SECTION 6.1 SUMMARY

- The purpose of a **confidence interval** is to estimate an unknown parameter with an indication of how accurate the estimate is and of how confident we are that the result is correct.

- Any confidence interval has two parts: an interval computed from the data and a confidence level. The interval often has the form

$$\text{estimate} \pm \text{margin of error}$$

- The **confidence level** states the probability that the method will give a correct answer. That is, if you use 95% confidence intervals, in the long run 95% of your intervals will contain the true parameter value. When you apply the method once (that is, to a single sample), you do not know if your interval gave a correct answer (this happens 95% of the time) or not (this happens 5% of the time).

- The **margin of error** for a level C confidence interval for the mean μ of a Normal population with known standard deviation σ, based on an SRS of size n, is given by

$$m = z^* \frac{\sigma}{\sqrt{n}}$$

Here z^* is obtained from the row labeled z^* at the bottom of **Table D**. The probability is C that a standard Normal random variable takes a value between $-z^*$ and z^*. The confidence interval is

$$\bar{x} \pm m$$

If the population is not Normal and n is large, the confidence level of this interval is approximately correct.

- Other things being equal, the margin of error of a confidence interval decreases as
  - the confidence level C decreases,
  - the sample size n increases, and
  - the population standard deviation σ decreases.
- The sample size n required to obtain a confidence interval of specified margin of error m for a population mean is

$$n = \left(\frac{z^* \sigma}{m}\right)^2$$

where z^* is the critical point for the desired level of confidence.

- A specific confidence interval formula is correct only under specific conditions. The most important conditions concern the method used to produce the data. Other factors such as the form of the population distribution may also be important. These conditions should be investigated *prior* to any calculations.

SECTION 6.2 SUMMARY

- A **test of significance** is intended to assess the evidence provided by data against a **null hypothesis** H_0 in favor of an **alternative hypothesis** H_a.

- The hypotheses are stated in terms of population parameters. Usually, H_0 is a statement that no effect or no difference is present, and H_a says that there is an effect or difference in a specific direction (**one-sided alternative**) or in either direction (**two-sided alternative**).

- The test is based on a **test statistic**. The ***P*-value** is the probability, computed assuming that H_0 is true, that the test statistic will take a value at least as extreme as that actually observed. Small *P*-values indicate strong evidence against H_0. Calculating *P*-values requires knowledge of the sampling distribution of the test statistic when H_0 is true.

- If the *P*-value is as small or smaller than a specified value α, the data are **statistically significant** at significance level α.

- Significance tests for the hypothesis H_0: $\mu = \mu_0$ concerning the unknown mean μ of a population are based on the ***z* statistic**:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

The z test assumes an SRS of size n, known population standard deviation σ, and either a Normal population or a large sample. *P*-values are computed from the Normal distribution (**Table A**). Fixed α tests use the table of **standard Normal critical values** (**Table D**).

SECTION 6.3 SUMMARY

- P-values are more informative than the reject-or-not result of a level α test. Beware of placing too much weight on traditional values of α, such as α = 0.05.

- Very small effects can be highly significant (small P), especially when a test is based on a large sample. A statistically significant effect need not be practically important. Plot the data to display the effect you are seeking, and use confidence intervals to estimate the actual values of parameters.

- On the other hand, lack of significance does not imply that H_0 is true, especially when the test has a low probability of detecting an effect.

- Significance tests are not always valid. Faulty data collection, outliers in the data, and testing a hypothesis on the same data that suggested the hypothesis can invalidate a test. Many tests run at once will probably produce some significant results by chance alone, even if all the null hypotheses are true.

SECTION 6.4 SUMMARY

- The **power** of a significance test measures its ability to detect an alternative hypothesis. The power to detect a specific alternative is calculated as the probability that the test will reject H_0 when that alternative is true. This calculation requires knowledge of the sampling distribution of the test statistic under the alternative hypothesis. Increasing the size of the sample increases the power when the significance level remains fixed.

- An alternative to significance testing regards H_0 and H_a as two statements of equal status that we must decide between. This **decision theory** point of view regards statistical inference in general as giving rules for making decisions in the presence of uncertainty.

- In the case of testing H_0 versus H_a, decision analysis chooses a decision rule on the basis of the probabilities of two types of error. A **Type I error** occurs if H_0 is rejected when it is in fact true. A **Type II error** occurs if H_0 is accepted when in fact H_a is true.

- In a fixed level α significance test, the significance level α is the probability of a Type I error, and the power to detect a specific alternative is 1 minus the probability of a Type II error for that alternative.

SECTION 7.1 SUMMARY

- Significance tests and confidence intervals for the mean μ of a Normal population are based on the sample mean $\bar{x}$ of an SRS. Because of the central limit theorem, the resulting procedures are approximately correct for other population distributions when the sample is large.

- The **standard error** of the sample mean is

$$\mathrm{SE}_{\bar{x}} = \frac{s}{\sqrt{n}}$$

- The standardized sample mean, or **one-sample *z* statistic**,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

has the $N(0, 1)$ distribution. If the standard deviation $\sigma/\sqrt{n}$ of $\bar{x}$ is replaced by the **standard error** $s/\sqrt{n}$, the **one-sample *t* statistic**

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

has the ***t* distribution** with $n - 1$ degrees of freedom.

