Introduction to Data Science

Unit 4 – Hypothesis and Inference

Preet Kanwal
Assistant Professor
Department of CSE
PESU Bangalore

Statistical Hypothesis

A statistical hypothesis is:

- A method of statistical inference.
- **An assumption about a population parameter.**

This assumption may or may not be true.

For example:

The statement “$\mu > 11$” is a hypothesis about the population mean μ.

We must perform a **Hypothesis Test** in order to determine how certain can we be that the Hypothesis is true.

Statistical Hypothesis

- The best way to determine whether a statistical hypothesis is true would be to examine the entire population.
- Since that is often impractical, researchers typically examine a random sample from the population.
- If sample data are not consistent with the statistical hypothesis, the hypothesis is rejected.

Hypothesis Test

Hypothesis tests are **statistical procedures** that evaluate two mutually exclusive statements about a population.

These two statements are called the null hypothesis and the alternative hypothesis. [**two types of statistical hypotheses**]

They are always statements about population attributes, such as the value of a parameter, the difference between corresponding parameters of multiple populations, or the type of distribution that best describes the population.

A hypothesis test **uses sample data to determine which statement is best supported by the data.**

Types of Statistical Hypotheses

There are two types of statistical hypotheses:

- **Null hypothesis:** The null hypothesis, denoted by H_0, is usually the hypothesis that sample observations result purely from chance(sampling error – random variation). **[Accepted Fact or No difference or No effect – It always says the sample is misleading]**

- **Alternative hypothesis.** The alternative hypothesis, denoted by H_1 or H_a, is the hypothesis that sample observations are influenced by some non-random cause **(investigator's belief).**

Guidelines to choose H_0

Null Hypotheis is chosen as something:

- Which the scientists or statisticians want to disprove or nullify.
- Or is an accepted fact.
- Lack of a difference : It just says with and without modification the results are the same(**no difference**). There's **no effect** of making a change.
- It says whatever the sample is saying or depicting is misleading.

Example : Identify H_0 and H_a

A coin was flipped 50 times, resulting in 40 Heads and 10 Tails.

Two interpretations of the sample:

1) Null Hypothesis : says that the effect indicated by the sample is due to Sampling variation and tha the coin is fair and unbaised. (**coin is given a benefit of doubt**)Hence, null hypothesis might be that half the flips would result in Heads and half, in Tails.

$$H_0 : P = 0.5$$

2) Alternative Hypothesis : says that the effect indicated by the sample is real and that the coin is not fair and biased. The alternative hypothesis might be that the number of Heads and Tails would be very different.

$$H_a : P \neq 0.5$$

Example 1 : Identify H_0 and H_a

A hypothesis is put forward that children who take vitamin C are less likely to become ill during flu season than those who do not. A hypothesis test is conducted where a sample group of children is given vitamin C for three months while another group is not.

Example 1 : Solution

1) Null Hypothesis : Taking vitamin C does not produce any effect.

2) Alternative Hypothesis : Taking Vitamic C produces a difference.

As it turns out, the alternative hypothesis is confirmed.

Example 2 : Identify H_0 and H_a

- An individual claims that, contrary to what is commonly accepted, he is telekinetic(the production of motion in objects without contact or other physical means.) and can read minds.

Example 2 : Solution

1) Null Hypothesis : He cannot read minds and is not telekinetic.

2) Alternative Hypothesis : He can read minds and is telekinetic.

Example 3 :Identify H_0 and H_a

- Accused of a grave crime, a person is placed under arrest and a trial is conducted.

Example 3 : Solution

1) Null Hypothesis : He is innocent, is challenged in court. [Defendant is given **benefit of doubt** - To give someone the benefit of the doubt means you assume that he or she is telling the truth until proven otherwise.].

2) Alternative Hypothesis : He is not innocent and has commited the crime.

Problem 1(a)

In each of the following situations, state the most appropriate null hypothesis regarding the population mean μ.

1) A new type of battery will be installed in heart pacemakers if it can be shown to have a mean lifetime is greater than eight years.

Problem 1(a) : Solution

1) A new type of battery will be installed in heart pacemakers if it can be shown to have a mean lifetime is greater than eight years.

$$H_0 : \boldsymbol{\mu} \leq 8$$

$$H_1 : \boldsymbol{\mu} > 8$$

Problem 1(b)

In each of the following situations, state the most appropriate null hypothesis regarding the population mean μ.

2) A new material for manufacturing tires will be used if it can be shown that the mean lifetime of the tires will be more than 60,000 miles.

Problem 1(b) : Solution

2) A new material for manufacturing tires will be used if it can be shown that the mean lifetime of the tires will be more than 60,000 miles.

$$H_0 : \boldsymbol{\mu} \leq 60{,}000$$

$$H_1 : \boldsymbol{\mu} > 60{,}000$$

Problem 1(c)

In each of the following situations, state the most appropriate null hypothesis regarding the population mean μ.

3) A quailty control inspector will recalibrate a flowmeter if the mean flow rate differs from 10 mL/s.

Problem 1(c) : Solution

3) A quailty control inspector will recalibrate a flowmeter if the mean flow rate differs from 10 mL/s.

$$H_0 : \boldsymbol{\mu} = 10$$

$$H_1 : \boldsymbol{\mu} \neq 10$$

Real world applications of Hypothesis Testing

1) Testing whether more men than women suffer from nightmares.
2) Evaluating the effect of the full moon on behavior.
3) Checking whether bumper stickers reflect car owner behavior.
4) Testing the claims of handwriting analysts.

Hypothesis Test

- In performing Hypothesis Test, we essentially put the null hypothesis on trial.
- We begin by assuming H_0 is true.
- The random sample provides the evidence.
- The Hypothesis test involves measuring the strength of disagreement between the sample and H_0 .
- Two methods :

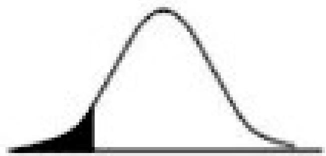
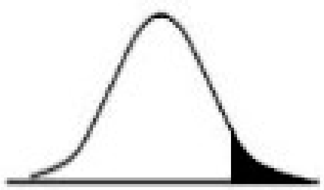
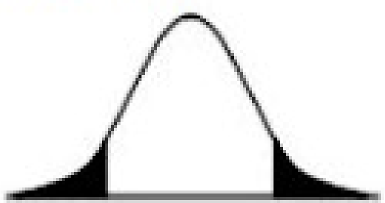
1) Traditional Method : Rejection region approach

2) P-value approach (Used in book – Commonly used)

Note

In any hypothesis test, we are calculating conditional probabilities based on the assumption that the null hypothesis is true.

Type of Hypothesis Test

	Type of Hypothesis Test	Alternate Hypotheis	Null Hypothesis
One – Tailed Tests(Values of the sample which cause rejection of H_0 fall only in one tail of the sample)	Left – tailed Test 	$<$	$>=$
	Right – tailed Test 	$>$	$<=$
	Two tailed test 	$\neq$	$=$

test-statistic

- A test statistic is a standardized value that is calculated from sample data during a hypothesis test.
- Used to determine whether to reject the null hypothesis.
- The test statistic **used to assess the strength of evidence against H_0** .
- Its observed value **changes randomly** from one random sample to a different sample.

Different Hypothesis Tests

- Different hypothesis tests use different test statistics based on the probability model assumed in the null hypothesis. Common tests and their test statistics include:

Hypothesis test	Test statistic
Z-test	Z-statistic
t-tests	t-statistic
ANOVA	F-statistic
Chi-square tests	Chi-square statistic

Steps in Performing a Hypothesis Test

1) Define H_0 and H_1.

2) Assume H_0 is true

3) Compute test statistic.

4) Compute P-value based on chosen α (significance level).

5) State a conclusion.

Unit Contents

Z – tests for:

1) Population mean μ of Large Samples. **(Section 6.1, 6.2)**
2) Difference between Two Population Means of Large Samples. **(Section 6.5)**
3) Population proportion of Large Samples. **(Section 6.3)**
4) Difference between Two Population proportions of Large Samples. **(Section 6.6)**
5) Paired Data. **(Section 6.8)**

t – tests for:

1) Paired Data. **(Section 6.8)**

Fixed level testing (Section 6.12)

Power of a Test (Section 6.13)

Distribution – Free tests (Section 6.9)

Chi – Squared test (Section 6.10)

A test that uses z-score as a test statistic is called z-test.

Z – tests for Population mean μ of Large Samples (Section 6.1)

Prof. Preet Kanwal

Example

An automobile engine emits oxides of Nitrogen(NO_x) at a mean rate of 100 mg/s.

- A modification to the engine design has been proposed that may reduce NO_x emissions.
- The new design will be put in production if it can be demostrated that the its mean emission rate is less than 100mg/s.
- A sample of 50 modified engines are built and tested. The sample mean NO_x emission is 92 mg/s and the sample SD is 21 mg/s.
- Define H_0 and H_1.

Null Hypothesis

Since the planned action(Putting modified engine design in production) is based on the criteria that the emission rates shold be lower than 100.

Null Hypothesis can be set as:

- Something that the manufacturers would like to disprove.
- Something that says the sample is misleading and might have come from the populaiton that has mean greater than 100. Hence,

$$\mathbf{H_0 : \mu \geq 100}$$

Alternate Hypothesis

- Opposite of Null Hypotheis.
- Says the effect indicated by the sample is real and accurately represents the whole population.

$$\mathbf{H_a : \mu < 100}$$

The question is: Is it plausible that this sample, with mean 92, could have come from a population whose mean is 100 or more?

We use Hypothesis tests to adress such questions

Null Distribution

It is the distribution of the statistic (X_bar) under the assumption that H_0 is true.

$$X\_bar \sim N(100,\ s^2 / n)$$

$$X\_bar \sim N(100,\ 2.97^2)$$

P - value

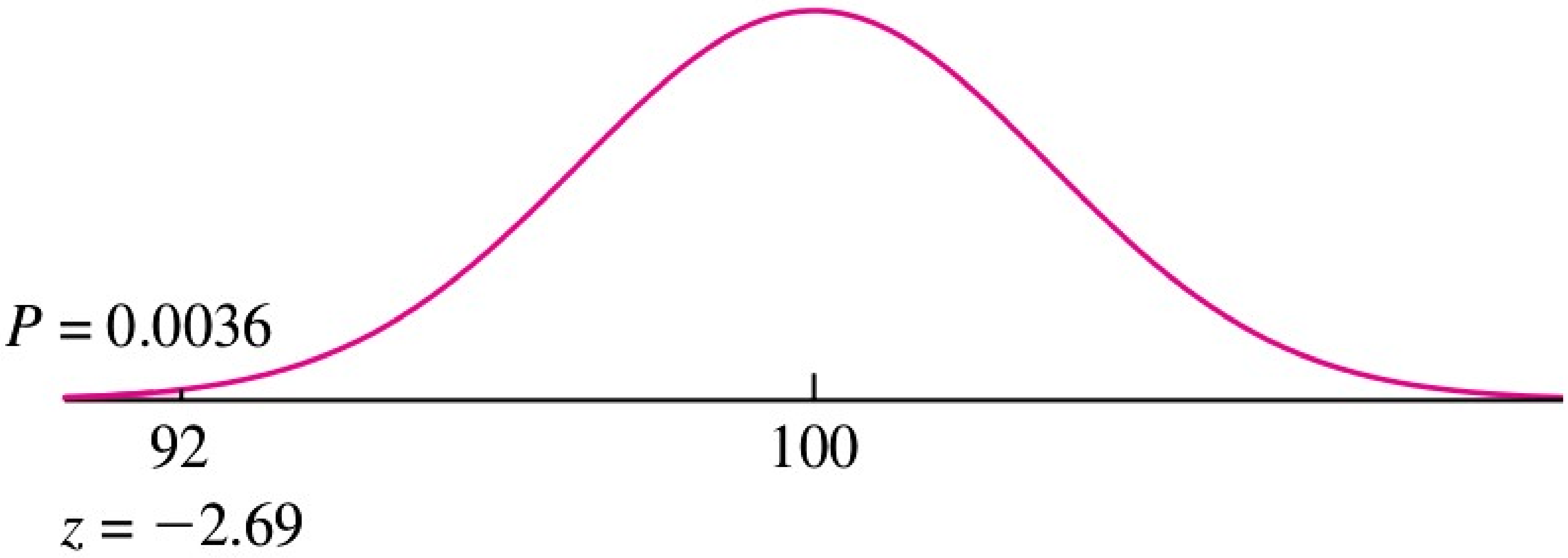
- P-value is the probability that number drawn from N(100, 2.97^2) distribution is less than or equal to 92.
- **Provides a quantitative measure of the plausibility of H_0.**
- This probability is determined by computing the z-score:

$$z = \frac{92 - 100}{2.97} = -2.69$$

- P(Z <= -2.69) = 0.0036

=> P – value for the test is 0.0036.

Result



P-value

- The p-value is defined as the probability of observing an extreme value of a statistic X_bar, since the value of X_bar could come out differently if the experiment were repeated, under the assumption that the null hypothesis is true.
- **The P-value Is Not the Probability That H_0 Is True : The truth or falsehood of H_0 cannot be changed by repeating the experiment.**
- It measures the strength of the evidence against the null hyposthesis.
- How likely is the effect observed in your sample data if the null hypothesis is true?
- It does not measure support for the alternative hypothesis.
- The p-value is the probability that the data could deviate from the null hypothesis as much as they did or more.

Low P-value

- Low P-values: **your data are unlikely with a true null.**

- The smaller value of p-value, tells the investigator that the **hypothesis under consideration may not adequately explain the observation.**

- A low P value suggests that your **sample provides enough evidence that you can reject the null hypothesis** for the entire population.

- When the P-value is small, then we can be confident that the t**rue value is really different.**

How low the P-value should be in order to reject H_0

- The smaller the P-value, the more certain we can be that H_0 is false.
- The larger the P-value, the more plausible H_0 becomes, but we can never be certain that H_0 is true.
- A rule of thumb suggests to reject H_0 whenever $P \leq 0.05$. While this rule is convenient, it has no scientific basis.

Problem 2

Fill in the blank:

In a test of $H_0 : \mu \geq 10$ versus $H_1 : \mu < 10$, the sample mean was X = 8 and the P-value was 0.04. This means that if $\mu = 10$, and the experiment were repeated 100 times, we would expect to obtain a value of X of 8 or less approximately _______ times.

i. 8

ii. 0.8

iii. 4

iv. 0.04

v. 80

Problem 2 : Solution

(iii) 4.

- The P-value is the probability, under the assumption that H_0 is true, of observing a result as extreme as or more extreme than that actually observed.
- Since P = 0.04, we expect to obtain a value of X less than or equal to 8 approximately 4 times in 100.

Significance Level

- The significance level, also denoted as alpha or α, is used as a **probability cutoff for making decisions about the null hypothesis.**
- Typical(Reasonable) values : 0.01, 0.05, 0.10.
- When a result has a small P-value, we say that it is "statistically significant."

Note: **There is no scientifically valid dividing line between plausibility and implausibility, so it is impossible to specify a "correct" significance level below which we should reject H_0**

Statistical Significance

Let α be any value between 0 and 1. Then, if $P \leq \alpha$,

- The result of the test is said to be statistically significant at the $100\alpha\%$ level.
- The null hypothesis is rejected at the $100\alpha\%$ level.
- When reporting the result of a hypothesis test, report the P-value, rather than just comparing it to 5% or 1%.

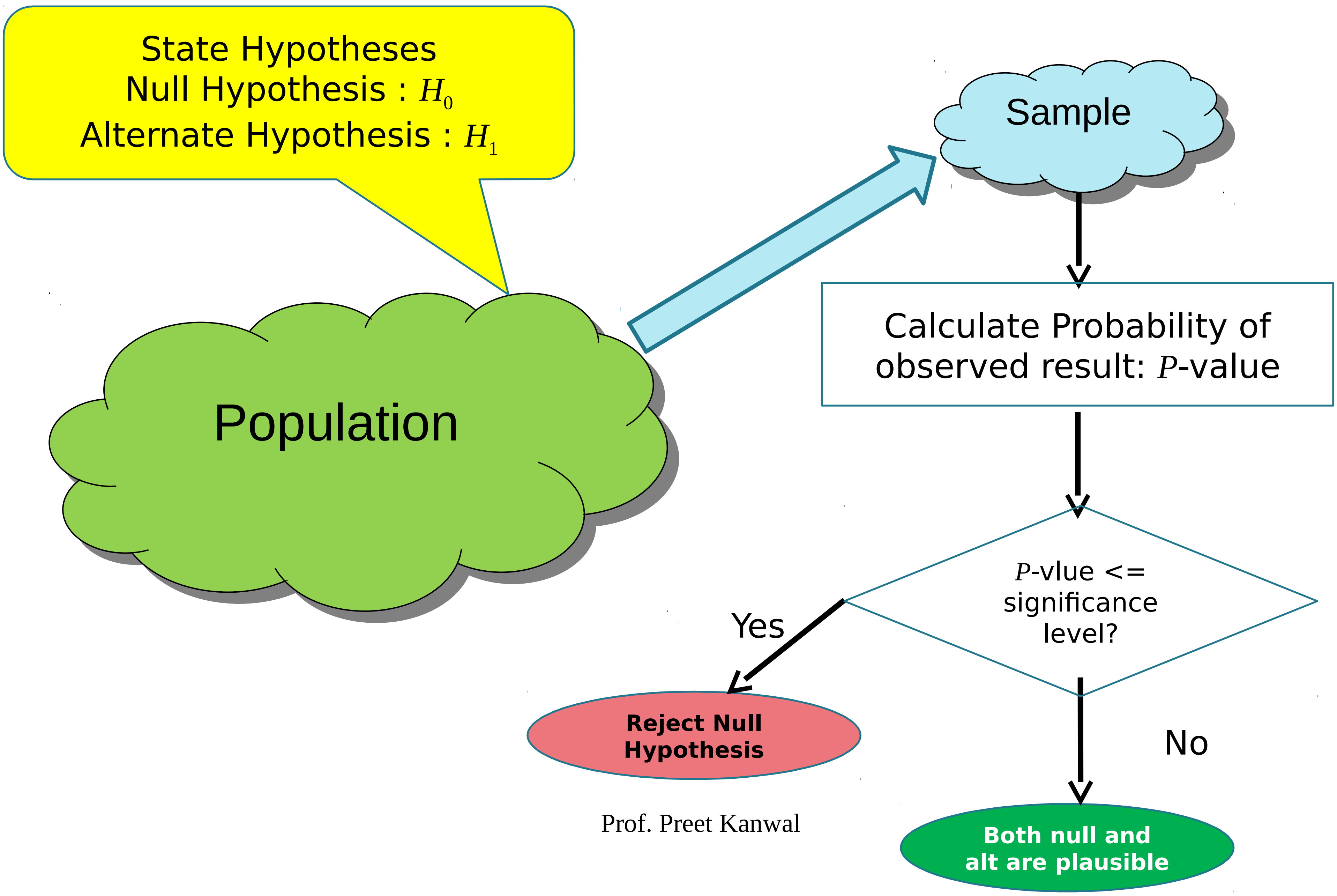
Conclusion in Hypothesis Test

1) Either is H_0 false and H_1 is true,

2) H_0 is plausible and therefore that H_1 is plausible as well.

