Assignment II

Pbm 1

In each of the following situations, state whether it is correctly stated hypothesis testing problem and why?

(i) $H_0: \mu = 25$, $H_1: \mu \neq 25$.

Yes, correctly stated, because the alternative hypothesis specifies values of μ that could be either greater or less than 25. It is called two-sided alternative hypotheses

(ii) $H_0: \sigma > 10$, $H_1: \sigma = 10$

No. Because the H_0 and H_0 should be contradiction to each other

(iii) $H_0: \bar{x} = 50$, $H_1: \bar{x} \neq 50$.

No. Hypothesis are always statements about the population or distribution under study, not statements about the sample

(iv) $H_0: P = 0.1$, $H_1: P = 0.5$

No. Values in both hypothesis are different

(v) $H_0: S = 30$, $H_1: S > 30$

No. Hypothesis are statements about the population or distribution under study, not statement about the sample.

Pbm 2

The college bookstore tells prospective students that the average cost of its textbooks is Rs. 52 with standard deviation of Rs. 4.50. A group of smart statistics students thinks that the average cost is higher. To test the book store's claim against their alternative, the students will select a random sample of size 100. Assume that the mean from their random sample is Rs 52.80. Perform a hypothesis test at the 5% level of significance and state your decision

$$\mu = 52, \quad \sigma = 4.5. \qquad \alpha = 0.05.$$

$$n = 100, \quad \bar{x} = 52.8.$$

H_0: Average cost of the textbooks = Rs. 52

H_1: Average cost of the textbooks > Rs. 52

$$Z_{test} = \frac{\bar{x} - \mu}{SE} \qquad SE = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{4.5}{10} = \underline{\underline{0.45}}$$

$$Z_{test} = \frac{52.8 - 52}{0.45} = \underline{\underline{1.778}}$$

$$Z_{(0.05)} = -1.64$$

$$Z_{test} > Z_{(0.05)}$$

We cannot reject the null hypothesis

Pbm-3

A certain chemical pollutant in the Genesee River has been constant for several years with mean $\mu = 34$ ppm and standard deviation $\sigma = 8$ ppm. A group of factory representatives whose companies discharge liquids into river is now claiming that they have lowered the average with improved filtration devices. A group of environmentalists will test to see if this is true at the 1% level of significance. Assume that their sample of size 50 gives a mean of 32.5 ppm. Perform a hypothesis test at the 1% level of significance and state your decision

$\mu = 34$ ppm , $\sigma = 8$ ppm $\qquad \alpha = 0.01$

$n = 50$, $\bar{x} = 32.5$

H_0: Average pollution level is equal to 34 ppm

H_1: Average pollution level is less than 34 ppm

$$S.E = \frac{8}{\sqrt{50}} = \underline{\underline{1.13}} \qquad SE = \frac{\sigma}{\sqrt{n}}$$

$$Z_{test} = \frac{\bar{x} - \mu}{SE} = \frac{32.5 - 34}{1.13}$$

$$\underline{\underline{Z_{test} = -1.327}}$$

$$Z_{(0.01)} = -2.33$$

$Z_{test} > Z_{(0.01)}$ We cannot reject the null hypothesis

Pbm 4

Based on population figures and other general information on the U.S population, suppose it has been estimated that, on average, a family of four in the U.S. spends about \$1135 annually on dental expenditures. Suppose further that a regional dental association wants to test to determine if this figure is accurate for their area of country. To test this, 22 families of 4 are randomly selected from the population in that area of the country and a log is kept of the family's dental expenditure for one year. The resulting data are given below. Assuming, that dental expenditure is normally distributed in the population, Use the data and the alpha of 0.5 to test the dental association's hypothesis

1008, 812, 1117, 1323, 1308, 1415, 831, 1021, 1287, 851
930, 730, 699, 872, 913, 944, 954, 987, 1695, 995,
1003, 994.

