HYPOTHESIS TESTING

1. Dean is concerned the grades have changed dramatically. The graduating mean GPA over 5 years is 2.75

$\mu_{pop} = 2.75$

Sample of 256 seniors find GPA is 2.85 with SD 0.65

$M_{SD} = 2.85 \qquad \sigma_{SD} = 0.65$

a) NULL HYPOTHESIS : THE GRADES (MEAN) HAVE CHANGED DRAMATICALY OVER LAST FIVE YEARS

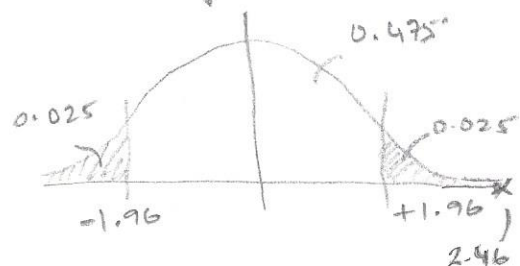ALTERNATIVE HYPOTHESIS : THE GRADES HAVE NOT CHANGED OVER LAST 5 YEARS

b) Std Error: $\sigma_{SD} = 0.65$

c) Since the Dean is worried the change and NOT exactly as 'reduced' or 'increased'

This scenario would have tails in both ends. (critical regions)

alpha level = 0.05 $[0.025, -0.025]$

[Refer to: statisticshowto.datasciencecentral.com/ what-is-an-alpha-level/]

d) Test the NULL HYPOTHESIS

For area under the curve of 0.475 Z score = 1.96

$$Z_{sample} = \frac{\bar{X} - \mu}{\sigma_{sample}} \Rightarrow \frac{2.85 - 2.75}{\frac{0.65}{\sqrt{256}}} = \frac{0.1}{\frac{0.65}{16}} = 2.46$$

When parameter is Sample mean $\bar{X}$ Std deviation $= S/\sqrt{n}$

2.46 is under the tail so GPA did change and so NULL HYPOTHESIS IS REJECTED.

2. College bookstore says average costs of text book is Rs 52

$M_{pop} = 52 \qquad \sigma_{pop} = 4.50$

To test the bookstore's claim, against their alternative, students select a random sample of size 100

$n = 100$

$M_{sample} = \bar{X} = 52.80$ Perform hypothesis test at 5% level of significance.

NULL HYPOTHESIS : AVG COST OF TEXT BOOK IS = Rs 52

ALTERNATE HYPOTHESIS : AVG COST OF TEXT BOOK IS > 52

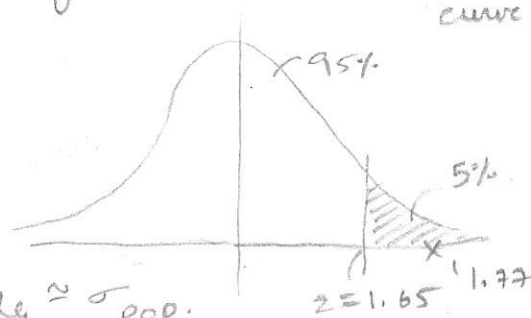Since hypothesis checks for = value the critical region lies both sides of the curve.

$\alpha = 0.05$ or 5%.

$Z_{score} = \dfrac{\overline{X} - \mu}{\sigma_{sample}/\sqrt{n}}$ $= \dfrac{52.80 - 52}{4.5/\sqrt{100}}$

$= \dfrac{0.80}{4.5/10} = \dfrac{0.8}{0.45} = 1.77$

95%

5%

$\sigma_{sample} \approx \sigma_{pop}$

$\sigma_{SD} = \dfrac{\sigma_{pop}}{\sqrt{N}}$

$z = 1.65$ 1.77

Z_{score} for 95% area under the curve = 1.65

Z falls value of the sample falls in the tail, so rejecting the NULL Hypothesis

So average book cost is greater than Rs 52.

ALTERNATE Hypothesis is ACCEPTED.

3. Chemical pollutant in river is constant $\mu = 34\,ppm$ $SD = 8\,ppm$.
The discharge has been lowered or $< 34\,ppm$.
Sample size of 50 gives a mean of $32.5\,ppm$. Perform a hypothesis test at 1% level of significance and state your decision.

$\mu = 34$ $\quad n = 50$
$\sigma = 8$ $\quad \overline{X} = 32.5$
          $\quad \sigma/\sqrt{n} = 8/\sqrt{50}$
          $\qquad\quad = 8/7.07$
          $\quad SE = 1.13$

NULL HYPOTHESIS: The discharge will be $= 34\,PPM$

ALTERNATE HYPOTHESIS: The discharge of chemicals is less than $34\,PPM$.

Since ALTERNATE HYPOTHESIS IS $< 34\,ppm$ (SAMPLE MEAN)

[0.01]
1%
$Z_S = -1.327$

-2.32

CRITICAL AREA IS ON THE LEFT SIDE OF THE CURVE.
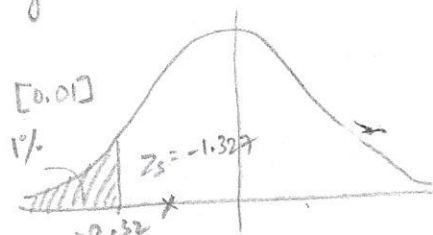
Z_{score} for 99% area under the curve $= -2.32$

$Z_{score} = \dfrac{32.5 - 34}{1.13} = \dfrac{-1.5}{1.13} = -1.327$ (Z SCORE IS WELL WITHIN THE AREA / NOT IN CRITICAL AREA)

NULL HYPOTHESIS IS TRUE. (DISCHARGE IS THE SAME)

4.   T score for Sample size 16. $\mu = 10$ $\overline{X} = 12$ $S = 1.5$ $N = 16$

$t\,score = \dfrac{\overline{X} - \mu}{S/\sqrt{n}} = \dfrac{12 - 10}{1.5/\sqrt{16}} = \dfrac{2}{1.5/4} = \dfrac{8}{1.5} = 5.33$

6. Sample of 100 voters are asked who would they vote. Does the data suggest all candidates are equally popular.

Higgins	Reardon	White	Charlton
41	19	24	16

[chi-square = 14.96, statistical significance 5%.]