- There is a t distribution for every positive **degrees of freedom *k***. All are symmetric distributions similar in shape to Normal distributions. The $t(k)$ distribution approaches the $N(0, 1)$ distribution as k increases.

- A level C **confidence interval for the mean** μ of a Normal population is

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

where t^* is the value for the $t(n-1)$ density curve with area C between $-t^*$ and t^*. The quantity

$$t^* \frac{s}{\sqrt{n}}$$

is the **margin of error**.

- Significance tests for H_0: $\mu = \mu_0$ are based on the t statistic. P-values or fixed significance levels are computed from the $t(n-1)$ distribution.

- A matched pairs analysis is needed when subjects or experimental units are matched in pairs or when there are two measurements on each individual or experimental unit and the question of interest concerns the difference between the two measurements.

- The one-sample procedures are used to analyze **matched pairs** data by first taking the differences within the matched pairs to produce a single sample.

- One-sample **equivalence testing** assesses whether a population mean μ is practically different from a hypothesized mean μ_0. This test requires a threshold δ, which represents the largest difference between μ and μ_0 such that the means are considered equivalent.

- The t procedures are relatively **robust** against non-Normal populations. The t procedures are useful for non-Normal data when $15 \leq n < 40$ unless the data show outliers or strong skewness. When $n \geq 40$, the t procedures can be used even for clearly skewed distributions.

SECTION 7.2 SUMMARY

- Significance tests and confidence intervals for the difference between the means μ_1 and μ_2 of two Normal populations are based on the difference $\bar{x}_1 - \bar{x}_2$ between the sample means from two independent SRSs. Because of the central limit theorem, the resulting procedures are approximately correct for other population distributions when the sample sizes are large.

- When independent SRSs of sizes n_1 and n_2 are drawn from two Normal populations with parameters μ_1, σ_1 and μ_2, σ_2 the **two-sample *z* statistic**

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

has the $N(0, 1)$ distribution.

- The **two-sample *t* statistic**

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

does *not* have a t distribution. However, good approximations are available.

- **Conservative inference procedures** for comparing μ_1 and μ_2 are obtained from the two-sample t statistic by using the $t(k)$ distribution with degrees of freedom k equal to the smaller of $n_1 - 1$ and $n_2 - 1$.

- **More accurate probability values** can be obtained by estimating the degrees of freedom from the data. This is the usual procedure for statistical software.

- An approximate level C **confidence interval** for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) \pm t* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Here, $t*$ is the value for the $t(k)$ density curve with area C between $-t*$ and $t*$, where k is computed from the data by software or is the smaller of $n_1 - 1$ and $n_2 - 1$. The quantity

$$t* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

is the **margin of error**.

- Significance tests for $H_0 : \mu_1 - \mu_2 = \Delta_0$ use the **two-sample *t* statistic**

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

The P-value is approximated using the $t(k)$ distribution where k is estimated from the data using software or is the smaller of $n_1 - 1$ and $n_2 - 1$.

- The guidelines for practical use of two-sample t procedures are similar to those for one-sample t procedures. Equal sample sizes are recommended.

- If we can assume that the two populations have equal variances, **pooled two-sample *t* procedures** can be used. These are based on the **pooled estimator**

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

of the unknown common variance and the $t(n_1 + n_2 - 2)$ distribution. We do not recommend this procedure for regular use.

SECTION 7.3 SUMMARY

- The **sample size** required to obtain a confidence interval with an expected margin of error no larger than m for a population mean satisfies the constraint

$$m \geq t^* s^* / \sqrt{n}$$

where t^* is the critical value for the desired level of confidence with $n - 1$ degrees of freedom, and s^* is the guessed value for the population standard deviation.

- The sample sizes necessary for a two-sample confidence interval can be obtained using a similar constraint, but guesses of both standard deviations and an estimate for the degrees of freedom are required. We suggest using the smaller of $n_1 - 1$ and $n_2 - 1$ for degrees of freedom.

- The **power** of the one-sample t test can be calculated like that of the z test, using an approximate value for both σ and s.

- The **power** of the two-sample t test is found by first finding the critical value for the significance test, the degrees of freedom, and the **noncentrality parameter** for the alternative of interest. These are used to calculate the power from a **noncentral *t* distribution**. A Normal approximation works quite well. Calculating margins of error for various study designs and conditions is an alternative procedure for evaluating designs.

- The **sign test** is a **distribution-free test** because it uses probability calculations that are correct for a wide range of population distributions.

- The sign test for "no treatment effect" in matched pairs counts the number of positive differences. The P-value is computed from the $B(n, 1/2)$ distribution, where n is the number of non-0 differences. The sign test is less powerful than the t test in cases where use of the t test is justified.

Section 8.1 Summary

- Inference about a population proportion p from an SRS of size n is based on the **sample proportion** $\hat{p} = X/n$. When n is large, $\hat{p}$ has approximately the Normal distribution with mean p and standard deviation $\sqrt{p(1-p)/n}$.