$\mu = \$1135$

$\sigma = ?$

$n = 22$

$\bar{x} = 1031.32.$

s (std deviation) $= 234.848$

H_0: Average expenditure is equal to \$1135

H_1: Average expenditure is not equal to \$1135

$$T_{test} = \frac{SM - PM}{SE} \qquad \alpha = 0.5$$

$$SE = \frac{s}{\sqrt{n}} = \frac{234.848}{\sqrt{22}} = \underline{50.07}$$

$$T_{test} = \frac{1031.32 - 1135}{50.07}$$

$$= \underline{\underline{-2.07}}$$

$$\left.\begin{matrix} T(0.25)/2 \\ df = 21 \end{matrix}\right\} = \pm 0.686.$$

$T_{test} < -0.686$. F_{test} ×

We can reject the null hypothesis

Pbm 5

In a report prepared by the Economic Research Department of the ~~mg~~ major bank the dpt manager maintains that the average annual family income on Metropolis is \$48,432. What do you conclude about the validity of the report if a random sample of 400 families shows and average income of \$48,574 with std deviation of 2000?

$\mu = 48,434$ $\qquad \alpha = 5\%$

$n = 400.$

$\bar{x} = 48,572$

$S = 2000$

H_0: Annual family income greater than or equal to 48,432

H_1: Annual family income less than or equal to 48,432

$$SE = \frac{2000}{\sqrt{400}} = 100$$

$$T_{test} = \frac{48,574 - 48,432}{100}$$

$$= 1.42$$

$$\left. \begin{array}{l} T_{(0.05)} \\ df = 399 \end{array} \right\} = 1.96 \qquad T_{test} < T_{(0.05)}$$

We cannot reject the null hypothesis

Pbm 6

Suppose that in past years the average price per square foot for warehouses in the United States has been \$32.28. A national real estate investor wants to determine whether that figure has changed now. The investor hires a researcher who randomly samples 19 warehouses that are for sale across the US and finds that the mean price per square foot is \$31.67, with std deviation of \$1.29. Assume that the prices of warehouses footage are normally distributed in population. If the researcher uses 5% level of significance, what statistical conclusion can be reached? What are the hypothesis

$\mu = 32.28$

$n = 19$ $\qquad \alpha = 5\%$

$\bar{x} = 31.67$

$s = 1.29$

H_0: Avg price per square foot is equal to \$32.28.

H_1: Average price per square foot is not equal to 32.28.

$$SE = \frac{1.29}{\sqrt{19}} = 0.296$$

$$T_{test} = \frac{31.67 - 32.28}{0.296} = -2.061$$

$$\left.\begin{array}{l} T_{(0.05)} \\ df = 18 \end{array}\right\} = 2.101 \quad \text{two tail test}$$

$$-2.101 < T_{test} < 2.01$$

We cannot reject the null hypothesis

Pbm 8

Find the t-score for a sample size of 16 taken from a population with mean 10 when the sample mean is 12 and the sample standard deviation is 1.5

$n = 16$

$\mu = 10$

$\bar{x} = 12$

$s = 1.5$

$$SE = \frac{1.5}{\sqrt{16}} = \cancel{0.384} = 0.374$$

$$T_{test} = \frac{\bar{x} - \mu}{SE}$$

$$T_{test} = \frac{12 - 10}{0.375} = 5.33$$

Pbm 9

Find the t-score below which we can expect 99% of sample means will fall if samples of size 16 are taken from a normally distributed population

$\alpha = 1\%$

$n = 16$

$$\left.\begin{matrix} t(0.01) \\ df = 15 \end{matrix}\right\} = \underline{\underline{2.947}}$$

Pbm 10

Of a random sample of size 25 drawn from a normal population gives a mean of 60 and a standard deviation of 4, find the range of t-scores where we can expect to find the middle 95% of all sample means. Compute the probability that $(-t_{0.05} < t < t_{0.10})$.

