**Assignment stats 2**

Q1. 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

Ans. 1. 𝐻0: 𝜇 = 25, 𝐻1: 𝜇 ≠ 25

It is correctly stated hypothesis as the null hypothesis contradicts the alternate hypothesis.

2. 𝐻0: 𝜎 > 10, 𝐻1: 𝜎 = 10

It is incorrectly stated hypothesis as the null hypothesis will always state as an equality claim.

However, when the alternate hypothesis is stated with the < sign, the implicit claim can be taken

as ≥ and when the alternate hypothesis is stated with the > sign, the implicit claim can be taken

as ≤.

3. 𝐻0: x = 50, 𝐻1: x ≠ 50

It is incorrectly stated hypothesis as the hypotheses are always statements about the population or

distribution under study and not statements about the sample.

4. 𝐻0: p = 0.1, 𝐻1: p = 0.5

It is incorrectly stated hypothesis as the values in both hypothesis is different.

5. 𝐻0: s = 30, 𝐻1: s > 30

It is incorrectly stated hypothesis as the hypotheses are always statements about the population or

distribution under study and not statements about the sample.

Q2. The college bookstore tells prospective students that the average cost of its

textbooks are Rs.52 with a standard deviation of Rs.4.50. A group of smart statistics

students think 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.

Ans. Population mean (μ) = 52

Population standard deviation (σ) = 4.50

Sample mean ( x ) = 52.80

Significance level (α) = 5% = 0.05

Sample size (n) = 100

H0: average cost of textbooks = 52

H1: average cost of textbooks > 52

Here, the sample is large and the population variance is known but since, we don’t know about the

normality of the data, we will use the Z-test from the table above.

Z(test) = x - μ

σ /

= (52.80 - 52)/0.45 = 1.78

Let’s find out the critical value at 5% significance level using the Critical value table.

Z (0.05) = 1.64 (since it is right tailed test).

We can clearly see that Z(test) > Z (0.05), that means our test value lie in the rejection region.

Thus, we can reject the null hypothesis i.e. the bookstore’s claim with 5% significance level.

**Using p-value to test the above hypothesis:**

p-value = P[Z>=1.78] = 1 - 0.9625 = 0.0375

i.e. p-value < 5% significance level and we are right in rejecting the null hypothesis.

Q3. 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. Population mean (μ) = 34

Population standard deviation (σ) = 8

Sample mean ( x ) = 32.5

Significance level (α) = 1% = 0.01

Sample size (n) = 50

H0: average chemical pollutant in the Genesee river = 34

H1: average chemical pollutant in the Genesee river < 34

Here, the sample is large and the population variance is known but since, we don’t know about the

normality of the data, we will use the Z-test from the table above.

Z(test) = x - μ

σ /

= (32.5 - 34)/0.8 = -1.33

Let’s find out the critical value at 1% significance level using the Critical value table.

Z (-0.01) = -2.33 (since it is left tailed test).

We can clearly see that Z(test) < Z (-0.01), that means our test value doesn’t lie in the rejection

region.

Thus, we can’t reject the null hypothesis with 1% significance level.

**Using p-value to test the above hypothesis:**

p-value = P[Z<=-1.33] = 1 - 0.9082 = 0.0918

i.e. p-value > 1% significance level and we are right in not rejecting the null hypothesis.

Q4. 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 0.05 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

Ans. Population mean (μ) = 1135

Sample mean ( x ) = 1031.32

Sample standard deviation(s) = 240.37

Significance level (α) = 5% = 0.05

Sample size (n) = 22

H0: average dental expenditure of a family = 1135

H1 average dental expenditure of a family < 1135

Here, the sample is small and the population variance is unknown and the sample is normally

distributed, we will use t-test.

t(test) = x - μ

s /

= (1031.32 - 1135)/51.25 = -2.02

Degree of Freedom = n-1 = 22-1 = 21

Let’s find out the critical value at 5% significance level using the t- value table.

t (-0.05) = -1.72 (since it is left tailed test).

