Hypothesis Assignments

Assignment on Hypothesis...

① UCLA DEAN GRADES

Seniors mean GPA over last 5 yrs = 2.75

Sample $n = 256$ seniors & their mean GPA = 2.85 with $\sigma_{sample} = 0.65$.

a) Null & alternative hypothesis for scenario?

Null : Grades remain same

Alternate :- Grades changed.

$\mu = 2.75$

$\bar{x} = 2.85$

$\sigma_s = 0.65$

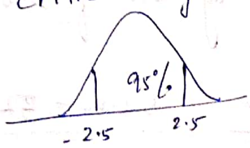
b) Standard error = $\sigma_{sample} = \frac{0.65}{\sqrt{256}} = \frac{0.65}{16}$

$= 0.04$

c) Alpha level of $.05$ = critical region of 5%

As it is 2-tailed test, hence the critical regions ~~are~~ ~~two~~ = 5%

$= \pm 1.96 \frac{\sigma}{\sqrt{n}}$



d) Test the null hypothesis.

$$Z = \frac{x - \mu}{\sigma} = \frac{2.85 - 2.75}{0.04}$$

$$= \frac{0.10}{0.04} = \frac{10}{4} = 2.5$$

- As it is inside the critical region i.e $> \pm 1.96$ hence we will reject this null hypothesis.

2) COLLEGE BOOKSTORE

Average cost of textbooks = Rs 52

S.D = Rs 4.50

Random sample of size = 100.

$\bar{x}$ (Avg mean of sample) = Rs 52.80

Null H_0 → Avg. cost of books being sold are not at a higher price.

Alternate H_1 → Avg cost of books is higher

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{52.80 - 52}{\frac{4.5}{\sqrt{100}}} = \frac{0.80 \times 10}{4.5}$$

$$= \frac{8}{4.5} = 1.77$$

For $\alpha = 0.05$, the critical value is 1.96 and z value obtained = 1.77.

As z value < 1.96, hence we accept null hypothesis that the books sold are not at a higher price.

3) CHEMICAL POLLUTANT

A certain chemical pollutant has been constant for yrs with mean = 34 ppm and SD = 8 ppm. A group of factory claimed that they have lowered the average with improved filtration devices. A group of environmentalists will test the hypothesis at 1% level of significance. Sample size = 50, mean of 32.5 ppm.

$\mu = 34$ ppm

$\sigma = 8$ ppm.

$n = 50$

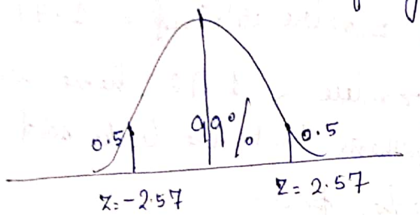
$\bar{x} = 32.5$ ppm.

$$Z = \frac{x - \mu}{\sigma/\sqrt{n}} = \frac{32.5 - 34}{8/\sqrt{50}}$$

$$= \frac{-1.5}{1.13}$$

$$= -1.327$$

Hypothesis at 1% level of significance



As z obtained value = $-1.32 > $ critical value -2.57, hence we accept H_0. Hence the claim that average mean is reduced with improved filtration mechanism is true.

④ 1-tailed test

Population proportion of traveler's check buyers who buy atleast \$2500 in checks when sweepstakes prizes are offered as atleast 10% higher than the population of such buyers when no sweepstakes are on:

Population 1 : With sweepstakes.

$n_1 = 300$

$x_1 = 120$

$\hat{p}_1 = \frac{120}{300} = 0.4$

Population 2 : No sweepstakes

$n_2 = 700$

$x_2 = 140$

$\hat{p}_2 = 0.20 = \frac{140}{700}$

$H_0 : p_1 - p_2 \leq 0.10$

$H_1 : p_1 - p_2 > 0.10$

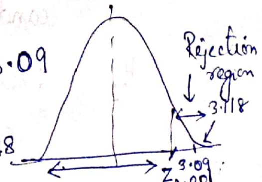
$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

$$= \frac{(0.40 - 0.20) - 0.10}{\sqrt{\frac{(0.40)\times(0.60)}{300} + \frac{(0.20)\times(0.80)}{700}}}$$

$$= \frac{0.10}{0.03207} = 3.118$$

100

Critical value of $z_{0.001} = 3.09$



As our calculated $z_{value} = 3.118$ lies in the critical region, hence we reject the null hypothesis.