Note: One can never conclude H_0 true and H_1 is false.

Fundamental Concept



Problem 3

For which P-value is the null Hypothesis more plausible?

a) $P = 0.05$

b) $P = 0.5$

Problem 3 : Solution

b) P = 0.5. The larger the P-value, the more plausible the null hypothesis.

Problem 4

True or false:

a. If we reject H_0 , then we conclude that H_0 is false.

b. If we do not reject H_0 , then we conclude that H_0 is true.

c. If we reject H_0 , then we conclude that H_1 is true.

d. If we do not reject H_0 , then we conclude that H_1 is false.

Problem 4 : Solution

(a) True.

(b) False. We conclude that H 0 is plausible.

(c) True.

(d) False. We conclude that H 0 is plausible, and therefore that H 1 is plausible as well.

Problem 5

If P = 0.01, which is the best conclusion?

I H_0 is definitely false.

II H_0 is definitely true.

III There is a 1% probability that H_0 is true.

IV H_0 might be true, but it's unlikely.

V H_0 might be false, but it's unlikely.

VI H_0 is plausible.

Problem 5 : Solution

(iv). A P-value of 0.01 means that if H_0 is true, then the observed value of the test statistic was in the most extreme 1% of its distribution. This is unlikely, but not impossible.

Problem 6

If P = 0.50, which is the best conclusion?

i. H_0 is definitely false.

ii. H_0 is definitely true.

Iii. There is a 50% probability that H_0 is true.

iv. H_0 is plausible, and H_1 is false.

v. Both H_0 and H_1 are plausible.

Problem 6 : Solution

- Option (v) A P-value of 0.50 means that if H_0 is true, then the observed value of the test statistic was in the most extreme 50% of its distribution. This is not at all unusual, so H_0 is plausible. We can never conclude that H_1 is false.

- So therefore we conclude that H_1 is plausible as well.

Problem 7

True or false: If P = 0.02, then

a. The result is statistically significant at the 5% level.

b. The result is statistically significant at the 1% level.

c. The null hypothesis is rejected at the 5% level.

d. The null hypothesis is rejected at the 1% level.

Problem 7 : Solution

(a) True. The result is statistically significant at any level greater than or equal to 2%.

(b) False. $P > 0.01$, so the result is not statistically significant at the 1% level.

(c) True. The null hypothesis is rejected at any level greater than or equal to 2%.

(d) False. $P > 0.01$, so the null hypothesis is not rejected at the 1% level.

Problem 8

George performed a hypothesis test.

- Luis checked George’s work by redoing the calculations.
- Both George and Luis agree that the result was statistically significant the 5% level, but they got different P-values.
- George got a P-value of 0.20 and Luis got a P-value of 0.02.

a) Is it possible that George’s work is correct? Explain.

b) Is it possible that Luis’s work is correct? Explain.

Problem 8 : Solution

(a) No. If the P-value is 0.20, then the result is not statistically significant at the 5% level.

(b) Yes. If the P-value is 0.02, then the result is statistically significant at the 5% level.

Problem 9

Let μ be the radiation level to which a radiation worker is exposed during the course of a year.

- The Environmental Protection Agency has set the maximum safe level of exposure at 5 rem per year.
- If a hypothesis test is to be performed to determine whether a work-place is safe, **which is the most appropriate null hypothesis:**

a) $H_0 : \mu \leq 5$

b) $H_0 : \mu \geq 5$, or

c) $H_0 : \mu = 5$?

Problem 9 : Solution

Option (b)

- Here, we wish to determine whether the work place is safe.

Hence,

H_1 = Place is safe, $H_1 : \mu < 5$

H_0 = Place is not safe, $H_0 : \mu \geq 5$ **(best choice)**

Problem 10

A machine that fills cereal boxes is supposed to be calibrated so that the mean fill weight is 12 oz. Let μ denote the true mean fill weight. Assume that in a test of the hypotheses,

$H_0 : \mu = 12$ versus $H_1 : \mu \neq 12$, the P-value is 0.30.

a. Should H_0 be rejected on the basis of this test?Explain.

b. Can you conclude that the machine is calibrated to provide a mean fill weight of 12 oz? Explain.

Problem 10 : Solution

(a) No. P = 0.30 is not small. Both the null and alternate hypotheses are therefore plausible.

(b) No, we cannot conclude that the null hypothesis is true. The alternate hypothesis is also plausible.

Problem 11

A method of applying zinc plating to steel is supposed to produce a coating whose mean thickness is no greater than 7 microns. A quality inspector measures the thickness of 36 coated specimens and tests;

$$H_0 : \mu \leq 7 \text{ versus } H_1 : \mu > 7$$

She obtains a P-value of 0.40. Since P > 0.05, she concludes that the mean thickness is within the specification. Is this conclusion correct? Explain.

Problem 11 : Solution

No, she cannot conclude that the null hypothesis is true, only that it is plausible.

Problem 12

Recently many companies have been experimenting with telecommuting, allowing employees to work at home on their computers. Among other things, telecommuting is supposed to reduce the number of sick days taken. Suppose that at one firm, it is known that over the past few years employees have taken a mean of 5.4 sick days. This year, the firm introduces telecommuting. Management chooses a simple random sample of 80 employees to follow in detail, and, at the end of the year, these employees average 4.5 sick days with a standard deviation of 2.7 days. Let μ represent the mean number of sick days for all employees of the firm.

a. Find the P-value for testing $H_0 : \mu \geq 5.4$ versus $H_1 : \mu < 5.4$.

b. Do you believe it is plausible that the mean number of sick days is at least 5.4, or are you convinced that it is less than 5.4? Explain your reasoning.

Problem 12(a) : Solution

X = 4.5, s = 2.7, n = 80. The null and alternate hypotheses are H 0 : μ ≥ 5.4 versus H 1 : μ < 5.4.

- z = (4.5 − 5.4)/(2.7/ sqrt(80)) = −2.98. Since the alternate hypothesis is of the form $\mu < \mu_0$, the P-value is the area to the left of z = −2.98.
- Thus P = 0.0014.

Problem 12(b) : Solution

- If the mean number of sick days were 5.4, the probability of observing a sample mean less than or equal to the observed value of 4.5 would be 0.0014.

- Since 0.0014 is a small probability, we are convinced that the mean number of sick days is less than 5.4.

Problem 13

The pH of an acid solution used to etch aluminum varies somewhat from batch to batch. In a sample of 50 batches the mean pH was 2.6, with a standard deviation of 0.3. Let μ represent the mean pH for batches of this solution.

a. Find the P-value for testing $H_0 : \mu \leq 2.5$ versus $H_1 : \mu > 2.5$.

b. Either the mean pH is greater than 2.5 mm, or the sample is in the most extreme _______% of its distribution.

Problem 13(a) : Solution

X = 2.6, s = 0.3, n = 50. The null and alternate hypotheses are H 0 : $\mu \leq 2.5$ versus H 1 : $\mu > 2.5$.

- $z = (2.6 - 2.5)/(0.3/ \text{sqrt}(50)) = 2.36$.
- Since the alternate hypothesis is of the form $\mu > \mu_0$, the P-value is the area to the right of $z = 2.36$.
- Thus $P = 0.0091$.

Problem 13(b) : Solution

(b) The P-value is 0.0091, so if H_0 is true then the sample is in the most extreme 0.91% of its distribution.

Problem 14

Lasers can provide highly accurate measurements of small movements. To determine the accuracy of such a laser, it was used to take 100 measurements of a known quantity. The sample mean error was 25 μm with a standard deviation of 60 μm. The laser is properly calibrated if the mean error is $\mu = 0$. A test is made of $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$.

a. Find the P-value.

b. Do you believe it is plausible that the laser is properly calibrated, or are you convinced that it is out of calibration? Explain your reasoning.

Problem 14(a) : Solution

$X = 25$, $s = 60$, $n = 100$. The null and alternate hypotheses are $H\ 0 : \mu = 0$ versus $H\ 1 : \mu = 0$.

- $z = (25 - 0)/(60/\ \mathrm{sqrt}(100)) = 4.17$.
- Since the alternate hypothesis is of the form $\mu = \mu\ 0$, the P-value is is the sum of the areas to the right of

  $z = 4.17$ and to the left of $z = -4.17$.
- Thus $P \approx 0$.

Problem 14(b) : Solution

- Since P-value is approx 0, we can reject H_0.

- That means If H_0 was true that is the mean error were 0, the probability of observing a sample mean as far from 0 as the value of 25 that was actually observed would be nearly 0.

- Therefore we are convinced that the mean error is not equal to 0. i.e. H_1 is true.

Fixed – Level Testing

Fixed – Level Testing

- There is no scientifically valid dividing line between plausibility and implausibility, so it is impossible to specify a "correct" significance level below which we should reject H_0.
- When possible, it is best simply to report the P-value, and not to make a firm decision whether or not to reject.
- **Sometimes, however, a decision has to be made.**
- If a decision is going to be made on the basis of a hypothesis test, there is no choice but to pick a cutoff point for the P-value.
- When this is done, the test is referred to as a **Fixed – Level Test.**

Summary

To conduct a fixed-level test:

- Choose a number α, where $0 < \alpha < 1$. This is called the significance level, or the level, of the test.
- Compute the P-value in the usual way.
- If $P \leq \alpha$, reject H_0. If $P > \alpha$, do not reject H_0.

Rejection Region approach for Hypothesis Test

Critical Point & Rejection Region

- In a fixed-level test, a critical point is a **value of the test statistic** that produces a P-value exactly equal to α.

- The region on the side of the critical point that leads to rejection is called the rejection region.

- The critical point itself is also in the rejection region.

Problem 1

A new concrete mix is being evaluated. The plan is to sample 100 concrete blocks made with the new mix, compute the sample mean compressive strength X , and then test,

$$H_0 : \mu \leq 1350 \text{ versus } H_1 : \mu > 1350$$

where the units are MPa.

- It is assumed from previous tests of this sort that the population standard deviation σ will be close to 70 MPa.
- **Find the critical point and the rejection region if the test will be conducted at a significance level of 5%.**

Problem 1 : Solution

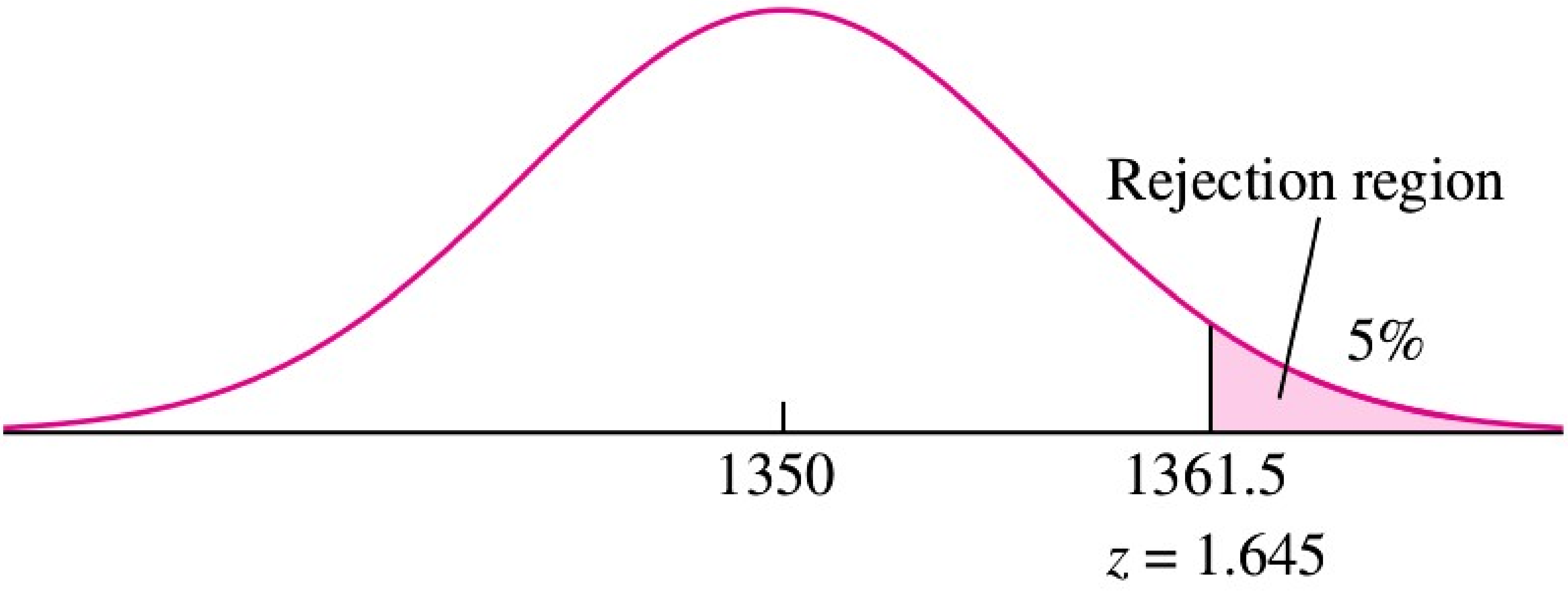
- $\alpha = 0.05$
- Since **$H_1 : \mu > 1350$**, its a right-tailed test.
- $z = 1.645$ **(Critical value)**
- Null distribution : X_bar ~ N(1350, 70^2/100)

  => X_bar= z * s/sqrt(n) + 1350

  => X_bar= 1.645 * 70/sqrt(100) + 1350

  => **X_bar = 1361.5 (Critical value can also be expressed this way)**
- The rejection region consists of all values of X_bar greater than or equal to 1361.5.

Problem 1 : Solution



Problem 2

- A process for a certain type of ore is designed to reduce the concentration of impurities to less than 2%.
- It is known that the standard deviation of impurities for processed ore is 0.6%.
- Let μ represent the mean impurity level, in percent, for ore specimens treated by this process.
- The impurity of 80 ore specimens is measured, and a test of the hypothesis **$H_0 : \mu \geq 2$ versus $H_1 : \mu < 2$** will be performed.

a. If the test is made at the 5% level, what is the rejection region?

b. If the sample mean impurity level is 1.85, will H_0 be rejected at the 10% level?

c. If the sample mean pH is 1.85, will H_0 be rejected at the 1% level?

d. If the value X_bar = 1.9 is a critical point, what is the level of the test?

Problem 2(a) : Solution

- Null distribution of X_bar:

  X_bar ~ N(2, 0.6^2/80)
- $\alpha = 0.05$
- Since **H1 : μ< 2**, its a left-tailed test.
- Z = -1.645 (Critical value)

  => X_bar = z * s/sqrt(n) + 2

  => X_bar= -1.645 * 0.6/sqrt(80) + 2

  => X_bar = 1.890
- Hence, H_0 will be rejected if X ≤ 1.890. The rejection region consists of all values of X less than or equal to 1.890.

Problem 2(b) : Solution without doing any calculation

- Critical value at 5% level of X_bar = 1.890

b. If the sample mean impurity level is 1.85, will H_0 be rejected at the 10% level?

- Since 1.85 < 1.890, **H_0** will be rejected at 5% level and hence will obviously be rejected at 10% level

Problem 2(b) : Solution using Rejection region approach

b. If the sample mean impurity level is 1.85, will H_0 be rejected at the 10% level?

$H_0 : \mu \geq 2$ versus $H_1 : \mu < 2$

- Null distribution of X_bar:

  X_bar ~ N(2, 0.6^2/80)

- $\alpha = 0.10$
- Z = −1.28 (Critical value)

  => X_bar = z * s/sqrt(n) + 2

  => X_bar= -1.28 * 0.6/sqrt(80) + 2

  => X_bar = 1.9141

- Since 1.85 < 1.9141 => **H_0 will be rejected at the 10% level**

Problem 2(b) : Solution using P-value approach

b. If the sample mean impurity level is 1.85, will H_0 be rejected at the 10% level?

$H_0 : \mu \geq 2$ versus $H_1 : \mu < 2$

- Null distribution of X_bar:

  $$\bar{X} \sim N(2, 0.6^2/80)$$

- $\alpha = 0.10$
- Finding z-score for 1.85
- $Z = (1.85 - 2)/ (0.6/\text{sqrt}(80)) = -2.24$
- => $P = P(Z < -2.24) = 0.0125$
- => $P < \alpha$ => **H_0 will be rejected at the 10% level**

Problem 2(c) : Solution using Rejection region approach

c. If the sample mean pH is 1.85, will H_0 be rejected at the 1% level?

$H_0 : \mu \geq 2$ versus $H_1 : \mu < 2$

- Null distribution of X_bar:

$$X\_bar \sim N(2, 0.6^2/80)$$

- A = 0.01
- Z = -2.33 (Critical value)

=> X_bar = z * s/sqrt(n) + 2

=> X_bar= -2.33 * 0.6/sqrt(80) + 2

=> X_bar = 1.8437

- Since 1.85 > 1.8437 => **H_0 will not be rejected at the 1% level**

Problem 2(c) : Solution using P-value approach

c) If the sample mean pH is 1.85, will H_0 be rejected at the 1% level?

$H_0 : \mu \geq 2$ versus $H_1 : \mu < 2$

- Null distribution of X_bar:

  $$X\_bar \sim N(2, 0.6^2/80)$$

- $\alpha = 0.01$
- Finding z-score for 1.85
- $Z = (1.85 – 2)/ (0.6/sqrt(80)) = -2.24$
- => P= P(Z < -2.24) = 0.0125
- => P > α => **H_0 will not be rejected at the 1% level**

Problem 2(d)

d. If the value 1.9 is a critical point, what is the level of the test?

$$H_0 : \mu \geq 2 \text{ versus } H_1 : \mu < 2$$

- Null distribution of X_bar:

$$X\_bar \sim N(2, 0.6^2/80)$$

- Finding z – score for 1.9
- Z = (1.9 – 2)/ (0.6/sqrt(80)) = -1.49
- => P(Z < -1.49) = 0.0681
- => level = α = 0.0681

Problem 3

A new braking system is being evaluated for a certain type of car.

- The braking system will be installed if it can be conclusively demonstrated that the stopping distance under certain controlled conditions at a speed of 30 mi/h is less than 90 ft.
- It is known that under these conditions the standard deviation of stopping distance is approximately 5 ft.
- A sample of 150 stops will be made from a speed of 30 mi/h.
- Let μ represent the mean stopping distance for the new braking system.

a) State the appropriate null and alternate hypotheses.