We can clearly see that t(test) < t (-0.05), means our test value lie in rejection region.

Thus, we can reject the null hypothesis with 5% significance level.

**Using p-value to test the above hypothesis:**

p-value = P[t<=-2.02] i.e. 0.025 < p < 0.05

i.e. p-value < 5% significance level and we are right in rejecting the null hypothesis.

Q5. 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?

Ans. Population mean (μ) = 48432

Sample mean ( x ) = 48574

Sample standard deviation(s) = 2000

Significance level (α) = 5% = 0.05

Sample size (n) = 400

H0: average annual family income= 48432

H1 average annual family income > 48432

Here, the sample is large and the population variance, normality of the sample is unknown

so we will use z-test.

Z(test) = x - μ

s /

= (48574 - 48432)/100 = 1.42

Let’s find out the critical value at 5% significance level using the Critical value table.

Z (0.05) = 1.64 (since it is right tailed test).

We can clearly see that Z(test) < Z (0.05), means our test value lie in the non-rejection region.

Thus, we can’t reject the null hypothesis with 5% significance level.

**Using p-value to test the above hypothesis:**

p-value = P[Z>=1.42] = 1 - 0.9222 = 0.0778

i.e. p-value > 5% significance level and we are right in not rejecting the null hypothesis.

Q6. 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?

Ans. Population mean (μ) = 32.38

Sample mean ( x ) = 31.67

Sample standard deviation(s) = 1.29

Significance level (α) = 5% = 0.05

Sample size (n) = 19

H0: average price per ft. for warehouses= 32.38

H1: average price per ft. for warehouses < 32.28

Here, the sample is small and the population variance is unknown and the sample is normally

distributed, we will use t-test.

t(test) = x - μ

s /

= (31.67 – 32.38)/100 = -2.44

Degree of Freedom = n-1 = 19-1 = 18

Let’s find out the critical value at 5% significance level using the t- value table.

t (-0.05) = -1.734 (since it is left tailed test).

We can clearly see that t(test) < t (-0.05), means our test value lie in the rejection region.

Thus, we can reject the null hypothesis with 5% significance level.

**Using p-value to test the above hypothesis:**

p-value = P[t<=-2.44] i.e. 0.01 < p < 0.025

i.e. p-value < 5% significance level and we are right in rejecting the null hypothesis.

Q7. Fill in the blank spaces in the table and draw your conclusions from it.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Acceptance region | Sample size | α (µ = 52, µ= 50.5) | β at µ = 52 | β at µ = 50.5 |
| 48.5 < x < 51.5 | 10 | 0.04, 1 | 0.96 | 0 |
| 48 < x < 52 | 10 | 0.5, 1 | 0.5 | 0 |
| 48.81 < x < 51.9 | 16 | 0.34, 1 | 0.66 | 0 |
| 48.42 < x < 51.58 | 16 | 0.04, 1 | 0.96 | 0 |

Case-1

When µ = 52, β > α i.e. probability of having the Type-II error is more as compared to Type-I error or we can say probability of not rejecting the null hypothesis when it is false is more as compared to probability of rejecting the null hypothesis when it is true.

Case-2

When µ = 50.5, β < α where α = 1, β = 0 i.e. probability of having the Type-I error is 100% as compared to Type-II error which is 0% or we can say probability of not rejecting the null hypothesis when it is false is 0% as compared to probability of rejecting the null hypothesis when it is true which is 100%.

Case-1 can still be acceptable but Case-2 can’t be applied in any situations as it can be dangerous.

Q8. 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.

Ans. Population mean (μ) = 10

Sample mean ( x ) = 12

Sample standard deviation(s) = 1.5

Sample size (n) = 16

t(test) = x - μ

s /

= (12–10)/0.375 = 5.33

i.e. t-score = 5.33

Q9. 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.

