MA5701: Statistical Methods

Chapter 3 : Principles of Inference

Kui Zhang, Mathematical Sciences

Example 3.1

- **Example 3.1 (see book)** - The National Center for Education Statistics reports that the year 2007 reading scores for fourth graders had a mean of **220.99** and a standard deviation of **35.73** (from 191,000 fourth grades). You believe that your school district is doing a great job of teaching reading that want to show that mean scores in your district would be higher than this national mean. You randomly select 50 fourth grader in your district, give the same exam, get 230.2 as the sample mean.
- Since your mean is **higher**, this seems to vindicate your belief.
- A critic points out that **you simply may have been lucky** in your sample.
- You can only afford to have this sample; how can you **take sampling variability into account** to explain your high score in your data?

Introduction

- There are two parts of statistical inference:
  - A **statement** about the value of that parameter.
  - A measure of the **reliability** (probability) of that statement.
- **Two different but related objectives of statistical inference**
  - **Tests of hypotheses** - hypothesize that parameters have some specific values or relationships and make decisions about that. Reliability is the probability that the decision is incorrect.
  - **Estimating parameters** - which is usually done in the form of an interval, and reliability is expressed as probability of true value is covered.
- In this chapter, we present basic principles of statistic inference.

Populations and Samples

- **Population –** its characteristic is generally described by the parameter.
  - Usually denoted by Greek letters, such as α, β, μ, etc.
  - It is generally unknown.
  - It is primary of interest of statistical inference and is estimated from sample.
  - In chapter 3, we developed models to describe “populations”.
- **Sample –** its characteristic is generally described by the statistics.
  - Usually denoted by alphabetic letters, such as x, y, z, p, etc.
  - It can be calculated from the data and is considered as known once the data is collected.
  - It is used to estimate the population parameter.

Examples

Population	Samples	Statistical Inference
Difference in elastic strengths of polymer yarn from two machines	Measure the strength of 10 yarns from each machine	Use statistics from the sample to infer the unknown parameters
Parameters	**Statistics**	
$Y_1 \sim N(\mu_1, \sigma_1^2)$ $Y_2 \sim N(\mu_2, \sigma_2^2)$	$\bar{y}_1, s_1^2$ $\bar{y}_2, s_2^2$	For example, use $\bar{y}_1$ as an estimate of μ_1
Parameters are constants, but unknown.	Statistics are random variables, for example, $\bar{Y}_1 \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$. Statistics are fixed once the sample is collected.	**Point Estimation**: $\hat{\mu}_1 = \bar{y}_1$ **Interval Estimation**: $\bar{y}_1 \pm t_{n-1,\alpha/2}\sqrt{\frac{s_1^2}{n_1}}$ **Hypothesis Testing**: $H_0: \mu_1 = \mu_2$

Hypothesis Testing - Overview

- **Statistical Hypothesis** – a statement of population parameters.
- **Test of hypothesis** – A procedure that enables us to agree and disagree with hypothesis using data from a sample.
- Hypothesis testing starts by making a set of two statements about the parameter. These two statements are exclusive and exhaustive, which means that one or the other statement must be true, but they cannot both be true.

Example 3.3 - Filling Peanuts Jars

- **Example 3.3 - Filling Peanuts Jars.** A company that packages salted peanuts in 8-oz. jars is interested in maintaining control on the amount of peanuts put in jars by one of its machines. Control is defined as averaging 8 oz. per jar and not consistently over or under filling the jars. To monitor this control, a sample of 16 jars is taken from the line at random time intervals and their contents weighed. The mean weight of peanuts in these 16 jars will be used to test the hypothesis that the machine is indeed working properly.

Example 3.3 - Filling Peanuts Jars

- **Example 3.3 – Filling Peanuts Jars.** From the sample of 16 jars, we find the sample mean is

$$\bar{y} = 7.89 \text{ oz.}$$

- We further assume that
  - The population standard deviation, σ, is known and $\sigma = 0.2$.
- In this chapter, we assume that the data is either normally distributed or the sample size is large enough so the Central Limit Theorem can be applied.

Hypothesis Testing - Overview

- **Example 3.3 – Filling Peanuts Jars.** The mean weight of peanuts in these 16 jars will be used to test the hypothesis that the machine is indeed working properly.
- We can solve this by the hypothesis testing.
  - If the machine is not working properly, we need to take some appropriate actions, which can be costly.
  - If the machine is working properly, then we can leave the process alone.

Hypothesis Testing - Overview

Example 3.3 – Filling Peanuts Jars.

- We make two complementary statements:
  - The average weight of peanuts in the Jar is 8 oz.
  - The average weight of peanuts in the Jar is different from 8 oz.
- We need to use data (here are weights from 16 samples) to make a decision about two statements.

Null Hypotheses

- **Definition 3.1 – Null hypothesis ($\boldsymbol{H_0}$) -** a statement about the values of one or more parameters. This hypothesis represents the status quo and is usually not rejected unless the sample results strongly imply that it is false.

Alternative Hypotheses

- **Definition 3.2 - Alternative hypothesis (H_a) -** a statement that contradicts the null hypothesis. This hypothesis is accepted if the null hypothesis is rejected. The alternative hypothesis is often called the ***research hypothesis*** because it usually implies that some action is to be performed, some money spent, or some established theory overturned.

Hypotheses – Example 3.3

- The null hypothesis is:

$$H_0 : \mu = 8$$

- If we fail to reject it, we do not need to do anything.
- The alternative hypothesis is:

$$H_a : \mu \neq 8$$

- If we reject the null hypothesis, we need to correct the process – thus we need strong evidence to do so.

