Inferences concerning Proportions

Estimations of Proportions:

The estimation of a proportion is the number of times, X, that an appropriate event occurs in n trails, occasions, or observations. The point estimator of the population proportion, itself, is usually the sample proportion $p = \frac{X}{n}$, namely, the proportion of the time that the event actually occurs.

Large sample confidence interval for proportion 'p' is

$$\frac{x}{n} - z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}} < p < \frac{x}{n} + z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$$

Where the degree of confidence is $(1-\alpha)100\%$.

The proportion of success $p = \frac{x}{n}$

Problem 1: If x = 36 of n = 100 persons interviewed are familiar with the tax incentives for installing certain energy-saving devices, construct a 95% confidence interval for the corresponding true proportion.

Solution: Given that x =36 and n = 100 and confidence is 95%

Then α=0.05 and $z_{\alpha/2} = z_{0.025} = 1.96$

Using above values in the confidence interval for p,

$$\frac{x}{n} - z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}} < p < \frac{x}{n} + z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$$

$$\frac{36}{100} - 1.96\sqrt{\frac{\frac{36}{100}\left(1-\frac{36}{100}\right)}{100}} < p < \frac{36}{100} + 1.96\sqrt{\frac{\frac{36}{100}\left(1-\frac{36}{100}\right)}{100}}$$

$$0.266 < p < 0.454$$

Note that the maximum error of estimate $E = z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$

Here $p = \frac{x}{n}$

Problem 2: In a sample survey conducted in a large city, 136 of 400 persons answered yes to the question of whether their city's public transportation is adequate. With 99% confidence, what can we say about the maximum error, if $\frac{x}{n} = \frac{136}{400} = 0.34$ is used as an estimate of the corresponding true proportion?

Solution: Since $\frac{x}{n} = \frac{136}{400} = 0.34$ and confidence is 99%, then α = 0.01 and $z_{\alpha/2} = z_{0.005} = 2.575$.

Then maximum error of estimate $E = z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$

$$E = 2.575\sqrt{\frac{(0.34)(0.66)}{400}} = 0.061$$

Note: To find the sample size when sample proportion is known is

$$n = p(1-p)\left[\frac{z_{\alpha/2}}{E}\right]^2 \quad \text{and}$$

When sample proportion p is unknown, $n = \frac{1}{4}\left[\frac{z_{\alpha/2}}{E}\right]^2$ (take p = 1/ 2)

Problem 3: What is the size of the smallest sample required to estimate an unknown proportion of customers who would pay for an additional service, to within a maximum error of 0.06 with at least 95% confidence?

Solution: From the given data maximum error of estimate E = 0.06,

Confidence is 95%, α= 0.05 and hence $z_{\alpha/2} = z_{0.025} = 1.96$

Then sample size $n = \frac{1}{4}\left[\frac{z_{\alpha/2}}{E}\right]^2$

$$=\frac{1}{4}\left[\frac{1.96}{0.06}\right]^2=266.77\approx 267$$

Problem 4: In a random sample of 200 claims filed against an insurance company writing collision insurance on cars, 84 exceeded \$3,500. Construct a 95% confidence interval for the true proportion of claims filed against this insurance company that exceed \$3,500, using the large sample confidence interval formula.

Solution: Given that x =84 and n = 200 and confidence is 95%

Then α=0.05 and $z_{\alpha/2}=z_{0.025}=1.96$

Using above values in the confidence interval for p,

$$\frac{x}{n}-z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}<p<\frac{x}{n}+z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$$

$$\frac{84}{200}-1.96\sqrt{\frac{\frac{84}{200}\left(1-\frac{84}{200}\right)}{200}}<p<\frac{84}{200}+1.96\sqrt{\frac{\frac{84}{200}\left(1-\frac{84}{200}\right)}{200}}$$
$$0.3516<p<0.4200$$

Problem 5: In a random sample of 200 claims filed against an insurance company writing collision insurance on cars, 84 exceeded \$3,500. What we say with 99% confidence about the maximum error if we use the sample proportion as an estimate of the true proportion of claim field against this insurance company.

