Chapter 7:
Inferences for Two Samples

Introduction

- In Chapters 5 and 6, we saw how to construct confidence intervals and perform hypothesis tests concerning a single mean or proportion.
- There are cases in which we have two populations, and we wish to study the difference between their means, proportions, or variance.
Motivating Example

- Suppose that a metallurgist is interested in estimating the difference in strength between two types of welds.
- She conducts an experiment in which a sample of 6 welds of one type has an average ultimate testing strength (in ksi) of 83.2 with a standard deviation of 5.2 and a sample of 8 welds of the other type has an average strength of 71.3 with a standard deviation of 3.1.
- It is easy to compute a point estimate for the difference in strengths. The difference between the sample means is $83.2 - 71.3 = 11.9$.
- To construct a CI, however, we need to know how to find a standard error and a critical value for this point estimate.
- To perform a hypothesis test to determine whether we can conclude that the mean strengths differ, we will need to know

Section 7.1: Large-Sample Inferences on Differences Between 2 Population Means

Set-Up:

Let $X$ and $Y$ be independent, with $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$. Then

$$X - Y \sim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2). \quad (7.1)$$
Let $X_1, ..., X_n$ be a large random sample of size $n_X$ from a population with mean $\mu_X$ and standard deviation $\sigma_X$, and let $Y_1, ..., Y_n$ be a large random sample of size $n_Y$ from a population with mean $\mu_Y$ and standard deviation $\sigma_Y$. If the two samples are independent, then a level $100(1-\alpha)$% CI for $\mu_X - \mu_Y$ is

$$(\bar{X} - \bar{Y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

(7.2)

When the values of $\sigma_X$ and $\sigma_Y$ are unknown, they can be replaced with the sample standard deviations $s_X$ and $s_Y$.

Example 7.1

The chemical composition of soil varies with depth. An article in *Communications in Soil Science and Plant Analysis* describes chemical analyses of soil taken from a farm in Western Australia. Fifty specimens were each taken at depths 50 and 250 cm. At a depth of 50 cm, the average NO$_3$ concentration (in mg/L) was 88.5 with a standard deviation of 49.4. At a depth of 250 cm, the average concentration was 110.6 with a standard deviation of 51.5. Find a 95% confidence interval for the difference in NO$_3$ concentrations at the two depths.
Hypothesis Tests on the Difference Between Two Means

- Now, we are interested in determining whether or not the means of two populations are equal to some specified value.
- The data will consist of two samples, one from each population.
- We will compute the difference of the sample means. Since each of the sample means follows an approximate normal distribution, the difference is approximately normal as well.

Hypothesis Test

- Let \( X_1, \ldots, X_n \) and \( Y_1, \ldots, Y_m \) be large (e.g., \( n_X \geq 30 \) and \( n_Y \geq 30 \)) samples from populations with mean \( \mu_X \) and \( \mu_Y \) and standard deviations \( \sigma_X \) and \( \sigma_Y \), respectively. Assume the samples are drawn independently of each other.
- To test a null hypothesis of the form \( H_0: \mu_X - \mu_Y \leq \Delta_0 \), \( H_0: \mu_X - \mu_Y \geq \Delta_0 \), or \( H_0: \mu_X - \mu_Y = \Delta_0 \).
- Compute the z-score:
  \[
  Z = \frac{(\bar{X} - \bar{Y}) - \Delta_0}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \tag{7.4}
  \]
- If \( \sigma_X \) and \( \sigma_Y \) are unknown they may be approximated by \( s_X \) and \( s_Y \).
Compute the $P$-value. The $P$-value is an area under the normal curve, which depends on the alternate hypothesis as follows.

- If the alternative hypothesis is $H_1: \mu_X - \mu_Y > \Delta_0$, then the $P$-value is the area to the right of $z$.
- If the alternative hypothesis is $H_1: \mu_X - \mu_Y < \Delta_0$, then the $P$-value is the area to the left of $z$.
- If the alternative hypothesis is $H_1: \mu_X - \mu_Y \neq \Delta_0$, then the $P$-value is the sum of the areas in the tails cut off by $z$ and $-z$.