Ans. Sample size(n) = 16

1 - α = 0.99

α = 0.01

Degree of Freedom = n - 1 = 16 - 1 = 15

t-score using the t-value table,

t(0.99) = -t(0.01) = -2.602

Q10. 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).

Ans. Sample mean ( x ) = 60

Sample standard deviation(s) = 4

Sample size (n) = 25

Significance level (α) = 5% = 0.05

Degree of Freedom = n - 1 = 25 - 1 = 24

Range of t-scores that lie in the middle 95% of all sample means can be found out with the help of

t-value table,

t-scores = (-2.064, 2.064)

P (- t(0.05) < t < t(0.10)) = - P ( t > t(0.05)) + P ( t < t(0.10))

= - P ( t > t(0.05)) + (1- P ( t > t(0.10)))

= - 0.05 + (1- 0.10)

= 0.85

Q11. 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

Ans. n1 = 1200 n2 = 800

x1 = 452 x2 = 523

s1 = 212 s2 = 185 Significance level(α) = 5% = 0.05

Let avg. no. of people travelling from Bangalore to Chennai = µ1

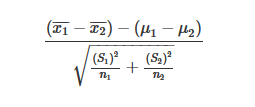
& avg. no. of people travelling from Bangalore to Hosur = µ2

H0: µ1 = µ2

H1: µ1 µ2

Since n1 and n2 are independent and both are large, we use the Z-test

Z(test) =



= 452 – 523 - 0/ 8.95 = -7.93

Since, it is a two tailed test, we need to find critical value for 2.5% on each tail.

Z (-2.5%) = -1.96 and Z (2.5%) = 1.96

We can clearly see that Z(test) < Z (-2.5%), means our test value lie in the rejection region.

Thus, we can reject the null hypothesis with 5% significance level.

**Using p-value to test the above hypothesis:**

p-value = 2\* P[Z<=-7.93] i.e. p = 0 i.e. can be rejected at any significance level

i.e. p-value < 5% significance level and we are right in rejecting the null hypothesis.

Q12. 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

Ans. n1 = 100 n2 = 100

x1 = 308 x2 = 254

s1 = 84 s2 = 67 Significance level(α) = 5% = 0.05

Let avg. no. of people preferring Duracell battery = µ1

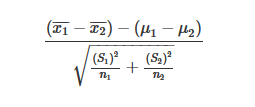
& avg. no. of people preferring Energizer battery = µ2

H0: µ1 = µ2

H1: µ1 µ2

Since n1 and n2 are independent and both are large, we use the Z-test

Z(test) =



= 308 - 254 - 0/ 10.74 = 5.02

Since, it is a two tailed test, we need to find critical value for 2.5% on each tail.

Z (-2.5%) = -1.96 and Z (2.5%) = 1.96

We can clearly see that Z(test) > Z (2.5%), means our test value lie in the rejection region.

Thus, we can reject the null hypothesis with 5% significance level.

**Using p-value to test the above hypothesis:**

p-value = 2\* P[Z>=5.02] = 2\*(1 - 1) = 0 i.e. can be rejected at any significance level

i.e. p-value < 5% significance level and we are right in rejecting the null hypothesis.

Q13. 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%

Ans. n1 = 14 n2 = 9 x1 = 0.317 x2 = 0.21

s1 = 0.12 s2 = 0.11 Significance level(α) = 5% = 0.05

Let avg. % increase when price of sugar is 27.50= µ1

& avg. % increase when price of sugar is 20 = µ2

H0: µ1 = µ2

H1: µ1 µ2

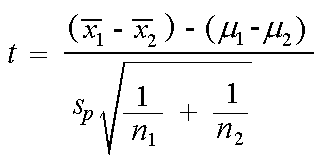
Since n1 and n2 are independent and both are small, we use the t-test

Pooled estimate (sp2) = (n1 - 1)s12 + (n2 - 1)s22  = 0.013

n1 + n2 - 2

Degree of Freedom = n1 + n2 - 2 = 14 + 9 - 2 = 21

t(test) =



= 0.317 - 0.21 - 0/0.05 = 2.14

Since, it is a two tailed test, we need to find critical value for 2.5% on each tail.

t (-2.5%) = -2.08 and t (2.5%) = 2.08

We can clearly see that t(test) > t (2.5%), means our test value lie in the rejection region.