Solution: we know that maximum error of estimate $E=z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$

Here sample size n = 200, x = 84 and confidence is 99%.

Hence α = 1- 0.99 = 0.01

Then $z_{\alpha/2}=z_{0.005}=2.575$

The maximum error of estimate is $E=z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$

$$= 2.575\sqrt{\frac{\frac{84}{200}\left(1-\frac{84}{200}\right)}{200}}$$

$$= 0.08987$$

Problem 6: In a random sample of 400 industrial accidents, it was found that 231 were due at least partially to unsafe working conditions. Construct a 99% confidence interval for the corresponding true proportion using the large sample confidence interval formula.

Solution: Large sample confidence interval for p is

$$\frac{x}{n} - z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}} < p < \frac{x}{n} + z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$$

Where the degree of confidence is (1-α)100%.

From the given data n = 400, x = 231, α = 1-0.99= 0.01, $z_{\alpha/2} = z_{0.005} = 2.575$

Then 99% confidence interval is

$$\frac{231}{400} - 2.575\sqrt{\frac{\frac{231}{400}\left(1-\frac{231}{400}\right)}{400}} < p < \frac{231}{400} + 2.575\sqrt{\frac{\frac{231}{400}\left(1-\frac{231}{400}\right)}{400}}$$

$$0.5775 - 0.0636 < p < 0.5775 + 0.0636$$

$$0.5139 < p < 0.6411$$

Problem 7: In a random sample of 90 sections of pipe in a chemical plant, 15 showed signs of serious corrosion. Construct a 95% confidence interval for the true proportion of pipe sections showing signs of serious corrosion, using the large sample confidence interval forumula.

Solution: Large sample confidence interval for p is

$$\frac{x}{n} - z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}} < p < \frac{x}{n} + z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$$

Where the degree of confidence is (1-α)100%.

From the given data n = 90,   x = 15, α = 1-0.95= 0.05, $z_{\alpha/2} = z_{0.025} = 1.96$

Then 99% confidence interval is

$$\frac{15}{90} - 1.96\sqrt{\frac{\frac{15}{90}\left(1-\frac{15}{90}\right)}{90}} < p < \frac{15}{90} + 1.96\sqrt{\frac{\frac{51}{90}\left(1-\frac{15}{90}\right)}{90}}$$

$$0.1667 - 0.0770 < p < 0.1667 + 0.0770$$

$$0.0897 < p < 0.2437$$

Hypothesis concerning One Proportion

Here we test the null hypothesis $p = p_0$ against one of the alternative hypothesis $p < p_0, p > p_0, or\ p \neq p_0$ with the use of the statistic

$$Z = \frac{X - np_0}{\sqrt{n\,p_0\left(1 - p_0\right)}}$$

Which is a random variable having approximately the standard normal distribution.

Critical region for Testing $p = p_0$	
Alternative Hypothesis	Reject null hypothesis if:
$p < p_0$	$Z < -z_\alpha$
$p > p_0$	$Z > z_\alpha$
$p \neq p_0$	$Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$

Problem 8: Transceivers provide wireless communication among electronic components of consumer products. Responding to a need for a fast, low-cost test of Bluetooth-capable transceivers, engineers developed a product test at the wafer level. In one set of trails with 60 devices selected from different wafer lots, 48 devices passed. Test the null hypothesis p = 0.70 against the alternative hypothesis p > 0.70 at the 0.05 level of significance.

Solution: Null hypothesis: p =0.70

Alternative hypothesis : p > 0.70

Level of significance α = 0.05

Then $z_{0.05}$=1.645

The null hypothesis is rejected if Z > 1.645 where the test statistic $Z = \dfrac{X - np_0}{\sqrt{np_0(1-p_0)}}$

From the given x =48, n =60, and p_0=0.70

Then $Z = \dfrac{X - np_0}{\sqrt{np_0(1-p_0)}} = \dfrac{48-60(0.70)}{\sqrt{60(0.70)(0.30)}} = 1.69$

Since z = 1.69 is greater than 1.645, the null hypothesis is rejected. So, we accept alternative hypothesis. That is p > 0.70 is accepted.