Fail to Reject Null Hypothesis

- When we reject the null hypothesis, we have strong evidence to support that the alternative hypothesis is true – so we accept the alternative hypothesis.
- When we fail to reject the null hypothesis, we just do not have strong evidence to support that the alternative hypothesis is true. In this case, we do not (and never) state to "accept the null hypothesis".

Possible Errors in Hypothesis Testing

- A **type I error** occurs when we incorrectly reject H_0, that is, when H_0 is true, and our sample-based inference procedure rejects it.
- A **type II error** occurs when we incorrectly fail to reject H_0, that is, when H_0 is not true, and our inference procedure fails to detect this fact.

	In the Population	
The Decision	H_0 **is True**	H_0 **is Not True**
H_0 **is Not Rejected**	Correct	Type II Error
H_0 **is Rejected**	Type I Error	Correct

A Few More Definitions

- The **rejection region** (also called the **critical region**) is the range of values of a sample statistic that will lead to rejection of the null hypothesis.
- α: denotes the probability of making a type I error;
- β : denotes the probability of making a type II error.
- $1 - \beta$: the power of test.
- In the hypothesis testing, we use a fixed small type I error and try to decrease the type II error (increase the power).

Example 3.3 – Filling Peanuts Jars

- **Example 3.3 – Filling Peanuts Jars.** We would like to use 16 samples to test if the weight is 8 oz.
- **Rejection Region of Example 3.3 –** the sample mean is much larger or less than 8 oz. But the exact rejection depends on α, the probability of making a type I error and variability of the data (or population).
- Here we assume that we reject the null hypothesis when $\bar{y} > 8.1$ or $\bar{y} < 7.9$.

Example 3.3 – Filling Peanuts Jars

- When the mean is 8 oz, there is a probability that $\bar{Y} > 8.1$ or $\bar{Y} < 7.9$, this probability is α, the probability of type I error. Here $\alpha = 0.0455$.
- When the mean is 8.15 oz (not 8 oz), there is a probability that $\bar{Y} < 8.1$ and $\bar{Y} > 7.9$, this probability is β, the probability of type II error. Here $\beta = 0.1587$.
- When the mean is 8.15 oz (not 8 oz), $1 - \beta$, is the probability that $\bar{Y} > 8.1$ or $\bar{Y} < 7.9$, this probability is the power. Here the power is $1 - \beta = 1 - 0.1587 = 0.8413$.

Type I Error, Type II Error, Sample Size

- For any fixed α, an increase in the sample size will cause a decrease in β.
- For a fixed sample size n, a decrease in α will cause an increase in β. Conversely, an increase in α will cause a decrease in β.
- An increase in the sample size will cause a decrease in both α and β.

Fail to Reject Null Hypothesis

- As we have mentioned, we do not “accept” the null hypothesis when we fail to reject the null hypothesis.
- We do not know the population parameter in most situations.
- Also, we use small α to control type I error. In most cases, we use $\alpha = 0.05$. We use smaller α if the rejection of null hypothesis has serious consequences.

Type I Error - Example

- **Example** – A drug company tests a new drug and must consider
  - Toxicity (side effects) – null hypothesis here is the drug is toxic, so what α should be used?
  - For toxicity, since the consequence is so severe if we reject the null hypothesis while the drug is toxic, we use a very small α such as $\alpha = 0.0001$ or less is common.

Type I Error – Example 3.3

- **Example 3.3 –** We would like to know if the weight of peanuts is 8 oz.
- **Example 3.3** – Should we use $\alpha = 0.05$ or $\alpha = 0.01$?
- **Conclusion**: we should use a smaller one, $\alpha = 0.01$.
- **Why?**

Type I Error – Example 3.3

- **Conclusion**: we should use a smaller one, $\alpha = 0.01$.
- Since we conduct the test often, we really are more concerned about the type I error than the type II error.
- If we make a type I error (reject the null when it is true), we need to search for the cause of a change when none is present. We are conducting this test so frequently, we run the risk of constantly searching for problems that do not exit. This can be costly.
- Since we are conducting this test so frequently, if we do not detect a true change on any sample (a type II error), we should pick it up later.

Type I Error, Type II Error, Rejection Region

If the null hypothesis is false, which of these statements characterizes a situation where the value of the test statistic falls in the rejection region?

1. A type I error has been committed.
2. A type II error has been committed.
3. Insufficient information has been given to make a decision.
4. The decision is correct.
5. None of the above is correct.

Type I Error, Type II Error, Rejection Region

If the value of the test statistic does not fall in the rejection region, the decision is:

1. Reject the null hypothesis.
2. Reject the alternative hypothesis.
3. Fail to reject the null hypothesis.
4. Fail to reject the alternative hypothesis.
5. There is insufficient information to make a decision.

Type I Error, Type II Error, Rejection Region

If the value of the test statistic falls in the rejection region, then:

1. We cannot commit a type I error.
2. We cannot commit a type II error.
3. We have proven that the null hypothesis is true.
4. We have proven that the null hypothesis is false.
5. None of the above is correct.

Five-Step Procedure for Hypothesis Testing

- **Step 1:** Specify appropriate H_0, H_a, and an acceptable level of α.
- **Step 2:** Define a sample-based test statistic based on H_0.
- **Step 3:** Find the rejection region for the specified H_a and α.
- **Step 4:** Collect the sample data and calculate the test statistic.
- **Step 5:** Make a decision to either reject or fail to reject H_0. This decision will normally result in a recommendation for action. Then interpret the results in the language of the problem.
- **Note:** It is imperative that the results be usable by the practitioner. Since H_a is of primary interest, this conclusion should be stated in terms of whether there was or was not evidence for the alternative hypothesis.

