Statistical Methods for the Biological, Environmental, and Health Sciences

STAT 007

Normal Probability Distributions

Chapter 6

Real Applications of the Normal Distribution and the Standard Normal Distribution

Sections 6-1 and 6-2

- In this sections we will:
  - Introduce the normal distribution and the standard normal distribution.
  - Identify their parameters: mean, variance, and standard deviation.

Normal Distribution

Example

Assume that pulse rates of adult males are normally distributed with a mean of 69.6 bpm and a standard deviation of 11.3 bpm.

a) Sketch the density curve of the distribution of the pulse rates of adult males.

b) Males with pulse rates greater than 100 bpm are considered to be at a high risk of stroke, heart disease, or cardiac death. Find the proportion of men that are at risk of stroke, heart disease, or cardiac death.

c) Normal pulse rates are generally considered to be between 60 bpm and 100 bpm. Find the proportion of men that have normal pulse rate.

d) Find the pulse rate that separates the highest 1% from the lowest 99%.

Practice

Look at the exercises at the end of Section 6-1 in page 229.

Specially, look at exercises:
1, 3, 9-12, 13-16, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37-40.

Look at the exercises at the end of Section 6-2 in page 238.

Specially, look at exercises:
1, 2, 3, 5-8, 9-12, 13-20, 24.

Sampling Distributions and Estimators

Section 6-3

- In this sections we will:
  - Introduce sampling distributions for statistics such as the sample mean, sample standard deviation, and sample proportion.
  - Define estimators and discuss what biased/unbiased estimators are.

Sampling Distribution of a statistic

- Consider a sample of size n of values $x_1, x_2, \ldots, x_n$.
- The sample mean, $\overline{x}$, is the statistic given by $\overline{x} = \frac{x_1+x_2+\ldots+x_n}{n}$.
- If the values of x_i are 0s or 1s, then the sample mean is equal to the sample proportion, denoted $\hat{p}$.
- The sample variance, s^2, is the statistic given by $s^2 = \frac{\sum(x_i-\overline{x})^2}{n-1}$.
- Our focus now is to study the distribution of the sample mean, $\overline{x}$, sample proportion, $\hat{p}$, and sample variance, s^2.
- Distribution of statistics are called sampling distributions.

Sampling Distribution of a statistic

Definition

The ***sampling distribution of a statistic*** *(sample proportion, $\hat{p}$, sample mean, $\overline{x}$, or sample variance, s^2) is the distribution of all values of the statistic when all possible samples of the same size n are taken from the same population.*

- Fact: Among the population of all adults, exactly 40% have brown eyes.
- In a survey of 1,000 adults, 42% of the subjects were observed to have brown eyes.
- Being so intrigued by this, 50,000 people became so enthusiastic that they each conducted their own individual survey of 1,000 randomly selected adults.
- Each of these 50,000 new surveyors reported the percentage that they found, with results such as 38%, 39%, and 43%.
- The authors obtained each of the 50,000 sample percentages, changed them to proportions, and then they constructed the histogram.

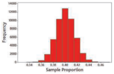

FIGURE 6-17 Histogram of 50,000 Sample Proportions

Sampling Distribution of a statistic

- **Behavior of the sample proportion,** $\hat{p}$:
  a) The distribution of sample proportions tends to approximate a normal distribution.
  b) Sample proportions *target* the value of the population proportion in the sense that the mean of all of the sample proportions $\hat{p}$ is equal to the population proportion p; the expected value of the sample proportion is equal to the population proportion.
- **Behavior of the sample mean,** $\overline{x}$:
  a) The distribution of sample means tends to be a normal distribution.
  b) The sample means *target* the value of the population mean. (That is, the mean of the sample means is the population mean. The expected value of the sample mean is equal to the population mean.)
- **Behavior of the sample variance,** s^2:
  a) The distribution of sample variances tends to be a distribution skewed to the right.
  b) The sample variances target the value of the population variance. (That is, the mean of the sample variances is the population variance. The expected value of the sample variance is equal to the population variance.)

