Problem Statement 2:

The college bookstore tells prospective students that the average cost of its textbooks is Rs. 52 with a standard deviation of Rs. 4.50. A group of smart statistics students thinks that the average cost is higher. To test the bookstore'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.

Solution:

$\mu = 52$, $\sigma = 4.50$, $n = 100$, $\alpha = 0.05$, $x = 52.80$

Hypothesis,

H0: $\mu = 52$

H1: $\mu \neq 52 = \mu < 52$ OR $\mu > 52$ (two tailed test)

S.E = 4.50/(100)^0.5 = 0.45

Z = (x - μ)/S.E = (52.8 – 52)/0.45 = 1.77

Z(test) = 1.77

$Z(\alpha) = Z(0.05) = \pm 1.64$

$- 1.64 < Z(\text{test}) < 1.64$

Thus we accept the H0 null hypothesis $\mu = 52$.

Problem Statement 3:

A certain chemical pollutant in the Genesee River has been constant for several years with mean $\mu =$ 34 ppm (parts per million) and standard deviation $\sigma = 8$ ppm. A group of factory representatives whose companies discharge liquids into the 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.

Solution:

$\mu = 34$, $\sigma = 8$, $n = 50$, $x = 32.5$, $\alpha = 0.01$

Hypothesis,

H0: $\mu = 34$

H1: $\mu < 34$

S.E = 8/(50)^0.5 = 1.13

Z = (x - μ)/S.E = (32.5 – 34)/1.13 = -1.32

Z(test) = -1.32

$Z(\alpha) = Z(0.01) = -2.32$

$Z(\text{test}) > Z(0.01)$

Thus we accept the H0 null hypothesis $\mu = 34$.

Problem Statement 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 an 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

Solution:

$\mu = 1135$, $\sigma = 234.84$, $n = 22$, $x = 1031.32$, $\alpha = 0.5$, $df = 22 - 1 = 21$

Hypothesis,

H0: $\mu = 1135$

H1: $\mu \neq 1135$

S.E = 234.84/(22)^0.5 = 50.06

t(score)= (x - μ)/S.E = (1031.32 – 1135)/50.06 = -2.07

t(score) = -2.07

$t(\alpha/2) = t(0.25)$ for one tailed with df 21 = ± 0.686

t(score) is not between ± 0.686.

Thus we reject the H0 null hypothesis.

Problem Statement 5:

In a report prepared by the Economic Research Department of a major bank the Department 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 a standard deviation of 2000?

Solution:

$\mu = 48432, \sigma = 2000, n = 400, \alpha = 0.1, x = 48574$

Hypothesis,

H0: $\mu = 48432$

H1: $\mu \neq 48432 = \mu < 48432$ OR $\mu > 48432$ (two tailed test)

$S.E = 2000/(400)^{\wedge}0.5 = 100$

$Z = (x - \mu)/S.E = (48574 – 48432)/100 = 1.42$

$Z(test) = 1.42$

$Z(\alpha/2) = Z(0.05) = \pm 1.64$

$- 1.64 < Z(test) < 1.64$

Thus we accept the H0 null hypothesis $\mu = 48432$.

Problem Statement 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 United States and finds that the mean price per square foot is \$31.67, with a standard deviation of \$1.29. assume that the prices of warehouse footage are normally distributed in population. If the researcher uses a 5% level of significance, what statistical conclusion can be reached? What are the hypotheses?

Solution:

$\mu = 32.28, \sigma = 1.29, n = 19, x = 31.67, \alpha = 0.5, df = 19 – 1 = 18$

Hypothesis,

H0: $\mu = 32.28$

H1: $\mu \neq 32.28$

$S.E = 1.29/(19)^{\wedge}0.5 = 0.2959$

t(score)= $(x - \mu)/S.E = (31.67 – 32.28)/0.2959 = -2.06$

t(score) = -2.06

$t(\alpha/2) = t(0.25)$ for on one tailed with df $18 = \pm 0.688$

t(score) is not between ± 0.688.

Thus we reject the H0 null hypothesis.

Problem Statement 7:

Fill in the blank spaces in the table and draw your conclusions from it.

Acceptance region	Sample size	α	β at $\mu = 52$	β at $\mu = 50.5$
$48.5 < x < 51.5$	10			
$48 < x < 52$	10			
$48.81 < x < 51.9$	16			
$48.82 < x < 51.58$	16			

Problem Statement 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.

Solution:

$\mu = 10$, $s = 1.5$, $n = 16$, $x = 12$

S.E = 1.5/(16)^0.5 = 0.375

t(score)= $(x - \mu)$/S.E = (12 – 10)/ 0.375 = 5.3

t(score) = 5.3

Problem Statement 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.

