Inference on Proportion of a Population

#.1 **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 exceeded $ 3,500, using the large sample confidence interval formula. Also test the hypothesis that the actual proportion of claims files that exceeded $3500 is not more than 0.5 at 5% level of significance.**

Solution

Let P denote true proportion of claims that exceeded $ 3,500.

Given N = 200 , X = 84

So, sample proportion is

$$\boldsymbol{p = \frac{X}{n} = \frac{84}{200} = 0.42}$$

Now large sample confidence interval is given by

$$P\left(p - Z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \le P \le p + Z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \right) = 1-\alpha$$

....

....

$$P(0.35 \le P \le 0.49) = 0.95$$

Test of hypothesis

Here, the null and alternative hypothesis are:

$$H_0: P = 0.5 \qquad H_1: P < 0.5$$

The test statistic is

$$Z_0 = \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}} = \frac{0.42 - 0.50}{\sqrt{\frac{0.50(1-0.50)}{200}}} = -2.29$$

It is left-tailed test, so critical value is $-Z_{0.05} = -1.64$.

Since $Z_0 < -1.64$, so the null hypothesis is rejected.

#.2 Arthritis is a painful, chronic inflammation of the joints. An experiment on the side effects of 440 arthritis patients using ibuprofen as pain reliever reported that 5% of them had different types of side effects. Construct 90% confidence interval for actual percentage of arthritis patients using ibuprofen as pain reliever who observed side effects of the medicine. The company producing the medicine claims that actual percentage of patients who had side effect of the medicine is less than 7%. Test the claim of the manufacturer at 5% level.

P(0.03291<=P<=0.06709)=0.1, Z_0= -1.644, rejected.

Error in Estimating Population Proportion

#.1 **In a random sample of 400 industrial accidents, it was found that 231 were due to unsafe working conditions. What can we say with 95% confidence about the maximum error in estimation of proportions of accidents due to unsafe working condition, if we use sample proportion as the estimate of population proportion?**

Solution-

$$E_{Max} = Z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}$$

Here sample proportion is

$$p = \frac{X}{n} = \frac{231}{400} = 0.5775$$

$$= Z_{\frac{0.05}{2}}\sqrt{\frac{0.5775(1-0.5775)}{400}} = 0.0877 = 8.77\%$$

Sample Size for Estimating Population Proportion

(I) When population proportion to be estimated is known

Expression-

$$n = \{P(1-P)\}\left(\frac{Z_{\frac{\alpha}{2}}}{E_{Max}}\right)^2$$

(II) When population proportion to be estimated is not known

Expression-

$$n = \frac{1}{4} \times \left(\frac{Z_{\frac{\alpha}{2}}}{E_{Max}} \right)^2$$

#.1 **What is the size of 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 the least 95% confidence?**

Solution-

$$n = \frac{1}{4}\left(\frac{1.96}{0.06}\right)^2 \cong 267$$

In above question, if it is known that the proportion to be estimated is at the least 0.75, what would be the sample size required?

$$n = \{0.75(1-0.75)\}\left(\frac{1.96}{0.06}\right)^2 \cong 200$$

Inference on Difference of Proportions of Two Populations

Statistic considered is

$$\frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}} \sim N(0,1)$$

Since values of P_1 and P_2 are unknown they are replaced by combined proportion of two samples given by

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

where n_1 and n_2 are sizes of the two samples and X_1 and X_2 are number of units possessing characteristic under study in the first and in the second sample.

So, statistic to be considered is

$$\frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} \sim N(0,1)$$

So, required confidence interval is

$$P\left(-Z_{\frac{\alpha}{2}} \leq \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} \leq Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

On re-arranging we have

$$P\left((p_1 - p_2) - Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} \leq P_1 - P_2 \leq (p_1 - p_2) + Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}\right)$$
$$= 1 - \alpha$$

Next to test $H_0: P = P_0$, the test statistic to be calculated is

$$Z_0 = \frac{(p_1 - p_2)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}$$

Problem

#.1

Photolithography plays an important role in manufacturing of IC made on thin silicon chips. Since too many rework operations were required in manufacturing process a quality improvement program was conducted.

Prior to the improvement program a sample of 200 units 26 required reworking. Following the training program the study of a new sample of 200 had only 12 units that needed rework.

(a) **Find a large sample 95% confidence interval for the true difference of the proportions of units that needed rework.**

(b) Is this sufficient evidence to conclude at the 0.01 level that the improvement program has been effective in reducing the rework.

Solution- In class-note.

Hints: **p1 = 0.13** **p2 = 0.06**

p (cap) = 0.095

se ….. = 0.029321 **Z value = 1.96**

P(0.01 <= P1 – P2 <= 0.127) = 0.95

Z0 = 2.39 (>1.64) reject.