$n = 25$

$\bar{x} = 60.$

$s = 4$

$df = 24$

$$P(-t_{0.05} < t < t_{0.10}) = 2.064 < t < 1.711$$

$$= \underline{\underline{0.353}}$$

Pbm 11

Two tailed test for difference between two population means

Is there evidence to conclude that the number of people travelling from Bangalore to Chennai is different from the number of people travelling from Bangalore to Hosur in a week, given the following

Population 1: Bangalore to Chennai $n_1 = 1200$.

$x_1 = 452$

$S_1 = 212$

Population 2: Bangalore to Hosur $n_2 = 800$.

$x_2 = 523$

$S_2 = 185$

S_1	S_2
$x_1 = 452$	$x_2 = 523$
$S_1 = 212$	$S_2 = 185$
$n_1 = 1200$	$n_2 = 800$.

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_A : \sigma_1^2 \neq \sigma_2^2$

$$F_{test} = \frac{S_A^2}{S_B^2} = \frac{44944}{34225} = 1.31$$

$df = \frac{1199}{799}$ $\alpha = 0.025$

$F_{test} > F_{crit}$ Means $\sigma_1^2 \neq \sigma_2^2$

H_0: People travelling from Bangalore to Chennai and Bangalore to Hosur are same

H_1: People travelling from Bangalore to Chennai and

Bangalore to Hosur are not same

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

$$t = \frac{452 - 523}{\sqrt{\frac{212^2}{1200} + \frac{185^2}{800}}}$$

$\alpha = 0.05$

$$= -7.926$$

$$df = \frac{\left(\frac{212^2}{1200} + \frac{185^2}{800}\right)^2}{\frac{1}{1199}(37.45)^2 + \frac{1}{799}(42.78)^2}$$

$$= 1860.89 \approx 1861$$

$t_{value}\,(0.025) = -1.96$ to $+1.96$.

$t_{test} < t_{value}$

We can reject the null hypothesis

Pbm 12

Is there evidence to conclude that the number of people preferring Duracell battery is different from the number of people preferring Energizer battery, given the following

Population 1 : Duracell

$n_1 = 100$

$x_1 = 308$

$s_1 = 84$

Population 2: Energizer

$n_2 = 100$

$x_2 = 254$

$s_2 = 67$

S_1	S_2
$x_1 = 308$	$x_2 = 254$
$s_1 = 84$	$s_2 = 67$
$n_1 = 100$	$n_2 = 100.$

$$F_{test} = \frac{84^2}{67^2} = 1.57.$$

$df = 99$ $\qquad \boxed{\sigma_1^2 \neq \sigma_2^2}$

$F_{value} = \pm 1.486$

Unpooled variance test

H_0: No: of people preferring duracell battery is equal to the no: of people preferring Energizer battery

H_1: No: of people preferring duracell battery is not equal to the no: of people preferring Energizer battery

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{308 - 254}{\sqrt{\frac{84^2}{100} + \frac{67^2}{100}}} = \underline{\underline{5.0257}}$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right) + \frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)}$$

$$= \frac{(70.56 - 44.89)^2}{\frac{70.56^2}{99} + \frac{44.89}{99}} = \underline{\underline{189}}$$

$t_{crit} = \underline{\pm 1.972}$

$t_{test} > t_{crit} \ (+1.972)$

We can reject the null hypothesis

Pbm 13

Pooled estimate of the population variance
Does the data provide sufficient evidence to conclude that average percentage increase in the price of sugar differs when it is sold at two different prices?

Population 1 : Price of sugar = Rs 27.50 $n1 = 14$

$X_1 = 0.317\%$

$S_1 = 0.12\%$

Population 2 : Price of sugar = Rs: 20:00 $n_2 = 9$.

$X_2 = 0.21\%$

$S_2 = 0.11\%$.

Ho: Average percentage increase is equal

H1: Average percentage increase is not equal.