Example 7.2

An article compares properties of welds made using carbon dioxide as a shielding gas with those of welds made using a mixture of argon and carbon dioxide. One property studied was the diameter of inclusions, which are particles embedded in the weld. A sample of 544 inclusions in welds made using argon shielding averaged 0.37 $\mu$m in diameter, with a standard deviation of 0.25 $\mu$m. A sample of 581 inclusions in welds made using carbon dioxide shielding averaged 0.40 $\mu$m in diameter, with a standard deviation of 0.26 $\mu$m. Can you conclude that the mean diameters of inclusions differ between the two shielding gases?
Section 7.2: Inference on the Difference Between 2 Proportions

- **Set-Up:**
  - Let $X$ be the number of successes in $n_X$ independent Bernoulli trials with success probability $p_X$, and let $Y$ be the number of successes in $n_Y$ independent Bernoulli trials with success probability $p_Y$, so that $X \sim Bin(n_X, p_X)$ and $Y \sim Bin(n_Y, p_Y)$.
  
- Define
  
  \[
  \tilde{n}_X = n_X + 2 \quad \quad \quad \tilde{n}_Y = n_Y + 2
  \]
  
  \[
  \tilde{p}_X = \frac{X + 1}{\tilde{n}_X} \quad \quad \quad \tilde{p}_Y = \frac{Y + 1}{\tilde{n}_Y}
  \]

\[ (\tilde{p}_X - \tilde{p}_Y) \pm z_{\alpha/2} \sqrt{\frac{\tilde{p}_X (1 - \tilde{p}_X)}{\tilde{n}_X} + \frac{\tilde{p}_Y (1 - \tilde{p}_Y)}{\tilde{n}_Y}} \quad (7.5) \]

- If the lower limit of the confidence interval is less than -1, replace it with -1.
- If the upper limit of the confidence interval is greater than 1, replace it with 1.
- There is a traditional confidence interval as well. It is a generalization of the one for a single proportion.
Example 7.4

Methods for estimating strength and stiffness requirements should be conservative in that they should overestimate rather than underestimate. The success rate of such a method can be measured by a probability of an overestimate. An article in *Journal of Structural Engineering* presents the results of an experiment that evaluated a standard method for estimating the brace force for a compression web brace. In a sample of 380 short test columns the method overestimated the force for 304 of them, and in a sample of 394 long test columns, the method overestimated the force for 360 of them. Find a 95% confidence interval for the difference between the success rates for long columns and short columns.

Hypothesis Tests on the Difference Between Two Proportions

- The procedure for testing the difference between two populations is similar to the procedure for testing the difference between two means.
- One of the null and alternative hypotheses are $H_0: p_x - p_y \geq 0$ versus $H_1: p_x - p_y < 0$. 
The test is based on the statistic $\hat{p}_X - \hat{p}_Y$.
- We must determine the null distribution of this statistic.
- By the Central Limit Theorem, since $n_X$ and $n_Y$ are both large, we know that the sample proportions for $X$ and $Y$ have an approximately normal distribution.

More on Proportions

- The difference between the proportions is also normally distributed.

- Let $\hat{p} = \frac{X+Y}{n_X+n_Y}$, then
  $$\hat{p}_X - \hat{p}_Y \sim N \left(0, \hat{p}(1-\hat{p}) \left(\frac{1}{n_X} + \frac{1}{n_Y}\right)\right) \quad (7.8)$$
Hypothesis Test

- Let $X \sim Bin(n_X, p_X)$ and $Y \sim Bin(n_Y, p_Y)$. Assume $n_X$ and $n_Y$ are large, and that $X$ and $Y$ are independent.

- To test a null hypothesis of the form $H_0: p_X - p_Y \leq 0$, $H_0: p_X - p_Y \geq 0$, and $H_0: p_X - p_Y = 0$.

- Compute
  \[ \hat{p}_X = \frac{X}{n_X}, \quad \hat{p}_Y = \frac{Y}{n_Y}, \quad \text{and} \quad \hat{p} = \frac{X + Y}{n_X + n_Y} \]

- Compute the $z$-score:
  \[ z = \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}(1 - \hat{p})(\frac{1}{n_X} + \frac{1}{n_Y})}} \quad (7.9) \]

P-value

Compute the $P$-value. The $P$-value is an area under the normal curve, which depends on the alternative hypothesis as follows:

- If the alternative hypothesis is $H_1: p_X - p_Y > 0$, then the $P$-value is the area to the right of $z$.

- If the alternative hypothesis is $H_1: p_X - p_Y < 0$, then the $P$-value is the area to the left of $z$.