Thus, we can reject the null hypothesis with 5% significance level.

**Using p-value to test the above hypothesis:**

p-value = P[t>=2.14] = 0.02 < p < 0.05

i.e. p-value < 5% significance level and we are right in rejecting the null hypothesis.

Q14. 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

Ans. n1 = 15 n2 = 12 x1 = 6598 x2 = 6870

s1 = 844 s2 = 669 Significance level(α) = 5% = 0.05

Let avg. sales before reduction of price = µ1

& avg. sales after reduction of price = µ2

H0: µ1 ≥ µ2

H1: µ1 < µ2

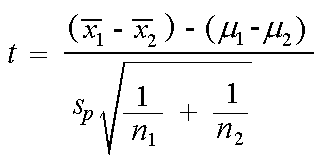
Since n1 and n2 are independent and both are small, we use the t-test

Pooled estimate (sp2) = (n1 - 1)s12 + (n2 - 1)s22  = 9972704 + 4923171

n1 + n2 - 2

Degree of Freedom = n1 + n2 - 2 = 15 + 12 - 2 = 25

t(test) =



= 6598 - 6870 - 0/298.96 = -0.9

Let’s find out the critical value at 5% significance level using the t- value table.

t (-0.05) = -1.708 (since it is left tailed test).

We can clearly see that t(test) > t (-0.05), means our test value lie in the non-rejection region.

Thus, we can’t reject the null hypothesis with 5% significance level.

**Using p-value to test the above hypothesis:**

p-value = P[t<=-0.9] = 0.15 < p < 0.20

i.e. p-value > 5% significance level and we are right in not rejecting the null hypothesis.

Q15. 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

Ans. n1 = 1000 n2 = 100 x1 = 53 x2 = 43

𝑝^1 = 0.53 𝑝^2= 0.53 Significance level(α) = 5% = 0.05

Let avg. bank share of car loan market in 1980= µ1

& avg. bank share of car loan market in 1985 = µ2

H0: µ1 = µ2

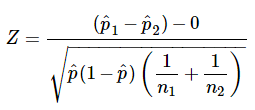
H1: µ1 ≠ µ2

Since n1 and n2 are independent and both are large, we use the z-test

= x1 + x2 = 0.08

n1 + n2

z(test) =



= 0 - 0- 0/0.0008 = 0

Since, it is a two tailed test, we need to find critical value for 2.5% on each tail.

z (-2.5%) = -1.96 and t (2.5%) = 1.96

We can see that t (0.05) < t(test) < t (-0.05), means our test value lie in the non-rejection region.

Thus, we can’t reject the null hypothesis with 5% significance level.

**Using p-value to test the above hypothesis:**

p-value = 2\*P[t>=0] = 2\*0.5 = 1 i.e. can be accepted at any significance level

i.e. p-value > 5% significance level and we are right in not rejecting the null hypothesis.

Q16. 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

1 = 0.40

Population 2: No sweepstakes n2 = 700

x2 = 140

2 = 0.20

Ans. n1 = 300 n2 = 700 x1 = 120 x2 = 140

1 = 0.40 2 = 0.20 Significance level(α) = 5% = 0.05

Let avg. buyers with sweepstakes = µ1

& avg. buyers without sweepstakes = µ2

H0: µ1 - µ2 ≤ 0.10

H1: µ1 - µ2 > 0.10

Since n1 and n2 are independent and both are large, we use the z-test

z(test)= 1 -2 -D

= 0.40 - 0.20 – 0.10/0.03207 = 3.11

Let’s find out the critical value at 5% significance level using the Critical value table.

z (0.05) = 1.64 (since it is right tailed test).

