

Problem Statement 1

Let X be random variable cost of textbooks in the college bookstore.

Since sample size is greater than 30, therefore the average cost $\bar{X} \sim N(52, \frac{4.5}{\sqrt{100}})$

$$H_0 : \mu = 52, H_1 : \mu > 52$$

Testing at 5% significant level and one tailed test, the 95% confidence interval of \bar{X} is given by

$$\begin{aligned} & \left(-\infty ; 52 + \varphi(0.95) * \frac{4.5}{\sqrt{100}} \right) \\ & = (-\infty ; 52.74018) \end{aligned}$$

Since mean from the sample Rs. 52.80 does not fall within the 95% confidence interval of \bar{X} ,

Rejected H_0 hence there is significant evidence that the average cost is higher than Rs. 52.80.

Used the following R code

```
> #Problem 1
> #
> #95% confidence interval
> upper_bound = qnorm(p = 0.05, mean = 52, sd = 4.5/sqrt(100), lower.tail = F)
> upper_bound
[1] 52.74018
```

Problem Statement 2

Let Y be random variable amount of chemical pollutant in the Genesee River .

Since sample size is greater than 30, $n = 50$, therefore the average amount of chemical pollutant $\bar{Y} \sim N(34, \frac{8}{\sqrt{50}})$

$$H_0 : \mu = 34, H_1 : \mu < 34$$

Testing at 1% significant level and one tailed test, the 99% confidence interval of \bar{Y} is given by

$$\begin{aligned} & \left(34 - \varphi(0.99) * \frac{8}{\sqrt{50}} ; \infty \right) \\ & = (31.36804; \infty) \end{aligned}$$

Since mean from the random sample 32.5ppm does fall within the 99% confidence interval of \bar{Y} ,

Do not ejected H_0 hence there is no significant evidence that the average chemical pollutant in the Genesee River has decreased.

Used the following R code

```
> #Problem 2
> #
> #95% confidence interval
```

```
> lower_bound = qnorm(p = 0.01, mean = 34, sd = 8/sqrt(50))
> lower_bound
[1] 31.36804
```

Problem Statement 3

Let W be random variable amount spend by a family of four annually on dental expenditures. Assuming, that dental expenditure is normally distributed but population standard deviation is unknown, therefore the average amount spend by a family of four annually

$\bar{W} \sim t$ distribution with $n - 1 = 21$ degrees of freedom

$$H_0 : \mu = 1135, H_1 : \mu \neq 1135$$

Testing at $\alpha = 0.5$ and a two tailed test, the 95% confidence interval is given by

$$\left(1135 - t_{0.975}(21) * \frac{s}{\sqrt{22}}; 1135 + t_{0.975}(21) * \frac{s}{\sqrt{22}} \right)$$

$$\text{where } s^2 = \frac{\sum_{i=1}^{22} (w_i^2 - \bar{w})}{22 - 1} \text{ and } \bar{w} = \frac{\sum_{i=1}^{22} w_i}{22}$$

Therefore $\bar{w} = 1031.318$ and $s = 240.3746$ and

Therefore confidence interval is (1028.424 ; 1241.576)

Since mean from the sample \$1031.32 does fall within the 95% confidence interval of \bar{W} ,

Do not ejected H_0 hence there is no significant evidence that the average *dental expenditure* of the country of interest is different from that of U.S.

Used the following R code

```
> #Problem 3
> #
> w = c(1008, 812, 1117, 1323, 1308, 1415, 831, 1021, 1287, 851, 930, 730,
+      699,
+      872, 913, 944, 954, 987, 1695, 995, 1003, 994)
>
> #Average amount of annually on dental expenditures from sample
> mean_w = mean(x = w)
> mean_w
[1] 1031.318
>
> #Sample standard deviation amount of annually on dental expenditures
> std_w = sqrt(sum((w - mean_w)^2)/(22-1))
> std_w
[1] 240.3746
>
> # 95% confidence interval
> lower_bound = 1135 + qt(p = 0.025, df = 21)*std_w/sqrt(22)
> lower_bound
[1] 1028.424
>
> upper_bound = 1135 + qt(p = 0.975, df = 21)*std_w/sqrt(22)
> upper_bound
[1] 1241.576
```

Problem Statement 4

Let X be random variable amount annual family income on Metropolis.

