STAB22

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

April 5, 2019

1 Recall

1.1 Normal distributions

Normal distribution: important class of densities, bell-shaped curves. Two parameters to describe a particular normal distribution fully are: the mean μ (center), and the standard deviation σ (spread).

The **standard normal distribution** is the normal distribution with mean $\mu = 0$, and standard deviation $\sigma = 1$.

Standardization: Any normal random variable can be transformed to a standard normal random variable via standardization. If X is a normal distribution with mean μ and standard deviation σ, then

$$Z = \frac{X - \mu}{\sigma}$$

1.1.1 Exercise 1

Assume the cholesterol levels of adult women can be described by a Normal model with a mean of 188 mg/dL and a standard deviation of 24.

a. Draw and label the Normal model.

b. What percent of adult women do you expect to have cholesterol levels over 200 mg/dL?

c. What percent of adult women do you expect to have cholesterol levels between 150 and 170 mg/dL?

d. Above what value are the highest 15% of women's cholesterol levels?

1.1.2 Exercise 2

A smelt is a type of food fish. Smelt lengths are normally distributed with mean 15 cm and standard deviation 1 cm. How long are the longest 10 percent of smelts?

(a) bigger than 10.14 cm

(b) bigger than 16.28 cm

(c) 10.14 cm

(d) less than 16.28 cm

(e) 16.28 cm

1.2 Bernoulli distributions

A **Bernoulli random variable** has exactly two possible outcomes. We typically label one of these outcomes a "success" and the other outcome a "failure". We may also denote a success by 1 and a failure by 0.

Example: there is a 70% chance that a student admitted in a graduate program attends the program. We introduce the random variable X such that:

X	$P(X = x)$
1	70%
0	30%

This random variable X is called Bernoulli random variable with parameter $p = 0.7$.

Question: Could you find out the mean for this Bernoulli distribution?

1.3 Binomial distributions

A **Binomial random variable** describes the probability of the number of successes in n independent **Bernoulli trials** with probability of a success p. Two parameters to describe a binomial distribution fully are:

- n (the number of trials)
- p (the probability of success at each trial)

Example: suppose that a university announced that it admitted 8 students for the following year's graduate program in music. There is a 40% chance that an admitted student will decide to accept the offer and attend this university. Admitted students accept the offer independently of each other.

The variable X = "Number of students that accept the offer", has a binomial distribution with parameters $n = 8$, $p = 0.4$.

The probability distribution of a Binomial random variable with parameters n and p. And use the Binomial table to find the following probabilities.

$P(X = 2) = 0.2029$,
$P(X = 5) = 0.1239$.

2 Normal approximation to the binomial distribution

2.1 Binomial distribution

Recall that for the Binomial distribution, it has mean $\mu = np$, and standard deviation $\sigma = \sqrt{np(1-p)}$.

2.2 Normal Approximation for Counts and 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:

X is approximately $N(np.\sqrt{np(1-p)})$,

$\hat{p}$ is approximately $N(p, \sqrt{\frac{p(1-p)}{n}})$

As a rule of thumb, we will use this approximation for values of n and p that satisfy: $np \geq 10$ and $n(1-p) \geq 10$.

2.3 Example

According to governmental data, 21% of American children under the age of six live in households with incomes less than the official poverty level. A study of learning in early childhood chooses a simple random sample of 300 children.

a. What is the mean number of children in the sample who come from poverty-level households? What is the standard deviation of this number?

b. Use the normal approximation to calculate the probability that at least 80 of the children in the sample live in poverty. Be sure to check that you can safely use the approximation.

c. What is the mean number of children in the sample who come from poverty-level households? What is the standard deviation of this number?

d. Use the normal approximation to calculate the probability that at least 80 of the children in the sample live in poverty. Be sure to check that you can safely use the approximation.

