

Question 1:-

In this question, we have the following data.

$n = 28$ (pigs)

Sample mean = 1

Sample Standard Deviation = 0.3

In R Code :-

```
sampleStandardDeviation <- 0.3  
sampleMean <- 1.0  
n <- 28
```

Let's define H_0 and H_1 as follows:-

$H_0 : \mu = 0.8$

$H_1 : \mu \neq 0.8$

Assumptions :-

1. The sample is SRS >> Simple and Random .
2. The sample is normally distributed.

With the given assumptions let's compute the test statistic as well as the corresponding P-Value as well using the R code.

```

sampleStandardDeviation <- 0.3
sampleMean <- 1.0
n <- 28

U_0 <- 0.8

tTest <- (sampleMean - U_0) / (s / sqrt(n))

c('test statistic',tTest)

pValue <- pt(tTest, df=n-1, lower.tail = FALSE) * 2

alpha <- 0.05

c('pValue',pValue)

if (pValue < alpha) {
  print("We Reject the Null Hypothesis")
} else
{
  print("We Accept the Null Hypothesis")
}

```

pValue = 0.0015

```

> sampleStandardDeviation <- 0.3
> sampleMean <- 1.0
> n <- 28
>
> U_0 <- 0.8
>
> tTest <- (sampleMean - U_0) / (s / sqrt(n))
>
> c('test statistic',tTest)
[1] "test statistic"    "3.52766841475279"
>
> pValue <- pt(tTest, df=n-1, lower.tail = FALSE) * 2
>
> alpha <- 0.05
>
> c('pValue',pValue)
[1] "pValue"           "0.00152120629911983"
>
> if (pValue < alpha) print("We Reject the Null Hypothesis") else print("We Accept the Null Hypothesis")
[1] "We Reject the Null Hypothesis"

```

Since pValue < alpha = 0.05, we reject the null hypothesis.

CONCLUSION

Since we reject the null hypothesis, the alternate hypothesis is that the Alcohol does indeed affect the mean response time.

Question 2:-

In this question, we have the following data.

$$\sigma_0 = 0.4$$

$$n = 10$$

$$X = 5.728, 5.731, 5.722, 5.719, 5.727, 5.724, 5.718, 5.726, 5.723, 5.722$$

IN R CODE :-

```
data <- c(5.728, 5.731, 5.722, 5.719, 5.727, 5.724, 5.718, 5.726, 5.723, 5.722)
n <- 10
s_0 <- 0.4
```

Sample mean :- \bar{x}

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

$$\begin{aligned}\bar{x} &= \frac{1}{10} \sum_{i=1}^{10} x_i \\ &= \underline{\underline{5.724}}\end{aligned}$$

Sample Standard Deviation ;

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$= \frac{1}{10-1} \sum_{i=1}^{10} (x_i - 5.724)^2$$

$$s^2 = 1.645 \times 10^{-5}$$

$$s = \underline{\underline{0.00405}}$$

Scanned with CamScanner

In R CODE :-

```
sampleMean <- mean(data)  
sampleMean  
  
sampleStandardDeviation <- sd(data)  
sampleStandardDeviation
```

```
> sampleMean <- mean(data)  
> sampleMean  
[1] 5.724  
>  
> sampleStandardDeviation <- sd(data)  
> sampleStandardDeviation  
[1] 0.004055175
```

Values found through manual calculation vs the R functions mean() and sd() are found to be equal.

1. HYPOTHESIS CONSTRUCTION

$$H_0: \sigma \geq 0.4$$

$$H_1: \sigma < 0.4$$

Assumptions :-

3. The sample is SRS >> Simple and Random .
4. The Population is normally distributed.

2. TEST STATISTICS

With the assumptions in minds, along with the Hypothesis revolving around σ , we use Chi-Square test statistic.

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi_{(n-1)}^2$$

$$= \frac{9 \times 0.00405^2}{0.4^2} \sim \chi_9^2$$

$$\chi_{\text{stat}}^2 = 0.000922$$

IN R CODE :-

```
chiSquareTestStatistic <- ((n-1)*sampleStandardDeviation**2)/(sigmaNOT**2)
c('chiSquareTestStatistic', chiSquareTestStatistic)
```

```
> chiSquareTestStatistic <- ((n-1)*sampleStandardDeviation**2)/(sigmaNOT**2)
> c('chiSquareTestStatistic', chiSquareTestStatistic)
[1] "chiSquareTestStatistic" "0.00092499999999952"
```

As we can see, both the computed values, using the formula in the R code as well as that computed manually agrees with one another.