Estimators: Unbiased and Biased

- An **estimator** is a statistic used to infer (or estimate) the value of a population parameter.
- An **unbiased estimator** is a statistic that *targets* the value of the corresponding population parameter in the sense that the sampling distribution of the statistic has a mean that is equal to the corresponding population parameter.
- Examples of unbiased estimators are:
  — sample proportion, $\hat{p}$, as an estimator of the population proportion, p.
  — sample mean, $\overline{x}$, as an estimator of the population mean, μ.
  — sample variance, s^2, as an estimator of the population variance, σ^2.
- Examples of biased estimators are:
  — sample median as an estimator of the population median.
  — sample range as an estimator of the population range.
  — sample standard deviation, s, as an estimator of the population standard deviation, σ.

Example (Estimators and sampling distributions)

The world has just began and there are only five families. The number of kids that these families have are $0, 1, 1, 2, 0$. So, the (population) probabilities of having 0, 1, or 2 kids are $0.4, 0.4$, and 0.2, respectively.

a) Find the (population) mean number of kids that families have.
b) Assume that now you don't know the population values anymore and that you randomly select five families and the number of kids they have are $1, 1, 0, 2, 1$. What statistic would you use to describe the mean number of kids that families have.
c) How are the words "population parameter", "population mean", "sample mean", "statistic", "estimator", "sampling distribution", "sampling distribution of the sample mean", and "biased/unbiased estimator", relate in this context?
d) Describe in words how to find the sampling distribution of the sample mean number of kids that a family has.
e) What is the mean of the sampling distribution of the sample mean.

Sampling Distributions

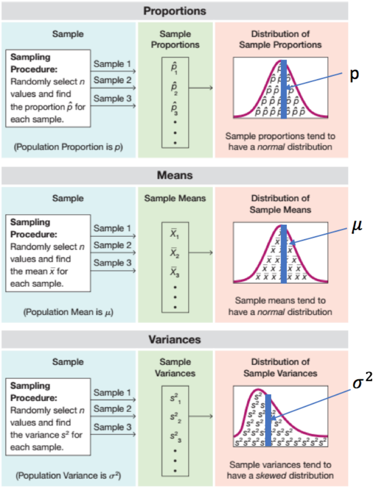

FIGURE 6-16 General Behavior of Sampling Distributions

Practice

Look at the exercises at the end of Section 6-3 in page 249.

Specially, look at exercises:
1-10.

The Central Limit Theorem

Section 6-4

- In this sections we will:
  - Introduce the central limit theorem (approximate distribution for sample means).
  - Motivate inferential statistics.

- The Central Limit Theorem allows us to use a normal distribution for some very meaningful and important applications.
- The Central Limit Theorem states the distribution of the sample mean, $\overline{x}$, regardless of the distribution of the random variables, x.
- If the random variables, x, follow a normal distribution, then the Central Limit Theorem provides the exact distribution of the sample mean, $\overline{x}$.
- For random variables, x, in general (continuous, discrete, skewed, whatever), then the Central Limit Theorem provides an approximation ($n > 30$) of the distribution of the sample mean, $\overline{x}$.

The Central Limit Theorem

Definition (Central Limit Theorem)

For all samples of the same size n with $n > 30$, the sampling distribution of $\overline{x}$ can be approximated by a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$.

Example

Assume that pulse rates of adult males are normally distributed with a mean of 69.6 bpm and a standard deviation of 11.3 bpm.

a) One adult male is randomly chosen from this population. Find the probability that he has a pulse rate greater than 76 bpm.

b) A random sample of 16 adult males is chosen from this population. Find the probability that their mean pulse rate is greater than 76 bpm.

c) Given that part b) involves a sample size that is not larger than 30, why can the central limit theorem be used?

The Central Limit Theorem

- **The Rare Event Rule for Inferential Statistics**: If, under a given *assumption*, the probability of a particular observed event is very small and the *observed event* occurs *significantly less* than or *significantly greater* than what we typically expect with that assumption, we *conclude* that the assumption is probably not correct.

Example (Introduction to Hypothesis Testing)

Assume that the population of human body temperatures has a mean of 98.6F, as is commonly believed. Also assume that the population standard deviation is 0.62F (based on data from University of Maryland researchers). A sample of size $n = 106$ is randomly selected.

a) Find the probability of getting a sample mean of 98.2F or lower. (The value of 98.2F was actually obtained from researchers).

b) How can this result be understood?

Practice

Look at the exercises at the end of Section 6-4 in page 258.

Specially, look at exercises:
1, 2, 4, 5, 7, 10, 13, 17, 20.