Since sample size is very large, n

= 400, therefore the average amount annual family income on Metropolis $\bar{X} \sim N(48432, \frac{2000}{\sqrt{400}})$

$$H_0 : \mu = 48\,432, H_1 : \mu \neq 48\,432$$

Testing at 5% significant level and two tailed test, the 95% confidence interval of \bar{X} is given by

$$\left(48432 - \varphi(0.975) * \frac{2000}{\sqrt{400}} ; 48432 + \varphi(0.975) * \frac{2000}{\sqrt{400}} \right) \\ = (48236; 48628)$$

Since mean from the sample \$48,574 does fall within the 95% confidence interval of \bar{W} ,

Do not ejected H_0 hence there is no significant evidence that suggest that the average annual family income on Metropolis is not \$48,432.

Used the following R code

```
> #Problem 4
> # 95% confidence interval
> lower_bound = qnorm(p = 0.025, mean = 48432, sd = 2000/sqrt(400))
> lower_bound
[1] 48236
>
> upper_bound = qnorm(p = 0.975, mean = 48432, sd = 2000/sqrt(400))
> upper_bound
[1] 48628
```

Problem Statement 5

Let Y be random variable price per square foot for warehouses in the United States.

Assuming that the prices of warehouse footage are normally distributed and population standard deviation is unknown, therefore the average price per square foot for warehouses

$\bar{Y} \sim t$ distribution with $n - 1 = 18$ degrees of freedom

$$H_0 : \mu = 32.28, H_1 : \mu \neq 32.28$$

Testing at $\alpha = 0.5$ and a two tailed test, the 95% confidence interval is given by

$$\left(32.28 - t_{0.975}(18) * \frac{s}{\sqrt{19}} ; 32.28 + t_{0.975}(18) * \frac{s}{\sqrt{19}} \right)$$

where $s = 1.29$ and $\bar{w} = 31.67$

Therefore confidence interval is (31.65824 ; 32.90176)

Since mean from the sample \$31.67 does fall within the 95% confidence interval of \bar{Y} ,

Do not reject H_0 hence there is no significant evidence to suggest that average price per square foot for warehouses in the United States has now changed.

Used the following R code

```
> #Problem 5
> # 95% confidence interval
> lower_bound = 32.28 + qt(p = 0.025,df = 18)*1.29/sqrt(19)
> lower_bound
[1] 31.65824
> upper_bound = 32.28 + qt(p = 0.975,df = 18)*1.29/sqrt(19)
> upper_bound
[1] 32.90176
```

Problem Statement 6

Assuming that

$X \sim N(\mu, 2.5)$ then $\bar{X} \sim N\left(50, \frac{2.5}{\sqrt{n}}\right)$ based on acceptance regions where n is the sample size

therefore for the acceptance Region ($a < \bar{x} < b$), $\alpha = P(\bar{X} < a) + P(\bar{X} > b)$ and

$$\text{where } \mu = c, \beta = P\left(\frac{a - c}{\frac{2.5}{\sqrt{n}}} < Z < \frac{b - c}{\frac{2.5}{\sqrt{n}}}\right)$$

Acceptance Region	Sample Size	α	B at $\mu = 52$	B at $\mu = 50.5$
$48.5 < \bar{x} < 51.5$	10	0.0576	0.2643	0.8923
$48 < \bar{x} < 52$	10	0.0114	0.5000	0.9703
$48.81 < \bar{x} < 51.19$	16	0.0569	0.0975	0.8618
$48.42 < \bar{x} < 51.58$	16	0.0115	0.2508	0.9576