Problem 9: A manufacturer of submersible pumps claims that at most 30% of the pumps require repairs within the first 5 years of operation. If a random sample of 120 of these pumps includes 47 which required repairs within the first 5 year, test the null hypothesis p = 0.30against the alternative hypothesis p > 0.30 at the 0.05 level of significance.

Solution: Null hypothesis: p =0.30

Alternative hypothesis : p > 0.30

Level of significance α = 0.05

Then $z_{0.05}$=1.645

The null hypothesis is rejected if Z > 1.645 where the test statistic $Z = \dfrac{X - np_0}{\sqrt{np_0(1-p_0)}}$

From the given x =47, n =120, and p_0=0.30

Then $Z = \dfrac{X - np_0}{\sqrt{np_0(1-p_0)}} = \dfrac{47-120(0.30)}{\sqrt{120(0.30)(0.70)}} = 2.1913$

Since z = 2.1913 is greater than 1.645, the null hypothesis is rejected. So, we accept alternative hypothesis. That is p > 0.30 is accepted.

Problem 10: The performance of a computer is observed over a period of 2 years to check the claim that the probability is 0.20 that its downtime will exceed 5 hours in any given week. Testing the null hypothesis p = 0.20 against the alternative hypothesis p ≠ 0.20, what can we conclude at the level of significance α= 0.05, if there we only 11 weeks in which the downtime of the computer exceeded 5 hours?

Solution: Null hypothesis: p =0.20

Alternative hypothesis : p ≠ 0.20

Level of significance α = 0.05

Then $z_{0.025}$=1.96

The null hypothesis is rejected if Z <-1.96 or Z>1.96 where the test statistic

$$Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$$

From the given x =11, n =104, and p_0=0.20

Then $Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} = \frac{11-104(0.20)}{\sqrt{105(0.20)(0.80)}} = -2.391$

Since z = -2.391 is less than -1.96, the null hypothesis is rejected. So, we accept alternative hypothesis. That is p ≠ 0.30 is accepted.

Problem 11: To check on an ambulance service's claim that at least 40% of its calls are life-threatening emergencies, a random sample was taken from its files, and it was found that only 49 of 150 calls were life-threatening emergencies. Can the null hypothesis p = 0.40 be rejected against the alternative hypothesis

P < 0.40 if the probability of a Type-I error is to be at most 0.01?

Solution: Null hypothesis: $p \geq 0.40$

Alternative hypothesis : p < 0.40

Level of significance α ≤0.05

Here we test the null hypothesis $p = 0.40$ against the alternative hypothesis p < 040 at the level of significance α= 0.05

Then $z_{0.05}=1.645$

The null hypothesis is rejected if Z <-1.645 where the test statistic $Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$

From the given x =49, n =150, and $p_0=0.40$

Then $Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} = \frac{49 - 150(0.40)}{\sqrt{150(0.40)(0.60)}} = -1.8333$

Since z = -1.8333 is less than -1.645, the null hypothesis is rejected. So, we accept alternative hypothesis. That is p <0.40 is accepted.

<u>Problem 12:</u> In a random sample of 600 cars making a right turn at a certain intersections, 157 pulled into the wrong lane. Test the null hypothesis that actually 30% of all drivers make this mistake at the given intersection, using the alternative hypothesis p ≠ 0.30 and the level of significance

(a) α = 0.05 (b) α = 0.01

<u>Solution:</u> Null hypothesis: $p = 0.30$

Alternative hypothesis : $p \neq 0.30$

Here we test the null hypothesis $p = 0.30$ against the alternative hypothesis p ≠ 0.30

The null hypothesis is rejected if Z < $-Z_{\alpha/2}$ or Z > $Z_{\alpha/2}$ where the test statistic $Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$

From the given x = 600, n =157, and $p_o = 0.30$

(i) at the level of significance α= 0.05

Then $z_{0.025}=1.96$

Then $Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} = \frac{157 - 600(0.30)}{\sqrt{600(0.30)(0.70)}} = -2.049$

Since z = -2.0491 is less than -1.96, the null hypothesis is rejected. So, we accept alternative hypothesis. That is p $\neq$0.30 is accepted.

