STAB22

Chong Chen
University of Toronto, Scarborough
Department of Computer and Mathematical Sciences

2019

1 Recall

1.1 Sampling Distribution of Sample Proportions

Sampling distribution, as well as Central limit theorem, can also be applied to sample proportions.

Draw an SRS of size n from a large population having population proportion p of successes. Let X be the count of successes in the sample and $\hat{p} = X/n$ be the sample proportion of successes. When n is large, the sampling distributions of these statistics are approximately Normal:

$\hat{p}$ follows the normal distribution $(p, \sqrt{\frac{p(1-p)}{n}})$.

1.2 Confidence Intervals for Proportions

Confidence interval for the population proportion p is given by

$$\hat{p} \pm z^* * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where standard error of the estimate

$$SE(\hat{P}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

also margin of error of the CI is

$$z^* * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where z^* is the critical value.

1.3 Interpretation

If repeated samples were taken and the 95% confidence interval computed for each sample, 95% of the resulting intervals would contain the true population proportion.

1.4 Handy table

Confidence level	z^*
90%	1.645
95%	1.960
99%	2.576

1.5 Exercise

Opinion poll with 1000 randomly sampled Canadians, 91% believe Canada's health care system better than US's. Find a 95% confidence interval for the population proportion of Canadians who believe Canada's health care system is better than US's.

2 Testing hypothesis about proportions

2.1 Preliminary Examples

University students will vote on proposal. You ask random sample of 400 students, out of which 250 are in favor.

Question: is the proportion of students that are in favor significantly different from 55%?

One possible approach: use the confidence interval: A 95% CI for the true proportion of students that are in favor is:

$$[0.577, 0.673]$$

Another approach: use statistical tests.

2.2 Introduction

- Hypothesis: In statistics, a hypothesis is a statement about some characteristics of a population.
- Test: A statistical significance test uses the data to summarize the evidence about a hypothesis.

2.3 The Structure of Statistical Tests

- Determine the null hypothesis and the alternative hypothesis

$$H_0 : p = p_0$$
$$H_A : p \neq p_0$$

- Calculate a test statistic

$$z^{obs} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 * (1 - p_0)}{n}}}$$

- Determine the p-value

  What would be the probability of observing a value at least as extreme as z^{obs} if the null hypothesis were true?

- Make a conclusion based on the p-value and the context of the problem

  1. Reject the null hypothesis if the p-value is small.

  2. How to decide whether the p-value small enough to reject H_0?

  $\rightarrow$ Reject H_0 if p-value less than the α you chose.

  NOTE: The "default" is $\alpha = 0.05$.

2.4 Exercise1

According to PetSmart, 63% of households own pets. Survey found 210 households out of a random sample of 300 own pets. Test if survey agrees with PetSmart (use $\alpha = 5\%$)

2.5 Exercise2

A test of significance with $\alpha = 0.05$ results in a P-value of 0.1. What do you conclude?

2.6 One-Sided Tests versus Two-Sided Tests

- One-sided test: only concerned with deviations from H_0 in one direction.
- Two-sided test: concerned with deviations from H_0 in either direction.
- If you are ever in doubt use a two sided-test.

2.7 Hypothesis Tests Errors

There are two types of errors in decisions:

- Type I Error: Rejection of H_0 when H_0 is true.
- Type II Error: Acceptance of H_0 when H_0 is false.

Table: Errors in Decisions

		Decision	
		Reject H_0	Do Not Reject H_0
Condition of H_0	H_0 True	Type I Error	Correct Decision
	H_0 False	Correct Decision	Type II Error

3 Examples

3.1

In a large mall a survey was taken. It was found that in a random sample of 45 women over the age of 25, 15 had children. Find the 90% confidence interval for the population proportion of women over the age of 25 in the mall who have children.

3.2

It is widely thought that there is a high incidence of disability among the homeless population. A random sample of 110 homeless people contained 84 who were disabled on one or more categories (such as psychiatric disability, medical disability etc.). Let p denote the proportion of all homeless people having one or more types of disability. Which of the following numbers is the closest to the value of the test statistic for testing the null hypothesis H_0: $p = 0.75$ against $H_A : p \neq 0.75$?

A. 0

B. 0.3

C. 0.6

3.3

According to an Ipsos Reid poll that sampled 1,005 Canadians, 59% Canadians oppose genetically modifying crops and animals to produce food. Is the proportion of Canadians that oppose genetically modifying crops and animals to produce food significantly *larger than* 50%?