- For large samples, the level C **margin of error of $\hat{p}$** is

$$m = z^* \mathrm{SE}_{\hat{p}}$$

  where the critical value z^* is the value for the standard Normal density curve with area C between $-z^*$ and z^*, and the **standard error of $\hat{p}$** is

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

- The level C **large-sample confidence interval** is

$$\hat{p} \pm m$$

  We recommend using this interval for 90%, 95%, and 99% confidence whenever the number of successes and the number of failures are both at least 10. When sample sizes are smaller, alternative procedures such as the **plus four estimate** of the population proportion are recommended.

- Tests of $H_0\!: p = p_0$ are based on the z **statistic**

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

  with P-values calculated from the $N(0, 1)$ distribution. Use this procedure when the expected number of successes, np_0, and the expected number of failures, $n(1-p_0)$, are both greater than 10.

- The **sample size** required to obtain a confidence interval of approximate margin of error m for a proportion is found from

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1-p^*)$$

  where p^* is a guessed value for the proportion and z^* is the standard Normal critical value for the desired level of confidence. To ensure that the margin of error of the interval is less than or equal to m no matter what $\hat{p}$ may be, use

$$n = \frac{1}{4}\left(\frac{z^*}{m}\right)^2$$

- Software can be used to determine the sample sizes for significance tests. Inputs include the significance level, the desired power, the null hypothesized value of p, and the alternative value of p.

Section 8.2 SUMMARY

- The **large-sample estimate of the difference in two population proportions** is

$$D = \hat{p}_1 - \hat{p}_2$$

where $\hat{p}_1$ and $\hat{p}_2$ are the sample proportions:

$$\hat{p}_1 = \frac{X_1}{n_1} \quad \text{and} \quad \hat{p}_2 = \frac{X_2}{n_2}$$

- The **standard error of *D*** is

$$\mathrm{SE}_D = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

- The level C **margin of error of *D*** is

$$m = z^*\mathrm{SE}_D$$

where z^* is the value for the standard Normal density curve with area C between $-z^*$ and z^*.

- The level C **large-sample confidence interval for *D*** is

$$D \pm m$$

We recommend using this interval for 90%, 95%, or 99% confidence when the number of successes and the number of failures in both samples are all at least 10. When sample sizes are smaller, alternative procedures such as the **plus four estimate of the difference in two population proportions** are recommended.

- Significance tests of $H_0\colon p_1 = p_2$ use the z **statistic**

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\mathrm{SE}_{D_p}}$$

with P-values from the $N(0, 1)$ distribution. In this statistic,

$$\mathrm{SE}_{D_p} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

and $\hat{p}$ is the **pooled estimate** of the common value of p_1 and p_2:

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

Use this test when the number of successes and the number of failures in each of the samples are at least 5.

- **Relative risk** is the ratio of two sample proportions:

$$\mathrm{RR} = \frac{\hat{p}_1}{\hat{p}_2}$$

Confidence intervals for relative risk are often used to summarize the comparison of two proportions.

Additional Examples
Chapters 6-8

CASE : Bids on Building Contracts

- *Cost Engineering* (Oct. 1988) reports on a study of the percentage difference between the lowest bid and the engineers estimate of the cost for building contracts to be approximately normal distributed with standard deviation approximately 25 percentage point (based on four bids).
- How many four bid contracts must be sampled in order to determine with 90% confidence the mean difference between lowest bid and the engineers estimate within 5 percentage points of the true value?

CASE : Passing Rates

- At *University of Nowhere*, the overall passing rate (the percentage of students earning a C or better) is approximately 80% (based on all courses offered).
- Some people claim, that mathematics courses are harder to pass than other courses.
- In a sample of 259 students, 67 failed a mathematical reasoning course.
- Is there statistical evidence to support the claim?
- In what range is the actual pass rate (at 95% confidence level)?

CASE : Rating the Quality of Hospital Maternity Care

- A health management firm wants to know which of two hospitals has the higher rating among former patients for such areas as service, food, cleanliness, staff friendliness, and the quality of the doctors.
- Does the data (based on a questionnaire administered to two samples of former patients) provide evidence to suggest that one hospital is superior to the other?

Hospital Rating Data

Hospital A			Hospital B		
81	86	73	89	55	59
77	90	91	64	37	58
75	62	98	35	57	65
74			68	42	71
			69	49	67

Ratings of two hospitals based on two samples of former patients
SOURCE : R.IMAN : A Data Based Approach to Statistics

Hospital Rating Data

Hospital A		
81	86	73
77	90	91
75	62	98
74		

Hospital B		
89	55	59
64	37	58
35	57	65
68	42	71
69	49	67

$\bar{x}_1 = 80.70 \quad s_1 = 10.65 \;\; n_1 = 10$

$\bar{x}_2 = 59.00 \quad s_2 = 14.19 \;\; n_2 = 15$

Ratings of two hospitals based on two samples of former patients
SOURCE : R.IMAN : A Data Based Approach to Statistics