Problem 3

$H_0 : \mu \geq 90$ versus $H_1 : \mu < 90$

b) Someone suggests rejecting H_0 if X_bar ≥ 89.4 ft. Is this an appropriate rejection region, or is something wrong? If this is an appropriate rejection region, find the level of the test. Otherwise explain what is wrong.

c). Someone suggests rejecting H_0 if X_bar ≤ 89.4 ft. Is this an appropriate rejection region, or is something wrong? If this is an appropriate rejection region, find the level of the test. Otherwise explain what is wrong.

Problem 3(b) : Solution

This is not an appropriate rejection region.

Since **$H_1 : \mu < 90$ => Left-tailed test**

- The rejection region should consist of values for X that will make the P-value of the test less than or equal to a chosen threshold level.
- Therefore the rejection region must be of the form

$$X \leq x_0$$

- This rejection region is of the form $X \geq \mathbf{x_0}$, and so it consists of values for which the P-value will be greater than some level.

Problem 3(c) : Solution

This is an appropriate rejection region.

Since **$H_1 : \mu < 90$ => Left-tailed test**

- Under H 0 , the z-score of 89.4 is

  $(89.4 - 90)/ 5/\text{sqrt}(150)= -1.47$.

- Since the alternate hypothesis is of the form

$$\mu < \mu_0$$

  level = $P(Z < -1.47)$

- Therefore the level is $\alpha = 0.0708$.

Relationship Between Hypothesis Tests & Confidence Intervals

Prof. Preet Kanwal

Introduction

- In a **hypothesis test** for a population mean μ, we specify a **particular value of μ** (the null hypothesis) and determine whether that value is plausible.

- In contrast, a **confidence interval** for a population mean μ can be thought of as the **collection of all values for μ** that meet a certain criterion of plausibility,specified by the confidence level $100(1 - \alpha)\%$.

Relationship

- The values contained within a **two-sided level** 100(1−α)% confidence interval for a population mean μ are precisely those values for which the P-value of a two-tailed hypothesis test will be greater than α.

  Example: the 95% confidence interval consists of precisely those values of μ whose P-values are greater than 0.05 in a hypothesis test.

- A **one-sided level** 100(1 − α)% confidence interval consists of all the values for which the P-value in a one-tailed test would be greater than α.

Confidence level and P-values

The confidence level is equivalent to (1 – α) level.

So, if your significance level is 0.05, the corresponding confidence level is 95%.

- If the P value is less than your significance (alpha) level, the hypothesis test is statistically significant.
- If the confidence interval does not contain the null hypothesis value, the results are statistically significant.
- If the P value is less than alpha, the confidence interval will not contain the null hypothesis value.

Example

- Let the P value = 0.031
- Significance level (α) = 0.05 => CL = 95%

- Since P < α , our results are statistically significant.

- Similarly, our 95% confidence interval [267, 394] does not include the null hypothesis mean of 260 and we draw the same conclusion.

Problem 1

A scientist computes the following use same data:

- 90% confidence interval to be (4.38, 6.02)
- 95% confidence interval to be (4.22, 6.18), and
- 99% confidence interval to be (3.91, 6.49).

Now **she wants to test H 0 : μ = 4 versus H 1 : μ ≠ 4.**

Regarding the P-value, which one of the following statements is true?

i. P > 0.10

ii. 0.05 < P < 0.10

iii. 0.01 < P < 0.05

iv. P < 0.01

Problem 1 : Solution

- 90% CI (4.38, 6.02) does not contain the value of null hypothesis (4). Hence, the result is statistically significant at 10%.

  => P-value < 0.10.

- 95% CI (4.22, 6.18) does not contain the value of null hypothesis (4). Hence, the result is statistically significant at 5%.

  =>P-value < 0.05.

- 99% CI (3.91, 6.49) contains the value of null hypothesis (4). This means the result is not statistically significant at 1%.

  => P-value > 0.01

- **Option (iii) : 0.01 < P < 0.05**

Problem 2

- The strength of a certain type of rubber is tested by subjecting pieces of the rubber to an abrasion test.
- **For the rubber to be acceptable, the mean weight loss μ must be less than 3.5 mg.**
- A large number of pieces of rubber that were cured in a certain way were subject to the abrasion test.
- A 95% upper confidence bound for the mean weight loss was computed from these data to be 3.45 mg.
- Someone suggests using these data to test

$H_0: \mu \geq 3.5$ versus $H_1: \mu < 3.5$.

a. Is it possible to determine from the confidence bound whether $P < 0.05$? Explain.

b. Is it possible to determine from the confidence bound whether $P < 0.01$? Explain.

Problem 2 : Solution

(a) Yes.

Since **3.5 > 3.45 (upper confidence bound)**

Quantities greater than the upper confidence bound will have P-values less than 0.05. **Therefore P < 0.05.**

(b) No, we would need to know the 99% upper confidence bound to determine whether P < 0.01.

Problem 3

A shipment of fibers is not acceptable if the mean breaking strength of the fibers is less than 50 N.

- A large sample of fibers from this shipment was tested, and a 98% lower confidence bound for the mean breaking strength was computed to be 50.1 N.
- Someone suggests using these data to test the hypotheses **$H_0: \mu \le 50$ versus $H_1: \mu > 50$.**

a. Is it possible to determine from the confidence bound whether $P < 0.01$? Explain.

b. Is it possible to determine from the confidence bound whether $P < 0.05$? Explain.

Problem 3 : Solution

(a) No, we would need to know the 99% lower confidence bound in order to determine whether $P < 0.01$.

(b) Yes, since 50 is less than the 98% lower confidence bound(50.1),

we know that $P < 0.02$. Therefore $P < 0.05$.

Statistical Significance vs. Practical Significance

Statistical Significance Is Not the Same as Practical Significance

- When a result has a small P-value, we say that it is “statistically significant.”
- Sometimes statistically significant results do not have any scientific or practical importance.

Example

- Assume that a process used to manufacture synthetic fibers is known to produce fibers with a mean breaking strength of 50 N.
- A new process, which would require considerable retooling to implement, has been developed.
- In a sample of 1000 fibers produced by this new method, the average breaking strength was 50.1 N, and the standard deviation was 1 N.
- **Can we conclude that the new process produces fibers with greater mean breaking strength?**

Example :Solution

Let μ be the mean breaking strength of fibers produced by the new process. We need to test,

$H_0 : \mu \leq 50$ versus $H_1 : \mu > 50$.

- In this way, if we reject H_0 , we will conclude that the new process is better.
- Under H0 , the sample mean X has a normal distribution with

  mean 50 and

  standard deviation 1/ sqrt(1000) = 0.0316.
- The z-score is:

  $z = (50.1 - 50) / 0.0316 = 3.16$
- P -value = $P(z > 3.16) = 0.0008$

Result is Statistically Significant

- The P-value is 0.0008. This is very strong evidence against H_0 .
- The new process produces fibers with a greater mean breaking strength.

Result is not Practically Significant

What practical conclusion should be drawn from this result?

- On the basis of the hypothesis test, we are quite sure that the new process is better.
- Would it be worthwhile to implement the new process? Probably not.
- The reason is that the difference between the old and new processes, although highly statistically significant, amounts to only 0.1 N.
- It is unlikely that this is difference is large enough to matter.

Conclusion

- Result can be statistically significant without being large enough to be of practical importance.
- The P-value does not measure practical significance.

Reason:

When the standard deviation is very small, even a small difference can be statistically significant.

Z-tests
for
Difference between Two Population Means of Large Samples.
(Section 6.5)

Prof. Preet Kanwal

Introduction

- We wish to determine whether the means of two populations are equal.
- The data will consist of two samples, one from each population.
- We will compute the difference of the sample means.
- If the difference is far from 0, we will conclude that the population means are different.
- If the difference is close to 0, we will conclude that the population means might be the same.

Null Distribution of $\overline{X} - \overline{Y}$

- The test is based on $\overline{X} - \overline{Y}$.
- Since both sample sizes are large, X and Y are both approximately normally distributed.
- Since the samples are independent, it follows that the null distribution of $\overline{X} - \overline{Y}$ is:

$$\overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y,\ \sigma_{\overline{X}}^2 + \sigma_{\overline{Y}}^2\right) = N\left(\mu_X - \mu_Y,\ \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}\right)$$

Z-tests for Difference between Two Population Means of Large Samples.

Let $X_1, \ldots, X_{n_X}$ and $Y_1, \ldots, Y_{n_Y}$ be *large* (e.g., $n_X > 30$ and $n_Y > 30$) samples from populations with means μ_X and μ_Y and standard deviations σ_X and σ_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: $z = \dfrac{(\overline{X} - \overline{Y}) - \Delta_0}{\sqrt{\sigma_X^2/n_X + \sigma_Y^2/n_Y}}$. If σ_X and σ_Y are unknown they may be approximated with s_X and s_Y, respectively.
- Compute the P-value. The P-value is an area under the normal curve, which depends on the alternate hypothesis as follows:

Alternate Hypothesis	***P*-value**
$H_1: \mu_X - \mu_Y > \Delta_0$	Area to the right of z
$H_1: \mu_X - \mu_Y < \Delta_0$	Area to the left of z
$H_1: \mu_X - \mu_Y \neq \Delta_0$	Sum of the areas in the tails cut off by z and $-z$

Problem 1

The article “Wired: Energy Drinks, Jock Identity, Masculine Norms, and Risk Taking” (K. Miller, Journal of American College Health, 2008:481–489) reports that in a sample of 413 male college students, the average number of energy drinks consumed per month was 2.49 with a standard deviation of 4.87, and in a sample of 382 female college students, the average was 1.22 with a standard deviation of 3.24.

Can you conclude that the mean number of energy drinks is greater for male students than for female students?

Problem 1 : Solution

(male students)

$X = 2.49$

$s_X = 4.87$

$n_X = 413$

(female students)

$Y = 1.22$

$s_Y = 3.24$

$n_Y = 382$

- **Can you conclude that the mean number of energy drinks is greater for male students than for female students?**

$$H_0 : \mu_X - \mu_Y \leq 0 \text{ versus } H_1 : \mu_X - \mu_Y > 0$$

Problem 1 : Solution

- Null Distribution$(0, 3.24^2/382 + 4.87^2/413)$

- test-statistic, $z = \dfrac{(2.49-1.22-0)}{\text{sqrt}(3.24^2/382 + 4.87^2/413)}$

  => $z = 4.36$.

- $P = P(Z > 4.36) \approx 0$
- Since $P < 0.05$, We can reject H_0 and accept H_1.
- Hence, we can conclude that the mean number of energy drinks is greater for male students than for female students

Problem 2

Two machines used to fill soft drink containers are being compared. The number of containers filled each minute is counted for 60 minutes for each machine. During the 60 minutes, machine 1 filled an average of 73.8 cans per minute with a standard deviation of 5.2 cans per minute, and machine 2 filled an average of 76.1 cans per minute with a standard deviation of 4.1 cans per minute.

Assuming that all necessary assumptions are met, perform a hypothesis test. Can you conclude that machine 2 is faster than machine 1?

Problem 2 : Solution

(machine 1)

$X = 73.8$

$s_X = 5.2$

$n_X = 60$

(machine 2)

$Y = 76.1$

$s_Y = 4.1$

$n_Y = 60$

- Can you conclude that machine 2 is faster than machine 1?

$$H_0 : \mu_X - \mu_Y \geq 0 \text{ versus } H_1 : \mu_X - \mu_Y < 0$$

Problem 2 : Solution

- test-statistic, $z = \dfrac{(73.8 - 76.1) - 0}{\text{sqrt}(5.2^2/60 + 4.1^2/60)}$

  => $z = -2.69$

- Since its a left-tailed test,

  $P = P(Z < -2.69) = 0.0036$

- Since $P < 0.05$ we reject H_0 and accept H_1.
- Hence, we conclude machine 2 is faster than machine 1.

Problem 3

The following MINITAB output presents the results of a hypothesis test for the difference $\mu_X - \mu_Y$ between two population means. Some of the numbers are missing.

```
Two-sample T for X vs Y

      N      Mean     StDev     SE Mean
X    78      23.3      (i)       1.26
Y    63     20.63     3.02       (ii)

Difference = mu (X) — mu (Y)
Estimate for difference: 2.670
95% CI for difference: (0.05472, 5.2853)
T-Test of difference = 0 (vs not =): T-Value = 2.03 P-Value = 0.045 DF = 90
```

a. Fill in the missing numbers for (i) and (ii).
b. The output presents a Student's t test. Compute the P-value using a z test. Are the two results similar?
c. Compute a 98% confidence interval for $\mu_X - \mu_Y$ based on the z statistic.

Problem 3 : Solution

(a) (i) StDev = (SE Mean) * $\sqrt{N}$ = 1.26 * $\sqrt{78}$ = 11.128

(ii) SE Mean = StDev / $\sqrt{N}$ = 3.02 /$\sqrt{63}$ = 0.380484

(b) $z = (23.3 - 20.63 - 0)/ \sqrt{(1.26^2 + 0.380484^2)} = 2.03$.

- Since the alternate hypothesis is of the form $\mu_X - \mu_Y \neq 0$, the P-value is the sum of the areas to the right of $z = 2.03$ and to the left of $z = -2.03$.

- **Thus P = 0.0212 + 0.0212 = 0.0424, and the result is similar to that of the t test.**

Problem 3 : Solution

98% confidence interval for $\mu_X - \mu_Y$:

$X = 23.3$ $s_X / \sqrt{n_X} = 1.26$

$Y = 20.63$ $s_Y / \sqrt{n_Y} = 0.380484$

$Z_{.01} = 2.33$

The 98% confidence interval is:

$23.3 - 20.63 \pm 2.33 \sqrt{(1.26^2 + 0.380484^2)}$

$= (-0.3967, 5.7367)$.

Z-tests
for
Population Proportions
of
Large Samples
(Section 6.3)

Prof. Preet Kanwal

Null distribution of p_hat

- Null distribution of p_hat,

$$\widehat{p} \sim N\left(p,\ \frac{p(1-p)}{n}\right)$$

- p is approximated by using p0 value.

Z-tests for Population Proportions of Large Samples

Let X be the number of successes in n independent Bernoulli trials, each with success probability p; in other words, let $X \sim \text{Bin}(n,\ p)$.

To test a null hypothesis of the form $H_0: p \le p_0$, $H_0: p \ge p_0$, or $H_0: p = p_0$, assuming that both np_0 and $n(1 - p_0)$ are greater than 10:

- Compute the z-score: $z = \dfrac{\widehat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$.
- Compute the P-value. The P-value is an area under the normal curve, which depends on the alternate hypothesis as follows:

Alternate Hypothesis	***P*-value**
$H_1: p > p_0$	Area to the right of z
$H_1: p < p_0$	Area to the left of z
$H_1: p \ne p_0$	Sum of the areas in the tails cut off by z and $-z$

Relationship with Confidence Intervals for a Proportion

A level $100(1-\alpha)\%$ confidence interval for a population mean μ contains those values for a parameter for which the P-value of a hypothesis test will be greater than α.

Problem 1

A random sample of 300 electronic components manufactured by a certain process are tested, and 25 are found to be defective.

- Let p represent the proportion of components manufactured by this process that are defective.
- The process engineer claims that $p \leq 0.05$.
- Does the sample provide enough evidence to reject the claim?

Problem 1 : Solution

p_hat = 25 /300 = 0.083

- The process engineer claims that $p \le 0.05$.

H_0 : $p \le 0.05$ versus H_1 : $p > 0.05$

- **Null distribution :**

p_hat ~ N(p , p(1 – p)/n)

- (0.083 – 0.05)

Z = --------------------

sqrt(0.05 (1 – 0.05)/ 300)

=> Z = 2.62

=> P = P(Z > 2.62) = 0.0044

Note that under the commonly used rule of thumb, we would reject H_0 and reject process engineer's claim. Hence, sample provides enough evidence.

Problem 2

The article "Application of Surgical Navigation to Total Hip Arthroplasty" (T. Ecker and S. Murphy, Journal of Engineering in Medicine, 2007:699–712) reports that in a sample of 113 people undergoing a certain type of hip replacement surgery on one hip, 65 of them had surgery on their right hip.

Can you conclude that frequency of this type of surgery differs between right and left hips?

Problem 2 : Solution

p_hat = 65/113 = 0.575

Can you conclude that frequency of this type of surgery differs between right and left hips?

H_0 : $p = 0.5$ versus H_1 : $p \neq 0.5$

$$z = \frac{(0.575 - 0.5)}{\text{sqrt}(0.5(1 - 0.5)/113)}$$

=> $z = 0.075/0.047 = 1.596 \approx 1.60$

Since its a two tailed test,

$P = P(Z < -1.60) + P(Z > 1.60) = 0.0548 + 0.0548 = 0.1096$

We cannot conclude that the frequency of this type of surgery differs between right and left hips.

Problem 3

During a recent drought, a water utility in a certain town sampled 100 residential water bills and found that 73 of the residences had reduced their water consumption over that of the previous year.

Can it be concluded that more than 60% of the residences in the town reduced their water consumption?

Problem 3 : Solution

p_hat = 73/100 = 0.73

Can it be concluded that more than 60% of the residences in the town reduced their water consumption?

H_0 : $p \le 0.60$ versus H_1 : $p > 0.60$

$$z = \frac{(0.73 - 0.60)}{\text{sqrt}(0.60(1 - 0.60)/100)}$$

=> $z = 2.65$

=> $P = P(Z > 2.65) = 0.0040$

Hence we can conclude that more than 60% of the residences in the town reduced their water consumption.

Problem 4

The following output presents the results of a hypothesis test for a population proportion p. Some of the numbers are missing. Fill in the numbers for (a) through (c). Test of $p \geq 0.7$ vs $p < 0.7$

Variable	X	N	Sample p	95% Upper Bound	Z-Value	P-Value
X	345	500	(a)	0.724021	(b)	(c)

Problem 4 : Solution

a) Sample p = p_hat = 345/500 = 0.690.

(b) The null and alternate hypotheses are

$$H_0 : p \geq 0.7 \text{ versus } H_1 : \mu < 0.7.$$

n = 500.

From part (a), p_hat = 0.690.

z = (0.690 − 0.700)/ sqrt(0.7(1 − 0.7)/500)

z = −0.49.

(c) Since the alternate hypothesis is of the form p < p 0 , the P-value is the area to the left of z = −0.49.

Thus P = P(Z < -0.49) = 0.3121.

Z-test
for
Difference between Two Population proportions of Large Samples.
(Section 6.6)

Prof. Preet Kanwal

Introduction

- The test is based on the statistic $\widehat{p}_X - \widehat{p}_Y$.
- Determining the null distribution of statistic $\widehat{p}_X - \widehat{p}_Y$

  By the Central Limit Theorem, since n_X and n_Y are both large,

$$\widehat{p}_X \sim N\left(p_X,\ \frac{p_X(1-p_X)}{n_X}\right) \qquad \widehat{p}_Y \sim N\left(p_Y,\ \frac{p_Y(1-p_Y)}{n_Y}\right)$$

Therefore,

$$\widehat{p}_X - \widehat{p}_Y \sim N\left(p_X - p_Y,\ \frac{p_X(1-p_X)}{n_X} + \frac{p_Y(1-p_Y)}{n_Y}\right)$$

Pooling

- "Pooling" is the name given to a technique used to obtain a **more precise estimate of the standard deviation of a sample statistic** by combining the estimates given by two (or more) independent samples.
- When performing tests (or calculating confidence intervals) for a difference of two means, we do not pool. In other statistical situations we may or may not pool.
- We always use a pooled estimate of the standard deviation (based on a pooled estimate of the proportion) when carrying out a hypothesis test whose null hypothesis is p1 = p2 (or equivalently, p1 - p2 = 0).
- If they are actually equal, then **combining them gives a better estimate of the standard error** than keeping them separate.

