### Problem Statement 1:

In each of the following situations, state whether it is a correctly stated hypothesis

testing problem and why?

1. H0: μ = 25, H1: μ ≠ 25

2. H0: σ > 10, H1: σ = 10

3. H0: x = 50, H1: x ≠ 50

4. H0: p = 0.1, H1: p = 0.5

5. H0: s = 30, H1: s > 30

1. Yes.

μ = 25 is the null hypothesis which is a statement that is assumed to be true.

μ ≠ 25 is the alternative hypothesis which is a statement that contradicts the null hypothesis.

Because the alternative hypothesis specifies that the value could be either < or > 25, it is a two-sided alternative hypothesis.

1. No.

H0 which is the null hypothesis, is supposed to contain “=” sign as it is always an equality claim. However, when the alternative hypothesis is stated with the < sign, the implicit claim is the null hypothesis can be taken as >= and when it is > sign, the null hypothesis can be taken as <=.

1. No.

Here, the hypothesis is given for a sample distribution. But this is incorrect. Hypotheses are always statements of the population not the sample.

1. No.

Value in both hypotheses is different. Alternate hypothesis does not exactly contradict null hypothesis.

1. No.

Here, the hypothesis is given for a sample distribution. But this is incorrect. Hypotheses are always statements of the population not the sample.

### 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 and state your decision.

Ans:

One-tailed hypothesis test

Hypothesis

H0: μ = Rs 52,

H1: μ > Rs 52

Significance level

𝜶 = 0.05

=> z score = 1.645 using z table

μ = 52

σ = 4.5

n = 100

x = 52.8

Z test

z = (x – μ) / (σ / √n)

= 52.8 - 52 / 4.5 / √100

= 8/ 4.5

= 1.78

If the z test falls in the rejection region, we can reject the null hypothesis.

But 1.78 > 1.645 i.e it falls into the rejection region.

Therefore, we reject null hypothesis, and accept the alternative hypothesis.

In other words, the average cost is indeed higher than Rs 52.

### Problem Statement 3:

A certain chemical pollutant in the Genesee River has been constant for several

years with mean μ = 34 ppm (parts per million) and standard deviation σ = 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.

Ans:

One-tailed hypothesis test

Hypothesis

H0: μ = 34 ppm,

H1: μ < 34 ppm

Significance level

𝜶 = 0.01

=> z score = -2.33 using z table

μ = 34

σ = 8

n = 50

x = 32.5

Z test

z = (x – μ) / (σ / √n)

= 32.5 - 34 / 8/ √50

= -1.33

-1.33 lies in the acceptance region.

Therefore, we cannot reject the null hypothesis.

We accept null hypothesis, i.e. we cannot believe the claim that the mean pollutant is < 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

Hypothesis

H0: μ = 1135,

H1: μ !=1135

Significance level

𝜶 = 0.5

Use t-test since sample size < 30 and population variance σ is not known.

μ = 1135

s = 240.3745859246388

n = 22

x= 1031.3181818181818

T statistic

t = (x – μ) / s / √n)

= 1031.3181818181818 -1135/ 240.375 / √22

= -2.02

Degrees of freedom = 22-1 = 21

Referring to t-table two tailed test we get P(0.5,21) = 2.080

-2.02 lies inside acceptance region of -2.08 < -2.02 < 2.08

Therefore, we accept the null hypothesis.

We accept null hypothesis, i.e. we believe the claim that the mean annual dental expenditure for a family of four is about $1135.

### 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?

Hypothesis

H0: μ = $48432,

H1: μ != $48432

Significance level

𝜶 = 0.5 (assume) => -1.96 < z < 1.96

μ = 48432

s = 2000

n = 400 ( much more than 30, use z test)

x= 48574

Z test

z = (x – μ) / (s / √n)

= 48574 - 48432 / 2000/ √400

= 1.42

-1.96< 1.42 < 1.96, lies in the acceptance region.

Therefore, we accept null hypothesis that average annual family income on Metropolis is $48,432.