(ii) at the level of significance α= 0.01

Then $z_{0.005}$=2.575

Then $Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} = \frac{157 - 600(0.30)}{\sqrt{600(0.30)(0.70)}} = -2.049$

Since z = -2.0491 is between -2.575 and 2.575l, the null hypothesis is accepted. That is p = 0.30 is accepted.

Conclusion: The null hypothesis is rejected with 95% confidence and accepted with 99% confidence.

HYPOTHESIS CONCERING SEVERAL PROPOPRTIONS

Many engineering problems concern a random variable that follows the binomial distribution. For example, consider a production process that manufactures items that are classified as either acceptable or defective. Modeling the occurrence of defectives with the binomial distribution is usually reasonable when the binomial parameter *p* represents the proportion of defective items produced. Consequently, many engineering decision problems involve hypothesis testing about *p*.

Suppose that we are interested in testing whether two or more binomial populations have the same parameter p. Let us consider k different binomial populations whose parameters are respectively p_1, p_2,…,p_k. Now we are interested in testing the null hypothesis $p_1 = p_2 = \ldots = p_k = p$ against the alternative hypothesis that these population proportions are not all equal. To perform a suitable large sample test of this hypothesis, we require independent random samples of size n_1, n_2,…,n_k from k different populations. The number of successes and failures in each of these k samples are given by the following table:

	Sample 1	Sample 2		Sample k	Total
Successes	x_1	x_2	…	x_k	x
Failures	$n_1 - x_1$	$n_2 - x_2$	…	$n_k - x_k$	n - x
Total	n_1	n_2	…	n_k	n

In the above table x represents the total number of successes, n – x represents the total number of failures and n the total number of trails. The entry in the cell belonging to the i^{th} row and j^{th} column is called the observed frequency o_{ij} with i= 1, 2 and j = 1, 2, …, k. Let us denote the observed proportion of success by $\overline{p}$. So, the value of $\overline{p}$ is given by $\overline{p} = \frac{x}{n}$.

Hence the expected number of successes and failures for the j^{th} sample are estimated by the following formulae:

$$e_{1j} = n_j \overline{p} == n_j \frac{x}{n}$$

$$e_{2j} = n_j (1 - \overline{p}) = n_j (1 - \frac{x}{n}) = n_j (\frac{n-x}{n})$$

The quantities e_{1j} and e_{2j} are called the expected cell frequencies for j =1, 2, …, k.

The test statistic for test concerning difference among proportions is given by

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{k} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

Decision: Reject the null hypothesis if the value of χ^2 exceeds χ^2_α with k – 1 degrees of freedom.

<u>Problem 13</u>: Samples of three kinds of materials subjected t extreme temperature changes, produced the results shown in the following table:

	Material A	Material B	Material C	Total
Crumbled	41	27	22	90
Remained intact	79	53	78	210

Total	120	80	100	300

Use the 0.05 level of significance to test whether, under the stated conditions, the probability of crumbing is the same for the three kinds of materials.

Solution:

Null Hypothesis, H_0: $p_1 = p_2 = p_3$

Alternative Hypothesis, H_1: p_1 , p_2 and p_3 are not all equal.