Pooled Proportion

However, the null hypothesis H_0 specifies that the population proportions are equal.

- Therefore we must estimate them both p_X and p_Y with a common value.
- The appropriate value is the pooled proportion,

$$\widehat{p} = \frac{X + Y}{n_X + n_Y}$$

Null distribution of $\widehat{p}_X - \widehat{p}_Y$

The null distribution of $\widehat{p}_X - \widehat{p}_Y$ is therefore estimated as,

$$\widehat{p}_X - \widehat{p}_Y \sim N\left(0,\ \widehat{p}(1-\widehat{p})\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)\right)$$

Where,

pooled proportion = $\widehat{p} = \dfrac{X+Y}{n_X + n_Y}$

Summary

Let $X \sim \text{Bin}(n_X,\ p_X)$ and let $Y \sim \text{Bin}(n_Y,\ p_Y)$. Assume that there are at least 10 successes and 10 failures in each sample, and that X and Y are independent.

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

- Compute $\widehat{p}_X = \dfrac{X}{n_X}$, $\widehat{p}_Y = \dfrac{Y}{n_Y}$, and $\widehat{p} = \dfrac{X+Y}{n_X+n_Y}$.
- Compute the z-score: $z = \dfrac{\widehat{p}_X - \widehat{p}_Y}{\sqrt{\widehat{p}(1-\widehat{p})(1/n_X + 1/n_Y)}}$.
- Compute the P-value. The P-value is an area under the normal curve, which depends on the alternate hypothesis as follows:

Alternate Hypothesis	***P*-value**
$H_1 : p_X - p_Y > 0$	Area to the right of z
$H_1 : p_X - p_Y < 0$	Area to the left of z
$H_1 : p_X - p_Y \neq 0$	Sum of the areas in the tails cut off by z and $-z$

Problem 1 - Description

A mobile computer network consists of computers that maintain wireless communication with one another as they move about a given area. A routing protocol is an algorithm that determines how messages will be relayed from machine to machine along the network, so as to have the greatest chance of reaching their destination.

An article compares the effectiveness of two routing protocols over a variety of metrics, including the rate of successful deliveries.

Problem 1

- Assume that using protocol A, 200 messages were sent, and 170 of them, or 85%, were successfully received.
- Using protocol B, 150 messages were sent, and 123 of them, or 82%, were successfully received.

Can we conclude that protocol A has the higher success rate?

Problem 1 - Solution

- Let X represent the number of messages successfully sent using protocol A.

$$X \sim Bin(200, p_X)$$

p_X_hat = 170/200 = 0.85

- Let Y represent the number of messages successfully sent using protocol B.

$$Y \sim Bin(150, p_Y)$$

p_Y_hat = 123/150 = 0.82

Problem 1 - Solution

Can we conclude that protocol A has the higher success rate?

$H_0 : p_X - p_Y \leq 0$ versus $H_1 : p_X - p_Y > 0$

- Pooled proportion,

  p_hat = (170 + 123) /(200 + 150) = 0.837

- Null distribution of $\widehat{p}_X - \widehat{p}_Y$

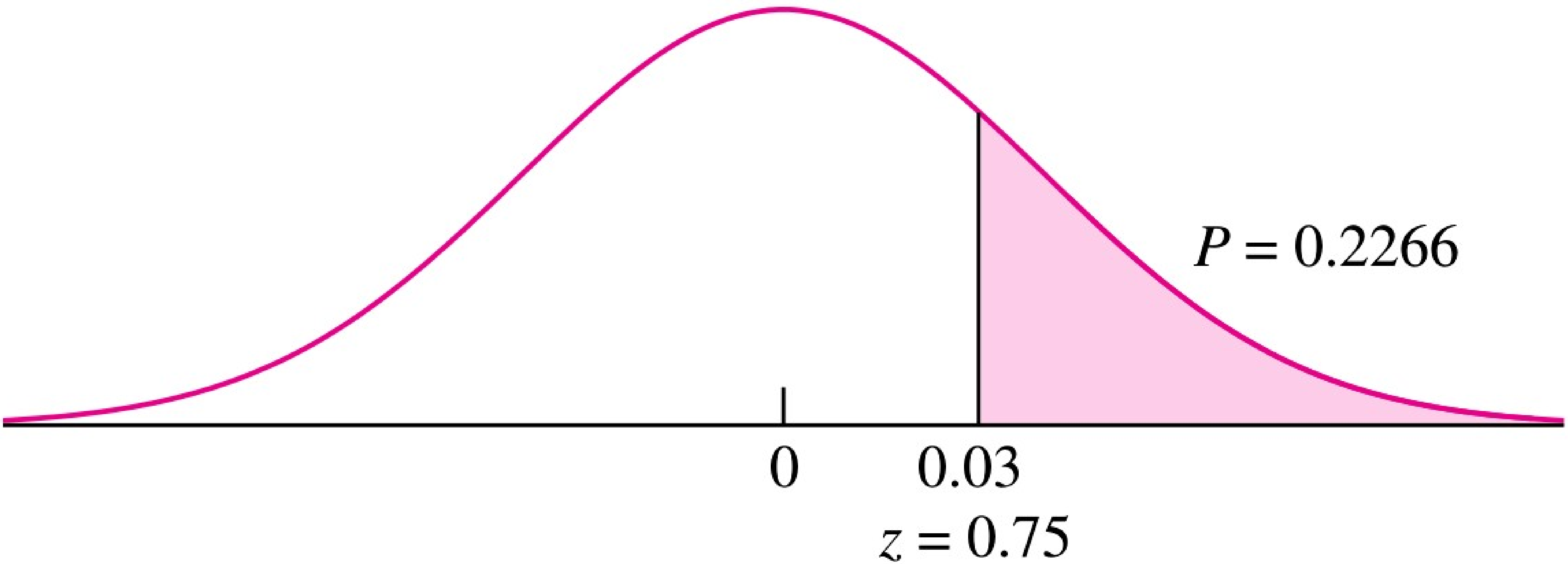
  $\widehat{p}_X - \widehat{p}_Y \sim N(0, 0.0399)$

- Test-statistic, z = (p_X_hat – p_Y_hat) – 0/SE

  => z = (0.85 – 0.82)/ 0.0399 = 0.75

- P = P(Z > 0.75) = 0.2266 => Both null and alt are plausible.

Problem 1 - Solution



Problem 2

- Two extrusion machines that manufacture steel rods are being compared.
- In a sample of 1000 rods taken from machine 1, 960 met specifications regarding length and diameter.
- In a sample of 600 rods taken from machine 2, 582 met the specifications.
- Machine 2 is more expensive to run, so it is decided that machine 1 will be used unless it can be convincingly shown that machine 2 produces a larger proportion of rods meeting specifications.

a. State the appropriate null and alternate hypotheses for making the decision as to which machine to use.

b. Compute the P-value.

c. Which machine should be used?

Problem 2 : Solution

- Let X represent rods manufactured using Machine 1 meeting specifications.

$$X \sim Bin(1000, p_X)$$

p_X_hat = 960/1000 = 0.96

- Let Y represent rods manufactured using Machine 2 meeting specifications.

$$Y \sim Bin(600, p_Y)$$

p_Y_hat = 582/600 = 0.97

- To show that machine 2 produces a larger proportion of rods meeting specifications, we need to test,

$$\mathbf{H_0 : p_X - p_Y \geq 0 \text{ versus } H_1 : p_X - p_Y < 0}$$

Problem 2 : Solution

- Pooled proportion,

  p_hat = (960 + 582)/(1000 + 600) = 0.96375
- Test-statistic:

$$z = \frac{0.960 - 0.970}{\sqrt{0.96375(1 - 0.96375)(1/1000 + 1/600)}} = -1.04.$$

- Since its left tailed test,

  P = P(Z < -1.04) = 0.1492
- Since P > 0.05 => We cannot reject **H_0.**
- Hence, machine 1 must be used, as we could not prove that machine 2 produces larger proportion of rods meeting specifications.

Problem 3

In a study conducted by the U.S. Department of Health and Human Services,

- A sample of 546 boys aged 6–11 was weighed, and it was determined that 87 of them were overweight.
- A sample of 508 girls aged 6–11 was also weighed, and 74 of them were overweight.

Can you conclude that the proportion of boys who are overweight differs from the proportion of girls who are overweight?

Problem 3 : Solution

- Let X represent no of boys that are overweight

$$X \sim Bin(546,\ p_X)$$

p_X_hat = 87/546 = 0.159

- Let Y represent no of girls that are overweight

$$Y \sim Bin(508,\ p_Y)$$

p_Y_hat = 74/508 = 0.146

- **Can you conclude that the proportion of boys who are overweight differs from the proportion of girls who are overweight?**

$$H_0 : p_X - p_Y = 0 \text{ versus } H_1 : p_X - p_Y \neq 0$$

Problem 3 : Solution

- Pooled proportion,

  p_hat =(87 + 74)/(546 + 508) = 0.15275

- Test-statistic:

$$z = \frac{0.159341 - 0.145669}{\sqrt{0.15275(1-0.15275)(1/546+1/508)}} = 0.62$$

- Since its two tailed test,

  P = P(Z < -0.62) + P(Z > 0.62)= 0.2676 + 0.2676 = 0.5352

- Since P > 0.05 => We cannot reject **$H_{0.}$**
- Hence cannot conclude that the proportion of boys who are overweight differs from the proportion of girls who are overweight

Tests with Paired Data (Section 6.8)

Prof. Preet Kanwal

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sample of ordered pairs whose differences $D_1, \ldots, D_n$ are a sample from a *normal* population with mean μ_D. Let s_D be the sample standard deviation of $D_1, \ldots, D_n$.

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 $t = \dfrac{\overline{D} - \mu_0}{s_D/\sqrt{n}}$.
- 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:

Alternate Hypothesis	***P*-value**
$H_1: \mu_D > \mu_0$	Area to the right of t
$H_1: \mu_D < \mu_0$	Area to the left of t
$H_1: \mu_D \neq \mu_0$	Sum of the areas in the tails cut off by t and $-t$

- If the sample is large, the D_i need not be normally distributed, the test statistic is $z = \dfrac{\overline{D} - \mu_0}{s_D/\sqrt{n}}$, and a z test should be performed.

Problem 1

Given the data, Can we conclude that the mean level of emissions is less for highway driving than for stop-and-go driving?

```
Paired T-Test and CI: StopGo, Highway

Paired T for StopGo - Highway

                N        Mean      StDev    SE Mean
StopGo          8     1413.75    850.780    300.796
Highway         8     1223.25    820.850    290.214
Difference      8      190.50    284.104    100.446

95% lower bound for mean difference: 0.197215
T-Test of mean difference = 0 (vs > 0):
```

Problem 1 : Solution

- The null and alternate hypotheses are

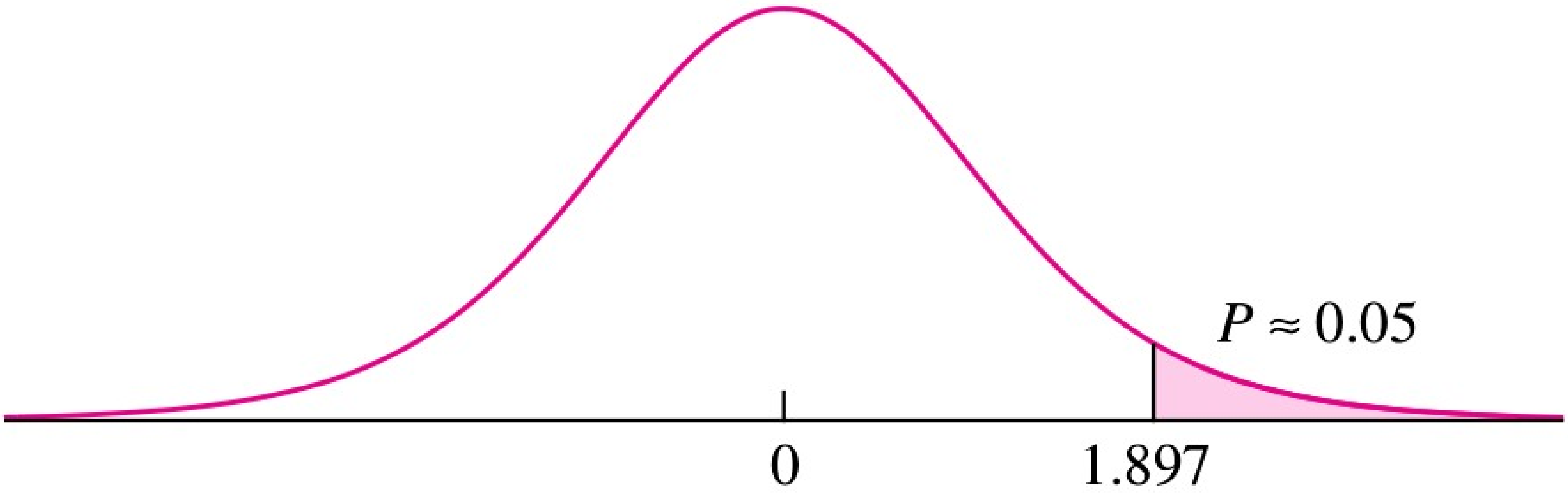
$$H_0: \mu_D \le 0 \quad \text{versus} \quad H_1: \mu_D > 0$$

- The test statistic is

$$t = \frac{\overline{D} - 0}{s_D/\sqrt{n}} = \frac{190.5 - 0}{284.1/\sqrt{8}} = 1.897$$

- => P = P(t > 1.897) ≈ 0.050
- Hence we can reject null hypothesis and conclude that mean level of emissions is less for highway driving than for stop-and-go driving.

Problem 1 : Solution



Problem 2

Two microprocessors are compared on a sample of six benchmark codes to determine whether there is a difference in speed. The times (in seconds) used by each processor on each code are given in the following table. **Can you conclude that the mean speeds of the two processors differ?**

	Code					
	1	2	3	4	5	6
Processor A	27.2	18.1	27.2	19.7	24.5	22.1
Processor B	24.1	19.3	26.8	20.1	27.6	29.8

Problem 2 : Solution

	Code					
	1	2	3	4	5	6
Processor A	27.2	18.1	27.2	19.7	24.5	22.1
Processor B	24.1	19.3	26.8	20.1	27.6	29.8
Differences	3.1	−1.2	0.4	−0.4	−3.1	−7.7

D_bar = −1.48333, s D = 3.6625, n = 6.

$$H_0 : \mu_D = 0 \text{ versus } H_1 : \mu_D \neq 0.$$

- t = (−1.48333 − 0)/(3.6625/ sqrt(6)) = −0.9921.
- P = P(t < -0.9921) + P(t > 0.9921) =

0.10 < P(t > 0.9921) < 0.25

=> P > 0.05, Hence we cannot reject null hypothesis & both null and alt are plausible.

Errors In Hypothesis Testing

Prof. Preet Kanwal

Introduction

- Since a fixed-level test results in a firm decision, there is a chance that the decision could be the wrong one.
- There are exactly two ways in which the decision can be wrong:
    - One can reject H_0 when it is in fact true. This is known as a type I error.
    - Or, one can fail to reject H_0 when it is false. This is known as a type II error.

Type I error (a "false positive")

- **When the null hypothesis is true and you reject it, you make a type I error.**
- The probability of making a type I error is α, which is the level of significance you set for your hypothesis test.
- An α of 0.05 indicates that you are willing to accept a 5% chance that you are wrong when you reject the null hypothesis.
- To lower this risk, you must use a lower value for α.
- However, using a lower value for alpha means that you will be less likely to detect a true difference if one really exists.

Problem 1

A hypothesis test is to be performed, and it is decided to reject the null hypothesis if $P \leq 0.10$. If H_0 is in fact true, what is the maximum probability that it will be rejected?

Problem 1 : Solution

The maximum probability of rejecting H_0 when true is the level $\alpha = 0.10$.

Type II error (a "false negative")

- **When the null hypothesis is false and you fail to reject it, you make a type II error.**
- The probability of making a type II error is β, which depends on the **power of the test**.
- You can decrease your risk of committing a type II error by ensuring your test has enough power.
- You can do this by ensuring your sample size is large enough to detect a practical difference when one truly exists.

Generally

- **Null Hypothesis : No effect**

- **Type I Error / False Positive**

  Detects an effect when there is no effect. **(False Alarm)**

- **Type II Error / False Negative**

  Fails to detect an effect, when it is actually present.