Step 1 - Specify $\boldsymbol{H_0}$, $\boldsymbol{H_a}$, and Choose $\boldsymbol{\alpha}$

- Note that α and β are inversely related – for a fixed sample size, we can reduce α only at the cost of increasing β.
- **Example 3.3 Filling Peanuts Jars –**

$$H_0: \mu = 8$$
$$H_a: \mu \neq 8$$
$$\alpha = 0.01$$

- We call this alternative as a two-sided alternative.

Three Different Alternatives

- In our course, the null hypothesis is always:

$$H_0\text{: parameter = a value.}$$

- For alternative hypothesis, we have
  - Not equal (two-sided alternative, two-tailed test) -

$$H_a\text{: parameter} \neq \text{a value}$$

  - Greater (one-sided alternative, right-tailed test):

$$H_a\text{: parameter} > \text{a value}$$

  - Less (one-sided alternative, left-tailed test):

$$H_a\text{: parameter} < \text{a value}$$

Three Different Alternatives

	Two-Tailed	Right-Tailed (Greater)	Left-Tailed (Less)
Hypotheses	$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu < \mu_0$
Key Words	Changed Different	Improved Greater Increased	Reduced Less Decreased
Rejection Region	Both Tails	Right Tails	Left Tails

Step 2 – Define Test Statistic

- **Definition 3.8 -** The **test statistic** is a statistic whose sampling distribution can be specified for both the null and alternative hypothesis (although sampling distribution when the alternative hypothesis is true may often be quite complex).
- For example, to test the hypothesis about the population mean and assume that the variance is known, we can use the sample mean.
- The sample mean: $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$ if the random sample from a normal distribution $N(\mu, \sigma^2)$ or by the central limit theorem.

Step 2 – Define Test Statistic

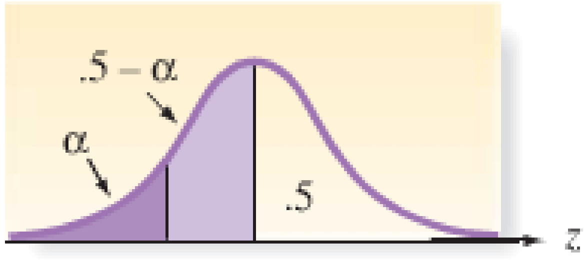
- **Example 3.3 – Filling Peanuts Jars.** We have 16 samples and further assume that the population standard deviation is 0.2 and the data follows a normal distribution.
- The test statistic is $\frac{\bar{Y}-\mu_0}{\sigma/\sqrt{n}} = \frac{\bar{Y}-8}{0.2/\sqrt{16}} \sim N(0,1)$ when $H_0: \mu = 8$ is true.
- The test statistic is $\frac{\bar{Y}-\mu_0}{\sigma/\sqrt{n}} = \frac{\bar{Y}-8}{0.2/\sqrt{16}}$, what is its distribution when $H_a: \mu = 8.2$ is true?

Step 3 – Determine Rejection Region

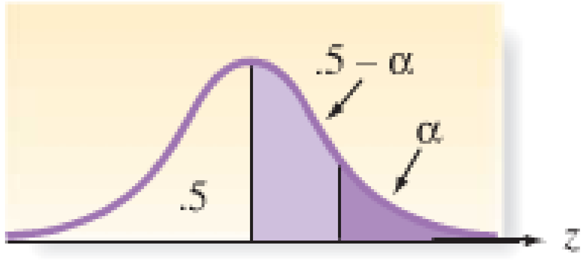
- **Definition 3.9 -** The **rejection region** comprises the values of the test statistic for which the researchers reject the null hypothesis (H_0).
- The probability of the test statistic falling in the rejection region is the specified α when the null hypothesis is true.
- **Decision Rule**:

  We reject the null hypothesis when the test statistic falls in the rejection region.

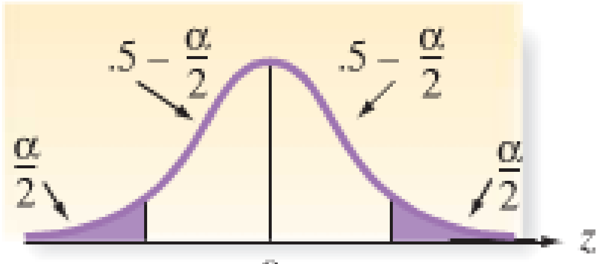
Step 3 – Determine Rejection Region



a. Form of H_a: <



b. Form of H_a: >



c. Form of H_a: ≠

Table 6.2 Rejection Regions for Common Values of α

	Alternative Hypotheses		
	Lower-Tailed	Upper-Tailed	Two-Tailed
$\alpha = .10$	$z < -1.28$	$z > 1.28$	$z < -1.645$ or $z > 1.645$
$\alpha = .05$	$z < -1.645$	$z > 1.645$	$z < -1.96$ or $z > 1.96$
$\alpha = .01$	$z < -2.33$	$z > 2.33$	$z < -2.575$ or $z > 2.575$

Step 3 – Determine Rejection Region

Example 3.3 - Filling Peanuts Jars –

$$H_0: \mu = 8$$
$$H_a: \mu \neq 8$$

$$\alpha = 0.01$$

- This is a two-sided (two-tailed) test.
- The rejection region is:

$$\left|\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}}\right| = \left|\frac{\bar{y}-8}{0.2/\sqrt{16}}\right| > z_{0.005} = 2.575$$

Step 4 and Step 5

- **Step 4:** Collect the sample data and calculate the test statistic.
- **Example 3.3 – Filling Peanuts Jars**, we have

$$z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} = \frac{7.89 - 8.0}{0.2/\sqrt{16}} = -2.2$$

- **Step 5:** Make a decision to either reject or fail to reject H_0.
- **Example 3.3**, we have $|z| = 2.2 < 2.575$, we fail to reject the null hypothesis.