Solution:

$1 - \alpha = 0.99$

$\alpha = 0.01$

df = n -1 = 15

t(score) = -2.602

Problem Statement 10:

If 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 $(-t0.05 < t < t0.10)$.

Solution:

Degree of freedom = n – 1 = 24 – 1 = 23

Confidence = 95%

Range of t(scores) from t table for above df and confidence is $\pm$2.069

-2.069 < t < 2.069

Probability that $(-t0.05 < t < t0.10)$ = 1 – 0.05 – 0.10 = 0.85

Problem Statement 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 n1 = 1200

x1 = 452

s1 = 212

Population 2: Bangalore to Hosur n2 = 800

x2 = 523

s2 = 185

Solution:

Hypothesis

H0: μ1 = μ2 => μ1 – μ2 = 0

H1: μ1 ≠ μ2 => μ1 – μ2 ≠ 0

S.E = ((s1)^2)/n1 * ((s2)^2)/n2 = (212^2)/1200 * (185^2)/800 = 8.95

Z = (x1 –x2) / SE = (452 – 523)/8.95 = -7.93

An extreme z-score in either tail of the distribution thus we reject the null hypothesis of no difference.

Problem Statement 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, $n1 = 100$, $x1 = 308$, $s1 = 84$

Population 2: Energizer, $n2 = 100$, $x2 = 254$, $s2 = 67$

Solution:

Hypothesis

H0: $\mu1 = \mu2 \Rightarrow \mu1 - \mu2 = 0$ => number of people preferring Duracell battery are not different from the number of people preferring Energizer battery

H1: $\mu1 \neq \mu2 \Rightarrow \mu1 - \mu2 \neq 0$ => number of people preferring Duracell battery is different from the number of people preferring Energizer battery

S.E = ((s1)^2)/n1 * ((s2)^2)/n2 = (84^2)/100 * (67^2)/100 = 115.45

Z = (x1 –x2) / SE = (308 – 254)/115.45 = 0.4677

Z(test) = 0.4677

Z(0.05/2) = Z(0.25) = 1.96

Using p-value

p-value = 2* P[Z>=|0.46|] = 2 * P[Z>=-0.46] = 2 * (1 - 0.6772) = 0.6456

we cannot reject the null hypothesis.

Problem Statement 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, x1 = 0.317%, s1 = 0.12%

Population 2: Price of sugar = Rs. 20.00, n2 = 9, x2 = 0.21%, s2 = 0.11%

Solution:

H0: $\mu1 = \mu2 \Rightarrow \mu1 - \mu2 = 0$ => average percentage increase in the price of sugar differs when it is not sold at two different prices

H1: $\mu1 \neq \mu2 \Rightarrow \mu1 - \mu2 \neq 0$ => average percentage increase in the price of sugar differs when it is sold at two different prices

Pooled variance = (SS1 + SS2) / (df1 + df1)

Pooled variance = ((df1)(s1^2) + (df2)(s2^2)) / (df1 + df2) = 0.03 / 111 = 0.012

$T = ((x1-x2) – 0) / (pooled\ variance * (1/n1 + 1/n2))^0.5$

$= (0.317 – 0.21) / (0.012 *(1/14+1/9))^0.5=2.15$

$T(score) = 2.15$

For $t(critical = 1\%)$ => for two tailed test $t(0.005, 21)$

$T(critical) = \pm 2.813$

$-2.813 < T(score) < 2.813$

Thus we accept the null hypothesis.

Problem Statement 14:

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

Population 1: Before reduction $n1 = 15$, $x1 =$ Rs. 6598 $s1 =$ Rs. 844

Population 2: After reduction $n2 = 12$, $x2 =$ RS. 6870, $s2 =$ Rs. 669

Solution:

Hypothesis

H0: $\mu 2 – \mu 1 <= 0$

H1: $\mu 2 – \mu 1 > 0$

$T = ((x2 – x1)-(u2-u1)) / ((((df1)(s1^2) + (df2)(s2^2)) / (df1 + df-2))*(1/n1 + 1/n2))^0.5$

$= 6870 – 6598 – 0 / ((14*844^2 + 11*669^2) / 15 + 12 – 2 (1/15 + 1/12))^0.5$

$= 272/298.96$

$T = 0.91$

Critical point $t(0.01) = 1.316$

Thus H0 may not be rejected at a low significance level of 1%.