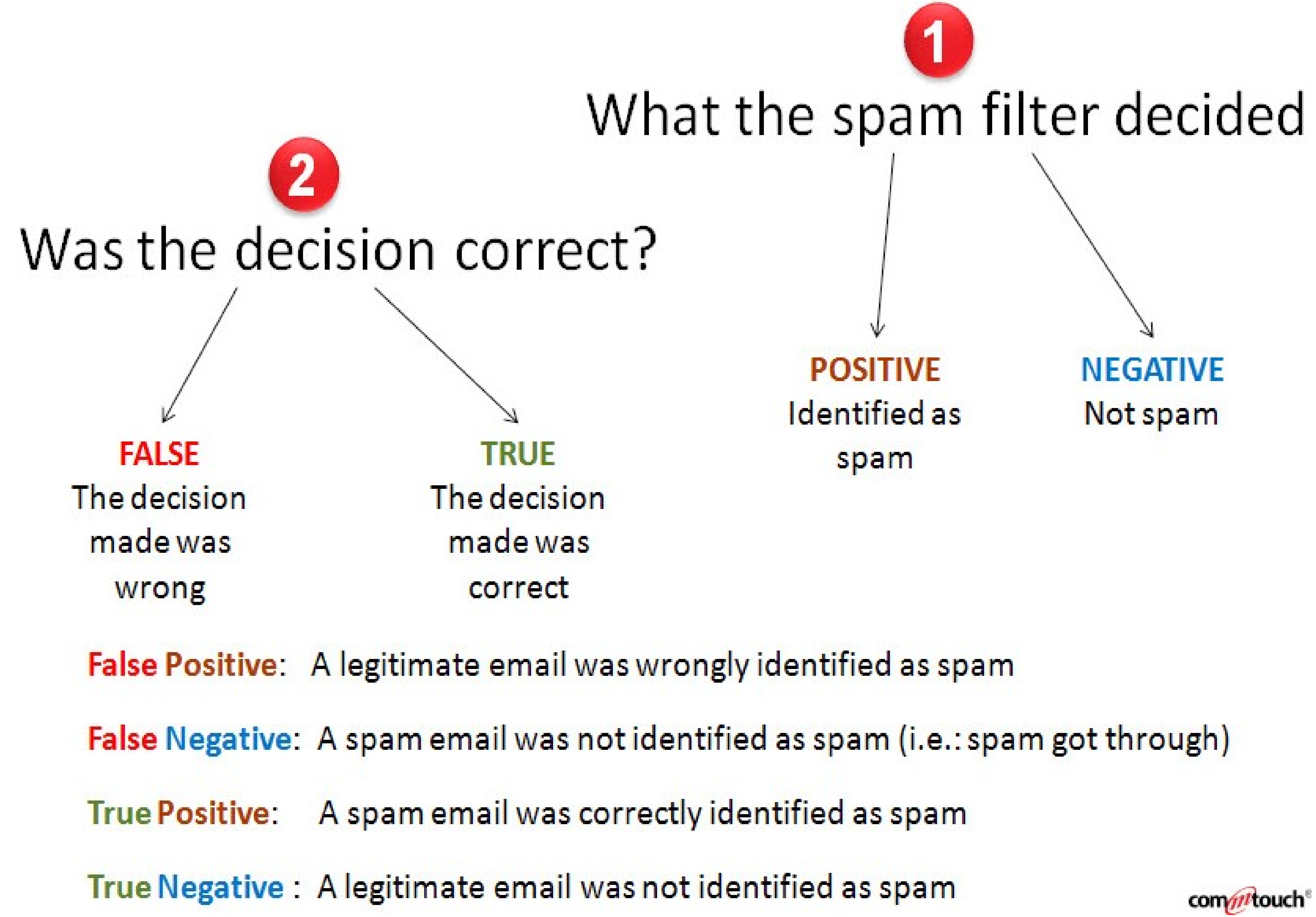
Example : Spam Filter

SPAM vs. HAM

- The statistics about an email filter's effectiveness are built upon the concept of false positives and false negatives.

H0: It is not a spam   vs.   H1: It is a spam

- A false positive may be an important message that ends up in your spam folder or, much worse, gets deleted.
- A false negative (spam that ends up in your inbox) is annoying.
- POPFile is an email classifier that happens to be very good at filtering spam.


1
What the spam filter decided
2
Was the decision correct?
POSITIVE
Identified as spam
NEGATIVE
Not spam
FALSE
The decision made was wrong
TRUE
The decision made was correct
False Positive: A legitimate email was wrongly identified as spam
False Negative: A spam email was not identified as spam (i.e.: spam got through)
True Positive: A spam email was correctly identified as spam
True Negative : A legitimate email was not identified as spam
commtouch

positive rate is, well, quite a bit more than annoying, if crucial messages go unread. Naturally, these companies continue to hone their spam-battling techniques, and now, Google has upped the ante with a new breed of artificial intelligence tools.

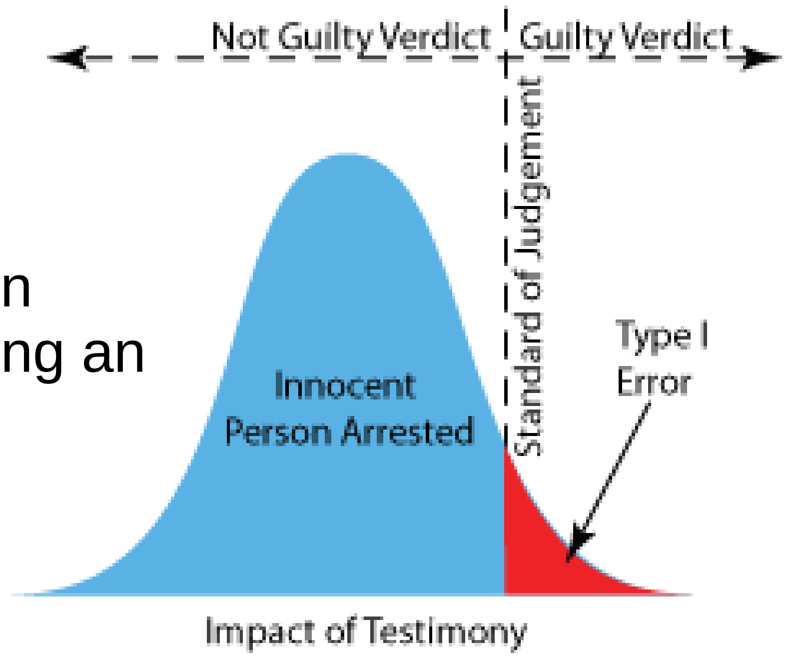
Three years after it last released Gmail's spam stats, Google says that its spam rate is down to 0.1 percent, and its false positive rate has dipped to 0.05 percent. The company credits the significant drop in large part to the introduction of brain-like "neural networks" into its spam filters that can learn to recognize junk mail and phishing messages by analyzing scads off the stuff across an enormous collection of computers.

"One of the great things about machine learning is that it adapts to changing situations." says John Rae-Grant, a senior product manager for Gmail, which Google says is now used by 900 million people across the globe. In other words, Gmail's spam filters don't just curb junk by applying pre-existing rules. They create new rules themselves as they go along.

Other Examples

Example 1

- **Null Hypothesis: Person is not guilty of the crime**



- **Type I Error / False Positive**

  Person is judged as guilty when the person actually did not commit the crime (convicting an innocent person)

- **Type II Error / False Negative**

  Person is judged not guilty when they actually did commit the crime (letting a guilty person go free)

Example 2

- **Null Hypothesis : A person is not Hiv – infected**

- **Type I Error / False Positive**

  A false positive result is an incorrect HIV-positive diagnosis in a person who is not actually infected.

- **Type II Error / False Negative**

  A false negative result is a failure to correctly identify an HIV-infected person as HIV-positive.

Example 3 : Quality Control

- **Null Hypothesis : Product satisfies the customer's specifications**

- **Type I Error / False Positive**

  Rejected the product, and was not sold to the customers(results in loss as a good product is rejected)

- **Type II Error / False Negative**

  Failed to reject the product : Accepted bad product and was sold to the customers(can result in losing the cutomer's and company's position is at stake)

Summary

Decision based on sample		Truth about the population: H_0 true	Truth about the population: H_a true
	Reject H_0	Type I error	Correct decision
	Accept H_0	Correct decision	Type II error

Problem 2

A test is made of the hypotheses:

$H_0: \mu \leq 10$ versus $H_1: \mu > 10$

For each of the following situations, determine whether the decision was correct, a type I error occurred, or a type II error occurred.

a. $\mu = 8$, H_0 is rejected.

b. $\mu = 10$, H_0 is not rejected.

c. $\mu = 14$, H_0 is not rejected.

d. $\mu = 12$, H_0 is rejected.

Problem 2 : Solution

$H_0 : \mu \leq 10$ versus $H_1 : \mu > 10$

a. $\mu = 8$, H_0 is rejected

Type I error. H_0 is true and was rejected.

b. $\mu = 10$, H_0 is not rejected

Correct decision. H_0 is true and was not rejected.

c. $\mu = 14$, H_0 is not rejected

Type II error. H_0 is false and was not rejected.

d. $\mu = 12$, H_0 is rejected

Correct decision. H_0 is false and was rejected.

Problem 3

- A new process is being considered for the liquefaction of coal. The old process yielded a mean of 15 kg of distillate fuel per kilogram of hydrogen consumed in the process.
- Let μ represent the mean of the new process. Following test is performed

$$H_0 : \mu \leq 15 \text{ versus } H_1 : \mu > 15$$

- The new process will be put into production if H_0 is rejected. Putting the **new process** into production is very **expensive.**
- Therefore it would be a **costly error** to put the new process into production if in fact it is no better than the old process.
- **Which procedure provides a smaller probability for this error, to test at the 5% level or to test at the 1% level?**

Problem 3 : Solution

- The 1% level.
- The error in question is rejecting H_0 when true, which is a type I error.
- When the level is smaller, it is less likely to reject H_0 , and thus less likely to make a type I error.

Power of a Test

Introducion

- In order for a test to be useful, a test must have **reasonably small probabilities of both type I and type II errors.**
- The type I error is kept small by choosing a small value of α as the significance level.
- Then the power of the test is calculated. If the power is large, then the probability of a type II error is small as well, and the test is a useful one.

Power of a Test

- The power of a test is the probability of rejecting H_0 when it is false.

$$\text{Power} = 1 - P(\text{type II error})$$

- Power calculations are generally done before data are collected.
- The purpose of a power calculation is to determine whether or not a hypothesis test, when performed, is likely to reject H_0 in the event that H_0 is false.

Computing Power

- In order to compute the power, it is necessary to specify a particular value of μ, because:

  power is different for different values of μ

- if μ is close to H_0 :

  the power will be small

  if μ is far from H_0 :

  the power will be large

- Compute the rejection region.
- Compute the probability that the test statistic falls in the rejection region if the alternate hypothesis is true. This is the power.

Alternate distribution

- The alternate distribution is obtained by shifting the null distribution to chosen value of μ.

- If the alternate mean is chosen **very close to the null mean**, the alternate curve will be almost identical with the null, and the power will be very **close to α.**

- If the alternate mean is **far from the null,** almost all the area under the alternate curve will lie in the rejection region, and the power will be **close to 1.**

How large the power must be for a test

- As with P-values, there is no scientifically valid dividing line between sufficient and insufficient power.

- In general, tests with power greater than 0.80 or perhaps 0.90 are considered acceptable, but there are no well-established rules of thumb.

Problem 1

Find the power of the 5% level test of

$H_0 : \mu \leq 80$ versus $H_1 : \mu > 80$

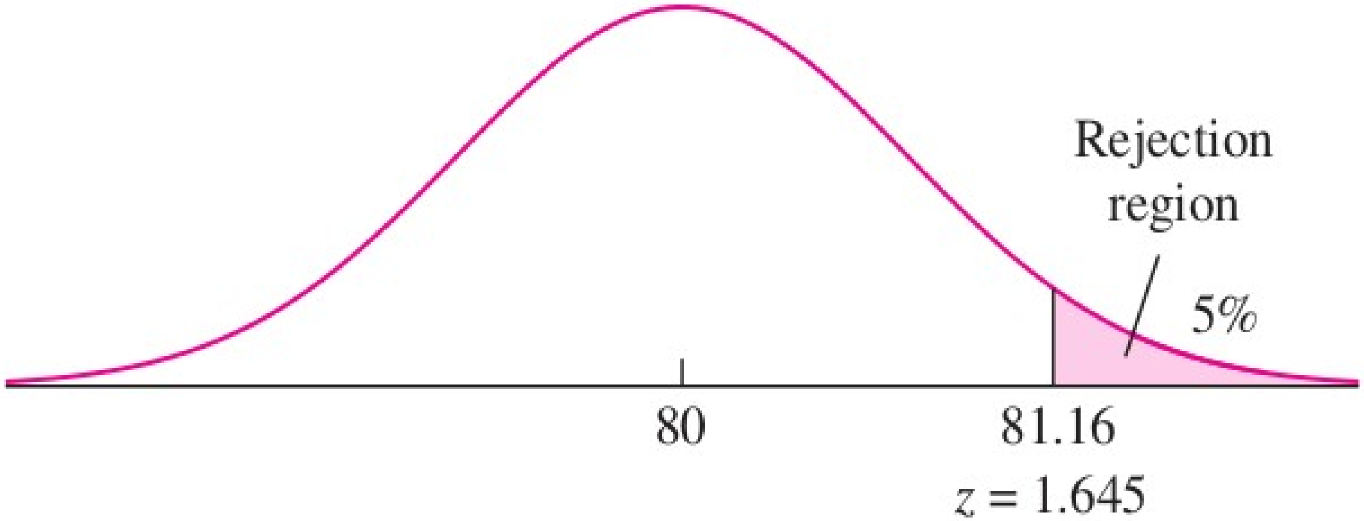
for the mean yield of the new process under the alternative $\mu = 81$, assuming $n = 50$ and $\sigma = 5$.

Critical Point

- Null distribution of X_bar :

X_bar ~ N(80, 5^2 /50)

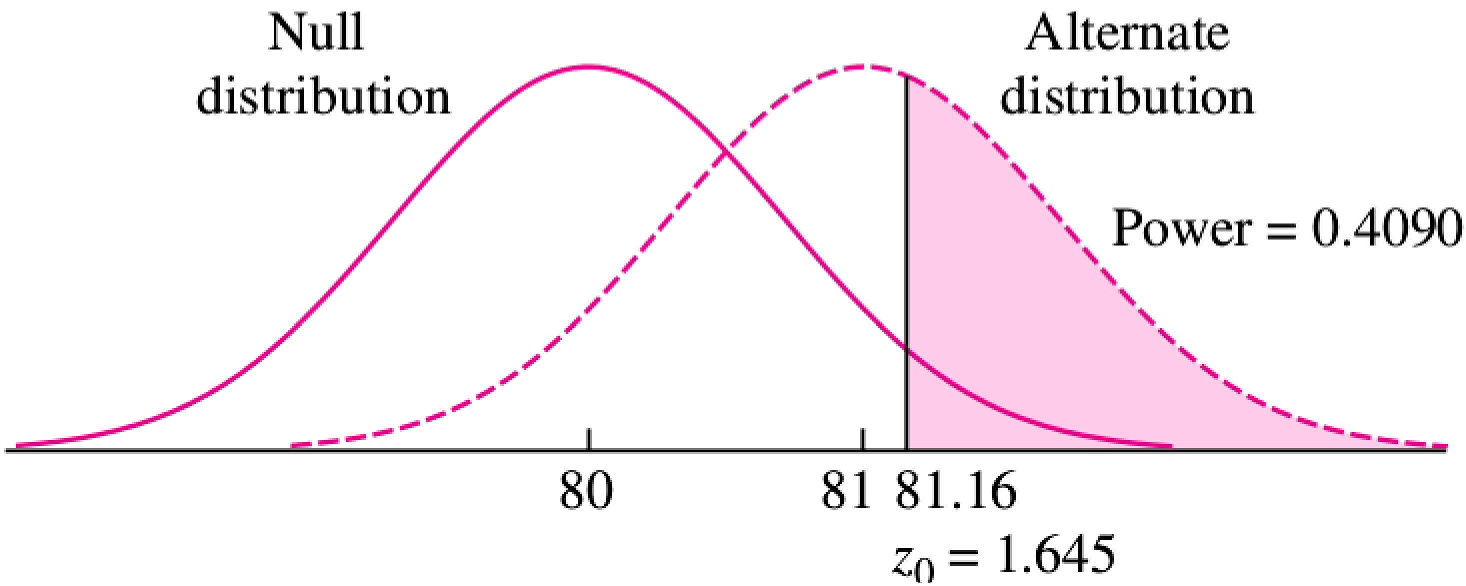
- The **critical point** has a z-score of 1.645, so its value is 80 + (1.645)(0.707) = 81.16.



Computing the power

- Power is the probability that X_bar will fall into the rejection region if the alternate hypothesis μ = 81 is true.
- ALternate distribution of X_bar :

X_bar ~ N(81, 5^2 /50)



$Z_1 = (81.16 - 81)/0.707 \Rightarrow z_1 = 0.23$

Conclusion

- A power of 0.4090 is very low.
- It means that if the mean yield of new process is actually equal to 81, there is only a 41% chance that the proposed experiment will detect the improvement over the old process and allow the new process to be put into production.
- It would be unwise to invest time and money to run this experiment, since it has a large chance to fail.

Problem 2

Find the power of the 5% level test of

$H_0 : \mu \leq 80$ versus $H_1 : \mu > 80$

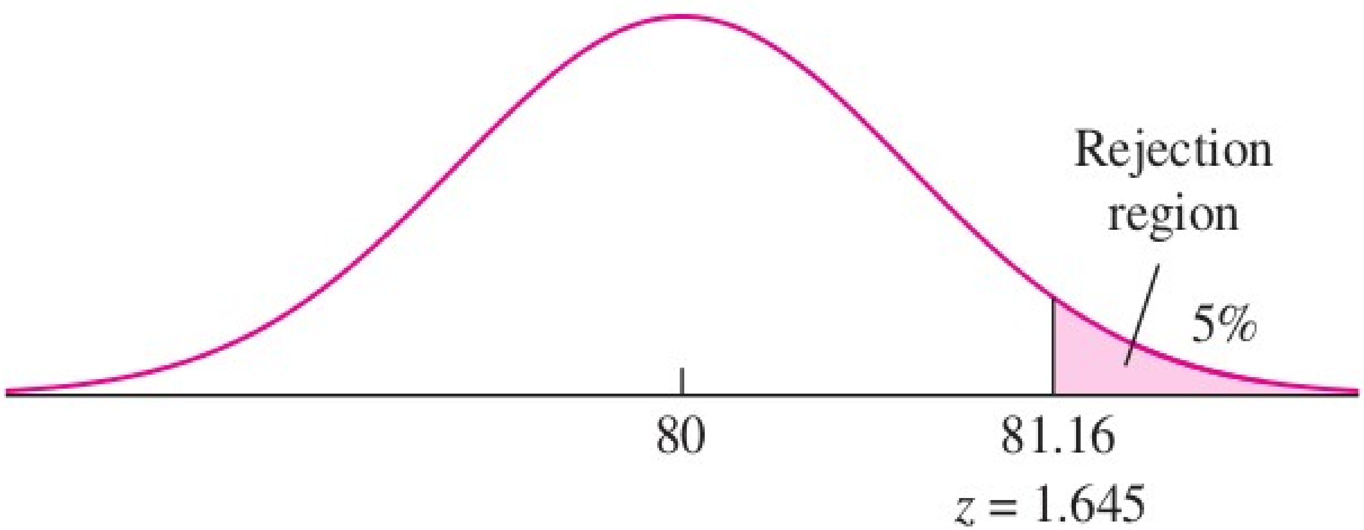
for the mean yield of the new process under the alternative $\mu = 82$, assuming $n = 50$ and $\sigma = 5$.