⑤ Voters

Sample of 100 voters, vote for 4 candidates.

Higgins : 41

Reardon : 19

White : 24

Charlton : 16.

Does the data suggest all are equally popular? Chisq mean = 14.96 & 3 df < 0.05.

Sol^n $\chi^2 = 14.96$

df = 3.

Applying Chisquared goodness of fit test,

Step-1 H_0 : There is no preference for any candidates

H_1 : There is preference for particular candidates

Step-2 $\alpha = 0.05$ Step-3 dof = 3

Step-4 $Z_{critical} = 7.81$

Step-5 Expected frequencies = $100/4 = 25$.

Observed	Expected	(O − E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
41	25	16	256	10.24
19	25	−6	36	1.44
24	25	−1	1	0.64
16	25	−9	36	3.24

$$\chi^2 = \Sigma\left(\frac{(O-E)^2}{E}\right) = 14.96 \text{ with } 3 \text{ dof}.$$

Step-6 As $\chi^2 = 14.96 > Z_{critical} = 7.81$ (0.05 with 3 df)

hence we reject null hypothesis.

Step-7: Hence it is proved that the voters do not ~~have~~ prefer the 4 candidates equally.

6 ANOVA TEST

15 trainees are assigned to 3 different types of approaches. Use 5% significance level for the hypothesis statistic.

	A_1	A_2	A_3
Test	86	90	82
	79	76	68
	81	88	73
	70	82	71
	84	89	81
Total	400	425	375
Mean Score	80	85	75

Sol^n **Step-1** ANOVA TEST

$H_0 : \mu_1 = \mu_2 = \mu_3$

H_1 : Not all μ_s are equal.

Step-2 : State α

$\alpha = 0.05$

Step-3 Degrees of freedom

$df_{between}$, df_{within} , df_{total}

Here (n) = how many ~~tests~~ persons are conducted for each test

$n = 5$.

(N) = total no. of individuals participating

$N = 15$

$df_{between} = a - 1 = 3 - 1 = 2$

a = how many factors

$df_{within} = N - a = 15 - 3 = 12$

$df_{total} = N - 1 = 15 - 1 = 14$

Step-4 State Decision Rule

To look up critical value for $\alpha = 0.05$, we need $df_{between} = 2$ & $df_{within} = 12$

From F-table, $F_{critical} = 3.8853$.

If $F > 3.8853$, reject H_0.

Step-5 Calculate F statistic

$SS_{between}$ (Sum of squares between)

$$= \frac{\sum(\sum a_i)^2}{n} - \frac{T^2}{N}$$

$$= \frac{400^2 + 425^2 + 375^2}{5} - \frac{(400+425+375)}{15}$$

$$= \frac{481250}{5} - \frac{1200}{15}$$

$$= 96250 - 80$$

$$= 96170$$

$$SS_{within} = \sum Y^2 - \frac{\sum(\sum a_i)^2}{n}$$

$$= 86^2+79^2+81^2+70^2+84^2 + 90^2+76^2+88^2+82^2+89^2 + 82^2+68^2+73^2+71^2+81^2 - (96250)$$

$$= 96698 - 96250$$

$$= 448.$$

$$SS_{total} = \sum Y^2 - \frac{T^2}{N}$$

$$= 96698 - 80$$

$$= 96{,}618$$

	SS	df	MS	F
Between	96170	2	48085	1288.10
Within	448	12	37.33	
Total	96618	14		

$$MS_{between} = SS_{betw} / df_{betw} = 96170/2 = 48085$$

$$MS_{within} = SS_{within} / df_{within} = 448/12 = 37.33$$

$$F = MS_{between} / MS_{within} = 1288.10$$

Step-6 State Results

If $F > 3.8853$, reject H_0.

