Problem Statement 1:

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 = 120

μ1 = 452

s1 = 212

Population 2: Bangalore to Hosur

n2 = 800

μ2 = 523

s2 = 185

Solution:

Given samples from two normal populations of size n1 and n2 with unknown means μ1 and μ2 and known standard deviations σ_1 and σ_2, the test statistic comparing the means is known as the two-sample z statistic

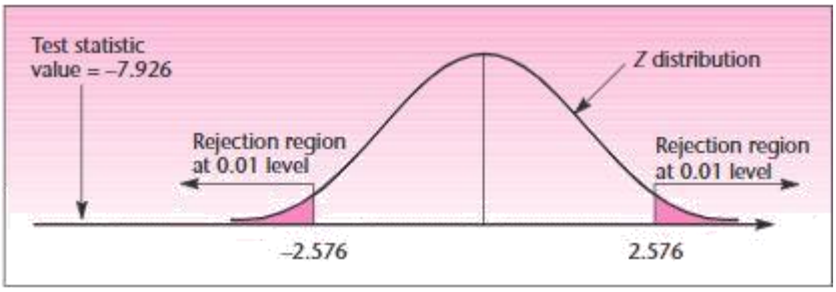
$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

T= (Observed difference in sample means) – (Hypothesized difference in population means)

Standard error

H0 = μ 1 - μ 2 =0

H1 = μ 1 - μ 2 ≠ 0



Z = (452 – 523) - 0

$\sqrt{}(212^2/120) – (185^2/800)$

=- 7.926

The computed value of the Z-statistic falls in the left-hand rejection region for any commonly used , and the p-value is very small.

We conclude that there is a statistically significant difference in the means of the population who are travelling between the Bangalore to Chennai and Housur.

Hence, We reject Null Hypothesis

Problem Statement 2:

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:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

H0 = μ 1 - μ 2 =0

H1 = μ 1 - μ 2 $\neq$ 0

Z = (308– 254) - 0

$\sqrt{}(84^2/100) – (67^2/100)$

=0.838

There is no proper evidence that people prefer Duracell over Energizer, hence we reject the null hypothesis.

Problem Statement 3:

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:

$H0 = \mu 1 - \mu 2 = 0$

$H1 = \mu 1 - \mu 2 \neq 0$

Significance level=5%

df=21 : critical value=2.080

$$t = \frac{(-0.317 - 0.210) - 0}{\frac{\sqrt{(13(0.12)^2 + (8)(0.11)^2}}{(1/14 + (1/9)\ 21}}$$

$= 2.15$

We reject the null hypothesis since, critical region should be <2.080

Statement 4:

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

$H0 = \mu 1 - \mu 2 = 0$

$H1 = \mu 1 - \mu 2 \neq 0$

Significance level=5%

df =21 : critical value=2.060

t = (6870 - 6598) - 0

$$\sqrt{(14(844)^2 + (11)(699)^2}$$

$$\overline{\qquad (1/15+ (1/12)\ 21}$$

= 0.91

This value of the statistic falls inside the non-rejection region for any usual level of significance. Hence, we accept the Null Hypothesis

Problem Statement 5:

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 ; p̂1 = 0.53

Population 2: 1985

n2 = 100; x2 = 43; p̂2= 0.53

Solution:

H0 = p 1 - p 2 =0

H1 = p 1 - p 2 ≠ 0

$$z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1 - \hat{p})(1/n_1 + 1/n_2)}}$$

=(0.53-0.43)/SQRT((0.48*(0.52)*(0.01+0.01)))

=1.41534

Problem Statement 6:

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 ; p̂1 = 0.40

Population 2: No sweepstakes

n2 = 700 ; x2 = 140 ; p̂2= 0.20

H0:p1-p2<=0.10

H1:p1-p2>0.10

z at 1% is 2.33

$$z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}$$

=((0.4-0.2)-0.1)/SQRT(((0.4*0.6)/300)+(0.2*0.8)/700)

=3.1180

So rejecting the null hypothesis

Problem Statement 7:

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 p̂− 1.