Level of significance, α = 0.05

$$\bar{p} = \frac{90}{300}$$

The expected frequencies for the cells are given as follows:

e_{11} = 120x90/300 = 36

e_{12} = 80x90/300 = 24

e_{13} = 90 – (36 + 24) = 30 (since, 36 + 24 + e_{13} =90)

e_{21} = 120 – 36 = 84 (since, 36 + e_{21} = 120)

e_{22} = 80 – 24 = 56

e_{23} = 100 – 30 = 70

Test statistic, $\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}$

$= (41 - 36)^2/36 + (27 - 24)^2/24 + (22 - 30)^2/30 + (79 - 84)^2/84$

$+ (53 - 56)^2/56 + (78 - 70)^2/70$

$= 4.575$

$$\chi^2_{0.05}(3-1)\,d.f = 5.991$$

Since $\chi^2 = 4.575$ does not exceeds $\chi^2_{0.05}$ (3 – 1) d.f = 5.991, we can't reject the null hypothesis at the 0.05 level of significance. Hence the probability of crumbling is the same for the three kinds of material.

Problem 14: Four methods are under development for making disks of a superconductivity material. Fifty disks are made by each method and they are checked for superconductivity when cooled with liquid nitrogen.

	Method 1	Method 2	Method 3	Method 4	Total

Superconductors	31	42	22	25	120
Failures	19	8	28	25	80
Total	50	50	50	50	200

Perform a chi square test with α = 0.05 to test whether the probability of superconductivity is the same for the four kinds of methods.

Solution:

Null Hypothesis, H_0: $p_1 = p_2 = p_3 = p_4$

Alternative Hypothesis, H_1: p_1 , p_2 , p_3 and p are not all equal.

Level of significance, α = 0.05

$\bar{p} = \frac{120}{200}$

The expected frequencies for the cells are given as follows:

e_{11} = 50x120/200 = 30

e_{12} = 50x120/200 = 30

e_{13} = 50x120/200 = 30

e_{14} = 50x120/200 = 30

e_{21} = 50 – 30 = 20

e_{22} = 50 – 30 = 20

e_{23} = 50 – 30 = 20

e_{24} = 50 – 30 = 20

Test statistic, $\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{4}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}$

$= (31-30)^2/30 + (42-30)^2/30 + (22-30)^2/30 + (25-30)^2/30$

$+ (19-20)^2/20 + (8-20)^2/20 + (28-20)^2/20 + (25-20)^2/20$

$= 19.50$

$\chi^2_{0.05}(4-1)\, d.f = 7.815$

Since $\chi^2 = 19.50$ exceeds $\chi^2_{0.05}$ (4 – 1) d.f = 7.815, we reject the null hypothesis at the 0.05 level of significance. Hence the probability of superconductivity is not the same for four methods.

Problem 15: The following data come from a study in which random samples of the employees of three government agencies were asked questions about their pension plan:

	Agency 1	Agency 2	Agency 3	Total
For the pension plan	67	84	109	260
Against the pension plan	33	66	41	140
Total	100	150	150	400

Use the 0.01 level of significance to test the null hypothesis that the actual proportions of employees favoring the pension plan are the same.

Solution:

Null Hypothesis, H_0: $p_1 = p_2 = p_3$

Alternative Hypothesis, H_1: p_1, p_2 and p_3 are not all equal.

Level of significance, $\alpha = 0.01$

$\bar{p} = \frac{260}{400}$

The expected frequencies for the cells are given as follows:

$e_{11} = 100x260/400 = 65$

$e_{12} = 150x260/400 = 97.5$

$e_{13} = 260 – (65 + 97.5) = 97.5$

$e_{21} = 100 – 65 = 35$

$e_{22} = 150 – 97.5 = 52.5$

$e_{23} = 150 – 97.5 = 52.5$

Test statistic, $\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(o_{ij} - e_{ij})^2}{e_{ij}}$

$= (67 - 65)^2/65 + (84 – 97.5)^2/97.5 + (109 – 97.5)^2/97.5$

$+ (33 - 35)^2/35 + (66 – 52.5)^2/52.5 + (41 – 52.5)^2/52.5$

$= 9.39$

$\chi^2_{0.01}(3-1)\, d.f = 9.210$

Since $\chi^2 = 9.39$ exceeds $\chi^2_{0.01}$ (3 – 1) d.f = 9.210, we have to reject the null hypothesis at the 0.01 level of significance. Hence the probability for favoring the pension plan by the three agencies is not the same.