- If the alternative hypothesis is $H_1: p_X - p_Y \neq 0$, then the $P$-value is the sum of the areas in the tails cut off by $z$ and -$z$. 
Example 7.5

Industrial firms often employ methods of “risk transfer”, such as insurance or indemnity clauses in contracts, as a technique of risk management. An article reports the results of a survey in which managers were asked which methods played a major role in the risk management strategy of their firms. In a sample of 43 oil companies, 22 indicated that risk transfer played a major role, while in a sample of 93 construction companies, 55 reported that risk transfer played a major role. Can we conclude that the proportion of oil companies that employ the method of risk transfer is less than the proportion of construction companies that do?

Section 7.3: Small-Sample Inferences on the Difference Between 2 Means

Let $X_1, \ldots, X_{n_X}$ be a random sample of size $n_X$ from a normal population with mean $\mu_X$ and standard deviation $\sigma_X$, and let $Y_1, \ldots, Y_{n_Y}$ be a random sample of size $n_Y$ from a normal population with mean $\mu_Y$ and standard deviation $\sigma_Y$. Assume that the two samples are independent. If the populations do not necessarily have the same variance, a level $100(1-\alpha)\%$ CI for $\mu_X - \mu_Y$ is

$$
\bar{X} - \bar{Y} \pm t_{v, \alpha/2} \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}
$$

The number of degrees of freedom, $v$, is given by (rounded down to the nearest integer)

$$
v = \frac{\left(\frac{s_X^2}{n_X}\right)^2 + \left(\frac{s_Y^2}{n_Y}\right)^2}{\frac{(s_X^2/n_X)^2}{n_X - 1} + \frac{(s_Y^2/n_Y)^2}{n_Y - 1}}
$$
Example 7.6

Resin-based composites are used in restorative dentistry. An article presents a comparison of the surface hardness of specimens cured for 40 seconds with constant power with that of specimens cured for 40 seconds with exponentially increasing power. Fifteen specimens were cured with each method. Those cured with constant power had an average surface hardness of 400.9 with a standard deviation of 10.6. Those cured with an exponentially increasing power had an average surface hardness of 367.2 with a standard deviation of 6.1. Find a 98% confidence interval for the difference in mean hardness between specimens cured by the two methods.

Another CI

Suppose we have the same set-up as before, but the populations are known to have nearly the same variance. Then a 100(1 − \( \alpha \))% CI for \( \mu_X - \mu_Y \) is

\[
\bar{X} - \bar{Y} \pm t_{n_X + n_Y - 2, \alpha/2} \cdot s_p \cdot \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}
\]

The quantity \( s_p^2 \) is the pooled variance, given by

\[
s_p^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}
\]
Hypothesis Test for Unequal Variance

- Let \( X_1, \ldots, X_n \) and \( Y_1, \ldots, Y_n \) be samples from normal populations with mean \( \mu_X \) and \( \mu_Y \) and standard deviations \( \sigma_X \) and \( \sigma_Y \), respectively. Assume the samples are drawn independently of each other.
- Assume that \( \sigma_X \) and \( \sigma_Y \) are not known to be equal.
- To test a null hypothesis of the form \( H_0: \mu_X - \mu_Y \leq \Delta_0 \), \( H_0: \mu_X - \mu_Y \geq \Delta_0 \), or \( H_0: \mu_X - \mu_Y = \Delta_0 \).
- Compute \( \nu \), rounded to the nearest integer.

\[
\nu = \frac{\left( \frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y} \right)^2}{\frac{s_X^2}{n_X}^2 \left( \frac{s_X^2}{n_X} \right) + \frac{s_Y^2}{n_Y}^2 \left( \frac{s_Y^2}{n_Y} \right) - 1} + \frac{n_Y - 1}{n_Y - 1} \]

Compute the test statistic \( t = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}} \) (7.13)

P-value

Compute the P-value. The P-value is an area under the Student’s t curve with \( \nu \) degrees of freedom, which depends on the alternate hypothesis as follows.
- If the alternative hypothesis is \( H_1: \mu_X - \mu_Y > \Delta_0 \), then the P-value is the area to the right of \( t \).
- If the alternative hypothesis is \( H_1: \mu_X - \mu_Y < \Delta_0 \), then the P-value is the area to the left of \( t \).
- If the alternative hypothesis is \( H_1: \mu_X - \mu_Y \neq \Delta_0 \), then the P-value is the sum of the areas in the tails cut off by \( t \) and \(-t\).
Example 7.7

Good website design can make Web navigation easier. An article presents a comparison of item recognition between two designs. A sample of 10 users using a conventional Web design averaged 32.3 items identified, with a standard deviation of 8.56. A sample of 10 users using a new structured Web design averaged 44.1 items identified, with a standard deviation of 10.09. Can we conclude that the mean number of items identified is greater with the new structured design?