S_1	S_2
$n_1 = 14$	$n_2 = 9$
$X_1 = 0.317$	$x_2 = 0.21$
$S_1 = 0.12$	$S_2 = 0.11$

$$Sp = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad df = n_1 + n_2 - 2$$

$$s_p = \sqrt{\frac{13 \times 0.12^2 + 8 \times 0.11^2}{21}}$$

$$= \underline{\underline{0.116}}.$$

$$t = \frac{0.317 - 0.21}{0.116 \times 0.425}$$

$$= \underline{\underline{2.17}}.$$

$t(0.05) = \pm 2.079.$

$df = 21$

$2.17 > \pm 2.079.$

We can reject the null hypothesis

Pbm 14

The manufacturers of compact disk players want to test whether a small price reduction is enough to increase the sales of their product. Is there evidence that the small price reduction is enough to increase sales of compact disk players

$n_1 = 15$

$x_1 = Rs. 6598$ $s_1 = Rs: 844$ } Population 1: Before reduction

Population 2: After Reduction

$n_2 = 12$

$x_2 = Rs\ 6870$

$s_2 = Rs\ 669$

s_1	s_2
$n_1 = 15$	$n_2 = 12$
$x_1 = 6598$	$x_2 = 6870$
$s_1 = 844$	$s_2 = 669$

$F_{test} = \underline{\underline{1.59}}$

$F_{crit}\ (0.05) = \underline{\underline{3.35}}$
df 14, 11

$F_T < F_{crit}$ $\quad \therefore \boxed{\sigma_1^2 = \sigma_2^2}$

$H_0 : \mu_2 - \mu_1 \leq 0$

$H_1 : \mu_2 - \mu_1 > 0.$

$$Sp = \sqrt{\frac{14 \times 844^2 + 11 \times 669^2}{25}}$$

$$= \underline{\underline{771.9}}$$

$$t_{test} = \frac{6870 - 6598}{771.9 \times \sqrt{\frac{1}{15} + \frac{1}{12}}}$$

$$= \underline{\underline{0.909}}$$

$\left.\begin{array}{l} t_{crit}\ (0.05) \\ df : 25 \end{array}\right\} = \underline{\underline{1.708}}$

$t_{value} < t_{crit}$

We cannot reject the null hypothesis

<u>Pbm 15</u>

Comparisons of two populations proportions when the hypothesized difference is zero. Carry out a two tailed test of the equality of bank's share of the car loan market in 1980s and 1995

Population 1 : with sweepstakes

$n_1 = 300.$

$x_1 = 120$

$\hat{p}_1 = 0.40.$

Population 2 : sweepstakes $n_2 = 700.$

$x_2 = 140$

$\hat{p}_2 = 0.20.$

Ho: Population proportion of traveler's check buyers (when sweepstakes prizes are offered) is 10% higher than the proportion of such buyers when no sweepstakes are on

H1: Population proportion of traveller's check buyers is less than or equal to the proportion of such buyers when sweepstakes are on

$H_0: \; P_1 - P_2 > 10$

$H_1: \; P_1 - P_2 \leq 0.$

$$\hat{p} = \frac{0.4 \times 300 + 0.2 \times 700}{1000}$$

$$= \underline{\underline{0.26}}$$

$$SE = \sqrt{0.26 \times 0.74 \times \frac{1}{300} + \frac{1}{700}}$$

$$= \underline{\underline{0.03}}$$

$$Z = \frac{(0.4 - 0.2 - 01}{0.03} = \frac{0.2 - 0.1}{0.3} = \underline{\underline{3.33}}$$

$$Z_{(0.05)} = \underline{\underline{-1.64}}$$

We cannot reject the null hypothesis

Population 1 : 1980.

