MATH F113
Probability and Statistics

Dr. Shivi Agarwal
Department of Mathematics

BITS Pilani
Pilani Campus

Chapter 9
Inferences on Proportions

This chapter will deal with inferences on proportions and hypothesis tests on them.

We will see how to employ the standard normal distribution to construct confidence intervals on ‘p’ and test hypotheses concerning its value for larger samples.

What is Sample Proportion?

Example:

- Suppose a student guesses at the answer on every question in a 300-question examination. If he gets 60 questions correct, then his proportion of correct guesses is 60/300=0.20.
- Thus, the proportion of correct guesses is simply the number of correct guesses divided by the total number of questions.

9.1 Estimating proportions:

Estimation of proportion should be done in the following situation:

We have a population of interest, a particular trait is being studied, and each member of the population can be classified as either having the trait or not (like, Bernoulli trials).

For this, we draw a random sample of size ‘n’ from the population, and associated with it is a collection of n independent random variables $X_{1,}$ X_2, ...X_n where

$$X_i = \begin{cases} 1 \text{ if the 'i' th member of the sample has the trait.} \\ 0 \text{ if the 'i' th member of the sample does not have the trait.} \end{cases}$$

In general,

$$X = \sum_{i=1}^{n} X_i$$

gives the number of objects in the sample with the trait and the statistic X/n gives the proportion of the sample with the trait.

Note that ***X*** is a binomial RV with parameters ***n*** (known) and ***p***.

Then the sample proportion which is a logical point estimator for p is given by

$$\hat{p} = \frac{X}{n} = \frac{\text{number in sample with the trait}}{\text{sample size}}$$

Confidence interval on 'p':

For this distribution of $\hat{p}$ must be determined. Assume sample size n is large.

This is nothing but the sample mean.

Therefore, by Central Limit theorem, $\hat{p}$ is approximately normally distributed with same mean as X_i's and variance equal to Var (X_i/n)

Since X_i is 1 when the object being sampled has the trait,

$P[\,X_i=1]=p$ and $P[\,X_i=0]=1-p$

x_i	1	0
$f(x_i)$	p	$1-p$

Also, it is easy to see that

$E[X_i]=1(p)+0(1-p)=p$

$E[X_i^2]=1^2(p)+0^2(1-p)=p$

$\mathrm{Var}\ X_i=E[X_i^2]-(E[X_i])^2=p-p^2=p(1-p).$

Therefore by CLT, $\hat{p}$ is approximately *normally distributed* with mean p and variance p(1-p)/n.

The random variable

$$(\hat{p}-p)/\sqrt{p(1-p)/n}$$

follows a standard normal distribution.

For a 100(1-α)% confidence interval

$$P[-z_{\alpha/2} \le (\hat{p}-p)/\sqrt{p(1-p)/n} \le z_{\alpha/2}] = 1-\alpha$$

Isolating p in the middle of the inequality,

$$P[\hat{p}-z_{\alpha/2}\sqrt{p(1-p)/n} \le p \le \hat{p}+z_{\alpha/2}\sqrt{p(1-p)/n}] = 1-\alpha$$

Based on the above expression, we get the confidence intervals for p

$$\hat{p} \pm z_{\alpha/2}\sqrt{p(1-p)/n}$$

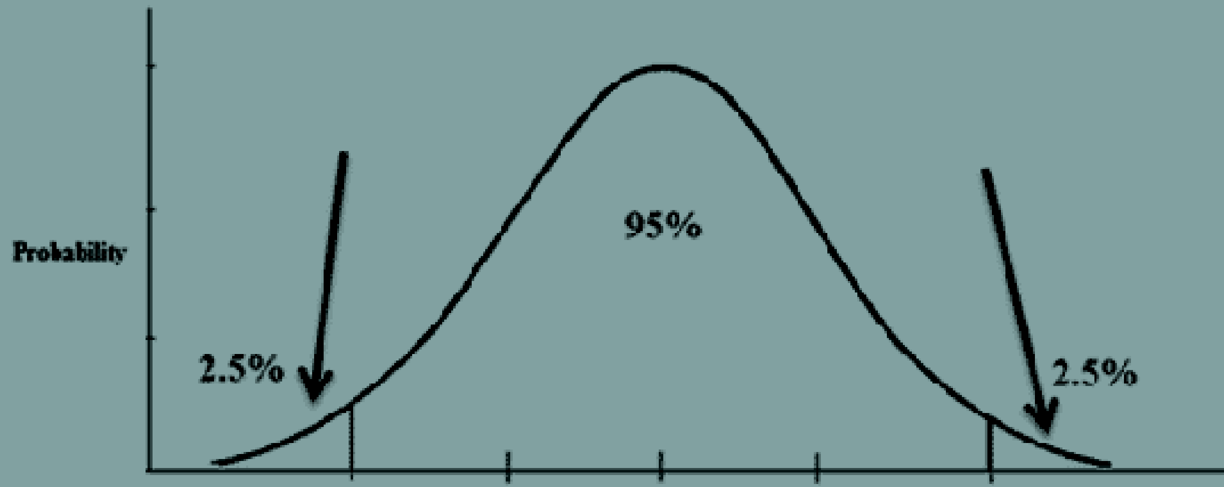
The above expression employs ‘p’ (which we don’t know). So replace it by $\hat{p}$

The confidence intervals become

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

Confidence Interval on p

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$



$$\alpha = 0.05 \qquad z_{0.05/2} = z_{0.025}$$

Sample size for estimating 'p':

We can be 100(1-α)% sure that $\hat{p}$ and p differ by at most d , where d is given by

$$d = z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

Sample size for estimating p, prior estimate $\hat{p}$ available

$$n = \frac{z^2_{\alpha/2}\,\hat{p}(1-\hat{p})}{d^2}$$

When prior estimate for p is not available, method for estimating n is based on the fact that $\hat{p}(1-\hat{p})$ can never be greater than 0.25. Hence replace this term by ¼ in the above formula.