Step 5 - Interpretation

- **Step 5:** Then interpret the results in the language of the problem.
- If we test statistic falls in rejection region, we reject the null hypothesis, we state that:

  At the (   )% significance level, the test statistic ($z =$) falls in the rejection region. Therefore, we reject the null hypothesis. The data provides sufficient evidence that (state the problem specified in the problem).

Step 5 - Interpretation

- **Step 5:** Then interpret the results in the language of the problem.
- If we test statistic does not fall in rejection region, we fail to reject the null hypothesis, we state that:

  At the (   )% significance level, the test statistic ($z =$) does not fall in the rejection region. Therefore, we fail to reject the null hypothesis. The data does not provide sufficient evidence that (state the problem specified in the problem).

Step 5 - Interpretation

- **Step 5:** Then interpret the results in the language of the problem.
- **Example 3.3 – Filling Peanuts Jars**, we state like this way:

  At the 1% significance level, the test statistic $z = -2.2$ does not fall in the rejection region. Therefore, we fail to reject the null hypothesis. The data dose not provide sufficient evidence that the filling process is out of control (or the machine is not working properly).

Step 5 – Interpretation

- When we reject the null hypothesis at the α level, we can also state that "that is statistically significant at the α level".
- When we fail to reject the null hypothesis at the α level, we can also state that "that is not statistically significant at the α level".
- For **Example 3.3** – Filling Peanuts Jars, we can state that "the difference between the weight of filling process and the weight of 8 oz is not statistically significant at the 0.01 level".
- For **Example 3.3** – Filling Peanuts Jars, we can also state that "the weight of filling process is not statistically significantly different from the weight of 8 oz at the 0.01 level".

Step 5 – Interpretation

- Therefore, when we state that “that is statistically significant at the α level”, we just mean that we reject the null hypothesis at the α level.
- Therefore, the true meaning of “that is statistically significant at the α level” is that if the null hypothesis is true, the corresponding statistic observed in the sample would occur by chance with probability of no more than α.

Hypothesis Testing - Interpretation

A research report states: The differences between public and private school seventh graders' attitudes toward minority groups was statistically significant at the α = 0.05 level. This means that:

1. It has been proven that the two groups are different.
2. There is a probability of 0.05 that the attitudes of the two groups are different.
3. There is a probability of 0.95 that the attitudes of the two groups are different.
4. If there is no difference between the groups, the difference observed in the sample would occur by chance with probability of no more than 0.05.
5. None of the above is correct.

P-Values – Observed Significance Level

- Problems with significance level and rejection region:
  - Users would have to specify an alpha and determine the corresponding rejection region for every test being requested.
  - The conclusion may be affected by very minor changes in sample statistics.
  - Only give a “discrete” but not “continuous” scale for rejecting the null hypothesis.
  - In many situations, we may need want to use our own “alpha”.
- Most statistical packages provide *p*-value, since what alpha that a researcher would like to use is generally unknown.

P-Values – Observed Significance Level

- **Definition 3.10 -** The ***p*-value** is the probability of committing a type I error if the actual sample value of the statistic is used as the boundary of the rejection region.
- The ***p*-value** is also the probability of observing a test statistic or more extreme.
- **Example 3.3 – Filling Peanut Jars** – The test statistic is -2.2, we have

$$p\text{-value} = 2\Pr(Z > |-2.2|) = 0.0278$$

P-Values – Observed Significance Level

- If the *p*-value $< \alpha$, we reject the null hypothesis.
- If the *p*-value $> \alpha$, we failed to reject the null hypothesis.
- Therefore, the *p*-value is therefore the smallest level of significance for which we would reject the null hypothesis with that sample.
- Consequently, the *p*-value is often called the “attained” or the “observed” significance level. It is also interpreted as an indicator of the weight of evidence against the null hypothesis.
- Smaller *p*-value indicates stronger evidence to reject the null hypothesis.

P-value Approach

To use the *p*-value to test hypothesis, you can:

- In Step 3, skip the calculation of rejection region.
- In Step 4, calculate the corresponding *p*-value based on H_a in Step 1.
- In Step 5, make a decision based on the *p*-value value and the significance level - α in Step 1. Interpret the results.

Calculation of *p*-value

- For two-sided test (H_a: $\mu \neq \mu_0$),

$$p\text{-value} = 2\Pr(Z > |z|)$$

- For upper-tailed (right-tailed) test (H_a: $\mu > \mu_0$),

$$p\text{-value} = \Pr(Z > z)$$

- For lower-tailed (left-tailed) test (H_a: $\mu < \mu_0$),

$$p\text{-value} = \Pr(Z < z)$$

- When the *p*-value $> \alpha$, we fail to reject the null hypothesis.
- When the *p*-value $< \alpha$, we reject the null hypothesis.

P-value Approach

In a hypothesis test the *p* value is 0.043. This means that we can find statistical significance at:

1. both the 0.05 and 0.01 levels
2. the 0.05 but not at the 0.01 level
3. the 0.01 but not at the 0.05 level
4. neither the 0.05 or 0.01 levels

P-value Approach

You are reading a research article that states that there is no significant evidence that the median income in the two groups differs, at $\alpha = 0.05$. You are interested in this conclusion but prefer to use $\alpha = 0.01$.

1. You would also say there is no significant evidence that the medians differ.
2. You would say there is significant evidence that the medians differ.
3. You do not know whether there is significant evidence or not, until you know the *p*-value.