We can see that z (0.05) > z(test) means our test value lie in the rejection region.

Thus, we can reject the null hypothesis with 5% significance level.

**Using p-value to test the above hypothesis:**

p-value = P[z>=3.11] = 0 i.e. can be rejected at any significance level

i.e. p-value < 5% significance level and we are right in rejecting the null hypothesis.

Q17. 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 – 1.

Ans.

|  |  |  |
| --- | --- | --- |
| Observed Frequency(O) | Expected Frequency(E) | (O-E)2 |
| 16 | 22 | 36 |
| 20 | 22 | 4 |
| 25 | 22 | 9 |
| 14 | 22 | 36 |
| 29 | 22 | 25 |
| 28 | 22 | 36 |
|  | Total | 110 |

H0: The die is unbiased

H1: The die is biased

ᵡcal2  = 2

= 110/ 22 = 5

Level of significance = 5% = 0.05

Degree of Freedom = - 1 = 6 - 1 = 5

Critical Value(ᵡ∝)2 = 11.0705

Since ᵡcal2 < (ᵡ∝)2

H0 is accepted

i.e. the die is unbiased

Q18. 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:

Observed Table

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

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”?

Ans. Expected value (Men who voted) = 6383\*4278/ 10000 = 2730

Expected value (Men who not voted) = 3617\*4278/ 10000 = 1547

Expected value (Women who voted) = 6383\*5722/ 10000 = 3652

Expected value (Women who not voted) = 3617\*5722/ 10000 = 2069

Expected Table

|  |  |  |  |
| --- | --- | --- | --- |
|  | Men | Women | Total |
| Voted | 2731 | 3652 | 6382 |
| Not Voted | 1547 | 2070 | 3617 |
| Total | 4278 | 5722 | 10000 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Observed Frequency(O) | Expected Frequency(F) | (O – E)2/ E |
| Men (Voted) | 2792 | 2731 | 1.3625 |
| Men (Not Voted) | 1486 | 1547 | 2.4053 |
| Women (Voted) | 3591 | 3652 | 1.0188 |
| Women (Not Voted) | 2131 | 2070 | 1.7975 |
| Total |  |  | 6.5841 |

H0: gender and voting are independent

H1: gender and voting are dependent

ᵡcal2  = 2

= 6.5841

Level of significance = 5% = 0.05

Degree of Freedom = (2 - 1) (2 - 1) = 1\*1 = 1

Critical Value(ᵡ∝)2 = 3.841

Since ᵡcal2 > (ᵡ∝)2

H0 is rejected

i.e. gender and voting are dependent

Q19. 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, < 0.05] .

Ans.

|  |  |  |
| --- | --- | --- |
| Observed Frequency(O) | Expected Frequency(E) | (O-E)2 |
| 41 | 25 | 256 |
| 19 | 25 | 36 |
| 24 | 25 | 1 |
| 16 | 25 | 81 |
|  | Total | 374 |

H0: Candidates are equally popular

H1: Candidates are not equally popular

ᵡcal2  = 2

= 374/ 25 = 14.96

Degree of Freedom = 3

Level of Significance ( < 0.05 and we take = 0.01

Critical Value(ᵡ∝)2 = 11.345

Since ᵡcal2 > (ᵡ∝)2

H0 is rejected

i.e. Candidates are not equally popular

Q20. 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: < 0.05].