$n_1 = 100$

$x_1 = 53$

$\hat{p}_1 = 0.53$

Population 2 : 1985

$n_2 = 100$

$x_2 = 43.$

$\hat{p}_2 = 0.43.$

$H_0 : \hat{P}_1 - \hat{P}_2 = 0$

$H_1 : \hat{P}_1 - \hat{P}_2 \neq 0.$

$$\hat{P} = \frac{\hat{P}_1 n_1 + \hat{P}_2 n_2}{n_1 + n_2} = \frac{0.53 \times 100 + 0.43 \times 100}{200}$$

$$= \underline{\underline{0.48}}$$

$$z = \frac{(\hat{P}_1 - \hat{P}_2) - 0}{SE}$$

$$SE = \sqrt{0.48 \times 0.52 \times \left(\frac{1}{100} + \frac{1}{100}\right)}$$

$$= \underline{\underline{0.07065}}$$

$$= \frac{0.53 - 0.43}{0.07065}$$

$$= \underline{\underline{1.415}}$$

$Z_{crit}\,(0.025) = \pm 1.96.$

We cannot reject the null hypothesis

Pbm 16

Carry out a one - tailed test to determine whether the population proportion of traveler's check buyers who buy at least $2500 in checks when sweepstakes prizes are offered as least 10% higher than the proportion of such buyers when no sweepstakes are on

Pbm 17

A die is thrown 132 times with the following results:
Number turned up: 1, 2, 3, 4, 5, 6.
Frequency: 16, 20, 25, 14, 29, 28
Is the die unbiased? Consider the degrees of freedom as $p^n - 1$

H_0: The die is unbiased

H_1: The die is biased

	O	E	$(O-E)^2$
1	16	22	36
2	20	22	4
3	25	22	9
4	14	22	64
5	29	22	49
6	28	22	36

$T = 132$

$E = \frac{132}{6} = 22$

$$\chi^2 = \frac{\Sigma(O-E)^2}{E}$$

$$= \frac{36+4+9+64+49+36}{22}$$

$$= 9$$

$$\left.\begin{array}{l} df = p^n - 1 \\ \quad = 3 \\ \alpha = 0.05 \end{array}\right\} \chi^2_{crit} = \underline{7.815}$$

$$\chi^2_{test} < \chi^2_{crit}$$

$\therefore$ The die is unbiased.

Pbm 18

In a certain town, there are about one million eligible voters. A simple random sample of 10,000 eligible voters was chosen to study the relationship between gender and participation in the last election. The results are summarized in the following 2×2 contingency table

	Men	Women
Voted	2792	3591
Not voted	1486	2131

We would want to check whether being a man or woman is independent having voted in the last election. In other words is "gender and voting independent"

H_0: Gender is independent of voting

H_1: Gender and voting are dependent

	Men	Women	Total
Voted	2792	3591	6383
not voted	1486	2131	3617
Total	4278	5722	10000

$$E(\text{Men who voted}) = \frac{6386 \times 4278}{10,000}$$

$$= \underline{\underline{2731.9308}}$$

$$E(\text{Women who voted}) = \frac{6383 \times 5722}{10000}$$

$$= \underline{\underline{3652.3526}}$$

$$E_{(\text{Men unvoted})} = \frac{3617 \times 4278}{10000}$$
$$= \underline{\underline{1547.3526}}$$

$$E_{(\text{Women unvoted})} = \frac{3617 \times 5722}{10000}$$
$$= \underline{\underline{2069.6474}}$$

$$\chi^2 = \frac{\sum (O-E)^2}{E} = 8$$

$$= \frac{(2792-2731)^2}{2731} + \frac{(3652-3591)^2}{3652} + \frac{(1486-1547)^2}{1547}$$

$$= \underline{\underline{4.7868}}$$

$$\chi^2_{(0.05)} = \underline{\underline{3.841}}.$$

$df = (2-1)(2-1) = 1.$

$\chi^2_{test} > \chi^2_{0.05}$ We reject the null hypothesis

Pbm.19

A sample of 100 voters are asked which of four candidates they would vote for in an election. The number supporting each candidate is given below.

Higgins	Reardon	White	Charlton
41	19	24	16.

Do the test suggest that all candidates are equally popular [chisquare = 14.96, with df 3, $P^{\wedge} < 0.05$]

H_0: All ~~the~~ candidates are equally popular

H_1: All candidates are not equally popular

$$\chi^2_{test} = 14.96.$$

$\chi^2_{(table)}$ with $df = 3$ $p < 0.05$

$$= \underline{\underline{7.815}}$$

$$\chi^2_{test} > \chi^2_{table}$$

We can reject the null hypothesis.