P-value Approach – Example 3.3

- **Example 3.3 – Filling Peanuts Jars.** Use the p-value approach to test

$$H_0: \mu = 8 \text{ versus } H_a: \mu \neq 8; \alpha = 0.01$$

- The p-value is:

$$p\text{–value} = 2 * \Pr(Z > |-2.2|) = 0.0278.$$

- The p-value = 0.0278 is greater than the significance level $\alpha = 0.01$. Therefore, we fail to reject the null hypothesis. The data dose not provide sufficient evidence that the filling process is out of control (or the machine is not working properly).

Example 3.4 – Aptitude Test

- **Example 3.4 – Aptitude Test.** An aptitude test has been used to test the ability of fourth graders to reason quantitatively. The test is constructed so that the scores are normally distributed with **a mean of 50** and **standard deviation of 10**. It is suspected that, with increasing exposure to computer-assisted learning, the test has become obsolete. That is, **it is suspected that the mean score is no longer 50**, **although σ remains the same**. This suspicion may be tested based on a sample of students who have been exposed to a certain amount of computer-assisted learning. A sample of **500 students** was collected and the **sample mean was 51.07**.

Example 3.4 – Aptitude Test

Example 3.4 – Aptitude Test. What information can we get?

- Sample size: $n = 500$, which is large enough so the central limit theorem can be applied to the sample mean.
- The sample mean is 51.07.
- A standard deviation of 10 is given – this should be used as the population standard deviation.
- We are interested in if the score is **different** from 50.
- Here we need the significance level, α, which is set as 0.05.

Uniformly Most Powerful Tests

- The test with higher power is more desirable so the test with the highest power should be used.
- For any specified alternative hypothesis, sample size, and level of significance, the test with the highest power is called a **"uniformly most powerful" (UMP)** test
- The construction of a power curve and the UMP test are not simple, and it becomes increasingly difficult for the applications in subsequent chapters.
- In our course, virtually all of the procedures we will be using provide uniformly most powerful tests, assuming that basic assumptions are met.
- Power calculations for more complex applications can be made easier through the use of computer programs. While there is no single program that calculates power for all hypothesis tests, some programs either have the option of calculating power for specific situations or can be adapted to do so.

Example 3.1 – As An Exercise

- **Example 3.1 (see book)** - The National Center for Education Statistics reports that the year 2007 reading scores for fourth graders had a mean of **220.99** and a standard deviation of **35.73** (from 191,000 fourth grades). You believe that your school district is doing a great job of teaching reading and want to show that mean scores in your district **would be higher than** this national mean. You randomly select **50 fourth graders** in your district, give the same exam, get **230.2** as the sample mean.

Example 3.1 – As An Exercise

Example 3.1 (see book) – Get to know some basics:

- A sample of 50 fourth grader is selected, so $n = 50$, which is large enough so the sample mean is has an approximate normal distribution even the score does not have a normal distribution.
- The sample mean is 230.2.
- A standard deviation of **35.73** is given – this should be used as the population standard deviation.
- We are interested in if the score in our district is **greater** than the national average.
- Here we also need the significance level, α, which is set as 0.05.

Point Estimator

- **Estimation** is a inferential procedure to use data from a sample to estimate the value of a parameter of the population.
- An ***estimator*** is a statistic used to estimate an unknown parameter of a population. We generally use $\hat{\theta}$ to represent an estimator of the arbitrary parameter.
- Examples of estimator:
  - Sample mean for population mean: $\hat{\mu}_1 = \bar{y}_1$.
  - Sample variance for population variance: $\hat{\sigma}_1^2 = s_1^2$.

Point Estimator - Example 3.3

Example 3.3 – Filling Peanuts Jars. A sample of 16 jars with peanuts was collected to see if the filling process is normal. So we are interested in two questions:

- What is the true current mean weight of peanuts in jars?
- Is the current mean weight different from the mean weight of 8 oz?

Point Estimator - Example 3.3

Example 3.3 -Filling Peanuts Jars. Suppose that the weight of peanuts follows a normal distribution with mean μ and variance σ^2.

- We may estimate the population mean using either sample mean or sample median.
- We can also estimate the population variance using either sample variance or some constant multiple of the interquartile range.
- The question here is: **which estimator should be used**?
- There are statistical theories about how to choose the best estimator. The estimators used in our book are generally "optimal".

Unbiased Estimator

- An ***unbiased estimator*** of an unknown parameter is one whose expected value is equal to the parameter of interest. In other words, if $E\left(\hat{\theta}\right) = \theta$, then $\hat{\theta}$ is unbiased estimator of θ.
- For example, $\bar{y}$ and s^2 from a random sample from $N(\mu, \sigma^2)$ are unbiased, since