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | Photograph | | |  |
| A | B | C | Total |
| 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 |

Observed Table

Ans. Expected Value (Photograph A by 5-6 years) = 40\*60 /200 = 12

Expected Value (Photograph B by 5-6 years) = 60\*60 /200 = 18

Expected Value (Photograph C by 5-6 years) = 100\*60 /200 = 30

Expected Value (Photograph A by 7-8 years) = 40\*70 /200 = 14

Expected Value (Photograph B by 7- 8years) = 60\*70 /200 = 21

Expected Value (Photograph C by 7-8 years) = 100\*70 /200 = 35

Expected Value (Photograph A by 9-10 years) = 40\*70 /200 = 14

Expected Value (Photograph B by 9-10 years) = 60\*70 /200 = 21

Expected Value (Photograph C by 9-10 years) = 100\*70 /200 = 35

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | Photograph | | |  |
| A | B | C | Total |
| Age of Child | 5 - 6 years | 12 | 18 | 30 | 60 |
|  | 7 - 8 years | 14 | 21 | 35 | 70 |
|  | 9 - 10 years | 14 | 21 | 35 | 70 |
| Total |  | 40 | 60 | 100 | 200 |

Expected Table

|  |  |  |  |
| --- | --- | --- | --- |
|  | Observed Frequency(O) | Expected Frequency(E) | (O-E)2/E |
| Photograph A by 5-6 years | 18 | 12 | 3 |
| Photograph B by 5-6 years | 22 | 18 | 0.89 |
| Photograph C by 5-6 years | 20 | 30 | 3.33 |
| Photograph A by 7-8 years | 2 | 14 | 10.29 |
| Photograph B by 7-8 years | 28 | 21 | 2.33 |
| Photograph C by 7-8 years | 40 | 35 | 0.71 |
| Photograph A by 9-10 years | 20 | 14 | 2.57 |
| Photograph B by 9-10 years | 10 | 21 | 5.77 |
| Photograph C by 9-10 years | 40 | 35 | 0.71 |
| Total |  |  | 29.6 |

H0: there isn’t a significant relationship between age and photograph

H1: there is a significant relationship between age and photograph

ᵡcal2  = 2

= 29.6

Degree of Freedom = 4

Level of Significance ( < 0.05 and we take = 0.01

Critical Value(ᵡ∝)2 = 13.277

Since ᵡcal2 > (ᵡ∝)2

H0 is rejected

i.e. there is a significant relationship between age and photograph

In this study, the preference of photograph A in the age group 7 - 8 years is very low as compared to

the photographs and also to the other age groups.

Q21. 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: < 0.05].

Observed Table

|  |  |  |  |
| --- | --- | --- | --- |
|  | Support | No support | Total |
| Conform | 18 | 40 | 58 |
| No Conform | 32 | 10 | 42 |
| Total | 50 | 50 | 100 |

Ans. Expected value (Support & Conform) = 58\*50 /100 = 29

Expected value (No Support & Conform) = 58\*50 /100 = 29

Expected value (Support & No Conform) = 42\*50 /100 = 21

Expected value (No Support & No Conform) = 42\*50 /100 = 21

Expected Table

|  |  |  |  |
| --- | --- | --- | --- |
|  | Support | No support | Total |
| Conform | 29 | 29 | 58 |
| No Conform | 21 | 21 | 42 |
| Total | 50 | 50 | 100 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Observed Frequency(O) | Expected Frequency(E) | (O-E)2/E |
| Support & Conform | 18 | 29 | 4.173 |
| No Support & Conform | 40 | 29 | 4.173 |
| Support & No Conform | 32 | 21 | 5.762 |
| No Support & No Conform | 10 | 21 | 5.762 |
| Total |  |  | 19.87 |

H0: there isn’t significant difference between the support and no support in true judgement

H1: there is a significant difference between the support and no support in true judgement

ᵡcal2  = 2

= 19.87

Degree of Freedom = 1

Level of Significance ( < 0.05 and we take = 0.01

Critical Value(ᵡ∝)2 = 6.635

Since ᵡcal2 > (ᵡ∝)2

H0 is rejected

i.e. there is a significant difference between the support and no support in true judgement

Q22. 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: < 0.01].