Used the following R code

```
> #Problem 6
> #
> # for acceptance region (48 ; 52)
> beta_1 = pnorm(q = 52,mean = 52,sd = 2.5/sqrt(10)) - pnorm(q = 48,mean = 52,sd = 2.5/sqrt(10))
> beta_1
[1] 0.4999998
>
> beta_2 = pnorm(q = 52,mean = 50.5,sd = 2.5/sqrt(10)) - pnorm(q = 48,mean = 50.5,sd = 2.5/sqrt(10))
> beta_2
[1] 0.9703275
>
> # for acceptance region (48.81 ; 51.19)
> alpha = pnorm(q = 48.81,mean = 50,sd = 2.5/sqrt(16)) + pnorm(q = 51.19,mean = 50,sd = 2.5/sqrt(16),lower.tail = F)
> alpha
[1] 0.05691018
>
```

```

> beta_1 = pnorm(q = 51.19,mean = 52,sd = 2.5/sqrt(16)) - pnorm(q = 48.81,
mean = 52,sd = 2.5/sqrt(16))
> beta_1
[1] 0.09748758
>
> beta_2 = pnorm(q = 51.19,mean = 50.5,sd = 2.5/sqrt(16)) - pnorm(q = 48.8
1,mean = 50.5,sd = 2.5/sqrt(16))
> beta_2
[1] 0.8617779
>
>
> # for acceptance region (48.42 ; 51.58)
> alpha = pnorm(q = 48.42,mean = 50,sd = 2.5/sqrt(16)) + pnorm(q = 51.58,
mean = 50,sd = 2.5/sqrt(16),lower.tail = F)
> alpha
[1] 0.01147144
>
> beta_1 = pnorm(q = 51.58,mean = 52,sd = 2.5/sqrt(16)) - pnorm(q = 48.42,
mean = 52,sd = 2.5/sqrt(16))
> beta_1
[1] 0.2507918
>
> beta_2 = pnorm(q = 51.58,mean = 50.5,sd = 2.5/sqrt(16)) - pnorm(q = 48.4
2,mean = 50.5,sd = 2.5/sqrt(16))
> beta_2
[1] 0.9575685

```

Problem Statement 7

$$\begin{aligned}
 t \text{ score} &= \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \\
 &= \frac{12 - 10}{\frac{1.5}{\sqrt{16}}} \\
 &= 5.3333
 \end{aligned}$$

Used the following R code

```

> #Problem 7
> t = (12-10)/(1.5/sqrt(16))
> t
[1] 5.333333

```

Problem Statement 8

$$\begin{aligned}
 t \text{ score} &= t_{0.99}(15) \\
 &= 2.6025
 \end{aligned}$$

Used the following R code

```

> #Problem 8
> t_score = qt(0.99,df = 15)
> t_score
[1] 2.60248

```

Problem Statement 9

Let Y be random variable number of sales of lunch packets per day.

Assuming, that the number of sales of lunch packets per day is normally distributed but population standard deviation

unknown, therefore the average number of sales of lunch packets per day

$\bar{Y} \sim t$ distribution with $n - 1 = 15$ degrees of freedom

$$H_0 : \mu = 300, H_1 : \mu > 300$$

Testing at $\alpha = 0.05$ and a two tailed test, the 95% confidence interval is given by

$$\left(-\infty ; 300 + t_{0.95}(15) * \frac{s}{\sqrt{16}} \right)$$

$$\text{where } s^2 = \frac{\sum_{i=1}^{16} (y_i^2 - \bar{y})}{16 - 1} \text{ and } \bar{y} = \frac{\sum_{i=1}^{16} y_i}{16}$$

Therefore $\bar{y} = 320$ and $s = 41.2666$ and

Therefore confidence interval is $(-\infty ; 405.3472)$

Since mean from the sample 320 does fall within the 95% confidence interval of \bar{Y} ,

Do not reject H_0 hence there is no significant evidence that suggest that the number of sales of lunch packets per day has increased.

Used the following R code

```
> #Problem 9
> #
> y = c(304, 367, 385, 386, 262, 329, 302, 292, 350, 320, 298, 258, 364, 294, 276, 333)
>
> #Average number of sales of lunch packets per day from sample
> mean_y = mean(x = y)
> mean_y
[1] 320
>
> #Sample standard deviation number of sales of lunch packets per day from sample
> std_y = sqrt(sum((y - mean_y)^2)/(16-1))
> std_y
[1] 41.26661
>
> # 95% confidence interval
> upper_bound = 300 + qt(p = 0.95, df = 15)*std_w/sqrt(16)
> upper_bound
[1] 405.3472
```