Statistical Methods for the Biological, Environmental, and Health Sciences

STAT 007

Normal Probability Distributions

Chapter 6

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.

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

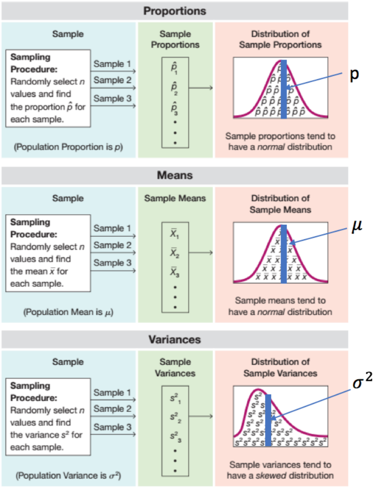


FIGURE 6-16 General Behavior of Sampling Distributions

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 a study).

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.

Assessing Normality

Section 6-5

- In this sections we will:
  - Determine whether the data set that we are analyzing comes from a normal distribution or not using graphs.

- Many times, statistical methods are developed on the assumption that the sample data are from a population having a normal distribution.

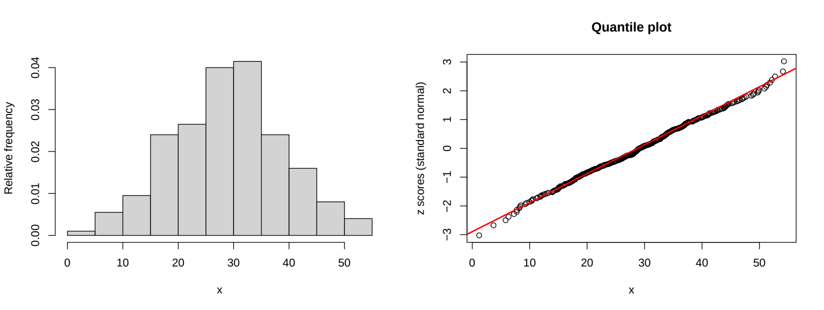
- Here we will present criteria for determining whether the normal distribution assumption is satisfied.

- We will consider the following three sequential criteria:
  1. visual inspection of a *histogram* to see if it is roughly bell-shaped.
  2. identify any *outliers*: if there are more than one, the data might not have a normal distribution.
  3. construct a *quantile plot*: if the pattern of points is reasonably close to a straight line the data has a normal distribution.

- A **normal quantile plot** is a graph of points (x, y) where each x value is from the original set of sample data, and each y value is the corresponding z score that is expected from the standard normal distribution.