Critical Point

- Null distribution of X_bar :

$$X\_bar \sim N(80, 5^2/50)$$

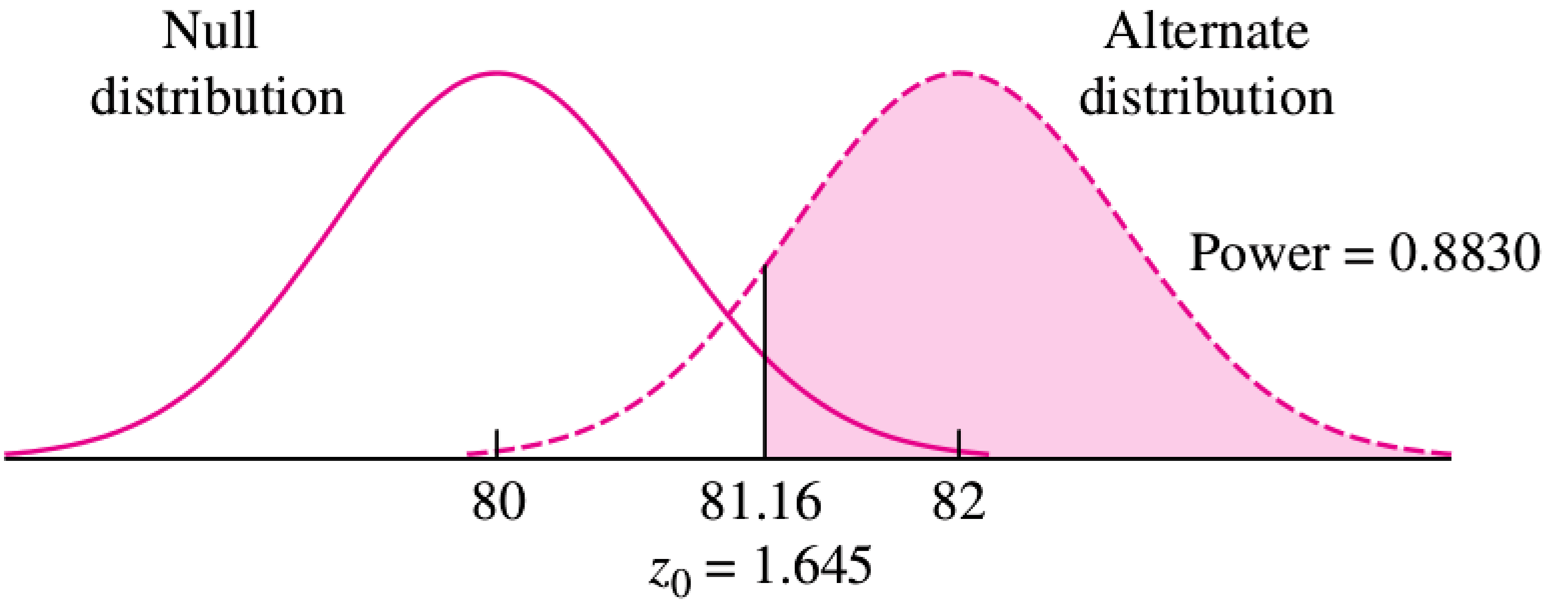
- The critical point has a z-score of 1.645, so its value is 80 + (1.645)(0.707) = 81.16.



Computing the power

- Power is the probability that X_bar will fall into the rejection region if the alternate hypothesis μ = 82 is true.
- Null distribution of X_bar :

X_bar ~ N(82, 5² /50)



$Z_1 = (81.16 - 82)/0.707 \Rightarrow z_1 = -1.19$

Conclusion

- A power of 0.8830 is high.
- It means that if the mean yield of new process is actually equal to 81, there is a 88.30% chance that the proposed experiment will detect the improvement over the old process and allow the new process to be put into production.
- It would be a wise decision to invest time and money to run this experiment, since it has a large chance to succeed.

Note:

When power is not large enough, it can be increased by increasing the sample size.

Problem 3

In testing the hypothesis:

$H0 : \mu \leq 80$ versus $H1 : \mu > 80$

regarding the mean yield of the new process, how many times must the new process be run so that a test conducted at a significance level of 5% will have power 0.90 against the alternative $\mu = 81$, if it is assumed that $\sigma = 5$?

Critical Point

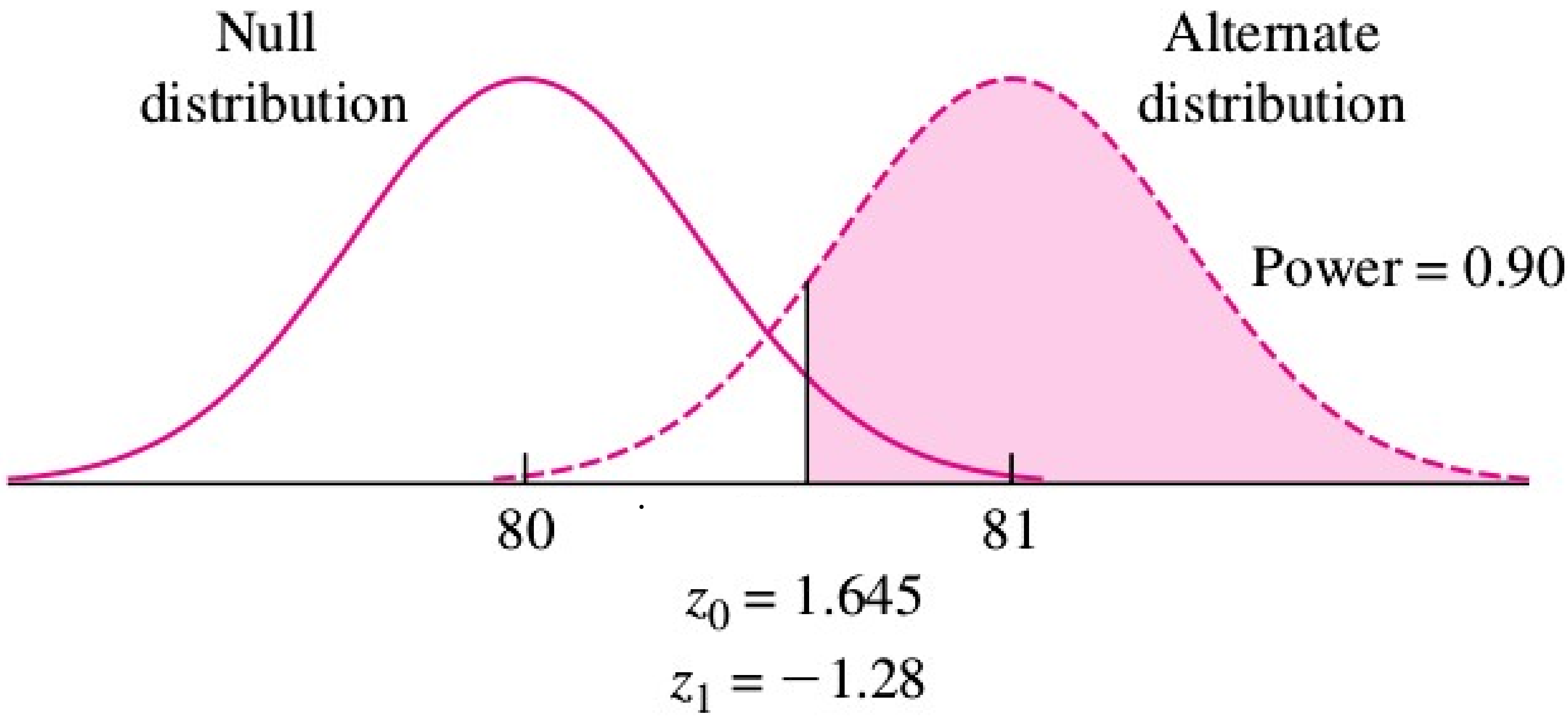
- Null distribution of X_bar :

$$\mathbf{X\_bar \sim N(80,\ 5^2/n)}$$

- The **critical point** has a z-score of 1.645, so its value is 80 + (1.645) * **5 / sqrt(n)**

Power

- The power of the test is the area of the rejection region under the alternate curve. This area must be 0.90.
- The critical point is therefore 81 − 1.28(5/ sqrt (n)) .



Solving for n

- Since there is only one critical point, these two expressions are equal. We therefore set them equal and solve for n.

$$80 + 1.645\left(\frac{5}{\sqrt{n}}\right) = 81 - 1.28\left(\frac{5}{\sqrt{n}}\right)$$

- n ≈ 214.
- The critical point is 80.56.

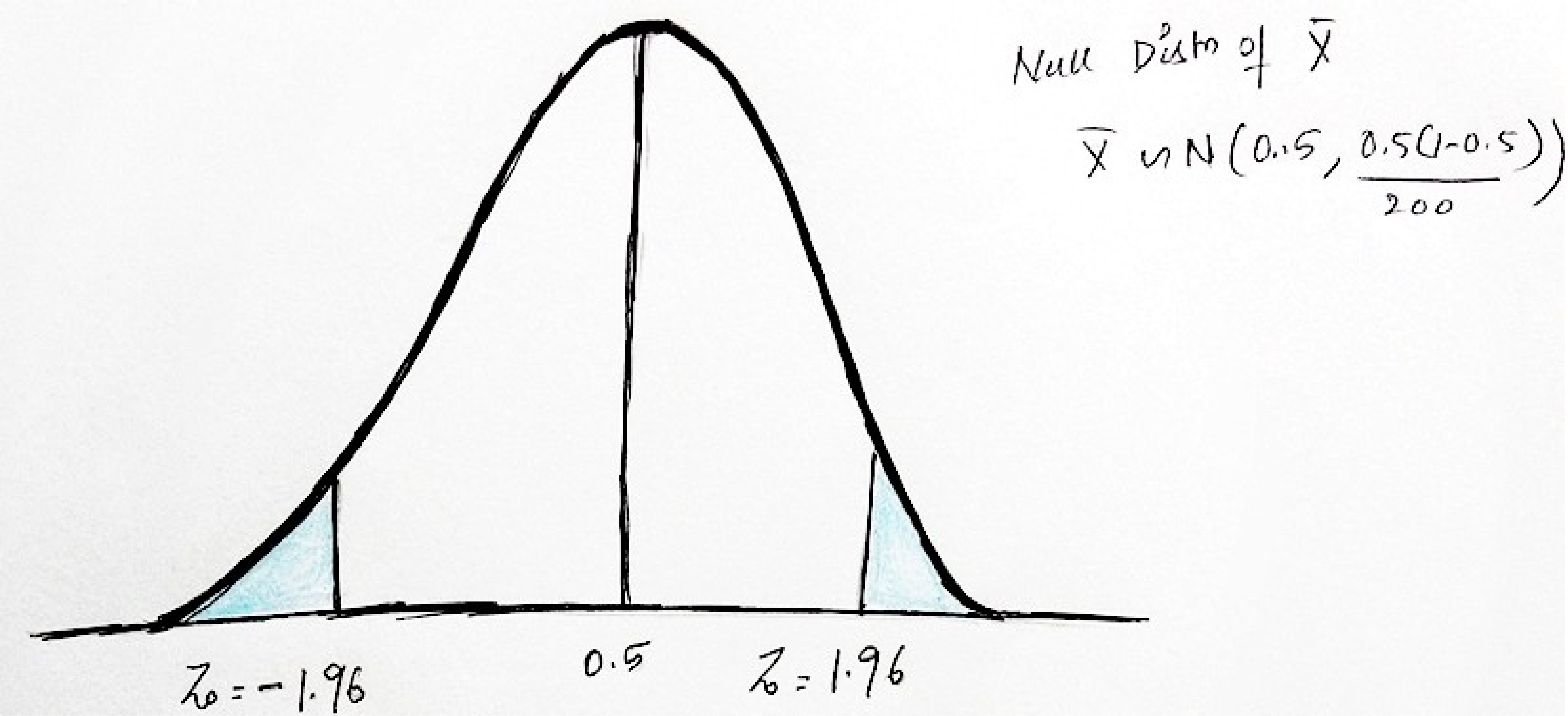
Problem 4

- A pollster will conduct a survey of a random sample of voters in a community to estimate the proportion who support a measure on school bonds.
- Let p be the proportion of the population who support the measure. The pollster will test:

$$H_0 : p = 0.50 \text{ versus } H_1 : p \neq 0.50$$

at the **5% level**. If 200 voters are sampled, **what is the power of the test if the true value of p is 0.55?**

Null Distribution



0.43 0.569

Critical Value

- Null distribution of p_hat:

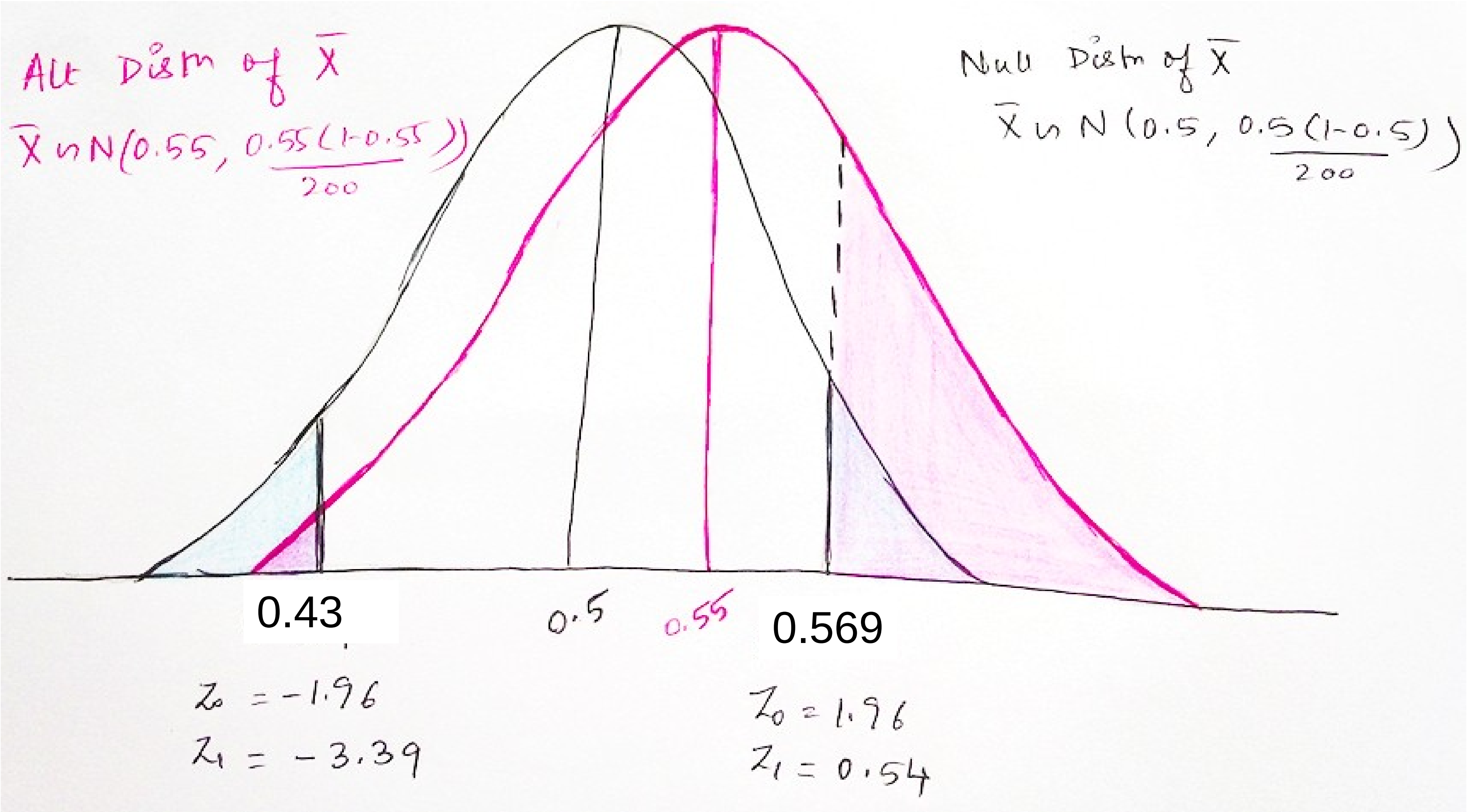
  p_hat ~ N(0.50, 0.50 (1 – 0.50)/200)

  p_hat ~ N(0.50, 0.25/200)

- Since its two-sided test, we reject H0 whenever (as α = 0.05):

Critical Value	
Z ≤ -1.96	Z ≥ 1.96
p_hat ≤ -1.96 * sqrt(0.25/200) + 0.5	p_hat ≥ 1.96 * sqrt(0.25/200) + 0.5
p_hat ≤ 0.43	p_hat ≥ 0.569

Alternate Distribution



Alternate Distribution

- Alternate distribution of p_hat:

  p_hat ~ N(0.55, 0.25/200)
- Computing power:

z- scores corresponding to Critical Value under the alt distribution	
Z = (0.43 – 0.55) / sqrt(0.25/ 200) Z = -3.39 P(Z < -3.39) = .0003	Z = (0.569 – 0.55)/ sqrt(0.25/200) Z = 0.54 P(Z > 0.54) = 0.2946
Power = 0.0003 + 0.2946 = 0.2949	

Problem 5

- A pollster will conduct a survey of a random sample of voters in a community to estimate the proportion who support a measure on school bonds.
- Let p be the proportion of the population who support the measure. The pollster will test:

$$H_0 : p = 0.50 \text{ versus } H_1 : p \neq 0.50$$

- at the **5% level**. **How many voters must be sampled so that the power will be 0.8 when the true value of p = 0.55?**

We can generally ignore the miniscule region associated with one of the tails

Critical Value under Null Distribution	
Z ≤ -1.96 p_hat ≤ -1.96 * sqrt(0.25/n) + 0.5 (Ignored)	Z = 1.96 p_hat = 1.96 * sqrt(0.25/n) + 0.5

Critical Value under Alternate Distribution	
Z = (0.43 – 0.55) / sqrt(0.25/ n) **Z = -3.39** **P(Z < -3.39) = .0003** **(ignored)**	**Since Power = 0.80** **P(Z > z) = 0.80 => z-score = -0.84** p_hat = -0.84 * sqrt(0.25/n) + 0.55

Setting the critical values equal

$1.96 * \sqrt{0.25/n} + 0.5 = -0.84 * \sqrt{0.25/n} + 0.55$

$\Rightarrow \sqrt{n} = (1.96 + 0.84) / 0.1 = 28$

$\Rightarrow n = 784$

Problem 6 – Part a

A tire company claims that the lifetimes of its tires average 50,000 miles. The standard deviation of tire lifetimes is known to be 5000 miles. You sample 100 tires and will test the hypothesis that the mean tire lifetime is at least 50,000 miles against the alternative that it is less. Assume, in fact, that the true mean lifetime is 49,500 miles.

a. State the null and alternate hypotheses. Which hypothesis is true?

Problem 6 – Part a : Solution

- **H0 : $\mu \geq 50{,}000$ vs. H1 : $\mu < 50{,}000$**
- Since True mean lifetime is 49,500 miles , this implies Alternate Hypothesis is true.

Problem 6 – Part b

A tire company claims that the lifetimes of its tires average 50,000 miles. The standard deviation of tire lifetimes is known to be 5000 miles. You sample 100 tires and will test the hypothesis that the mean tire lifetime is at least 50,000 miles against the alternative that it is less. Assume, in fact, that the true mean lifetime is 49,500 miles.

$H0 : \mu \geq 50{,}000$ vs. $H1 : \mu < 50{,}000$

b. If the test is made at the 5% level, what is the power?