Solution:

Problem:

Number turned up: 1, 2, 3, 4, 5, 6

Frequency: 16, 20, 25, 14, 29, 28

Hypothesis:

h0:dice is biased

h1:dice is unbiased

df : 5

Significance level:5% : 11.07

freq	EF(np)	(Obs-ef)2	X2cal
16	22	36	9
20	22	4	
25	22	9	
14	22	64	
29	22	49	
28	22	36	
		198	

Since the calculated value is less than observed value, so fail to reject null Hypothesis

---

Problem Statement 8:

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
Not voted	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:

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

Calculation of Chi square	
(o-E)^2	Chi
1.37848	6.660455899
2.432633	
1.030607	
1.818736	

Workings:		
Expected	2730.6474	3652.353
DF:(2-1)*(2-1)=1	1547.3526	2069.647

Inference: Since p-value is between 0 and 1 % so we reject the null hypothesis

---

Problem Statement 9:

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, $\hat{p} < 0.05$].

Solution:

DF=(4-1)=3		
~~Significance level=5 %~~		
Chi table value=7.82		
Workings		
Expected(np)	(O-E)^2	(o-E)^2/E
25	256	10.24
	36	1.44
	1	0.04
	81	3.24
		14.96

Inference: Our obtained Chi-Square value 14.96 >7.82, and so we conclude that people do not prefer candidate equally

---

Problem Statement 10:

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: $\hat{p} < 0.05$].

###		**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:

A	B	C	Overall
18	22	20	60
2	28	40	70
20	10	40	70
40	60	100	200

Workings : DF=(3-1)*(3-1)=4							
Expected							
	Table			(o-E)^2/E			chi
12	18		30	3	0.8888889	3.3333333	29.6
14	21		35	10.285714	2.3333333	0.7142857	
14	21		35	2.5714286	5.7619048	0.7142857	

Inference: Critical value at .01 significance level and df 4 is 13.28 which is less than our calculated value. Hence, we find no significant relationship.

Problem Statement 11:

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 between the "support" and "no support" conditions in the frequency with which individuals are likely to conform? [Chi-Square = 19.87, with 1 df: p̂< 0.05].

Solution:

	support	no support	Overall
confirm	18	40	58
not confirm	32	10	42
	50	50	100

Calculations: DF = 1

Expected		(|o-E|-0.5)^2		(|o-E|-0.5)^2 /E		chi-Square
29	29	110.25	110.25	3.801724	3.801724	18.1
21	21	110.25	110.25	5.25	5.25	

Inference: Calculated chi square value is higher than chi square table value at .01 significance level at df 1(6.63) ,hence, this concludes that there is statistical difference between support and no support situation.

Problem Statement 12:

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: p̂< 0.01].

##	**Height**	
	Short	Tall
Leader	12	32
Follower	22	14
Unclassifiable	9	6

Solution:

Short	Tall	Overall
12	32	44
22	14	36
9	6	15
43	52	95

Calculations: for df:(3-1)*(2-1)=2

Expected Table		(o-E)^2/E		Chi
19.92	24.08	3.146233	2.601693	10.712
16.29	19.71	1.997579	1.651844	
6.79	8.21	0.719706	0.595142	

<u>Inference</u>: Since the calculated value of Chi-square value 10.712 is higher than critical value at df 2 at significance level .01.

Hence we conclude that there is a significant relationship between height and leadership.

Problem Statement 13:

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 or separated	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:

Married	Widowed	Never Married	Overall
679	103	114	896
63	10	20	93
42	18	25	85
784	131	159	1074

Calculations:

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

At significance level 1% =13.28

Expected Table			(o-E)^2/E			Chi Square
654.0633	109.2886	132.648	0.950731	0.361858	2.621596	31.61
67.88827	11.34358	13.76816	0.351978	0.159138	2.820703	
62.04842	10.36778	12.5838	6.477829	5.618435	12.25084	

Inference: Since the calculated value is way greater than at 1%, we conclude there is some dependency between marital and job

---