NPTEL DATA SCIENCE FOR ENGINEERS

ASSIGNMENT 3

1. Sum of the deviations about mean is

Solution: c

The sum of the deviations from the mean is zero. This will always be the case as it is a property of the sample mean, i.e., the sum of the deviations below the mean will always equal the sum of the deviations above the mean.

2. The mode of the normal distribution is

Solution: a

In normal distribution the mode and the median are the same as the mean, the mean for normal distribution is represented by μ.

3. For the positively skewed distribution the outliers or extreme values will lie in _____

Solution: b

In statistics, a positively skewed (or right-skewed) distribution is a type of distribution in which most values are clustered around the left tail of the distribution while the right tail of the distribution is longer. So the outliers or extreme values will lie in right tail of the distribution

4. The range of t distribution is

Solution: b

t distribution ranges from -∞ to ∞. It is one of the properties of t distribution.

5. Statistical test power is denoted by

Solution: c

β = probability of a Type II error, known as a "false negative"

1-β = probability of a "true positive", which is correctly rejecting the null hypothesis. "1-β" is also known as the power of the test.

6. If type I error is decreasing

Solution: b

Type II error, also known as a "false negative": the error of not rejecting a null hypothesis when the alternative hypothesis is the true state of nature. If type I error “false positive “is decreasing vice versa Type II error will increase.

Download the data set *“seatbelts.csv”.* Load the data set in to your R workspace and answer the questions 7 to 10.

The description of the dataset is given below:

This assessment is based on the “Road Casualties in Great Britain 1969-84” inbuilt R data set, which is called “Seatbelts”.

The ‘Seatbelts’ data set in R is a multiple time-series data set that was commissioned by the Department of Transport in 1984 to measure differences in deaths before and after front seatbelt legislation was introduced on 31st January 1983. It provides monthly total numerical data on a number of incidents including those related to death and injury in Road Traffic Accidents (RTA’s). The data set starts in January 1969 and observations run until December 1984.

Variable name	Description
DriversKilled	Number of car drivers killed
drivers	Total number of drivers
front	Number of front-seat passengers killed or seriously injured.
rear	Number of rear-seat passengers killed or seriously injured.
kms	Total number of distances driven
PetrolPrice	Petrol price
VanKilled	number of van (‘light goods vehicle’) drivers killed

law	0/1: was the law in effect that month?

7. Average number of car drivers killed after the law was in effect is ____?
   Hint: Use the function filter from "dplyr" package to subset the dataset

Solution: c

Average number of car drivers killed after the law was in effect is 100.

```
> data_law_1=filter(data,law!=0)
> summary(data_law_1$DriversKilled)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   60.0    85.0    92.0   100.3   119.0   154.0
```

8. The standard deviation of drivers killed before the law came into effect is _____?
   Hint: Use the function filter from "dplyr" package to subset the dataset

Solution: b

The standard deviation of drivers killed before the law came into effect is 24.260.

```
> data_law_0=filter(data,law==0)
> sd(data_law_0$DriversKilled)
[1] 24.26088
```

9. Calculate the variance for the variables "front" and "rear" and choose the correct option.

Solution: b:Variance of front seat passengers is greater than variance of rear seat passengers.

```
> var(data$front)
[1] 30659.65
> var(data$rear)
[1] 6905.977
```

10. Maximum kms driven by the driver is ___?

Solution: a

Maximum kms driven by the driver is 21626.

```
> summary(data$kms)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   7685   12685   14987   14994   17203   21626
```