# Statistical Data Analysis Problem sheet 6 Solution - Group 9

*Exercise 1*

1. Here given,
   significance level

$$\alpha = 0.05 \; for \; observed$$
$$\bar{x} = 1.5 \; with \; n = 15$$

   Where,

$$\sigma^2 = 4 \; and$$
$$H_0 : \theta = 1 \; versus \; H_1 : \mu \neq 1$$

Here, the test statistic is

$$\frac{\bar{x} - \theta_0}{\sigma / \sqrt{n}}$$

So, the test reject the Null Hypothesis if

$$\left| \frac{\bar{x} - \theta_0}{\sigma / \sqrt{n}} \right| \geq t_{\alpha/2, n-1}$$

where n-1 is the degrees of freedom and α is the significance level. So, we get

$$\left| \frac{\bar{x} - \theta_0}{\sigma / \sqrt{n}} \right| \geq t_{0.025, 14}$$

Now if we use the values( We are using σ value for the *S*), our test statistic

$$z = \left| \frac{1.5 - 1}{2 / \sqrt{15}} \right| = 0.968$$

Now, from the critical value chart, we get

$$t_{0.025, 14} = 1.761$$

Since,

$$|z| = 0.968 < t_{0.025, 14} = 1.761$$

We fail to reject the Null Hypothesis. We do not have enough evidence at 5% significance level.

2. A power function depends on the null hypothesis.

- Here we will define some functions to get the power for different sample size given the standard deviation is σ = 2, Θ = 1.

```
from math import *
def phi(x):
    #'Cumulative distribution function for the standard normal distribution given the sigma = 2, theta =1'
    return (1.0 + erf(x / 2*sqrt(2.0))) / 2.0
```

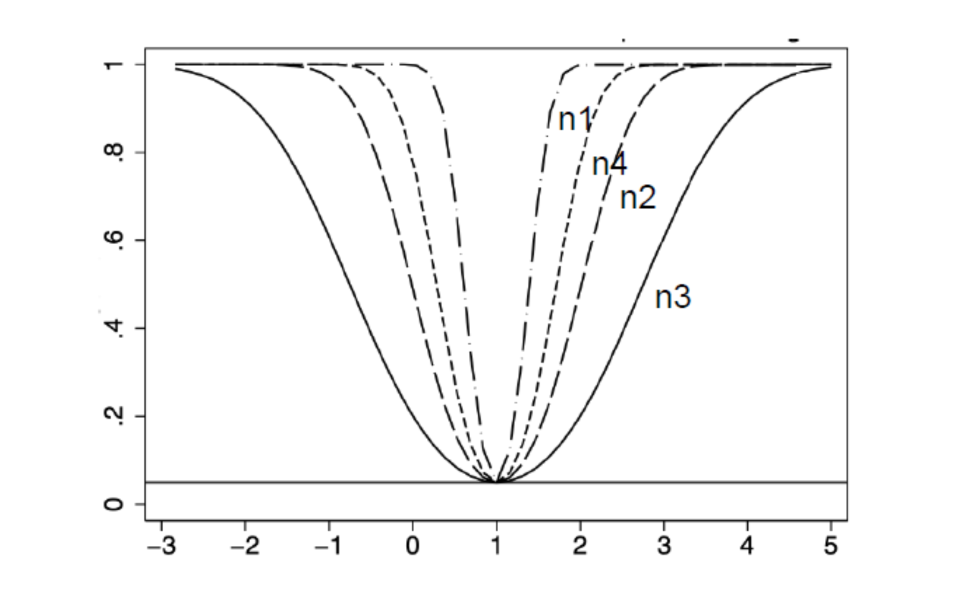
```
import scipy.stats
def critv(n):
  #critical value calculator for signifiance level 0.05
  res = scipy.stats.t.ppf(q=1-0.05/2,df=(n-1))
  return res

```

```
def powerfunc(n):
  #Power function
  return (1-phi(critv(n)-1*sqrt(n))+phi(-critv(n)-1*sqrt(n)))
```

```
#for different sample sizes, the power for the given sigma
samplist = [5,15,30,100]
for i in samplist:
  powerv = powerfunc(i)
  print(f"Power for sample size {i} = {powerv}")
```

```
Power for sample size 5 = 0.29446875765867175
Power for sample size 15 = 0.9580235145565433
Power for sample size 30 = 0.9997004217282532
Power for sample size 100 = 0.9999999999999994
```

Here we can see that when the sample size increases, the power function gets narrower. Thus we can say that the outermost function is associated with the least sample size $n_3$ = 5, and so on till the innermost function is associated with the largest sample size $n_1$ =100.