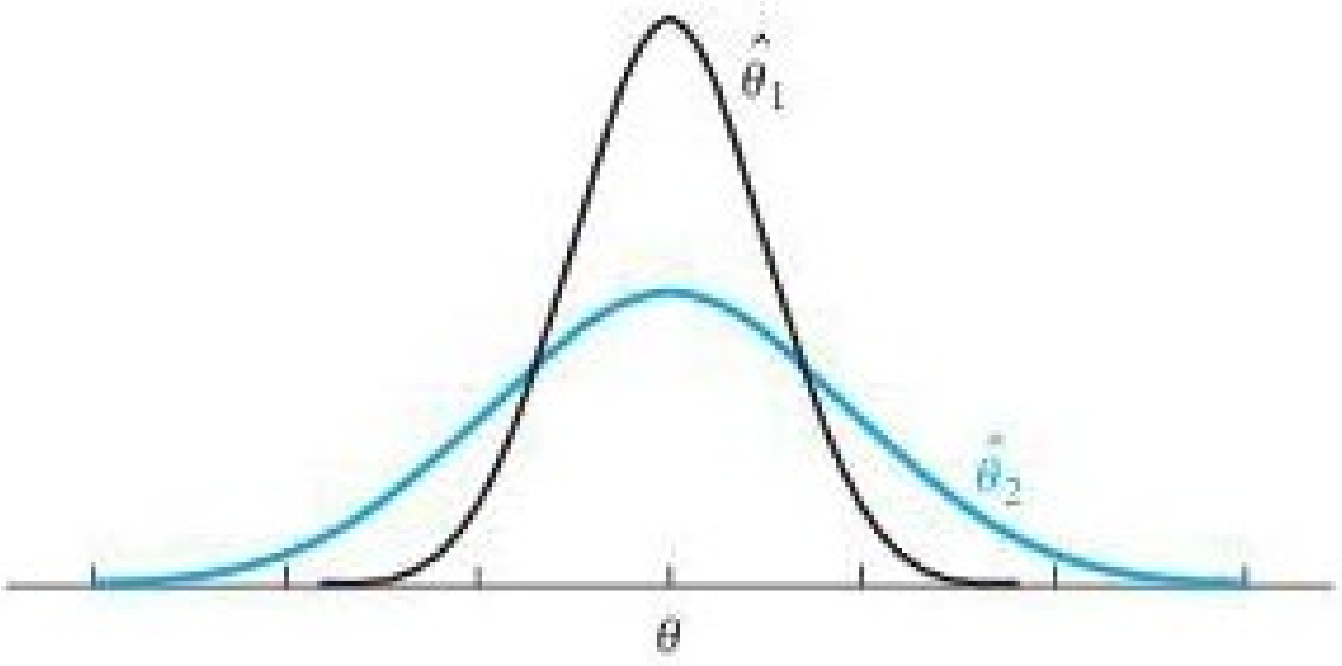
$$E(\bar{Y}) = \mu \text{ and } E(S^2) = \sigma^2.$$

- However, sample standard deviation is a biased estimator of population standard deviation, because

$$E(S) \neq \sigma$$

Precision of Unbiased Estimators

- An estimator is ***more precise*** if its sampling distribution has a smaller standard error (or variance).



Sample Mean from a Normal Population

For a random sample from $N(\mu, \sigma^2)$, the sample mean is the "best" estimator for the population mean μ, because:

1. The sample mean is an unbiased estimator of the population mean.
2. Among all unbiased estimator of the population mean, the sample mean has the smallest variance (Need some sophisticated statistical theory to prove this statement).

Point Estimators

- The sample mean is a *point estimator* since we use this specific value to estimate the parameter. If the random sample is from $N(\mu, \sigma^2)$, then $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$.
- We have $\Pr(\bar{Y} = \mu) = 0$.
- Thus, we know that this **point estimate has no chance of being correct**.
- Thus, we prefer to use interval estimators.

Interval Estimators – Variance Known

- We assume the random sample is from a normal distribution, or the central limit theorem holds, we have $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$.
- The two-sided interval estimate of the population mean is given by

$$\left(\bar{y} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{y} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

Interval Estimators – Variance Known

For interval estimate: $\left(\bar{y} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{y} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$

- Here, $1 - \alpha$ is the **coverage probability** (**confidence level**).
- We call it as $100(1 - \alpha)\%$ confidence interval for μ.
- We can use any α. But in general, we use $\alpha = 0.05$ or $\alpha = 0.01$.

Interval Estimators – Variance Known

For interval estimate: $\left(\bar{y} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{y} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$

- $z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ is the **margin of error** (**Definition 3.13)** which is defined as one-half the width of a confidence interval.
  - It increases with increased confidence coefficient (decreased α).
  - It increases with decreased sample size.
  - It increases with increased population variance.

Interval Estimators – Variance Known

- We have:

$$1-\alpha = \Pr(-z_{\alpha/2} < \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2})$$

$$= \Pr(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{Y}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

$$= \Pr(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu-\bar{Y} < z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

$$= \Pr(\bar{Y}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{Y}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

Interval Estimators – Interpretation

- The $100(1-\alpha)\%$ (or just use $1-\alpha$) confidence interval of μ is

$$\left(\bar{y} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{y} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

- Therefore, **we are $\mathbf{100(1-\alpha)\%}$ confident that the population mean $\boldsymbol{\mu}$ is inside of this interval**.
- We can state: **the probability that the interval generated by this way covers the true population mean $\boldsymbol{\mu}$ is $\mathbf{1-\alpha}$.**

Interval Estimators – Interpretation

- The $100(1-\alpha)\%$ confidence interval of μ is

$$\left(\bar{y} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{y} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

- You **can not** state: **the probability that the population mean μ is inside of this interval is $1-\alpha$ since μ is a constant, not a random variable.**

- For a given sample, you **can not** state that: **the probability that the interval covers the true μ is $1-\alpha$ since after the confidence interval is calculated, it either covers or does not cover the true μ.**

Interval Estimators – Examples/Exercises

- **Example 3.3 – Filling Peanuts Jars.** From the sample of 16 jars, we find the sample mean is $\bar{y} = 7.89$. The population standard deviation, σ, is known and $\sigma = 0.2$.
- Find 99% confidence interval of the population mean.
- **Solution:** The 99% confidence interval of population mean is:

$$\left(\bar{y} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = \left(7.89 \pm 2.5758 * \frac{0.2}{\sqrt{16}}\right) = (7.761, 8.019)$$

Interval Estimators – Examples/Exercises

- **Example 3.3 – Filling Peanuts Jars.** From the sample of 16 jars, we find the sample mean is $\bar{y} = 7.89$. The population standard deviation, σ, is known and $\sigma = 0.2$.
- Find 95% confidence interval of the population mean.
- **Solution:** The 95% confidence interval of population mean is:

$$\left(\bar{y} \pm z_{\frac{0.05}{2}} \frac{\sigma}{\sqrt{n}}\right) = \left(7.89 \pm 1.96 * \frac{0.2}{\sqrt{16}}\right) = (7.792, 7.988)$$

Interval Estimators – Interpretation

Interpretation of 95% confidence interval, $(7.792, 7.988)$. Which statement is correct:

1. $\Pr(7.792 < \mu < 7.988) = 0.95$.
2. 95% of all weights between 7.792 and 7.988.
3. We sampled 95% of all weights.
4. We know that $7.792 < \mu < 7.988$.
5. We are 95% confident that the true population mean is between 7.792 and 7.988.