Thus sample size for estimating p, no prior estimate available:

$$n = \frac{z^2_{\alpha/2}}{4d^2}$$

If another range for p is provided maximize p(1-p) over that range.

Q 9. Consider the function $g(\hat{p}) = \hat{p}(1-\hat{p})$

(a) Find $g'(\hat{p})$

(b) Find the critical point for g.

(c) Find $g''(\hat{p})$, and use this to argue that

g assumes its maximum value at the critical point.

(d) What is the maximum value asumed by the function g?

9.2 Hypothesis testing on proportions:

Let p_o be the null value of p.

Then

$H_o: p = p_o$	$H_o: p = p_o$	$H_o: p = p_o$
$H_1: p > p_o$	$H_1: p < p_o$	$H_1: p \neq p_o$
Right-tailed	Left-tailed	Two-tailed test

$$\text{Test statistic for testing } H_0 : p = p_0$$

$$Z = (\hat{p} - p_0) / \sqrt{p_0(1 - p_0) / n}$$

The test statistic is a logical choice because it compares the unbiased estimator of p with the null value of p_o.

For a right-handed test, *H_o is rejected in favor of H_1 if the observed value of the test statistic is a large positive number.*

For a left-handed test, *H_o is rejected in favor of H_1 if the observed value of the test statistic is a large negative number.*

Hypothesis Testing on Proportions

Test statistic for testing $H_0: p = p_0$

$$Z = (\hat{p} - p_0) \Big/ \sqrt{\frac{p_0(1-p_0)}{n}}$$

- Rejection Rule: P –Value Approach

  Reject H_0 if P –value $\leq \alpha$ *(0.05)*

- Rejection Rule: Critical Value Approach

  - $H_1: p > p_0$ — Reject H_0 if $z \geq z_\alpha$
  - $H_1: p < p_0$ — Reject H_0 if $z \leq -z_\alpha$
  - $H_1: p \neq p_0$ — Reject H_0 if $z \leq -z_{\alpha/2}$ or $z \geq z_{\alpha/2}$

Hypothesis Testing on Proportions

Note that the test statistic is a logical choice because it compares the unbiased estimator of p with the null value of p_o.

Test statistic for testing $H_0 : p = p_0$

$$Z = \left(\hat{p} - p_0\right) \Big/ \sqrt{\frac{p_0\left(1 - p_0\right)}{n}}$$

Steps of Hypothesis Testing

Step 1. Develop the null and alternative hypotheses of proportion; Calculate the observed value of test statistic.

Step 2. Specify the level of significance α.

Step 3. Collect the sample data and compute the test statistic.

Step 4. Based on α, identify critical values (use the value of test statistic to compute the P-value).

Step 5. Reject H_0 if the calculated test statistic value falls in the the rejection region (Reject H_0 if P-value $\leq \alpha$).

Exercises (Section 9.1)

2. A study of electromechanical protection devices used in electrical power systems showed that of 193 devices that failed when tested, 75 were due to mechanical part failures.

a) Find a point estimate for p, the proportion of failures that are due to mechanical failures.

b) Find a 95% confidence interval on p.

c) How large a sample is required to estimate p to within 0.03 with 95% confidence.

Section 9.2

10. A poll of investment analysts taken earlier suggests that a majority of these individuals think that the dominant issue affecting the future of the solar energy industry is falling energy prices. A new survey is being taken to see if this is still the case. Let p denote the proportion of investment analysts holding this opinion.

a) Set up the appropriate null and alternative hypothesis.

b) When the survey is conducted, 59 of the 100 analysts sampled agreed that the major issue is falling energy prices. Is this sufficient to allow us to reject H_o? Explain based on P value of the test.

c) Interpret your results in the context of this problem.

15. A battery operated digital pressure monitor is being developed for use in calibrating pneumatic pressure gauges in the field. It is thought that 95 % of the readings it gives lies within 0.01 lb/in² of the true reading. In a series of 100 tests, the gauge is subjected to a pressure of 10,000 lb/in². A test is considered to be a success if the reading lies within 10,000± 0.01 lb/in². We want to test

$$H_0: p = 0.95$$

$$H_1: p \neq 0.95$$

At the $\alpha = 0.05$ level.

a) What are the critical points for the test.

b) When the data are gathered, it is found that 98 out of 100 readings were successful. Can H_0 be rejected at $\alpha=0.05$ level? What type of error are you now subject?

33. A survey of mining companies is to be conducted to estimate p, the proportion of companies that anticipate hiring either graduating seniors or experienced engineers during the coming year.

a) How large a sample is required to estimate p within 0.04 with 94% confidence?

b) A sample of size 500 yields 105 companies that plan to hire such engineers. Find the point estimate of p. Also find 94% confidence interval for p.