Hence, χ^2 Test Statistic = 0.00092

3. DECISION

Alternative hypothesis is $H_1: \sigma < 0.4$, i.e left tailed test. Therefore, rejection region lies in the left side.

CRITICAL VALUE APPROACH

With the help of χ^2 - distribution table, we calculate for $X_{critical}$:-

χ^2 critical= 3.33

USING R :-

```
chiSquareCriticalValue <- (qchisq(0.05, df=n-1, lower.tail=TRUE))
c('chiSquareCriticalValue', chiSquareCriticalValue)
```

```
> chiSquareCriticalValue <- (qchisq(0.05, df=n-1, lower.tail=TRUE))
> c('chiSquareCriticalValue', chiSquareCriticalValue)
[1] "chiSquareCriticalValue" "3.32511284306681"
```

The R function and the manually calculated solutions give us approximately the same values.

χ^2 critical= 3.32

p - value approach

We calculate p-value using R, because the value of $\chi^2_{\text{stat}} = 0.00092$ is too small and precise to locate on the Chi-Square distribution table.

p-value using R:-

```
p_Value<-pchisq(q=chiSquareTestStatistic, df=n-1, lower.tail=TRUE)
c('pValue', p_Value)
```

```
> p_Value<-pchisq(q=chiSquareTestStatistic, df=n-1, lower.tail=TRUE)
> c('pValue', p_Value)
[1] "pValue"           "1.87923767405513e-17"
```

$$p_Value = 1.8792 \times 10^{-17}$$

4. CONCLUSION

1. $\chi^2_{\text{stat}} < \chi^2_{\text{critical}}$, tstat lies in the rejection region and further we reject H_0 .
2. $p_value < \alpha = 0.05$, we reject the H_0

Using 1 and 2, we reject the null hypothesis.

Yes, the new method must be adopted, since standard deviation will drop below 0.4.

Question 3 :-

In this question. The Data is given as follows.

$n = 25$

Number of smokers = 11

Number of non smokers = 14

Smokers Blood Pressure :- 124, 134, 136, 125, 133, 127, 135, 131, 133, 125, 118.

Non-smokers Blood Pressure :- 130, 122, 128, 129, 118, 122, 116, 127, 135, 120, 122, 120, 115, 123

```
n<-25  
dataSmokers <- c(124, 134, 136, 125, 133, 127, 135, 131, 133, 125, 118)  
numberOfSmokers <- 11  
dataNonSmokers <- c(130, 122, 128, 129, 118, 122, 116, 127, 135, 120, 122, 120, 115, 123)  
numberOfNonSmokers <- 14
```

Sample mean of smokers & non-smokers :-

\bar{x}_s & \bar{x}_n :-

$$\bar{x}_s = \frac{1}{n} \sum_{i=1}^n x_{si}$$

$$= \frac{1}{11} \sum_{i=1}^{11} x_{si}$$

$$= \underline{\underline{129.181}}$$

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_{ni}$$

$$= \frac{1}{14} \sum_{i=1}^{14} x_{ni}$$

$$= \underline{\underline{123.357}}$$

Sample standard deviation of smokers
and non smokers :-

$$s_s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{si} - \bar{x}_s)^2$$

$$s_s^2 = 32.763$$

$$s_s = 5.723$$

$$s_N^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ni} - \bar{x}_N)^2$$

$$s_N^2 = \cancel{32} \quad 32.862$$

$$s_N = 5.732$$

Scanned with CamScanner

```
SmokersMean <- mean(dataSmokers)
SmokersMean

NonSmokersMean <- mean(dataNonSmokers)
NonSmokersMean

SmokersStandardDeviation <- sd(dataSmokers)
SmokersStandardDeviation

NonSmokersStandardDeviation <- sd(dataNonSmokers)
NonSmokersStandardDeviation
```

```

> SmokersMean <- mean(dataSmokers)
> SmokersMean
[1] 129.1818
>
> NonSmokersMean <- mean(dataNonSmokers)
> NonSmokersMean
[1] 123.3571
>
> SmokersStandardDeviation <- sd(dataSmokers)
> SmokersStandardDeviation
[1] 5.723953
>
> NonSmokersStandardDeviation <- sd(dataNonSmokers)
> NonSmokersStandardDeviation
[1] 5.732594

```

The values found manually calculated as well as those computed using the mean and sd functions of R are found to be same.