Interval Estimators – Interpretation

The blood pressure of 100 patients from hospitable is collected. Based on the data, the 95% confidence interval of mean blood pressure of patients in a hospital is from 95.1 to 143.6, which of the following statements is correct:

1. The true mean blood pressure of patients in that hospital is between 95.1 and 143.6.
2. About 95% of patients in that hospital will have the blood pressure in the interval from 95.1 to 143.6.
3. The probability of true mean blood pressure of patients in that hospital in the interval from 95.1 to 143.6 is 0.95.
4. None of the above is correct.

Interval Estimators – Exercises

- **Example 3.4 – Aptitude Test.** A sample of 500 students was collected and the sample mean was 51.07. The population standard deviation is 10.
- Find 99% confidence interval of the population mean.
- **Solution:** In Class Exercise.

Hypothesis Testing and Confidence Interval

- There is a direct relationship between hypothesis testing and confidence interval estimation (for two-tailed hypothesis tests)
  - A hypothesis test for $H_0\colon\ \mu = \mu_0$ against $H_1\colon\ \mu \neq \mu_0$ will be rejected at a significance level of α if μ_0 is not in the (1- α) confidence interval for μ.
  - Any value of μ inside the (1- α) confidence interval will not be rejected by an α -level significance test.

Hypothesis Testing and Confidence Interval

- **Example 3.3** – 95% CI is (7.792, 7.988), we reject H_0: $\mu = 8$ at the 0.05 level of significance.
- **Example 3.4** – 99% CI is (49.92, 52.22), we do not reject H_0: $\mu = 50$ at the 0.01 level of significance.
- For one-tailed test, we need one-sided confidence interval.

One-Sided Interval Estimators When Variance Known

- The $100(1-\alpha)\%$ upper (one-sided) confidence interval of μ is given by

$$\left(-\infty, \bar{y} + z_{\alpha}\frac{\sigma}{\sqrt{n}}\right)$$

- The $100(1-\alpha)\%$ lower (one-sided) confidence interval of μ is given by

$$\left(\bar{y} - z_{\alpha}\frac{\sigma}{\sqrt{n}}, \infty\right)$$

One-Sided Interval Estimators - Example 3.1

Example 3.1 – Some basics:

- A sample of 50 fourth grader is selected with a sample mean of 230.2.
- The population standard deviation is **35.73**.
- We are interested in if the score in our district is **greater** than the national average, which is 220.99, at $\alpha = 0.05$.
- We need to calculate 95% lower confidence interval.
- **Solution:** $\left(\bar{y} - z_{\alpha}\frac{\sigma}{\sqrt{n}}, \infty\right) = \left(230.2 - 1.645 * \frac{35.73}{\sqrt{50}}, \infty\right) = (221.89, \infty)$

One-Sided Interval Estimators - Example 3.3

- **Example 3.3 – Filling Peanuts Jars.** From the sample of 16 jars, we find the sample mean is

$$\bar{y} = 7.89 \text{ oz.}$$

- We further assume that
  - The population standard deviation, σ, is known and $\sigma = 0.2$.
- Now we are more interested in knowing if $H_a: \mu < 8$.

One-Sided Interval Estimators - Example 3.3

- **Example 3.3 – Filling Peanuts Jars.**
- Five steps to perform the test. Assume $\alpha = 0.05$.
- Calculate the p-value.
- Calculate the appropriate one-sided confidence interval.

One-Sided Interval Estimators - Example 3.3

- **Step 5:** Then interpret the results in the language of the problem.
- **Example 3.3 – Filling Peanuts Jars (Slide 78)** , we state like this way:

  At the 5% significance level, the test statistic $z = -2.2$ falls in the rejection region. Therefore, we reject the null hypothesis. The data provides sufficient evidence that the weight of peanuts from the filling process is less than the weight of 8 oz.

One-Sided Interval Estimators - Example 3.3

- **Step 5:** Then interpret the results in the language of the problem.
- **Example 3.3 – Filling Peanuts Jars (Slide 78)** , we state like this way:

  At the 5% significance level, the p-value is 0.0139 and less than 0.05. Therefore, we reject the null hypothesis. The data provides sufficient evidence that the weight of peanuts from the filling process is less than the weight of 8 oz.

One-Sided Interval Estimators - Example 3.3

- **Step 5:** Then interpret the results in the language of the problem.
- **Example 3.3 – Filling Peanuts Jars (Slide 78)** , we state like this way:

  The 95% upper confidence interval of the mean weight of peanuts in jars is $(-\infty, 7.972)$, which does not contain the weight of 8 oz. Therefore, we reject the null hypothesis at 5% significance level. The data provides sufficient evidence that the weight of peanuts from the filling process is less than the weight of 8 oz.

Sample Size for Interval Estimation

- We will skip this part. Please refer to the textbook if you want to learn this part.

Probability of Type II Error

- There are many reasons for concerning the probability of the type II error, for example:
  - The probability of making a type II error may be so large that the test may not be useful.
  - Because of the trade-off between α and β, we may find that we may need to increase α in order to have a reasonable value for β.
  - Sometimes we have a choice of testing procedures where we may get different values of β for a given α.