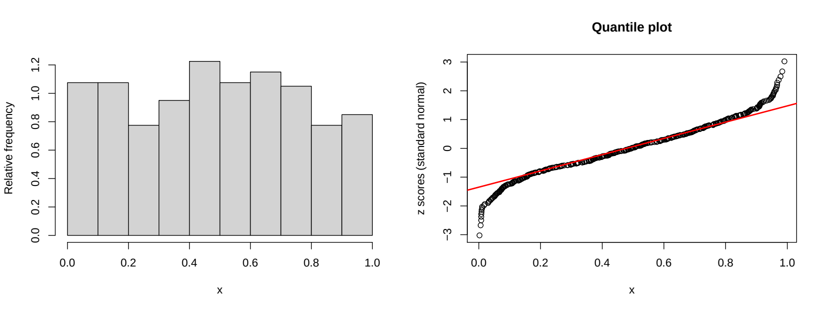
Assessing Normality

Example



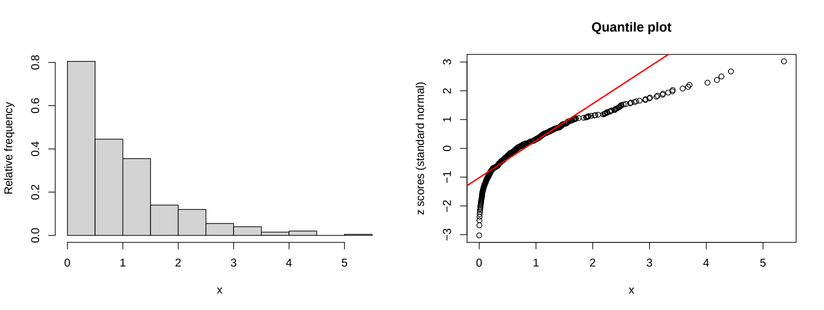
Assessing Normality

Example



Assessing Normality

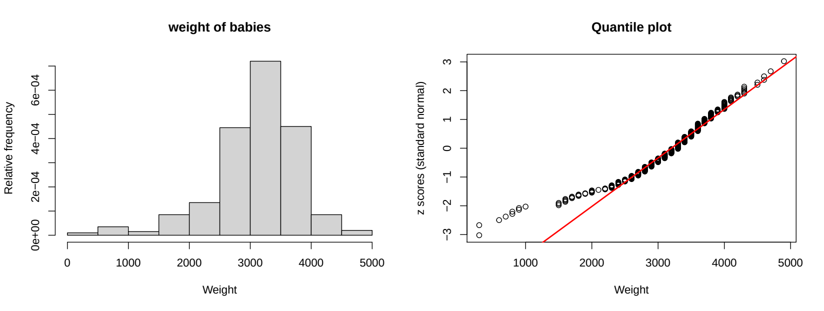
Example



Assessing Normality

Example

The following plots describe the weight of 400 newborn babies. The data contains about 8 outliers.



Practice

Look at the exercises at the end of Section 6-5 in page 266.

Specially, look at exercises:
1, 2, 3, 4, 5-8.

Normal as Approximation to Binomial

Section 6-6

- In this sections we will:
  - Present a method for approximating probabilities from the Binomial distribution using the normal distribution.
  - Introduce a continuity correction when computing this approximated probabilities.

- We will introduce a method for using a normal distribution as an approximation to a binomial probability distribution.
- In order to do this approximation, a few requirements need to be satisfied (details in next slide).
- Because the Binomial probability distribution is discrete and the normal approximation is continuous, we use a "continuity correction" so that a whole number x is represented by the interval $x - 0.5$ and $x + 0.5$.
- Rationale for using the normal approximation:
  - a) the sampling distribution of the sample proportion tends to approximate a normal distribution
  - b) Histograms of data from the Binomial distribution tend to have a bell shape.

Procedure

a) (Requirement 1) The sample is a *simple random sample* of size n from a population in which the proportion of successes is p, **or** the sample is the result of conducting n independent trials of the *binomial experiment* in which the probability of success is p.

b) (Requirement 2) $np \geq 5$ **and** $n(1-p) \geq 5$.

c) If requirements 1 and 2 are satisfied, then the probability distribution of the random variable x can be approximated by a normal distribution with parameters $\mu = np$ and $\sigma = \sqrt{np(1-p)}$.

d) When using the normal approximation, adjust the discrete whole number x by using a "continuity correction". The following table gives some examples:

Statement about *discrete* value	Area of the *continuous* normal distribution
at least x (includes x and above)	to the right of $x - 0.5$
more than x (doesn't include x)	to the right of $x + 0.5$
at most x (includes x and below)	to the left of $x + 0.5$
fewer than x (doesn't include x)	to the left of $x - 0.5$
exactly x	between $x - 0.5$ and $x + 0.5$

Example

In one of Mendel's famous hybridization experiments, he expected that among 580 offspring peas, 145 (or 25%) of them would be yellow, but form the experiment he actually got 152 yellow peas. Assume that Mendel's 25% is correct.

a) What requirements need to be met to use the normal distribution as an approximation to the binomial distribution?

b) What are the parameters of the normal distribution?

c) Find the probability of observing exactly 152 yellow peas.

d) The observed value of 152 yellow peas is considered significantly high if the probability of observing 152 or more yellow peas is less than 0.05 (unlikely). Is the observed value significantly high?

e) Use the rule of thumb to determine whether the observed 152 yellow peas is a significantly high value.

Practice

Look at the exercises at the end of Section 6-6 in page 274.

Specially, look at exercises:
1, 2, 3, 4, 5-8, 13-16.