3 Sampling distributions

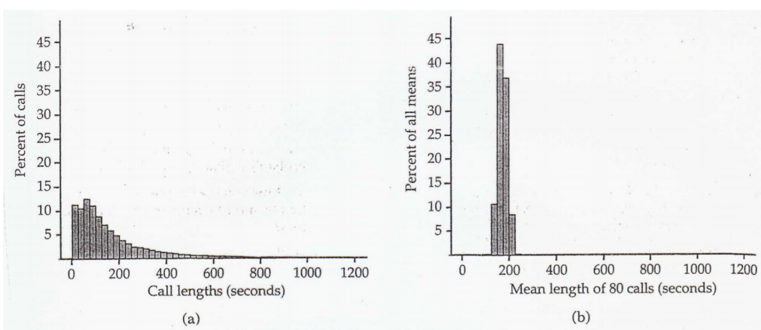
3.1 Sampling Distribution of Sample Mean

Sample means are among the most common statistics, and we are often interested in their sampling distribution.

The figure below shows:

(a) the distribution of lengths of all customer service calls received by a bank in a month $\mu = 170$.

(b) the distribution of the sample means $\bar{x}$ for 500 random samples of size 80 from this population.



The sample mean $\bar{x}$ from a sample or an experiment is an estimate of the mean of the μ underlying population.

Let X_1, X_2, ..., X_n be taken from a population with mean μ and standard deviation σ.

Then

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

REMARK: **Central limit theorem**

Draw a Simple Random Sample of size n from any population with mean μ and finite standard deviation σ. When n is large enough, the sampling distribution of the sample mean $\bar{X}$ follows the normal distribution $(\mu, \frac{\sigma}{\sqrt{n}})$.

3.2 Sampling Distribution of Sample Proportions

Sampling distribution, as well as Central limit theorem, can also be applied to sample proportions. And the idea is similar to *section 2.2*.

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 $(p, \sqrt{\frac{p(1-p)}{n}})$.

4 Confidence intervals for proportions

An interval estimate is an interval of numbers (around the point estimate) which is highly likely to contain the parameter value. Interval estimate is also known as **confidence interval**.

Confidence interval for the population proportion p is given by

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

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

$$z_\alpha * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

α is the confidence level

5 Exercises

5.1

In a large population, 60% of the voters support a particular party. If a random sample of 300 voters were selected from this population, what would be the approximate probability that more than 175 of them would support this party?

(a) 0.35
(b) 0.72
(c) 0.96
(d) 0.51
(e) 0.60

5.2

In which of the following cases is the normal distribution a reasonable approximation of the sampling distribution of a proportion?

(I) The probability of an airline flight arriving on time is 90%. The airline operates more than 500 flights per day. The normal distribution is a reasonable approximation of the sampling distribution of the proportion of flights arriving on time in a simple random sample of 30 flights of this airline.

(II) In a large country 0.2% of the population is affected by a disease. The normal distribution is a reasonable approximation of the sampling distribution of the proportion of individuals affected by the disease in a simple random sample of 2000 individuals in this country.

(III) A company has 150 workers. The proportion of workers who smoke in this company is 40%. The normal distribution is a reasonable approximation of the sampling distribution of the proportion of workers who smoke in a simple random sample of 30 workers of this company.

(a) The normal distribution is a reasonable approximation of the sampling distribution of a proportion only in case (I).

(b) The normal distribution is a reasonable approximation of the sampling distribution of a proportion only in case (II).

(c) The normal distribution is a reasonable approximation of the sampling distribution of a proportion only in case (III).

(d) The normal distribution is a reasonable approximation of the sampling distribution of a proportion in at least two of the cases (I), (II), (III).

(e) The normal distribution is a reasonable approximation of the sampling distribution of a proportion in none of the three cases (I), (II), (III).

5.3

The weight of Canada Post letters is normally distributed with a mean of 2 oz. and a standard deviation of 0.5 oz.

If I select at random 1000 Canada Post letters, what is the probability that at least 200 will weigh less than 1 oz.?

(a) 0.0000
(b) 0.0046
(c) 0.1040
(d) 0.8960
(e) 1.0000