Here, $F_{actual} = 1288.10 > F_{critical} = 3.88$,

hence H_0 reject.

Step-7 $F(2, 12) = 1288.10$, $\alpha < 0.05$

the three methods were significantly different in terms of teaching.

7) SCHOOL NURSE

Nurse thinks that the average height of 7^{th} graders has increased. The average height of a 7^{th} grader 5 yrs ago = 145 cm with SD of 20 cm. She takes a random sample of 200 students & finds average height = 147 cm. Conduct one-tail hypothesis using .05 significance level.

$\mu = 145$ cm

$\sigma = 20$ cm

$n = 200$

$\bar{x} = 147$ cm

$$z_{score} = \frac{147-145}{\frac{20}{\sqrt{200}}} = \frac{2}{1.41} = 1.418$$

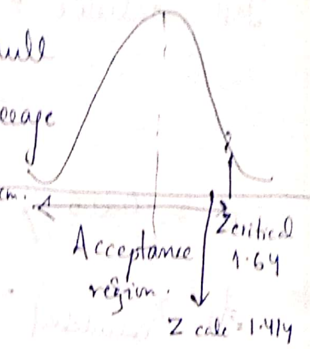
H_0 : Avg height of 7^{th} graders ~~has increased~~ is ≤ 145

H_1 : Avg height of 7^{th} graders > 145.

For $\alpha = 0.05$, $z_{critical} = 1.64$

The calculated $z_{score} = 1.414 < z_{critical} = 1.64$

Hence we accept the null hypothesis where the average height of 7^{th} graders ≤ 145 cm.



(8) FARMER

Planting technique to increase the yield of his plants.

Average no. of pods in pea = 145

SD = 100 pods.

After new planting techniques introduced average no. of pods in pea = 147 with random sample n. What is the hypothesis & test statistic?

Suppose random sample $n = 144$.

H_0 : $\mu \leq 145$

H_1 : $\mu > 145 = 147$

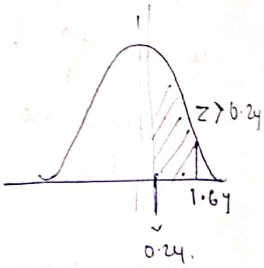
Scanned with CamScanner

Test statistic $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{147 - 145}{\frac{100}{\sqrt{144}}} \approx 0.2$

$\alpha = 0.05 \Rightarrow z_{critical} = 1.64$

As $z_{calculated} = 0.24 < z_{critical} = 1.64$, hence we accept H_0 which says new planting technique did not yield more pea pods.

$P(z > 0.24)$



$z_{0.24} = 0.5948$

$P(z > 0.24) = 1 - 0.5948$
$= 0.41$
$= 41\%$

As p value $0.41 > 0.05$ (significance), hence we don't reject null hypothesis.

⑨ PIZZA

Buy mozzarella cheese in a 4.5 pound
Mean of the cheese = 72 ounces.
Sample of 7 measurements =
70, 69, 73, 68, 71, 69, 71.

Are these differences due to chance or the distributor is giving less cheese?

a) State the hypothesis.

a) $H_0 : \mu = 72$

$H_1 : \mu \neq 72.$

b) Test statistic

$$\bar{x} = \frac{70 + 69 + 73 + 68 + 71 + 69 + 71}{7} = 70.143$$

Standard deviation $= 1.676$

Two-tailed t-test

$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{70.143 - 72}{\frac{1.676}{\sqrt{7}}} = -2.93$$

c) Would the null hypothesis be rejected at 10% level, 5% level & 1% level?

From python, &

ttest, p-value = ttest_1samp(weight_sample, 72)

p-value = 0.0262.

As p-value of weight sample $0.0262 < 0.1$ & 0.05, hence the null Hypothesis is rejected for 10% & 5%. However $0.0262 > 0.01$, so $\alpha = 1\%$ H_0 is not rejected.