Hypothesis concerning two proportions

This is a particular case of several proportions with k = 2. In this case we proceed as per the given below procedure:

Null Hypothesis, H_0: $p_1 = p_2$

Alternative Hypothesis, H_1 : $p_1 < p_2$ (or) $p_1 > p_2$ (or) $p_1 \neq p_2$

Given x_1, x_2, n_1 and n_2

Level of significance = α

Test Statistic, $Z = \dfrac{\dfrac{X_1}{n_1} - \dfrac{X_2}{n_2}}{\sqrt{\bar{p}(1-\bar{p})}\sqrt{(\dfrac{1}{n_1} + \dfrac{1}{n_2})}}$ with $\bar{p} = \dfrac{X_1 + X_2}{n_1 + n_2}$

Critical Regions for testing the Null Hypothesis, H_0: $p_1 = p_2$

Alternate Hypothesis	Reject null hypothesis if
$p_1 < p_2$	$Z < -z_\alpha$
$p_1 > p_2$	$Z > z_\alpha$
$p_1 \neq p_2$	$Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$

Finally, we have to write the decision that either accepting the null hypothesis or rejecting the null hypothesis.

Also the (1 – α)100% large sample confidence interval for the difference of two proportions is given by

$$\frac{x_1}{n_1} - \frac{x_2}{n_2} \pm z_{\alpha/2}\sqrt{\frac{\frac{x_1}{n_1}\left(1-\frac{x_1}{n_1}\right)}{n_1} + \frac{\frac{x_2}{n_2}\left(1-\frac{x_2}{n_2}\right)}{n_2}}$$

Problem 16: A study shows that 16 of 200 tractors produced on one assembly line required extensively adjustments before they could be shipped. While the same was true for 14 of 400 tractors produced on another assembly line. At the 0.01 level of significance, does this support the claim that the second production line does superior work? Also construct a 95% confidence interval for $p_1 - p_2$.

Solution:

Null Hypothesis, H_0: $p_1 = p_2$

Alternative Hypothesis, H_1 : $p_1 > p_2$

Given $x_1 = 16$, $x_2 = 14$, $n_1 = 200$ and $n_2 = 400$

Level of significance, $\alpha = 0.01$

Test Statistic, $$Z = \frac{\frac{X_1}{n_1} - \frac{X_2}{n_2}}{\sqrt{\bar{p}(1-\bar{p})}\sqrt{\left(\frac{1}{200} + \frac{1}{400}\right)}} \quad with\ \bar{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

$$= \frac{\frac{16}{200} - \frac{14}{400}}{\sqrt{0.05(1-0.05)}\sqrt{\left(\frac{1}{200} + \frac{1}{400}\right)}} \quad with\ \bar{p} = \frac{16+14}{200+400} = 0.05$$

$$= \frac{0.045}{\sqrt{(0.0475)(0.0075)}}$$

$= 2.384$

From table 3, $Z_{0.01} = 2.33$

Critical Regions for testing the Null Hypothesis, H_0: $p_1 = p_2$

Alternate Hypothesis	Reject null hypothesis if
$p_1 > p_2$	$Z > z_\alpha$

Decision: Since Z = 2.384 exceeds $Z_{0.01} = 2.33$, we have to reject the null hypothesis. So, accept the alternative hypothesis. That is the true proportion of tractors requiring extensive adjustments is greater for first assembly line than for the second.

A 95% confidence interval for the difference of two proportions is given by

$$\left(\frac{x_1}{n_1}-\frac{x_2}{n_2}\right)\pm z_{\alpha/2}\sqrt{\frac{\frac{x_1}{n_1}\left(1-\frac{x_1}{n_1}\right)}{n_1}+\frac{\frac{x_2}{n_2}\left(1-\frac{x_2}{n_2}\right)}{n_2}}$$

$$\Rightarrow \quad ((0.08-0.035)\pm z_{0.025}\sqrt{\frac{0.08(1-0.08)}{200}+\frac{0.035(1-0.035)}{400}}$$

