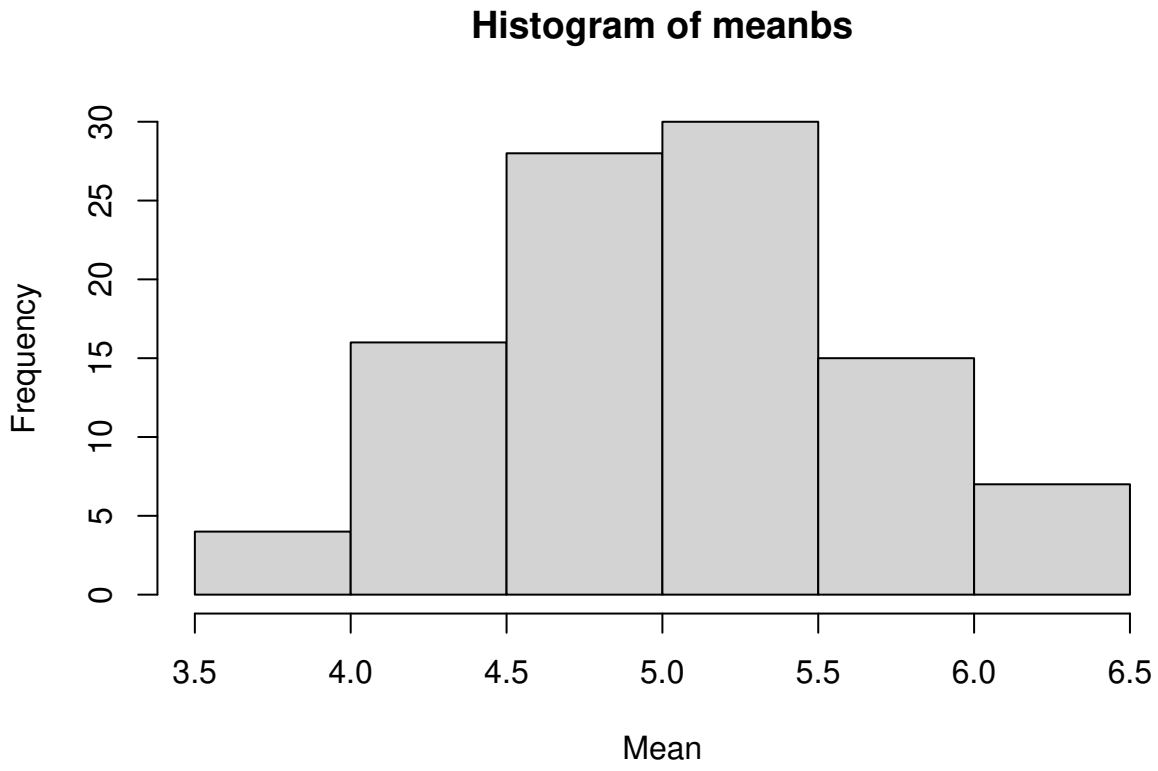
```
set.seed(100)
N <- 100
meanbs <- c()
for (i in 1:N){
  r <- sample(poisson, replace = T)
  meanbs[i] <- mean(r)
}
quantile(meanbs,c(0.025,0.975))
```

```
##   2.5%  97.5%
## 3.9000 6.2525
```

```
meanbs[1:5]
```

```
## [1] 5.4 4.3 5.1 5.8 4.8
```

```
 hist(meanbs,xlab="Mean",ylab="Frequency",
main="Histogram of meanbs")
```



g) The x axis is the means obtained after bootstrapping. y axis has represents frequency. And we can see that the histogram is centered around mean of 5 which is equal to actual λ and the estimate which is 5.1

h) Steps:

1) Decide the number of Samples N (which is 100 in our case) to create.

2) create a large number 100 of 'replicate' data sets by sampling with replacement from the original data set of size 10.

3) The 100 parameter estimates obtained from the bootstrapped samples can then be used to give an estimate of the sampling variability.

4) A bootstrap estimate of the 95% confidence interval is given by the v2.5% and 97.5% quantiles of the bootstrapped sampling distribution.

Confidence Intervel for the mean of the poisson distribution 3.9000, 6.2525

Q2)

Ho : $\mu_o = \mu_n$

H1 : $\mu_o > \mu_n$

a) The data is unpaired because it is related to two different categories, which are organic disorders and non organic disorders.

```
organic_disorders <- c(17.53, 20.60, 17.62, 28.93, 27.10)
non_organic_disorders <- c(5.59, 14.76, 13.32, 12.45, 12.79)

var(organic_disorders)
```

```
## [1] 28.63163
```

```
var(non_organic_disorders)
```

```
## [1] 12.75927
```

```
var.test(organic_disorders,non_organic_disorders,alt="two.sided")
```

```
##
##  F test to compare two variances
##
## data:  organic_disorders and non_organic_disorders
## F = 2.244, num df = 4, denom df = 4, p-value = 0.453
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##   0.2336384 21.5524357
## sample estimates:
## ratio of variances
##           2.243987
```

b) the variance of the two groups are different, with ratio of variances is 2.243 which is not equal to 1.

```
t.test(organic_disorders,non_organic_disorders,
alternative="greater",unpaired=T)
```

```
##
##  Welch Two Sample t-test
##
## data:  organic_disorders and non_organic_disorders
## t = 3.6751, df = 6.9744, p-value = 0.003981
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  5.119939      Inf
## sample estimates:
## mean of x mean of y
##    22.356    11.782
```

```
qt(0.95,6)
```

```
## [1] 1.94318
```

c) We use unpaired t-test and we get a p value of 0.003981 which is less than 0.05 hence we reject the null hypothesis. The assumption we use is, the data has a normal distribution.

d) A non parametric test is preferable as we dont need to assume the distribution of the sample and non parametric tests are better at handling outliers.

```
wilcox.test(organic_disorders,non_organic_disorders,
alternative="greater",unpaired=T)
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  organic_disorders and non_organic_disorders
## W = 25, p-value = 0.003968
## alternative hypothesis: true location shift is greater than 0
```

e) We use a wilcox test and the p value we get is 0.003981 which is less than 0.05 hence we reject the null hypothesis. mean percentage excretion of organic disorders is greater than the non organic disorders. Since it is a non parametric test we dont assume anything.

Conclusion: we can conclude from above tests that mean percentage excretion in organic disorders in greater than non organic disorders.

References: Maths503 notes, and workshop solutions.

Feedback comments

Qn 1: 9

-1b, -1c, -1g,

Qn 2: 7

-1b

Index of comments