### 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?

Hypothesis

H0: μ = $32.28,

H1: μ != $32.28

Significance level

𝜶 = 0.5 (assume)

Referring to t-table two tailed test we get P(0.5,18) = 2.101

μ = 32.28

s = 1.29

n = 19 ( less than 30, use t test)

x= 31.67

T statistic

t = (x – μ) / (s / √n)

= (31.67 – 32.28) / 1.29 / √19)

= -0.61 / 1.29/4.36

= -0.61 / 0.2958 = -2.06

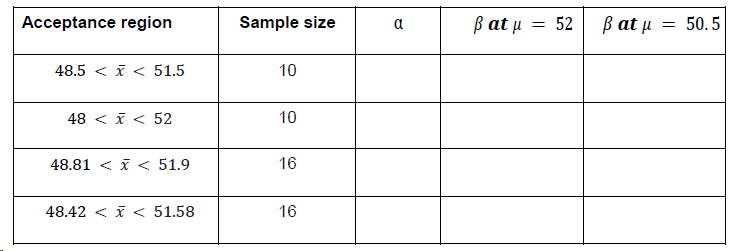
-2.06 lies in the acceptance region. -2.101 < -2.06 < 2.101.

There we accept the null hypothesis that the price per square foot for warehouses is $32.28.

### 

### Problem Statement 7:

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



α (Alpha) is the probability of Type I error in any hypothesis test–incorrectly rejecting the null hypothesis.

β (Beta) is the probability of Type II error in any hypothesis test–incorrectly failing to reject the null hypothesis. (1 – β is power).

Df = 10 - 1 = 9

For a t-distribution, variance = df / df - 2

= 9 / 7

s = sqrt(9/7) = sqrt (1.285714286) = 1.133893419 = approx. 1.13

t = (x – μ) / (s / √n)

1. 48.5 < x < 51.5

Given μ = 52

=> 48.5 - 52 / 1.13 / sqrt(10) < t-score < 51.5-52 / 1.13/sqrt(10)

=> -3.5/1.13/3.16 < t < 0.5 / 1.13/3.16

=> -9.7876 < t < 1.398

β = P(t1.4) - P (t-9.79) = 0.1 - 0.0005 = 0.0995

Given μ = 50.5

=> 48.5 - 50.5 / 1.13 / sqrt(10) < t-score < 51.5-50.5 / 1.13/sqrt(10)

=> -2/1.13/3.16 < t < 1 / 1.13/3.16

=> -5.6 < t < 2.8

β = P(t2.8) - P (t-5.6) = approx. 0.01 - 0.0005 = 0.0095

To calculate α, Assume μ = 50

=> 48.5 - 50 / 1.13 / sqrt(10) < t-score < 51.5-50 / 1.13/sqrt(10)

=> -4.19 < t < + 4.19 -> acceptance region

so , α = (approx.) = 0.002 ( referring to 2 tailed t table)

1. 48 < x < 52

Given μ = 52

=> 48 - 52 / 1.13 / sqrt(10) < t-score < 52-52 / 1.13/sqrt(10)

=> -4/1.13/3.16 < t < 0

=> -11.18< t < 0

β = P(t0) - P (t-11.18) = approx. 0.5- 0.0005 = 0.4995

Given μ = 50.5

=> 48 - 50.5 / 1.13 / sqrt(10) < t-score < 52-50.5 / 1.13/sqrt(10)

=> -2.5/1.13/3.16 < t < 1.5 / 1.13/3.16

=> -2.5/0.36 < t < 1.5 / 0.36

=> -6.94 < t < 4.16

β = P(t4.16) - P (t-6.94) = approx. 0.001 - 0.0005 = =5×10−4

Assume μ = 50

=> 48 - 50 / 1.13 / sqrt(10) < t-score < 52-50 / 1.13/sqrt(10)

=> -5.55 < t < + 5.55 -> acceptance region

so , α = (approx.) = 0.001 ( referring to 2 tailed t table)

Parts III. and IV.