Power and Sample Size

- **Definition 3.11 -** The **power** of a test is the probability of correctly rejecting the null hypothesis when it is false.
- Therefore, we have power = $1 - \beta$.
- More reasons for concerning the probability of the type II error (or the power):
  - Sometimes we have a choice of testing procedures where we may get different values of β for a given α. So we want a test with the largest power.
  - We may need to calculate sample size to ensure the study will have enough power.
- The calculation of power and sample size can be very difficult in some situations.

Power Calculation

Use the following steps to calculate the power for $\mu = \mu_1 (\mu_1 \neq \mu_0)$:

- **Step 1 –** Find the rejection region, R. We have $\Pr\left(\frac{\bar{Y}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \in R\right) = \alpha$.
- **Step 2 –** Rewrite the formula - $\Pr(\bar{Y} \in R^*) = \alpha$. Note that under the null hypothesis $\mu = \mu_0$, $\bar{Y} \sim N(\mu_0, \frac{\sigma^2}{n})$.
- **Step 3 –** The power is $\Pr(\bar{Y} \in R^*)$. Note that under the alternative hypothesis $\mu = \mu_1$, $\bar{Y} \sim N(\mu_1, \frac{\sigma^2}{n})$.

Power Calculation – Example

- **Example:** Assume that a random sample of size 25 is to be taken from a normal population with μ = 10 and σ = 2. The value of μ, however, is not known by the person taking the sample.
- **Problem 1 (Used as An Example)** Suppose the person wanted to test $H_0: \mu = 10.6$ against $H_a: \mu < 10.6$. Compute the power for $\alpha = 0.05$ and $\mu = 10$.
- **Problem 2 (Used as An Exercise):** Suppose the person wanted to test $H_0: \mu = 10.6$ against $H_a: \mu \neq 10.6$. Compute the power for $\alpha = 0.05$ and $\mu = 10$.

Sample Size for Hypothesis Testing

- Sample size determination must satisfy:
  - Required Confidence - level of significance (α) for hypothesis testing
  - Error – the probability of type II error (β) or power for hypothesis testing
  - The difference, called δ (delta), between the hypothesized value and the specified value ($\delta = \mu_a - \mu_0$)
  - The population variance or standard deviation.
- If the population standard deviation is unknown, we can use the following empirical rule:
  - Standard deviation is estimated by the range divided by 4.

Sample Size for Hypothesis Testing

- One-tailed Tests - $n = \sigma^2(z_\alpha + z_\beta)^2/\delta^2$
- Two-tailed Tests - $n = \sigma^2(z_{\alpha/2} + z_\beta)^2/\delta^2$
- **Example 3.6 -** $H_0: \mu = 35; H_1: \mu > 35;\ \alpha = 0.05; \beta = 0.10; \delta = 37 - 35 = 2$

Sample Size for Hypothesis Testing - Example

- **Example 3.6 -** In a study of the effect of a certain drug on the behavior of laboratory animals, a research psychologist needed to determine the appropriate sample size. The study was to estimate the time necessary for the animal to travel through a maze under the influence of this drug.
- We know: (1) $\alpha = 0.05$; (2) $\beta = 0.10$; (3) $\mu_0 = 35$; (4) $\mu_1 = 37$; (5) an anticipated range of times of 15 to 60 seconds.
- **Solution:** $\delta = 37 - 35 = 2$ and $\sigma = \frac{60-15}{4} = 11.25$. So sample size is

$$n = \frac{11.25^2 * (z_{0.05} + z_{0.10})^2}{2^2} = \frac{11.25^2 * (1.64485 + 1.28155)^2}{2^2} = 271$$

Assumptions about Normality

- Generally, are based on normal distribution
- Robust methods have been developed when normality is not satisfied
  - Generally they have wider confidence intervals and/or lower power
- Two principles to develop robust methods
  - Trimming, which consists of discarding a small pre-specified portion of the most extreme observations and making appropriate adjustments to the test statistics.
  - Nonparametric methods, which avoid dependence on the sampling distribution by making strictly probabilistic arguments (often referred to as distribution-free methods).

Statistical Significance versus Practical Significance

- We can have a statistically significant result that has no practical implications.

Example 3.7

- **Example 3.7.** In the January/February 1992 *International Contact Lens Clinic* publication, there is an article that presented the results of a clinical trial designed to determine the effect of defective disposable contact lenses on ocular integrity (Efron and Veys, 1992).
- This is double blind experiment: 29 samples who wore a defective lens in one eye and a nondefective one in the other. Neither the research officer nor the subject knew which eye wore the defective lens.

Example 3.7 – Statistical Significance

- The study indicated that a significantly greater ocular response was observed in eyes wearing defective lenses in the form of corneal epithelial microcysts (among other results) - p value is 0.04. With a level of significance of 0.05, the conclusion would be that the defective lenses resulted in more microcysts being measured.

Example 3.7 – No Practical Significance

- The study reported a mean number of microcysts for the eyes wearing defective lenses as 3.3 and the mean for eyes wearing the nondefective lenses as 1.6.
- In an invited commentary, Dr. X points out that the observation of fewer than 50 microcysts per eye requires no clinical action other than regular patient follow up.
- We may use the following hypothesis test: $H_a: |d| \geq 50$.

Chapter Summary

- Steps to conduct hypothesis tests:
  - State the hypotheses and the significance level
  - Collect data and compute test statistics
  - Make a decision to confirm or deny hypothesis.
- Steps to calculate confidence interval:
  - Identify the parameter and the confidence level
  - Collect data and compute the statistics for the confidence interval
  - Interpret the interval in the context of the situation.
- The distinction between null and alternative hypotheses can be difficult in some situations.