Observed Table

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

Ans. Expected value (Leader is short) = 44\*43 /95 = 19.92

Expected value (Leader is tall) = 44\*52 /95 = 24.08

Expected value (Follower is short) = 36\*43 /95 = 16.29

Expected value (Follower is tall) = 36\*52 /95 = 19.71

Expected value (Unclassifiable is short) = 15\*43 /95 = 6.79

Expected value (Unclassifiable is tall) = 15\*52 /95 = 8.21

Expected Table

|  |  |  |  |
| --- | --- | --- | --- |
|  | Height | |  |
| Short | Tall | Total |
| Leader | 19.92 | 24.08 | 44 |
| Follower | 16.29 | 19.71 | 36 |
| Unclassifiable | 6.79 | 8.21 | 15 |
| Total | 43 | 52 | 95 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Observed Frequency(O) | Expected Frequency(E) | (O-E)2/E |
| Leader is short | 12 | 19.92 | 3.146 |
| Leader is tall | 32 | 24.08 | 2.602 |
| Follower is short | 22 | 16.79 | 1.998 |
| Follower is tall | 14 | 19.71 | 1.652 |
| Unclassifiable is short | 9 | 6.79 | 0.72 |
| Unclassifiable is tall | 6 | 8.21 | 0.595 |
| Total |  |  | 10.712 |

ᵡcal2  = 2

= 10.712

H0: there isn’t relationship between height and leadership qualities

H1: there is a relationship between height and leadership qualities

Degree of Freedom = 2

Level of Significance ( = 0.01 and we take = 0.01

Critical Value(ᵡ∝)2 = 9.21

Since ᵡcal2 > (ᵡ∝)2

H0 is rejected

i.e. there is a relationship between height and leadership qualities

Q23. 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.)

Observed Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Married | Widowed, divorced or separated | 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 |

Ans. Expected value - Men (married & employed) = 784\*896/ 1074 = 654

Expected value - Men (widowed/divorced/separated & employed) = 131\*896/ 1074 = 109

Expected value - Men (never married & employed) = 159\*896/ 1074 = 133

Expected value - Men (married & unemployed) = 784\*93/ 1074 = 68

Expected value - Men (widowed/divorced/separated & unemployed) = 131\*93/ 1074 = 11

Expected value - Men (never married & unemployed) = 159\*93/ 1074 = 14

Expected value - Men (married & not in labor force) = 784\*85/ 1074 = 62

Expected value - Men (widowed/divorced/separated & not in labor force) = 131\*85/ 1074 = 10

Expected value - Men (never married & not in labor force) = 159\*85/ 1074 = 13

Expected Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Married | Widowed, divorced or separated | Never married | Total |
| Employed | 654 | 109 | 133 | 896 |
| Unemployed | 68 | 11 | 14 | 93 |
| Not in labor force | 62 | 11 | 12 | 85 |
| Total | 784 | 131 | 159 | 1074 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Observed Frequency(O) | Expected Frequency(E) | (O – E)2/ E |
| Men (married & employed) | 679 | 654 | 0.95 |
| Men (widowed/divorced/separated & employed) | 103 | 109 | 0.33 |
| Men (never married & employed) | 114 | 133 | 2.71 |
| Men (married & employed) | 63 | 68 | 0.37 |
| Men (widowed/divorced/separated & employed) | 10 | 11 | 0.09 |
| Men (never married & employed) | 20 | 14 | 2.57 |
| Men (married & employed) | 42 | 62 | 6.45 |
| Men (widowed/divorced/separated & employed) | 18 | 11 | 4.45 |
| Men (never married & employed) | 25 | 12 | 14.08 |
| Total |  |  | 32 |

ᵡcal2  = 2

= 32

H0: there isn’t relationship between marital and labor force status

H1: there is a relationship between marital and labor force status

Degree of Freedom = (3 - 1) (3 - 1) = 2\*2 = 4

Level of Significance ( = 0.05

Critical Value(ᵡ∝)2 = 9.488

Since ᵡcal2 > (ᵡ∝)2

H0 is rejected

i.e. there is a relationship between marital and labor force status