Hypothesis Test with Equal Variance

- Let $X_1, ..., X_{n_X}$ and $Y_1, ..., Y_{n_Y}$ be samples from normal populations with mean $\mu_X$ and $\mu_Y$ and standard deviations $\sigma_X$ and $\sigma_Y$, respectively. Assume the samples are drawn independently of each other.
- Assume that $\sigma_X$ and $\sigma_Y$ are known to be equal.
- To test a null hypothesis of the form $H_0: \mu_X - \mu_Y \leq \Delta_0$, $H_0: \mu_X - \mu_Y \geq \Delta_0$, or $H_0: \mu_X - \mu_Y = \Delta_0$.
- Compute $s_p = \sqrt{\frac{(n_X-1)s_X^2 + (n_Y-1)s_Y^2}{n_X + n_Y - 2}}$
- Compute the test statistic $t = \frac{(\bar{X} - \bar{Y}) - \Delta_0}{s_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$
**P-value**

Compute the $P$-value. The $P$-value is an area under the Student’s $t$ curve with $v$ degrees of freedom, which depends on the alternate hypothesis as follows.

- If the alternative hypothesis is $H_1: \mu_X - \mu_Y > \Delta_0$, then the $P$-value is the area to the right of $t$.
- If the alternative hypothesis is $H_1: \mu_X - \mu_Y < \Delta_0$, then the $P$-value is the area to the left of $t$.
- If the alternative hypothesis is $H_1: \mu_X - \mu_Y \neq \Delta_0$, then the $P$-value is the sum of the areas in the tails cut off by $t$ and $-t$.

---

**Section 7.4: Inference for Paired Data**

**Set-Up:**
Consider paired data. An example is tread wear on tires. A manufacturer wishes to compare the tread wear of tires made of a new material with that of tires made of a conventional material. One tire of each type is placed on each front wheel of 10 front-wheel-drive automobiles. The choice as to which type of tire goes on the right wheel and which goes on the left is made with the flip of a coin. Each car is driven for 40,000, a measurement of tread wear is then made on each tire. The measurements are not independent, since the tires are on the same car.
More on Paired Data

Let \((X_1, Y_1), \ldots, (X_n, Y_n)\) be sample pairs. Let \(D_i = X_i - Y_i\). Let \(\mu_X\) and \(\mu_Y\) represent the population means for \(X\) and \(Y\), respectively. We wish to find a CI for the difference \(\mu_X - \mu_Y\). Let \(\mu_D\) represent the population mean of the differences, then \(\mu_D = \mu_X - \mu_Y\). It follows that a CI for \(\mu_D\) will also be a CI for \(\mu_X - \mu_Y\).

Now, the sample \(D_1, \ldots, D_n\) is a random sample from a population with mean \(\mu_D\), we can use one-sample methods to find CIs for \(\mu_D\).
Confidence Interval

Let $D_1, ..., D_n$ be a small random sample ($n < 30$) of differences of pairs. If the population of differences is approximately normal, then a level 100(1-$\alpha$)% CI for $\mu_D$ is

$$\bar{D} \pm t_{n-1,\alpha/2} \frac{s_D}{\sqrt{n}}$$

(7.14)

If the sample size is large, a level 100(1-$\alpha$)% CI for $\mu_D$ is

$$\bar{D} \pm z_{\alpha/2} \sigma_D$$

(7.15)

In practice, $\sigma_D$ is approximated with $s_D / \sqrt{n}$.

Hypothesis Tests with Paired Data

- We present a method for testing hypotheses involving the difference between two population means on the basis of such paired data.
- If the sample is large, the $D_i$ need not be normally distributed. The test statistic is

$$Z = \frac{\bar{D} - \mu_0}{s_D / \sqrt{n}}$$

and a $z$-test should be performed.
Hypothesis Test

- Let \((X_1, Y_1), \ldots, (X_n, Y_n)\) be sample of ordered pairs whose differences \(D_1, \ldots, D_n\) are a sample from a normal population with mean \(\mu_D\).
- To test a null hypothesis of the form \(H_0: \mu_D \leq \mu_0\), \(H_0: \mu_D \geq \mu_0\), or \(H_0: \mu_D = \mu_0\).
- Compute the test statistic
  \[
t = \frac{\bar{D} - \mu_0}{S_D / \sqrt{n}}
  \]