	Expected	Observed	$x^2 (O-E)^2/E$
Higgins	25	41	10.24
Reardon	25	19	1.44
White	25	24	0.04
Charlton	25	16	3.24
			14.96

$x^2 = 14.96$

[5% (α)]
[7.81]
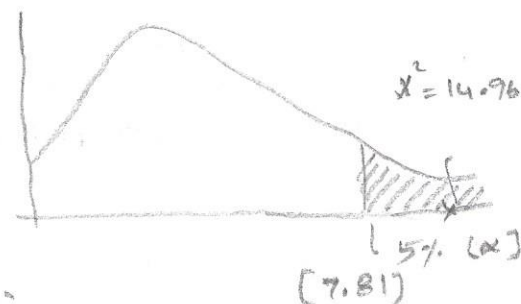
degree of freedom = 4-1 = 3

The chi square value falls under the critical region

So the ALTERNATE HYPOTHESIS ' THE CANDIDATES ARE NOT EQUALLY POPULAR' IS PROVED.

7. ANALYSIS OF VARIANCE (ANOVA) METHOD TO BE USED (NOT PART OF THE STATISTICS CURICULLUM)

5. PROBLEM IS BASED ON 'DIFFERENCE OF MEANS', SO CAN BE IGNORED - Bredhri's Comments.

6. Sample of 100 voters are asked for voting preferences in a election.

8. ALT HYPOTHESIS: Nurse thinks average height of 7th graders has increased.
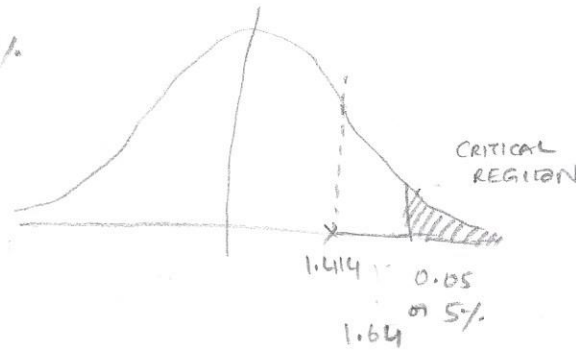
Five years ago $\mu = 145$ $\sigma = 20$

Sample $N = 200$ Z score for the area of 5%
    $\overline{X} = 147$ $= -1.65$
    $S = 20/\sqrt{200}$

CRITICAL
REGION

1.414   0.05
         or 5%
1.64

$$Z_{score}^{Sample} = \frac{147-145}{2/\sqrt{2}} = \frac{2}{2/\sqrt{2}} = \sqrt{2} = 1.414$$

The Z sample is well within the NULL region and not in the critical region
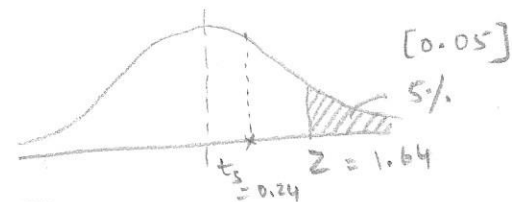
so NULL HYPOTHESIS IS PROVED

9. FARMER IS trying a planting technique which will increase the yield on pea plants.

    $\mu = 145$ pods
    $\sigma = 100$ pods

NULL HYPOTHESIS: AVG PODS = 145
ALT HYPOTHESIS: AVG PODS > 145

[0.05]
5%

$t_s = 0.24$ $Z = 1.64$

Sample:
    $\overline{X} = 147$
    $n = 144$
    $S = \frac{100}{\sqrt{n}}$

$$t_{score}^{sample} = \frac{147-145}{100/\sqrt{n}} = \frac{2}{100/\sqrt{n}} = \frac{2\sqrt{144}}{100} = \frac{12}{50} = \frac{12}{50} = 0.24$$

If the significance statistically is 5% or 0.05

5% Area $\Rightarrow Z_{score} = 1.64$ [THE TSCORE IS WELL WITHIN ACCEPTANCE AREA]

so NULL HYPOTHESIS IS PROVED, THERE IS NO SIGNIFICANT INCREASE IN AVG NO OF PODS IN A PLANT.

10. 4.5 lb slab will yield 72 ounces of cheese. $\mu = 72$
    Sample size $n = 7$ measures are $[70, 69, 73, 68, 71, 69, 71]$ $\overline{X} = 70.142$
    Are the difference due to chance or distribution is giving less. $S^2 = 2.408; S = 1.55$

    b) test statistic $t = \frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{70.142 - 72}{1.55/\sqrt{7}}$

    a) NULL HYPOTHESIS: CHANGE IS DUE TO CHANCE
       ALTERNATE: BEEN GIVEN LESS.
       at 10% SIGNIFICANT LEVEL

-1.28
-1.65
-2.32
-3.171
1% 5% 10%

$$t = \frac{-1.858}{1.55/\sqrt{7}} \Rightarrow -3.171$$

    10% = -1.28   NULL HYPOTHESIS FAILED
    5% = -1.65   NULL HYPOTHESIS FAILED
    1% = -2.32   NULL HYPOTHESIS FAILED