1. HYPOTHESIS CONSTRUCTION

$$H_0 : \mu S = \mu N$$

$$H_1 : \mu S \neq \mu N$$

With $\mu_0 = \mu_S - \mu_N$ we can also rewrite this as:

$$H_0 : \mu_0 = 0$$

$$H_1 : \mu_0 \neq 0$$

Assumptions:-

1. SRS :- Both the samples are simple and random
2. Both are independent
3. σ_N & σ_S are unknown and unequal.
4. Both are from Normal Distribution.

2. TEST STATISTICS

With the given assumptions, lets compute the t-distribution test statistic.

$$t_{\text{stat}} = \frac{(\bar{x}_S - \bar{x}_N) - (\mu_S - \mu_N)}{\sqrt{\frac{s_S^2}{n_S} + \frac{s_N^2}{n_N}}} \sim t_{df}$$

$$\sqrt{\frac{s_S^2}{n_S} + \frac{s_N^2}{n_N}}$$

$$= \frac{(129.181 - 123.357) - (0)}{\sqrt{\frac{32.763}{11} + \frac{32.682}{14}}}$$

$$t_{\text{stat}} = 2.5236$$

$$df = \frac{(A + B)^2}{\frac{A^2}{n_1-1} + \frac{B^2}{n_2-1}}$$

$$A = \frac{s_1^2}{n_1} \quad B = \frac{s_2^2}{n_2}$$

$$\text{with } A = 2.978$$

$$B = 2.347$$

$$df = \frac{(2.978 + 2.347)^2}{\frac{(2.978)^2}{10} + \frac{(2.347)^2}{13}}$$

$$= \frac{5.325^2}{.886 + .923}$$

$$df = \underline{\underline{21.66}}$$

 Scanned with CamScanner

Let's Calculate the test statistic using the R Code Functions :-

```
muNull <- 0
testStatistic<-((SmokersMean-NonSmokersMean)-muNull)/sqrt(((SmokersStandardDeviation^2)/numberOfSmokers)+((NonSmokersStandardDeviation^2)/numberOfNonSmokers))
c('test statistic', testStatistic)
```

```
testTest <- t.test(x=dataSmokers, y=dataNonSmokers, paired=FALSE, var.equal=FALSE,
                     alternative="two.sided")
|
testStatistic <- testTest
c('t_stat_inbuilt', testStatistic)
```

```
> muNull <- 0
> testStatistic<-((SmokersMean-NonSmokersMean)-muNull)/sqrt(((SmokersStandardDeviation^2)/numberOfSmokers)+((NonSmokersStandardDeviation^2)/numberOfNonSmokers))
> c('test statistic', testStatistic)
[1] "test statistic"    "2.52393122373416"
```

```
> testStatistic <- testTest
> c('t_stat_inbuilt', testStatistic)
[[1]]
[1] "t_stat_inbuilt"
```

```
$statistic
      t
2.523931
```

So we see that the manually computed values as well as the value computed through the R code, both the hard coded formula as well as the inbuilt formula is the same.

Let's calculate the value for degrees of freedom as well using R code.

```
degreesOfFreedom <- testTest$parameter  
c('degreesOfFreedom', degreesOfFreedom)
```

```
> degreesOfFreedom <- testTest$parameter  
> c('degreesOfFreedom', degreesOfFreedom)  
           df  
"degreesOfFreedom" "21.6359081281093"  
> |
```

We find that the manually computed as well as the R code values match for df values as well.

CRITICAL VALUE APPROACH

Using the R code, let's calculate the tcritical value:-

```
tCriticalValue <-(qt(p=0.025, df=21.66, lower.tail=FALSE))
c('tCriticalValue',tCriticalValue)
```

```
> tCriticalValue <-(qt(p=0.025, df=21.66, lower.tail=FALSE))
> c('tCriticalValue',tCriticalValue)
[1] "tCriticalValue"    "2.07576214430401"
>
```

p - value approach

Using the R code, let's calculate the p-value:-

```
pValue <- 2*pt(q=testStatistic, df=21.66, lower.tail=FALSE)
c('pValue', pValue)
```

```
pValueFunc <- testTest$p.value
c('p_value', pValueFunc)
```

```
> pValue <- 2*pt(q=testStatistic, df=21.66, lower.tail=FALSE)
> c('pValue', pValue)
[1] "pValue"           "0.0194602422878936"
>
> pValueFunc <- testTest$p.value
> c('p_value', pValueFunc)
[1] "p_value"           "0.0194697631350557"
```

4. CONCLUSION

1. Since $t_{stat} > t_{critical}$, t_{stat} lies in the rejection region, hence we reject H_0 .
2. Since $pValue < \alpha$, hence we reject H_0 .

From both 1 and 2 we reject the null hypothesis.

Mean of Blood Smokers and the Mean of Non Smokers are not the same.