P-value

Compute the P-value. The P-value is an area under the Student’s t curve with \(n - 1\) degrees of freedom, which depends on the alternate hypothesis as follows.
- If the alternative hypothesis is \(H_1: \mu_D > \mu_0\), then the P-value is the area to the right of \(t\).
- If the alternative hypothesis is \(H_1: \mu_D < \mu_0\), then the P-value is the area to the left of \(t\).
- If the alternative hypothesis is \(H_1: \mu_D \neq \mu_0\), then the P-value is the sum of the areas in the tails cut off by \(t\) and \(-t\).
Section 7.5: The F Test for Equality of Variances

- Sometimes it is desirable to test a null hypothesis that two populations have equal variances.
- In general, there is no good way to do this.
- In the special case where both populations are normal, there is a method available.
- Let $X_1, ..., X_m$ be a simple random sample from a $N(\mu_1, \sigma_1^2)$ population, and let $Y_1, ..., Y_n$ be a simple random sample from a $N(\mu_2, \sigma_2^2)$ population.
- Assume that the samples are chosen independently.
- The values of the means are irrelevant, we are only concerned with the variances.

Hypotheses

- Let $s_1^2$ and $s_2^2$ be the sample variances.
- Any of three null hypothesis may be tested. They are
  \[ H_0: \frac{\sigma_1^2}{\sigma_2^2} \leq 1, \quad \text{or} \quad \sigma_1^2 \leq \sigma_2^2 \]
  \[ H_0: \frac{\sigma_1^2}{\sigma_2^2} \geq 1, \quad \text{or} \quad \sigma_1^2 \geq \sigma_2^2 \]
  \[ H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1, \quad \text{or} \quad \sigma_1^2 = \sigma_2^2 \]
The test statistic is the ratio of the two sample variances:
\[ F = \frac{s_1^2}{s_2^2} \] (7.16)

When \( H_0 \) is true, we assume that \( \sigma_1^2 / \sigma_2^2 = 1 \), or equivalently \( \sigma_1^2 = \sigma_2^2 \). When \( s_1^2 \) and \( s_2^2 \) are, on average, the same size, \( F \) is likely to be near 1.

When \( H_0 \) is false, we assume that \( \sigma_1^2 > \sigma_2^2 \). When \( s_1^2 \) is likely to be larger than \( s_2^2 \), and \( F \) is likely to be greater than 1.

Statistics that have an \( F \) distribution are ratios of quantities, such as the ratio of two variances.

The \( F \) distribution has two values for the degrees of freedom: one associated with the numerator, and one associated with the denominator.

The degrees of freedom are indicated with subscripts under the letter \( F \).

Note that the numerator degrees of freedom are always listed first.

A table for the \( F \) distribution is provided (Table A.7 in Appendix A).
The null distribution of the $F$ statistic is $F_{m-1,n-1}$.

The number of degrees of freedom for the numerator is one less than the sample size used to compute $s_1^2$, and the number of degrees of freedom for the denominator is one less than the sample size used to compute $s_2^2$.

Note that the $F$ test is sensitive to the assumption that the samples come from normal populations.

If the shapes of the populations differ much from the normal curve, the $F$ test may give misleading results.

The $F$ test does not prove that two variances are equal. The basic reason for the failure to reject the null hypothesis does not justify the assumption that the null hypothesis is true.

Example 7.9

In a series of experiments to determine the absorption rate of certain pesticides into skin, measured amounts of two pesticides were applied to several skin specimens. After a time, the amounts absorbed (in $\mu$g) were measured. For pesticide A, the variance of the amounts absorbed in 6 specimens was 2.3, while for pesticide B, the variance of the amounts absorbed in 10 specimens was 0.6. Assume that for each pesticide, the amounts absorbed are a simple random sample from a normal population. Can we conclude that the variance in the amount absorbed is greater for pesticide A than for pesticide B?
# Summary

- **Large Sample Inference on the Difference Between Two Population Means**
  - Confidence Intervals
  - Hypothesis Tests
- **Inferences on The Difference Between Two Proportions**
  - Confidence Intervals
  - Hypothesis Tests
- **Small Sample Inferences on the Difference Between Two Means**
  - Confidence Intervals
  - Hypothesis Tests
- **Inferences Using Paired Data**
- **The \( F \) Test for Equality of Variance**