Df = 16- 1 = 15

For a t-distribution, variance = df / df - 2

= 15 / 13

s = sqrt(15/13) = sqrt (1.154) = 1.133893419 = approx. 1.074

1. 48.81 < x < 51.9

Given μ = 52

=> 48.81- 52 / 1.074 / sqrt(16) < t-score < 51.9-52 / 1.074 / sqrt(16)

=> -3.19/0.2685 < t < -0.1/0.2685

=> -11.88 < t < -0.372

β = P(t-0.372) - P (t-11.88) = approx. 0.35

Given μ = 50.5

=> 48.81- 50.5 / 1.074 / sqrt(16) < t-score < 51.9-50.5 / 1.074 / sqrt(16)

=> -1.69/0.2685 < t < 1.4/0.2685

=> -6.3 < t < 5.21

β = P(t5.21) - P (t-6.3) = approx. 0.001

Assume μ = 50.355

=> 48.81 - 50.355 / 1.074 / sqrt(16) < t-score < 51.9-50.355 / 1.074/sqrt(16)

=>+5.75 < t < + 5.75 -> acceptance region

so , α = (approx.) = 0.001 ( referring to 2 tailed t table)

1. 48.42 < x < 51.58

### 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.

μ = 10

s = 1.5

n = 16 ( less than 30, use t test)

x= 12

t = (x – μ) / (s / √n)

T-score = 12-10 / 1.5 √4 = 2/1.5 x 2 = ⅔ = 0.67

### 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.

1 - 𝜶 = 0.99 =>

𝜶 = 0.01

Df = n-1 = 16-1 = 15

T (0.99) = - T (0.01) = -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).

s = 4

n = 25 ( less than 30, use t test)

x= 60

Significance level

𝜶 = 1-0.95 = 0.05

Df = 25 -1 = 24

From t-table for two tailed test

T(0.05,24) = 2.064

Range of t-scores where we can expect to find

the middle 95% of all sample means = -2.064 <t < 2.064

Compute the probability that (−𝑡0.05 <𝑡<𝑡0.10).

From t-table

t(0.10) = t(0.10,24) = 1.318

-t(0.05) = - t(0.05,24) = -1.711

Prob(−𝑡0.05 <𝑡<𝑡0.10) = Prob(-1.711 < t < 1.318) = 1 - 0.05 - 0.1 = 1 - 0.15 = 0.85

= 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

Hypotheses:

H0: μ1 – μ2 = 0, which is the same as H0: μ1 = μ2

Ha: μ1 – μ2 ≠ 0, which is the same as Ha: μ1 ≠ μ2

Significance level (assume)

𝜶 = 0.05

T = (452-523)

------------------------------------

√(212)^2 / 1200 + (185)^2 / 800

T score = -71 / (31.45 + 42.78) = -71 / 74.23

= -0.956

P(|t| > = 0.956 ) = 0.339361

> 0.05

=> We will fail to reject the null hypothesis.

This suggests that the number of people travelling from Bangalore to Chennai is not different from the number of people.

### 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

Hypotheses:

H0: μ1 – μ2 = 0, which is the same as H0: μ1 = μ2

Ha: μ1 – μ2 ≠ 0, which is the same as Ha: μ1 ≠ μ2

Significance level (assume)

𝜶 = 0.05

T = (308-254)

------------------------------------

√(84)^2 / 100 + (67)^2 / 100

= 54 / 10.75 = 5.023

The p-value is < .00001.

=> We will reject the null hypothesis.

This suggests that there is evidence to conclude that the number of people preferring Duracell battery is different from the number of people preferring Energizer battery.