3. Here given,

$$\alpha = 0.05,\ emperical\ mean\ \bar{x} = 1.5\ and\ sample\ size\ n = 15$$

- We have to determine the confidence interval for Θ given the values and here we will use σ as the S.
- We can derive the confidence interval for Θ when σ is known.

  Here

$$t_{1-\alpha/2;n-1}\ denote\ the\ (1-\alpha/2)\ quantile\ of\ the\ t\ distribution\ with\ n-1\ degrees\ of\ freedom,$$
$$P(T_{n-1} < t_{1-\alpha/2;n-1}) = 1-\alpha/2, 0<\alpha<1$$

  Then using the symmetry of the t distribution, we have

$$1-\alpha = P\bigg(-t_{1-\alpha/2;n-1} < \frac{\bar{x}-\theta}{\sigma/\sqrt{n}} < t_{1-\alpha/2;n-1}\bigg)$$
$$= P\bigg(-t_{1-\alpha/2;n-1} \times \frac{\sigma}{\sqrt{n}} < \bar{x}-\theta < t_{1-\alpha/2;n-1} \times \frac{\sigma}{\sqrt{n}}\bigg)$$
$$= P\bigg(\bar{x}-t_{1-\alpha/2;n-1} \times \frac{\sigma}{\sqrt{n}} < \theta < \bar{x}+t_{1-\alpha/2;n-1} \times \frac{\sigma}{\sqrt{n}}\bigg)$$

  Then (1 - α) = 95% confidence interval for Θ is:

$$CI_{1-0.05}(\theta) = \bigg(\bar{x}-t_{1-\alpha/2;n-1} \times \frac{\sigma}{\sqrt{n}}\ ,\ \bar{x}+t_{1-\alpha/2;n-1} \times \frac{\sigma}{\sqrt{n}}\bigg)$$
$$CI_{0.95}(\theta) = \bigg(1.5 - t_{0.975;14} \times \frac{2}{\sqrt{(15)}}\ ,\ 1.5 + t_{0.975;14} \times \frac{2}{\sqrt{(15)}}\bigg)$$

  From the t-table we can find the 0.975 quantile of the t distribution with 14 degrees of freedom, which is 2.145.

  Now we get

$$CI_{0.95}(\theta) = (0.392, 2.608)$$

  Thus, we are 95% confident that Θ is between 0.392 and 2.608.

4. If we increase the sample size, the confidence interval will change.

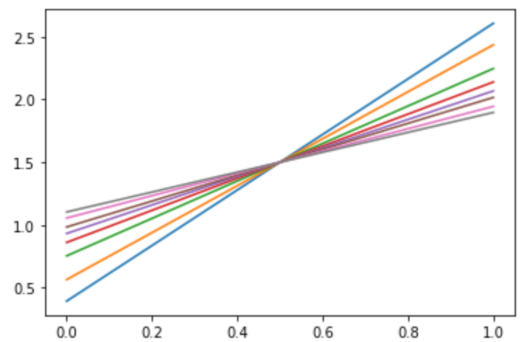
- Let the emprical mean and variance is unchanged.
- Let us find 95% confidence interval for different sample size to see the change.

```
#defining a 95% confidence interval finding function
def confint(x):
```

```
    resL = 1.5 - critv(x)*(2/sqrt(x))
    resR = 1.5 + critv(x)*(2/sqrt(x))
    return (resL, resR)
```

```
# creating an increasing sample size list
import matplotlib.pyplot as plt
sampleSize = [15,20,30,40,50,60,80,100]
intervals = []
for i in sampleSize:
  intervs = confint(i)
  intervals.append(intervs)
  print(f"For sample size = {i}, the 95% confidence level is : {intervs}")
  plt.plot(intervs)
plt.show()
```

```
For sample size = 15, the 95% confidence level is : (0.3924369168707169, 2.607563083129283)
For sample size = 20, the 95% confidence level is : (0.5639711871601811, 2.436028812839819)
For sample size = 30, the 95% confidence level is : (0.7531877264838003, 2.2468122735161997)
For sample size = 40, the 95% confidence level is : (0.8603689716395658, 2.139631028360434)
For sample size = 50, the 95% confidence level is : (0.9316062897552536, 2.0683937102447465)
For sample size = 60, the 95% confidence level is : (0.9833452152569502, 2.01665478474305)
For sample size = 80, the 95% confidence level is : (1.054921802463506, 1.945078197536494)
For sample size = 100, the 95% confidence level is : (1.1031566096982635, 1.8968433903017365)
```




As we can see, with the increasing sample size, the confidence interval is getting tighter which gives us better precision with a smaller margin of error.