$$\Rightarrow \quad (0.08-0.035)\pm 1.96\sqrt{\frac{0.08(1-0.08)}{200}+\frac{0.035(1-0.035)}{400}}$$

$$\Rightarrow \quad 0.045\pm 1.96\sqrt{0.000368+0.00008443}$$

$$\Rightarrow \quad 0.003< p_1-p_2 <0.087$$

<u>Problem 17</u>: Photolithography plays a central role in manufacturing integrated circuits made on thin disks of silicon. Prior to a quality improvement program, too many rework operations were required. In a sample of 200 units, 26 required reworking of the photolithographic step. Following training in the use of pareto charts and other approaches to identify significant problems, improvements were made. A new sample of size 200 had only 12 that needed rework. Is this sufficient evidence at the 0.01 level of significance that the improvements have been effective in reducing rework?

Solution:

Null Hypothesis, H_0: $p_1 = p_2$

Alternative Hypothesis, H_1 : $p_1 > p_2$

Given $x_1 = 26$, $x_2 = 12$, $n_1 = 200$ and $n_2 = 200$

Level of significance, $\alpha = 0.01$

Test Statistic, $Z=\dfrac{\frac{X_1}{n_1}-\frac{X_2}{n_2}}{\sqrt{\bar{p}(1-\bar{p})}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ with $\bar{p}=\dfrac{X_1+X_2}{n_1+n_2}$

$$=\frac{\frac{26}{200}-\frac{12}{200}}{\sqrt{0.095(1-0.095)}\sqrt{\left(\frac{1}{200}+\frac{1}{200}\right)}} \quad with\ \bar{p}=\frac{26+12}{200+200}=0.095$$

$$=\frac{0.07}{(0.2932)(0.1)}$$

$= 2.3873$

$Z_{0.01} = 2.33$

Critical Regions for testing the Null Hypothesis, H_0: $p_1 = p_2$

Alternate Hypothesis	Reject null hypothesis if
$p_1 > p_2$	$Z > z_\alpha$

Decision: Since Z = 2.3873 exceeds $Z_{0.01}$ = 2.33, we have to reject the null hypothesis. So, accept the alternative hypothesis.

<u>Problem 18</u>: The owner of a machine shop must decide which of two snack-vending machines to install in his shop. If each machine is tested for 250 times and the first machine fails to work(neither delivers the snack nor returns the money) 13 times and the second machine fails to work 7 times, test at the 0.05 level of significance whether the difference between the corresponding sample proportions is significant.

Solution:

Null Hypothesis, H_0: $p_1 = p_2$

Alternative Hypothesis, H_1 : $p_1 \neq p_2$

Given $x_1 = 13$, $x_2 = 7$, $n_1 = 250$ and $n_2 = 250$

Level of significance, $\alpha = 0.05$

Test Statistic, $$Z = \frac{\frac{X_1}{n_1} - \frac{X_2}{n_2}}{\sqrt{\bar{p}(1-\bar{p})}\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})}} \quad with \ \bar{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

$$= \frac{\frac{13}{250} - \frac{7}{250}}{\sqrt{0.04(1-0.04)}\sqrt{\left(\frac{1}{250} + \frac{1}{250}\right)}} \quad with \ \bar{p} = \frac{13+7}{250+250} = 0.04$$

$$= \frac{0.024}{(0.1959)(0.0.0844)}$$

$= 1.369$

$Z_{0.025} = 1.96$

Critical Regions for testing the Null Hypothesis, H_0: $p_1 = p_2$

Alternate Hypothesis	Reject null hypothesis if
$p_1 \neq p_2$	$Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$

Decision: Since Z = 1.369 does not exceed $Z_{0.025} = 1.96$, we can't reject the null hypothesis. So, accept the null hypothesis.

Home work: A study showed that 64 of 180 persons who saw a photocopying machine advertised during the telecast of a baseball game and 75 of 180 other persons who saw it advertised on a variety show remembered the brand name 2 hours later. Use the 0.05 level of significance whether the difference between the corresponding sample proportions is significant?