### 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%

Hypotheses:

H0: μ1 – μ2 = 0, which is the same as H0: μ1 = μ2

Ha: μ1 – μ2 ≠ 0, which is the same as Ha: μ1 ≠ μ2

Df = 14 + 9 -2 = 21

Pooled variance formula:



N = 14, M = 9

Sx = 0.12 x 27.5 = 3.3

Sy = 0.11 x 20 = 2.2

Sp^2 = (13)(3.3)^2 + (8)(2.2)^2

---------------------------------

14+9 - 2

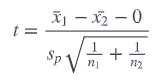
= 141.57 + 77.44

----------------------

21

= 10.43

Sp = 3.23

T score = 

X1 = 0.317 x 27.5 = 8.72

X2 = 0.21 x 20 = 4.2

T = 4.52

----------

3.23 ( sqrt (0.0714 + 0.1111) )

= 4.52

----------

3.23 (0.43)

= 4.52 / 1.39

T = 3.25

Significance level (assume)

𝜶 = 0.05

The p-value is .003831 < 0.05

=> We will reject the null hypothesis.

This suggests that the data provides sufficient evidence to conclude that the average percentage increase in the price of sugar differs when it is sold at two different prices.

### 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

Df = n1 + n2 - 2

Df = 15 + 12 -2 = 25

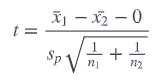
Hypotheses:

H0: μ2 – μ1 <= 0

Ha: μ2 – μ1 > 0, which is the same as Ha: μ1 ≠ μ2

Pooled variance formula:





T score = 0.91

Significance level (assume)

𝜶 = 0.05

Acceptance region (one tailed test) : -1.708 < t < 1.708

0.91 falls in the acceptance region.

=> We will accept the null hypothesis. This means that sligh price reduction is not enough to increase sales of compact disk players.

### 

### 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

Hypotheses:

H0: p1 – p2 = 0, which is the same as H0: p1 = p2

Ha: p1 – p2 ≠ 0, which is the same as Ha: p1 ≠ p2

P = x1 + x2

-----------

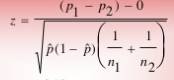
N1 + n2

= 53 + 43

-----------

1000 + 100

= 96 / 1100 = 0.087



Z - score = 0

Z(0.05) = 1.645

0 < 1.645. It is in the acceptance region.

Therefore, we will accept the null hypothesis. I.e. banks’ share in terms of proportion of the car loan market in 1980 and 1995 is equal among the two years.

### 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 = 0.40

Population 2: No sweepstakes n2 = 700

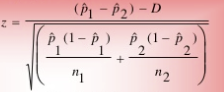
x2 = 140

𝑝 2= 0.20

Hypotheses:

H0: p1 – p2 <= 0.10, which is the same as H0: p1 = p2

Ha: p1 – p2 > 0.10, which is the same as Ha: p1 ≠ p2



Z = 3.118

Z(0.05) = 1.645

Z(0.001) = 3.09

=> Null hypothesis is rejected at any common level of significance.

I.e. 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.

### 

### 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 .

Null Hypothesis: The die is unbiased.

On the basis of hypothesis that the die is unbiased, we expect each number to turn up,

132/6=22 times

Perform chi-squared test:

|  |  |  |  |
| --- | --- | --- | --- |
| O | E | (O-E)2 | (O-E)2 / E |
| 16 | 22 | 36 | 1.63636364 |
| 20 | 22 | 4 | 0.18181818 |
| 25 | 22 | 9 | 0.40909091 |
| 14 | 22 | 64 | 2.90909091 |
| 29 | 22 | 49 | 2.22727273 |
| 28 | 22 | 36 | 1.63636364 |
|  |  |  |  |
|  |  | Sum | 9 |

Df = n-1 = 6-1 =5

Confidence level = 5%

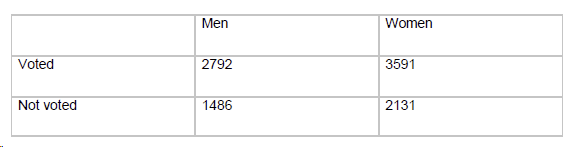
Chi-square(table) = 11.07

9 < 11.07

=> Accept null hypothesis i.e. die is unbiased.

### 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:



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”?