Problem Statement 15:

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

Population 1: 1980

$n1 = 1000$, $x1 = 53$, $p\ 1 = 0.53$

Population 2: 1985

$n2 = 100, x2 = 43, p\ 2 = 0.53$

Solution:

Hypothesis

$H0 \Rightarrow p1 - p2 = 0$

$H1 \Rightarrow p1 - p2 \neq 0$

$p = x1 + x2 / n1 + n2 = 53+43/1000+100 = 0.48$

$Z = (p1 - p2) - 0/(p^* (1-p)^*(1/n1+1/n2))$

$= (0.53 - 0.53) - 0 / (0.48 * 0.52 * (1/1000 + 1/100))$

$= 0.10/0.0706 = 1.42$

$Z(score) = 1.42$

Z (critical at 10%=0.1)

For two tailed test 0.05

$Z(0.05) = 1.645$

$-1.645 < Z\ (score) < 1.645$

Thus we accept the H0 null hypothesis at significance level 10%

Problem Statement 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 at least 10% higher than the proportion of such buyers when no sweepstakes are on.

Population 1: With sweepstakes

$n1 = 300, x1 = 120, p\ 1 = 0.40$

Population 2: No sweepstakes

$n2 = 700, x2 = 140, p\ 2 = 0.20$

Solution:

$H0 \Rightarrow p1 - p2 <= 0.10$

$H1 \Rightarrow p1 - p2 > 0.10$

$Z = ((p1 - p2) - D) / ((p1*(1 - p1)/ n1) + (p2 * (1 - p2)/n2))$^0.5

$Z = (0.40 - 0.20) - 0.10 / (0.40 * 0.60 / 300 + 0.20 * 0.80 / 700)$^0.5

$Z = 0.10/ 0.0320$

Z = 3.11

Z(critical = 0.001) = 3.09

Z score goes out of critical region

Thus we reject the null hypothesis.

Problem Statement 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 $p - 1$.

Solution:

H0 => the die is unbiased

H1 => the die is biased

Expected frequency for each number 132/6 = 22

Degree of freedom = 6 – 1 = 5

Observed	Expected	(O –E)^2
16	22	36
20	22	4
25	22	9
14	22	64
29	22	49
28	22	36

$x^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$ = 198 / 22 = 9

$x^2 = 9$

x^2(critical at 0.05 with df 5)= 11.070

$x^2 < x^2\ (critical)$

Thus we accept the H0 null hypothesis.

Problem Statement 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 2X2 (read two by two) contingency table:

	Men	Women
Voted	2792	3591
Didn't vote	1486	2131

We would want to check whether being a man or a woman (columns) is independent of having voted in the last election (rows). In other words, is “gender and voting independent”?

Solution:

H0 = gender is independent of Voting

H1 = gender and Voting are dependent

Observed

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

Expected

	Men	Women	Total
Voted	6383 * 4278 / 10000	6383 * 5722 / 10000	
Not voted	3617 * 4278/10000	3617 * 5722 / 10000	
Total			

	Men	Women	Total
Voted	2731	3652	6383
Not voted	1547	2070	3617
Total	4278	5722	10000

$x^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$

c11 := (2792-2731)^2/2731 = 1.3625

c12 := (3591-3652)^2/3652 = 1.0188

c21 := (1486-1547)^2/1547 = 2.4053

c22 := (2131-2070)^2/2070 = 1.7975

$x^2 = \text{c11} + \text{c12} + \text{c21} + \text{c22}$

$x^2 = 6.5841$

degrees of freedom is (2-1)(2-1)=1*1=1

Let’s see the critical value using d.o.f 2 and significance 5%:

Critical chi = 3.841

$x^2 >$ Critical chi. Thus, we will reject the null hypothesis.

Problem Statement 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 data suggest that all candidates are equally popular? [Chi-Square = 14.96, with 3 df, p 0.05].

Solution:

H0 => all candidates are equally popular

H1 => all candidates are not equally popular

Expected frequencies are therefore 41+19+24+16 = 100/4 = 25 per candidate.

Observed	41	19	24	16
Expected	25	25	25	25
O – E	16	-6	-1	-9
(O – E)^2	256	36	1	81
(O – E)^2 / E	10.24	1.44	0.04	3.24

$x^2 = 10.24 + 1.44 + 0.04 + 3.24$

$x^2 = 14.96$

The critical value of Chi-Square for a 0.05 significance level and 3 d.f. = 7.82

$x^2 > Chi\ critical$

Thus we reject H0 null hypothesis

Problem Statement 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 between age and photograph preference? What is wrong with this study? [Chi-Square = 29.6, with 4 df: $p < 0.05$].

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

Solution:

H0 => there is a significant relationship between age and photograph preference

H1 => there is no significant relationship between age and photograph preference.