Problem 6 – Part b : Solution

H0 : $\mu \geq 50{,}000$ vs. H1 : $\mu < 50{,}000$

- Null distribution of X_bar:

$$\bar{X} \sim N(50{,}000\ ,\ 5000^2 / 100)$$

- $\alpha = 0.05$
  - Since its a left tailed test, z = -1.645
  - Critical value: -1.645 * 5000/10 + 50,000 = 49177.5
- Alternate Distribution of X_bar:

$$\bar{X} \sim N(49{,}500\ ,\ 5000^2 / 100)$$

  - Z = (49177.5 – 49500)/ (5000/10) = -0.645 = -0.65
  - Power = P(Z < -0.65) = 0.2578.

Problem 6 – Part c

A tire company claims that the lifetimes of its tires average 50,000 miles. The standard deviation of tire lifetimes is known to be 5000 miles. You sample 100 tires and will test the hypothesis that the mean tire lifetime is at least 50,000 miles against the alternative that it is less. Assume, in fact, that the true mean lifetime is 49,500 miles.

c. At what level should the test be conducted so that the power is 0.80?

Problem 6 – Part c : Solution

- Alternate Distribution of X_bar:

  $$X\_bar \sim N(49{,}500\ ,\ \ 5000^2 / 100)$$

- Power = 0.80 => P(Z < z) = 0.80 => z = 0.84
- Critical value:

  0.84 * 5000/10 + 49500 = 49, 920

- Null distribution of X_bar:

  $$X\_bar \sim N(50,\ 000\ ,\ 5000^2 / 100)$$

- Z = (49920 – 50000)/ 500 = -0.16
- Significance level = P(Z < -0.16) = 0.4364.

Problem 6 – Part d

A tire company claims that the lifetimes of its tires average 50,000 miles. The standard deviation of tire lifetimes is known to be 5000 miles. You sample 100 tires and will test the hypothesis that the mean tire lifetime is at least 50,000 miles against the alternative that it is less. Assume, in fact, that the true mean lifetime is 49,500 miles.

d. It is decided to reject H 0 if the sample mean is less than 49,400. Find the level and power of this test.

Problem 6 – Part d : Solution

- **Null distribution of X_bar:**

$$X\_bar \sim N(50{,}000 , 5000^2 / 100)$$

- Critical value = 49,400
- Z = (49400 – 50000)/500 = -1.2
- **Significance level = P(Z < -1.2) = 0.1151**
- **Alternate Distribution of X_bar:**

$$X\_bar \sim N(49{,}500 , 5000^2 / 100)$$

- Z = (49400 – 49500)/500 = −0.20
- **Power = P(Z < -0.20) = 0.4207**

Problem 6 – Part e

A tire company claims that the lifetimes of its tires average 50,000 miles. The standard deviation of tire lifetimes is known to be 5000 miles. You sample 100 tires and will test the hypothesis that the mean tire lifetime is at least 50,000 miles against the alternative that it is less. Assume, in fact, that the true mean lifetime is 49,500 miles.

e. You are given the opportunity to sample more tires. How many tires should be sampled in total so that the power is 0.80 if the test is made at the 5% level?

Problem 6 – Part e : Solution

H0 : μ ≥ 50,000 vs. H1 : μ < 50,000

- Null distribution of X_bar:

  $$X\_bar \sim N(50,000 \;,\; 5000^2 / n)$$

- α = 0.05
  - Since its a left tailed test, z = -1.645
  - Critical value: -1.645 * 5000/sqrt(n) + 50,000
- Alternate Distribution of X_bar:

  $$X\_bar \sim N(49,500 \;,\; 5000^2 / n)$$

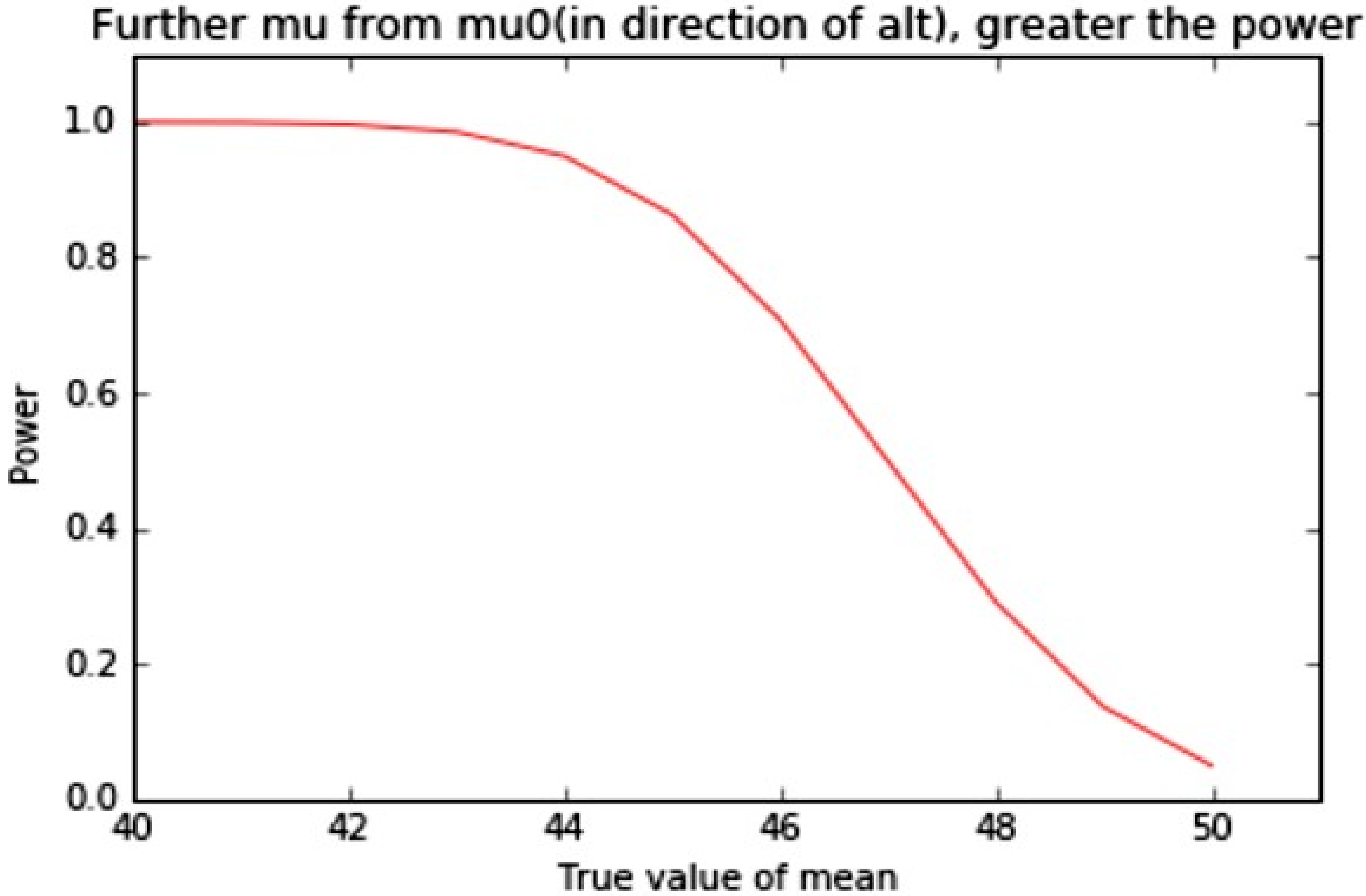
  - Power = P(Z < z) = 0.80 => z = 0.84
  - Critical value = 0.84 * 5000/sqrt(n) + 49500
- Setting the critical values equal:

  -1.645 * 5000/sqrt(n) + 50,000 = 0.84 * 5000/sqrt(n) + 49500

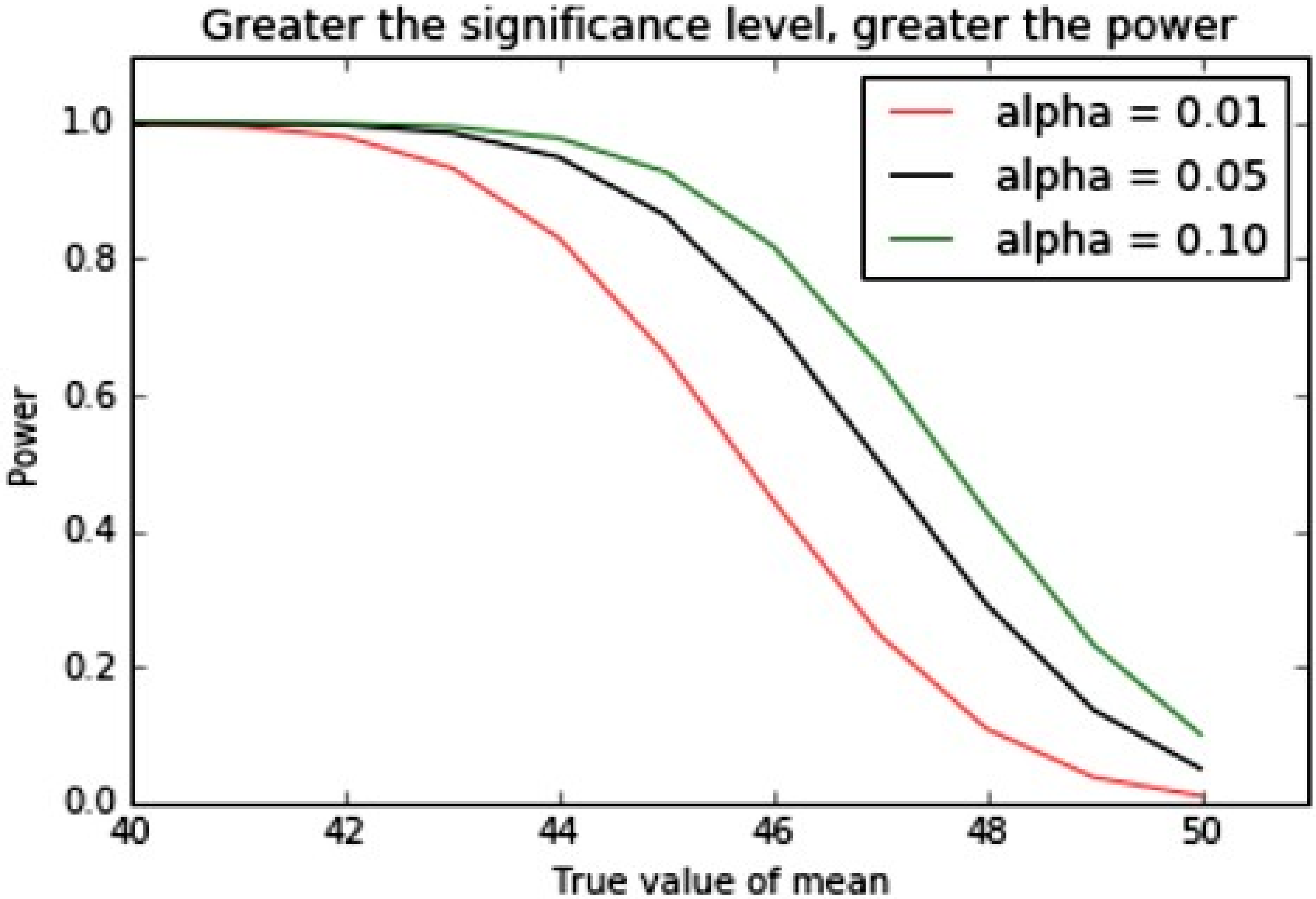
  => n = 618

Factors affecting Power of a test

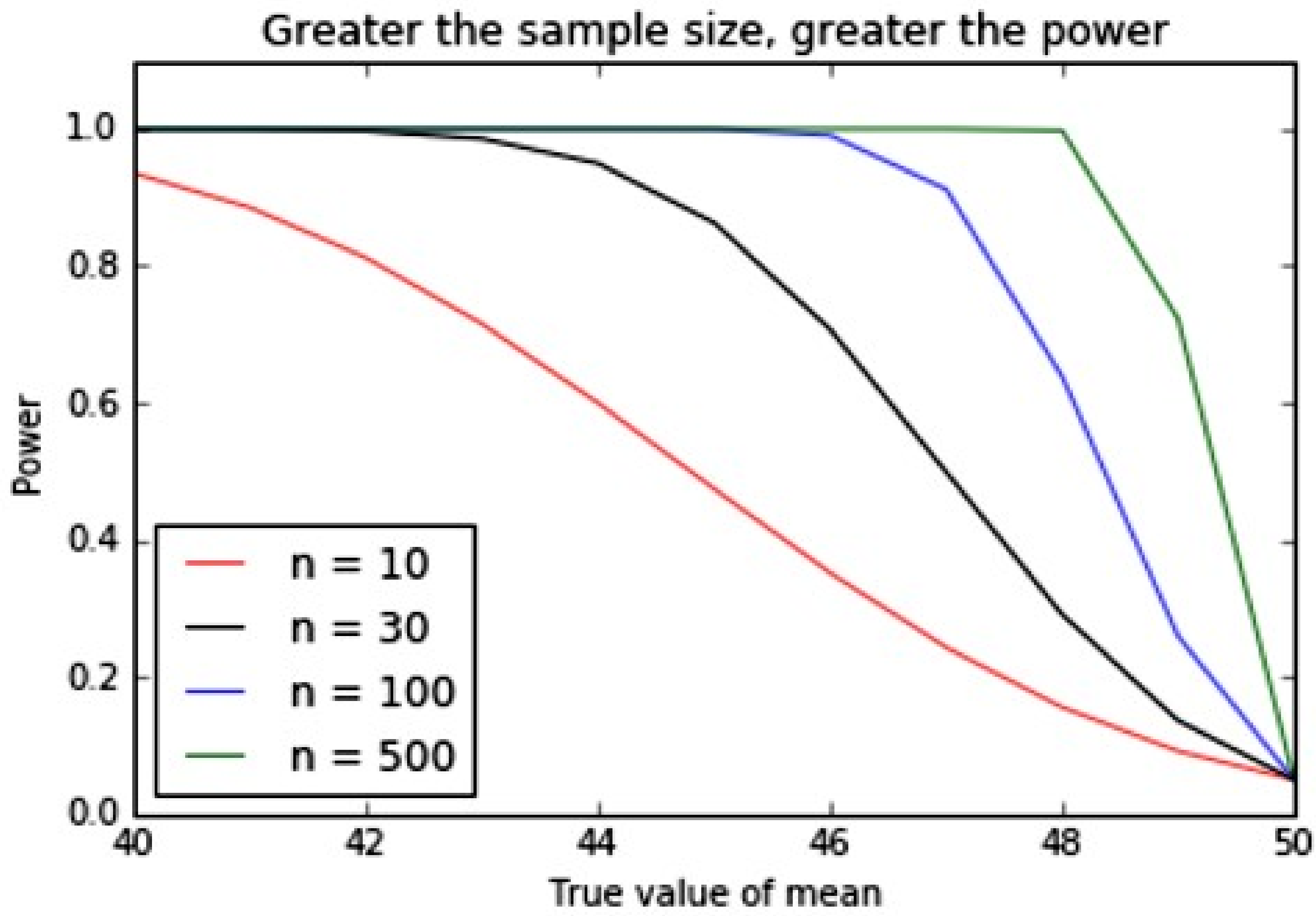
H0 : μ ≥ 50, H1: μ < 50, α = 0.05, n =30, σ = 10



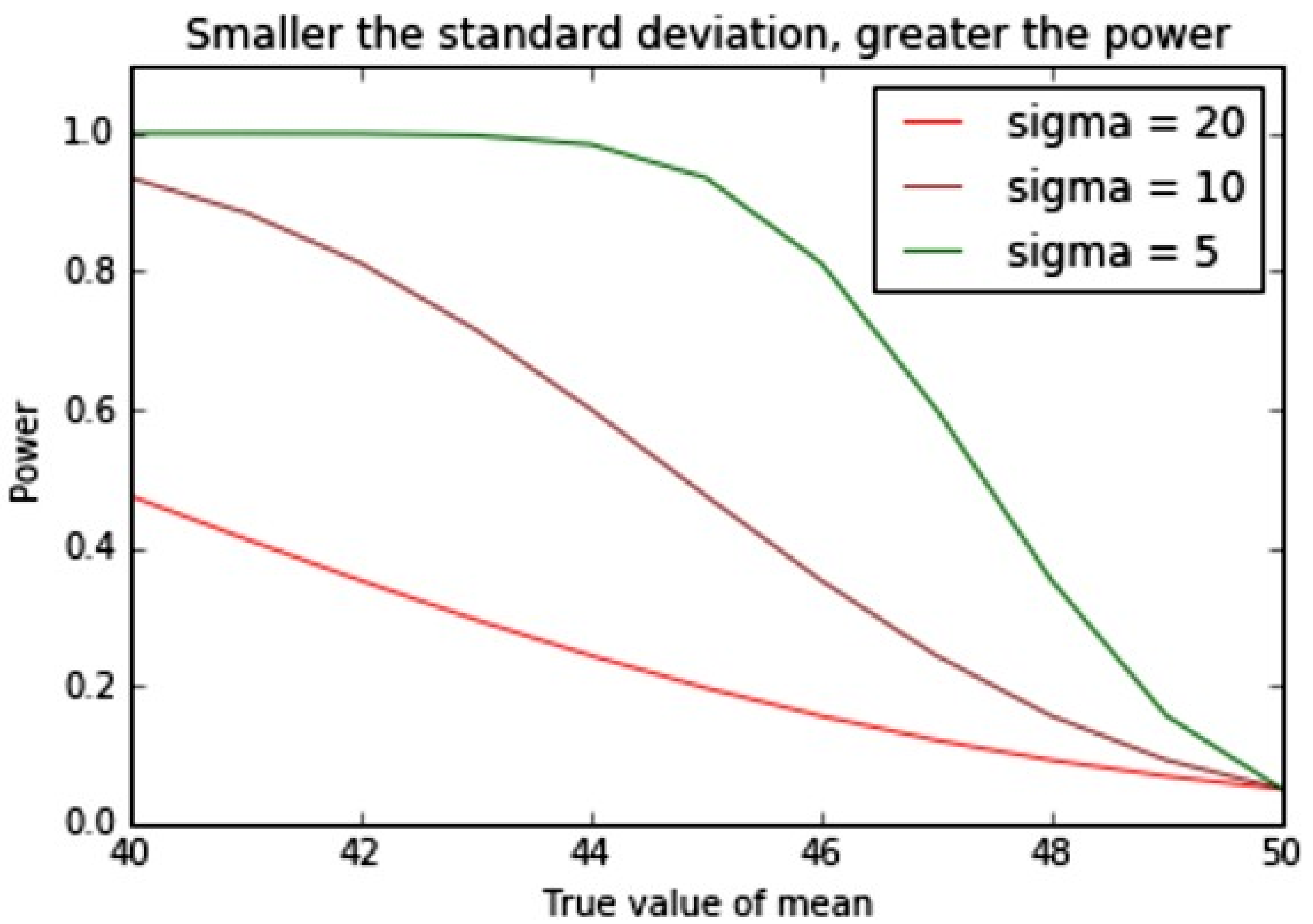
H0 : μ ≥ 50, H1: μ < 50, n =30, σ = 10



H0 : $\mu \geq 50$, H1: $\mu < 50$, $\alpha = 0.05$, $\sigma = 10$



H0 : μ ≥ 50, H1: μ < 50, α = 0.05, n =30



Problem 7

A power calculation has shown that if $\mu = 8$, the power of a test of $H_0 : \mu \geq 10$ versus $H_1 : \mu < 10$ is 0.90. If instead $\mu = 7$, **which one of the following statements is true?**

i. The power of the test will be less than 0.90.

ii. The power of the test will be greater than 0.90.

iii. We cannot determine the power of the test without knowing the population standard deviation σ .