•Null hypothesis – sex is independent of voting

•Alternative – sex and voting are dependent

Observed

|  |  |  |  |
| --- | --- | --- | --- |
|  | Men | Women |  |
| Voted | 2792 | 3591 | 6383 |
| Not Voted | 1486 | 2131 | 3617 |
|  | 4278 | 5722 | 10000 |

Expected = To find the values, do row count x column count / total grand sum (10000)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Men | Women |  |
| Voted | 2731 | 3652 | 6383 |
| Not Voted | 1547 | 2070 | 3617 |
|  | 4278 | 5722 | 10000 |

Chi-square = Sq(O – e) / e

(2792 – 2731)2 = 1.36

2731

(3591 – 3652)2 = 1.0

3652

Etc . X2 = 1.4+1.0+2.4+1.8 = 6.6

Degrees of freedom 2 – 1 = 1

Taking confidence interval = 0.05

chi-square(table) = 3.84

6.6 > 3.84 => We reject the null hypothesis.

The data supports the hypothesis that sex and voting are dependent in this town.

### 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:



Do the data suggest that all candidates are equally popular? [Chi-Square = 14.96, with 3 df, 𝑝 < 0.05 .

Goodness-of-Fit test:

The null hypothesis is that all candidates are equally popular.

Expected frequencies are therefore 100/4 = 25 per candidate.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| O | 41 | 19 | 24 | 16 |
| E | 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 |

Adding together the last row gives us our value of c2 :

(O - E)2

å ----------------- = 10.24+ 1.44 + 0.04 + 3.24 = 14.96, with 4 - 1 = 3 degrees of freedom.

E

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

14.96 > 7.82 => We reject the null hypothesis.

The data suggests that voters do not prefer the four candidates equally.

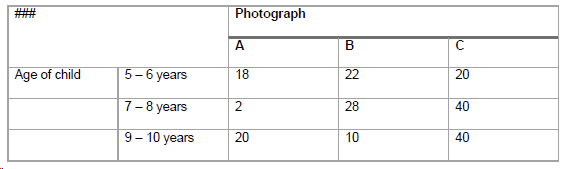
### Problem Statement 20:

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].



Calculate totals:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| photograph: |  |  |  |  |
| age of child: | A: | B: | C: | row totals: |
| 5-6 years | 18 | 22 | 20 | 60 |
| 7-8 years | 2 | 28 | 40 | 70 |
| 9-10 years | 20 | 10 | 40 | 70 |
| column totals: | 40 | 60 | 100 | 200 |

Expected frequency

(row total \* column total)

E = --------------------------------------

grand total

For each cell of the above table, this gives us:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| O: | 18 | 22 | 20 | 2 | 28 | 40 | 20 | 10 | 40 |
| E: | 12 | 18 | 30 | 14 | 21 | 35 | 14 | 21 | 35 |

Next, work out (O - E):

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (O-E): | 6 | 4 | -10 | -12 | 7 | 5 | 6 | 11 | 5 |

Square each of these, to get (O - E)2 :

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (O - E)2: | 36 | 16 | 100 | 144 | 49 | 25 | 36 | 121 | 25 |

Divide each of the above numbers by E, to get (O - E)2 / E:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (O - E)2  ----------  E | 3 | 0.89 | 3.33 | 10.29 | 2.33 | 0.71 | 2.57 | 5.76 | 0.71 |

Chi-squared is the sum of these:

c2 = 29.60 (verified)

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

Chi-Square (table) for a 0.001 significance level and 4 d.f. is 18.46.

29.6 > 18.46

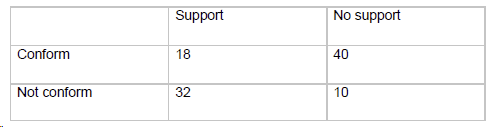
=>We accept the alternative hypothesis, i.e.there is a significant relationship between age of child and photograph preference.

### Problem Statement 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.



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].