Observed

		Photograph			Total
		A	B	C	
Age of child	5-6 years	18	22	20	60
	7-8 years	2	28	40	70
	9-10 years:	20	10	40	70
Total		40	60	100	200

Expected

$$E = \frac{\text{(row total * column total)}}{\text{grand total}}$$

		Photograph		
		A	B	C
Age of child	5-6 years	60*40/200	60*60/200	60*100/200
	7-8 years	70*40/200	70*60/200	70*100/200
	9-10 years:	70*40/200	70*60/200	70*100/200

		Photograph		
		A	B	C
Age of child	5-6 years	12	18	30
	7-8 years	14	21	35
	9-10 years:	14	21	35

O – E

A	B	C
6	4	-10
-12	7	5
6	- 11	5

(O – E)^2 / E

A	B	C
3	0.89	3.33
10.29	2.33	0.71
2.57	5.76	0.71

x^2= 29.60

d.f. = (rows - 1) * (columns - 1) = 2 * 2 = 4.

Chi critical for a 0.001 significance level and 4 d.f. = 18.46

$x^2 > Chi\ critical$

Thus we reject the null hypothesis.

Problem Statement 21:

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

	Support	No Support
Conform	18	40
Not Conform	32	10

Is there a significant difference between the "support" and "no support" conditions in the frequency with which individuals are likely to conform? [Chi-Square = 19.87, with 1 df: $p < 0.05$].

Solution:

H0 => there is a significant difference between the "support" and "no support"

H1 => there is no significant difference between the "support" and "no support"

Df = (2-1)(2-1) = 1

Observed values

	Support	No Support	Total
Conform	18	40	58
Not Conform	32	10	42
Total	50	50	100

Expected

	Support	No Support	Total
Conform	58 * 50 / 100	58 * 50 / 100	58
Not Conform	42 * 50 / 100	42 * 50 / 100	42
Total	50	50	100

	Support	No Support
Conform	29	29
Not Conform	21	21

O – E

	Support	No Support
Conform	-11	11
Not Conform	11	-11

(O – E)^2 / E

	Support	No Support
Conform	4.17	4.17
Not Conform	5.76	5.76

$x^2 = 19.86$

Chi critical for a 0.001 significance level and 1 d.f. = 10.828

$x^2 > Chi\ critical$

Thus we reject the null hypothesis.

Problem Statement 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 relationship between height and leadership qualities? [Chi-Square = 10.71, with 2 df: $p < 0.01$]

	Height	
	Short	Tall
Leader	12	32
Follower	22	14
unclassifiable	9	6

Solution:

H0 => there is no relationship between height and leadership qualities

H1 => there is a relationship between height and leadership qualities

Observed values

	Height		Total
	Short	Tall	
Leader	12	32	44
Follower	22	14	36
unclassifiable	9	6	15
Total	43	52	95

Expected Values

	Height	
	Short	Tall
Leader	44 *43 /95 = 19.92	44*52/95 = 24.08
Follower	36*43/95 = 16.29	36*52/95 = 19.71
unclassifiable	15*43/95 = 6.79	15*52/95 = 8.21

(O – E)^2 / E

	Height	
	Short	Tall
Leader	3.146	2.602
Follower	1.998	1.652
unclassifiable	0.720	0.595

$x^2 = 10.712$

Chi critical for a 0.01 significance level and 2 d.f. = 9.210

$x^2 > Chi\ critical$

Thus we reject the null hypothesis.

Problem Statement 23:

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

	Married	Widowed/Divorced	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.)

Solution:

H0 => Men of different marital status seem to have different distributions of labor force status

H1=> this is just chance variation

DF = (3-1)(3-1) = 4

Expected Values

	Married	Widowed/Divorced	Never Married	Total
Employed	679	103	114	896
Unemployed	63	10	20	93
Not in labor force	42	18	25	85
Total	784	131	159	1074

Expected

	Married	Widowed/Divorced	Never Married
Employed	896*784/1074	869*131/1074	896*159/1074
Unemployed	93*784/1074	93*131/1074	93*159/1074
Not in labor force	85*784/1074	85*131/1074	85*159/1074

	Married	Widowed/Divorced	Never Married
Employed	654	109	133
Unemployed	68	11	14
Not in labor force	62	10	12

(O – E)^2 / E

	Married	Widowed/Divorced	Never Married
Employed	0.96	0.33	2.71
Unemployed	0.37	0.09	2.57
Not in labor force	6.45	6.4	1.08

$x^2 = 20.96$

Chi critical for a 0.01 significance level and 4 d.f. = 13.277

$x^2 > Chi\ critical$

Thus we reject the null hypothesis.