Pbm 20

Children of three ages are asked to indicate their preference for three photographs of adults. Do the data suggest that there is a significant relationship b/w age and photograph preference? what is wrong with this study? [Chi-square = 29.6 with 4 df: $p < 0.05$]

	Photograph		
Age of Child	A	B	C
5 - 6 years	18	22	20
7 - 8 years	2	28	40
9 - 10 years.	20	10	40.

H_0: Age and photograph preference have relationship

H_1: Age and photograph preference have no relationship

$$\chi^2_{test} = 29.6$$

$df = 4$

$(n_1-1)(n_2-1) = (3-1)\times(3-1)$

$= 4$

$p^{n} < 0.05$

$\chi^2_{table} = \underline{\underline{9.488}}$.

We can reject the null hypothesis

Pbm 21

A study of conformity using the Asch paradigm involved two conditions: One where one confederate supported the true judgement and another where no confederate gave the correct response.

	Support	No Support
Conform	18	40
Not conform	32	10

Is there a significant difference b/w the "support" and "no support" conditions in the frequency with which individuals are likely to conform (chi-square = 19.87, with 1 df: $p^{\wedge} < 0.05$)

$\chi^2_{test} = 19.87$.

$df = 1$

H_0: There is no significant difference between the "support" and "no support"

H_1: There is significant difference between the "support" and "no support"

$\chi^2_{test} = 19.87$

$df = 1$

$\chi^2_{table} = 0.001.$

Since $\chi^2_{test} > \chi^2_{table}$, the null hypothesis has been rejected

Pbm - 22

We want to test whether short people differ with respect to their leadership qualities (Genghis khan, Adolf Hitler and Napoleon were all stature - deprived, and how many midget MP's are there?) The following table shows the frequencies with which 43 short people and 52 tall people were categorized as "leaders" "followers" or as "Unclassifiable". Is there a relation-ship between height and leadership qualities

[chi-Square = 10.71, with 2df : $p < 0.01$]

	Height short	Tall.
Leader	12	32
follower	22	14
Unclassifier	9	6

H_0: height and leadership qualities are related to each other

H_1: height and leadership qualities are not related to each other.

$$\chi^2_{test} = 10.71$$

$$df = (3-1) \times (2-1)$$

$$= 2$$

$$p < 0.01 \quad (\alpha = 0.01)$$

$$\chi^2_{table} = \underline{9.210}$$

$$\chi^2_{test} > \chi^2_{table}$$

Since $\chi^2_{test} > \chi^2_{table}$, ~~height and~~ we can conclude that height and leadership qualities are not related

pbm - 23

Each respondent in the current population Survey of March 1993 was classified as employed, unemployed or outside the labour force. The results for men in California age 35 - 44 can be cross tabulated by marital status, as follows

	Married	Widowed, divorced or seperated	Never Married
Employed	679	103	114
Unemployed	63	10	20
Not in labor force	42	18	25

Men of different marital status seem to have different distributions of labor force status. Or is this just chance variation (you may assume the table results from a simple random sample)

H_0: Men of different marital status have different distribution of labour force status

H_1: Its just a chance variation

	Married	widow/divorced/ seperated	Never married	T
Employed	679	103	114	896
Unemployed	63	10	20	93
Not in labour	42	18	25	83
Total	784	131	159	1074 G.T

O	679	103	114	63	10	20	42	18	25
E	654	109	132	68	11	14	62	10	12
$(O-E)^2$	625	36	324	25	1	36	400	64	169
$\frac{(O-E)^2}{E}$	0.96	0.33	2.45	0.37	0.09	2.57	6.45	6.4	14.08.

$$\chi^2_{test} = \frac{\Sigma(O-E)^2}{E} = 33.7$$

$$\chi^2_{0.05} \quad df = 4 = 9.488.$$

Since $\chi^2_{test} > \chi^2_{0.05}$, we can reject the null hypothesis