Given

|  |  |  |  |
| --- | --- | --- | --- |
|  | support | no support | row totals: |
| conform: | 18 | 40 | 58 |
| not conform: | 32 | 10 | 42 |
| column totals: | 50 | 50 | 100 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| O: | 18 | 40 | 32 | 10 |
| E: | 29 | 29 | 21 | 21 |

Next, work out (O - E):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| (|O-E|): | 11 | 11 | 11 | 11 |

Square each of these, to get (O - E)2 :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| (|O-E|)2: | 121 | 121 | 121 | 121 |

Divide each of the above numbers by E, to get (O - E)2 / E:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| (O - E)2  -----------  E | 4.17 | 4.17 | 5.76 | 5.76 |

Chi-squared is the sum of these:

c2 = 19.86

d.f. = (rows - 1) \* (columns - 1) = 1 \* 1 = 1.

Chi-Square (table with significance level 0.01) = 6.63

19.86 > 6.63 => We reject the null hypothesis.

i.e. there is a significant difference between the "support" and "no support" conditions, in terms of the frequency with which individuals conformed.

### Problem Statement 22:

We want to test whether short people differ with respect to their leadership qualities

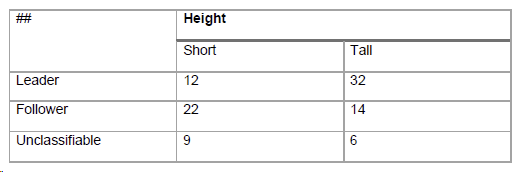
(Genghis Khan, Adolf Hitler and Napoleon were all stature-deprived, and how many midget

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]*.*



Null hypothesis is that there is no relationship between height and leadership qualities.

Expected frequencies are in brackets:

|  |  |  |  |
| --- | --- | --- | --- |
| **height:** |  |  |  |
|  | **short** | **tall** | **row totals:** |
| **leader:** | **12 (19.92)** | **32 (24.08)** | **44** |
| **follower:** | **22 (16.29)** | **14 (19.71)** | **36** |
| **unclassifiable:** | **9 (6.79)** | **6 (8.21)** | **15** |
| **column totals:** | **43** | **52** | **95** |

Chi-Square = 3.146 + 2.602 + 1.998 + 1.652 + 0.720 + 0.595 = 10.712, with 2 d.f.

Chi-Square (table at significance level 0.01) = 9.21

10.712 > 9.21

=> We reject null hypothesis.

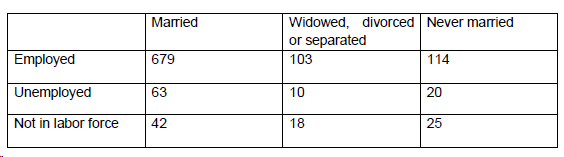
We accept alternative hypothesis that there seems to be a relationship between height and leadership qualities.

### 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:



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.)

Null hypothesis - men of different marital status do not have different distributions of labor force status

Alternative hypothesis - men of different marital status have different distributions of labor force status

Observed

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

Expected

|  |  |  |  |
| --- | --- | --- | --- |
|  | Married | Widowed/Divorced/Separated | Never Married |
| Employed | 655 | 110 | 133 |
| Unemployed | 68 | 12 | 14 |
| Not in labor force | 61 | 9 | 12 |

|  |  |  |  |
| --- | --- | --- | --- |
| O | E | (O-E) square | (O-E) square /E |
| 679 | 655 | 576 | 0.879389313 |
| 63 | 68 | 25 | 0.367647059 |
| 42 | 61 | 361 | 5.918032787 |
| 103 | 110 | 49 | 0.445454545 |
| 10 | 12 | 4 | 0.333333333 |
| 18 | 9 | 81 | 9 |
| 114 | 133 | 361 | 2.714285714 |
| 20 | 14 | 36 | 2.571428571 |
| 25 | 12 | 169 | 14.08333333 |
|  |  |  |  |
|  |  | Chi-square | 36.31290466 |

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

Chi-Square (table with significance level 0.01) = 13.28

36.312 > 13.28

=> We reject null hypothesis.

This means that men of different marital status have different distributions of labor force status.