Problem 7 : Solution

(ii). Since 7 is farther from the null mean of 10 than 8 is, the power against the alternative $\mu = 7$ will be greater than the power against the alternative $\mu = 8$.

Summary

Power increases as:

- α increases
- n Increases
- σ decreases
- The true value of μ gets further from μ_0(in the direction of alternative hypothesis)

Chi-square Distribution

Chi-square Distribution

- Continuous distibution
- If Z ~ N(0 , 1) then,

$$\chi 2 \; = Z^2$$

  with degree of freedom = 1

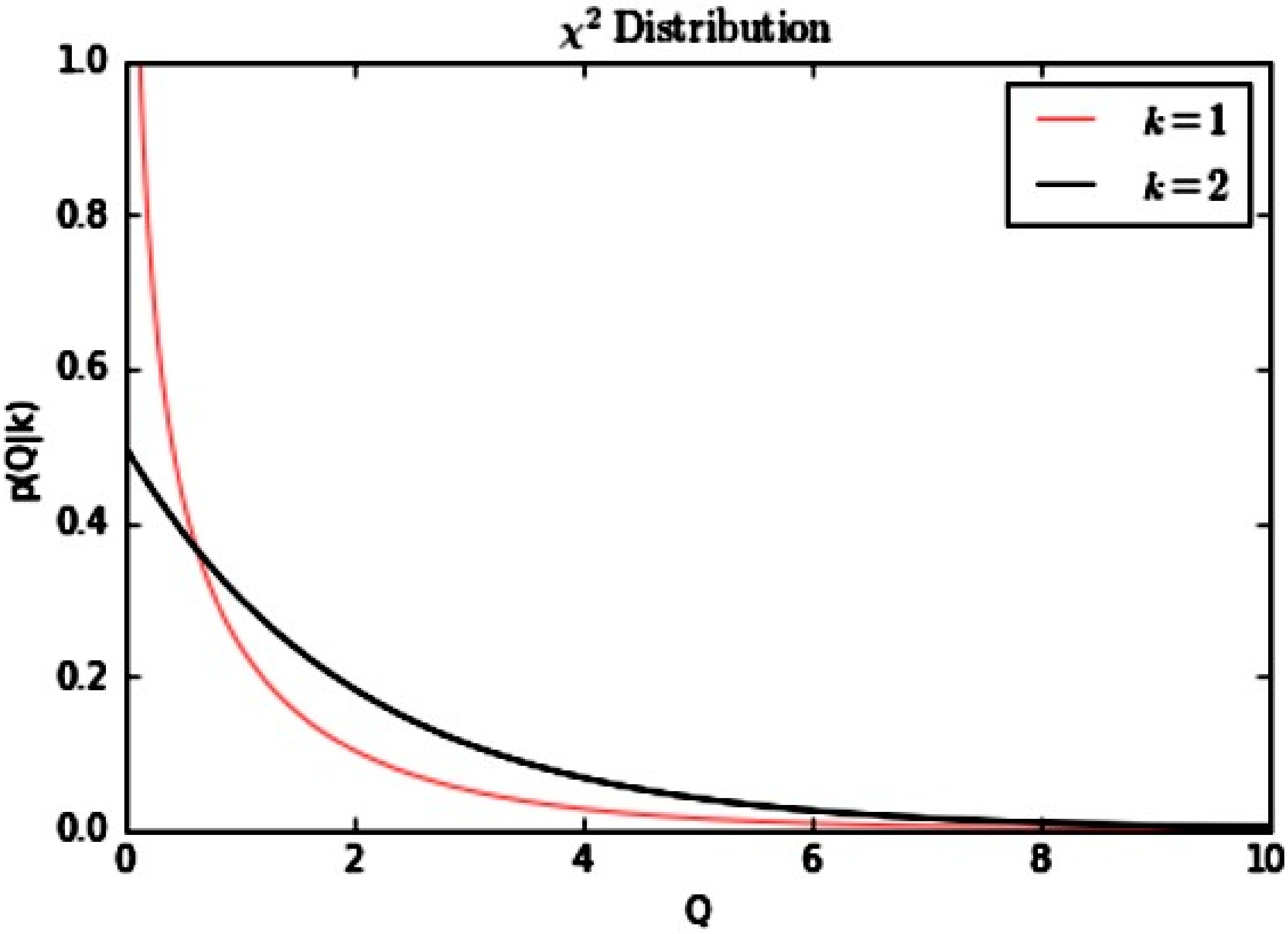
- If Z_1, Z_2, ... Z_k are standard Normal random variables then:

$$\chi 2 \; = (Z_1)^2 + \; (Z_2)^2 \; + ... + (Z_k)^2$$

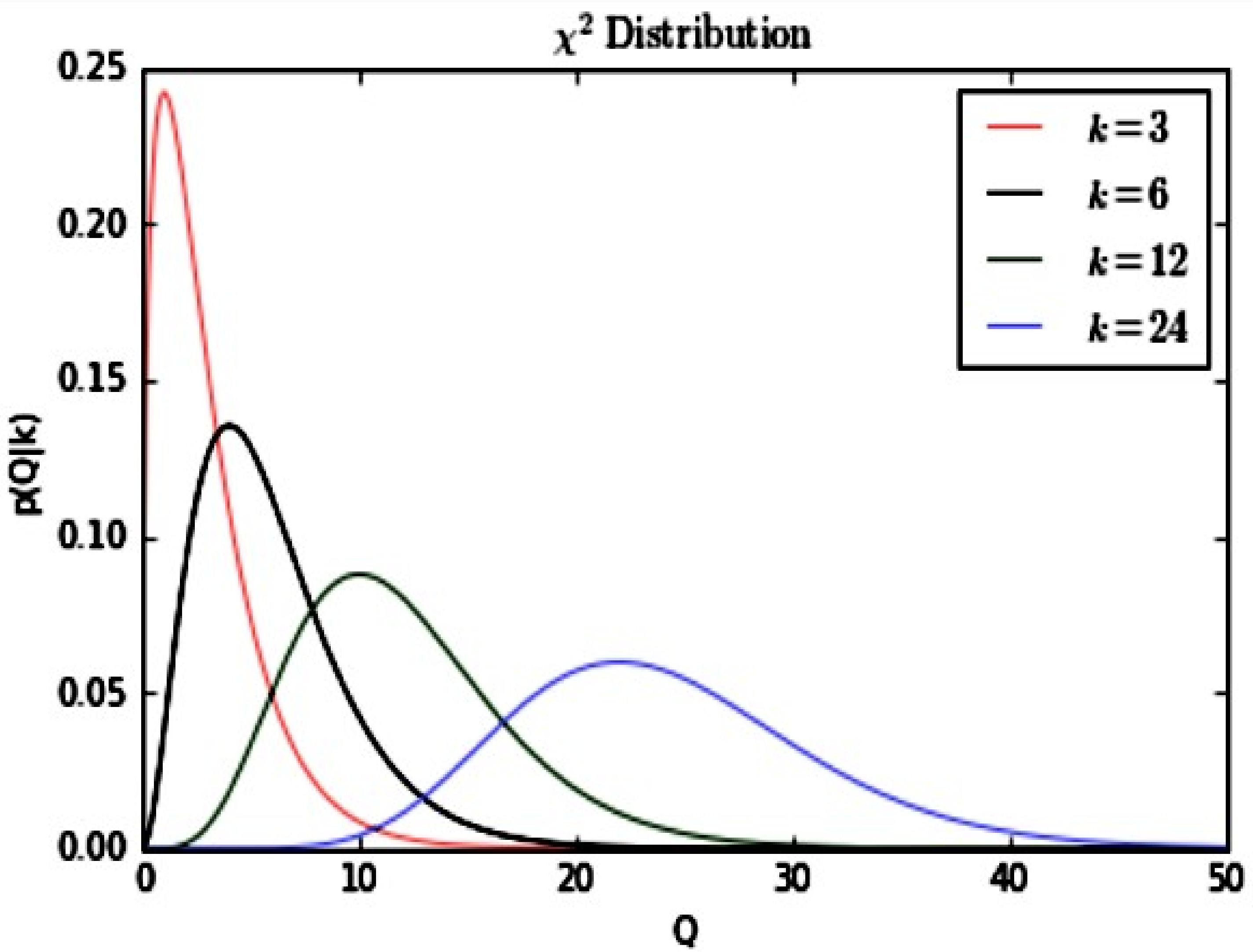
  is a chi square distribution with k degrees of freedom

- Mean = k
- Variance = k^2
- Chi square curve with k degrees of freedom reaches a peak at k – 2.

Pdf Plotted using scipy.stats.chi2



Pdf Plotted using scipy.stats.chi2



Chi-square Test

Chi-square Test

- The technique to analyze a discrete outcome uses what is called a chi-square test.
- Chi-squared test is **used to assess the following types of comparison**:
  - A **test of goodness of fit** establishes whether an observed frequency distribution differs from a theoretical distribution.
  - A **test of homogeneity** used to determine whether frequency counts are distributed identically across different populations.
  - A **test of independence** used to determine whether the two categorical variables are independent of each other.

Conditions to be met

The test procedure described in this lesson is appropriate when the following conditions are met:

- The sampling method is simple random sampling.
- The variables under study are (each) categorical.
- The expected frequency count for each cell of the table is at least 5.

Procedure of Chi-squared Test

- State appropriate null and alternate Hypothesis.
- Calculate the chi-squared test statistic, χ^2 .
- Determine the degrees of freedom, df, of that statistic.
- Compute P-value using Chi-square distribution using χ^2 and df value.
- If P-value <= 0.05 , reject Null Hypothesis.

Chi-Square Goodness of Fit Test

Chi-Square Goodness of Fit Test

- The test is applied when you have **one categorical variable** from a single population.
- It is used to determine whether sample data are consistent with a hypothesized distribution.
- As the name indicates, the idea is to assess whether the distribution of responses in the sample "fits" a specified population distribution.
- For a chi-square goodness of fit test, the hypotheses take the following form:
  - H0: The data are consistent with a specified distribution.
  - Ha: The data are not consistent with a specified distribution.

Problem 1

A 6-sided dice is thrown 60 times.

- The number of times it lands with 1, 2, 3, 4, 5 and 6 face up is 5, 8, 9, 8, 10 and 20, respectively.
- Is the dice biased, according to the Pearson's chi-squared test at a level of:
  - 95%, and
  - 99%?

Problem 1 : Solution

Category	Observed	Expected(N * p)
1	5	10
2	8	10
3	9	10
4	8	10
5	10	10
6	20	10
Total	60	60

Chi-square Statistic χ 2

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(5-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(9-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(10-10)^2}{10} + \frac{(20-10)^2}{10}$$

= 2.5 + 0.4 + 0.1 + 0.4 + 0 + 10

= 13.4

- The number of degrees of freedom is n − 1 = 5

P-value and Conclusion

Degrees of freedom	Probability less than the critical value				
	0.10	0.05	0.025	0.01	0.001
5	9.236	11.070	12.833	15.086	20.515
α	0.90	0.95	0.975	0.99	0.999

- As the chi-squared statistic of 13.4 exceeds this critical value, we reject the null hypothesis and conclude that the die is biased at 95% significance level.

- At 99% significance level, the critical value is 15.086. As the chi-squared statistic does not exceed it, we fail to reject the null hypothesis and thus conclude that there is insufficient evidence to show that the die is biased at 99% significance level.

Problem 2

Acme Toy Company prints baseball cards. The company claims that 30% of the cards are rookies, 60% veterans, and 10% are All-Stars.

Suppose a random sample of 100 cards has 50 rookies, 45 veterans, and 5 All-Stars. Is this consistent with Acme's claim? Use a 0.05 level of significance.

Problem 2 : Solution

Category	Observed	Expected(N * p)
Rookies	50	30
Veterans	45	60
All - Stars	5	10
Total	100	100

Chi-square Statistic χ 2

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(50 - 30)^2}{30} + \frac{(45 - 60)^2}{60} + \frac{(5 - 10)^2}{10}$$

$= 13.33 + 3.75 + 2.50$

$= 19.58$

- The number of degrees of freedom is $n - 1 = 2$

P-value and Conclusion

- P(X2 > 19.58) < 0.005
- Since the P-value is less than the significance level (0.05), we can reject the null hypothesis.

Chi-Square Test of Homogeneity

Chi-Square Test of Homogeneity

- The test is applied to a single categorical variable from two different populations.
- It is used to determine whether frequency counts are distributed identically across different populations.
- For example:
  - In a survey of TV viewing preferences, we might ask respondents to identify their favorite program.
  - We might ask the same question of two different populations, such as males and females.
  - We could use a chi-square test for homogeneity to determine whether male viewing preferences differed significantly from female viewing preferences.

Null and Alternate Hypothesis

- Suppose that data were sampled from r populations, and assume that the categorical variable had c levels. At any specified level of the categorical variable, the null hypothesis states that each population has the same proportion of observations. Thus,

$$H_0: P_{\text{level 1 of population 1}} = P_{\text{level 1 of population 2}} = \ldots = P_{\text{level 1 of population r}}$$

$$H_0: P_{\text{level 2 of population 1}} = P_{\text{level 2 of population 2}} = \ldots = P_{\text{level 2 of population r}}$$

$$\ldots$$

$$H_0: P_{\text{level c of population 1}} = P_{\text{level c of population 2}} = \ldots = P_{\text{level c of population r}}$$

- The alternative hypothesis (Ha) is that at least one of the null hypothesis statements is false.

Problem 1

- In a study of the television viewing habits of children, a developmental psychologist selects a random sample of 300 first graders - 100 boys and 200 girls. Each child is asked which of the following TV programs they like best: The Lone Ranger, Sesame Street, or The Simpsons. Results are shown in the **contingency table** below.

	Viewing Preferences			Row total
	Lone Ranger	Sesame Street	The Simpsons	
Boys	50	30	20	100
Girls	50	80	70	200
Column total	100	110	90	300

- Do the boys' preferences for these TV programs differ significantly from the girls' preferences? Use a 0.05 level of significance.

Stating the hypotheses.

- **Null hypothesis:** The null hypothesis states that the proportion of boys who prefer the Lone Ranger is identical to the proportion of girls. Similarly, for the other programs. Thus,

$$H_0: P_{\text{boys who prefer Lone Ranger}} = P_{\text{girls who prefer Lone Ranger}}$$
$$H_0: P_{\text{boys who prefer Sesame Street}} = P_{\text{girls who prefer Sesame Street}}$$
$$H_0: P_{\text{boys who prefer The Simpsons}} = P_{\text{girls who prefer The Simpsons}}$$

- **Alternative hypothesis:** At least one of the null hypothesis statements is false.

Expected Frequency Counts

$E_{ij} = E_{r,c} = (n_r * n_c) / n$

- here, n_r – row total and n_c - column total
- n is the over all total

Cell	n_r	n_c	n	$E_{r,c} = (n_r * n_c) / n$	$E_{1,1}$
1, 1	100	100	300	(100 * 100) / 300	33.3
1, 2	100	110	300	(100 * 110) / 300	36.7
1, 3	100	90	300	(100 * 90) / 300	30.0
2, 1	200	100	300	(200 * 100) / 300	66.7
2, 2	200	110	300	(200 * 110) / 300	73.3
2, 3	200	90	300	(200 * 90) / 300	60.0

Note:

Both row totals and column totals in the observed table must be the same as the row and column totals, respectively, in the expected table.

Chi-square Statistic χ 2

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Cell	$O_{i,j}$	$E_{i,j}$	$(O_{i,j} – E_{i,j})^2 / E_{i,j}$	$(O_{i,j} – E_{i,j})^2 / E_{i,j}$
1, 1	50	33.3	$(50 – 33.3)^2$ / 33.3	8.38
1, 2	30	36.7	$(30 – 36.7)^2$ / 36.7	1.22
1, 3	20	30.0	$(20 – 30)^2$ / 30	3.33
2, 1	50	66.7	$(50 – 66.7)^2$ / 66.7	4.18
2, 2	80	73.3	$(80 – 73.3)^2$ / 73.3	0.61
2, 3	70	60.0	$(70 – 60)^2$ / 60	1.67
			Total	19.39

P-value and Conclusion

- DF = (r - 1) * (c - 1) = (2 - 1) * (3 - 1) = 2
- The P-value is the probability that a chi-square statistic having 2 degrees of freedom is more extreme than 19.39.
- P(X2 > 19.39) < 0.005
- **Conclusion** :Since the P-value (0.0000) is less than the significance level (0.05), we reject the null hypothesis.

Problem 2

- An article presents results of a survey of adults with diabetes. Each respondent was catego rized by gender and income level. The numbers in each category (calculated from percentages given in the article) are presented in the following table.

	Poor	Near Poor	Low Income	Middle Income	High Income
Men	156	77	253	513	604
Women	348	152	433	592	511

- Can you conclude that the proportions in the various income categories differ between men and women?

Contingency Table

	Poor	Near Poor	Low income	Middle income	High income	Row total
Men	156	77	253	513	604	1603
Women	348	152	433	592	511	2036
Column total	504	229	686	1105	1115	3639

Expected Frequency Counts

$$E_{ij} = E_{r,c} = (n_r * n_c) / n$$

Cell	n_r	n_c	n	$E_{r,c} = (n_r * n_c) / n$	$E_{1,1}$
1, 1	1603	504	3639	(1603 * 504) / 3639	222.01
1, 2	1603	229	3639	(1603 * 229) / 3639	100.88
1, 3	1603	686	3639	(1603 * 686) / 3639	302.19
1, 4	1603	1105	3639	(1603 * 1105) / 3639	486.76
1, 5	1603	1115	3639	(1603 * 1115) / 3639	491.16
2, 1	2036	504	3639	(2036 * 504) / 3639	281.99
2, 2	2036	229	3639	(2036 * 229) / 3639	128.12
2, 3	2036	686	3639	(2036 * 686) / 3639	383.81
2, 4	2036	1105	3639	(2036 * 1105) / 3639	618.24
2, 5	2036	1115	3639	(2036 * 1115) / 3639	623.84

Summary

		Poor	Near Poor	Low income	Middle income	High income	Row total
Men	Observed	156	77	253	513	604	1603
	Expected	**222.01**	**100.88**	**302.19**	**486.76**	**491.16**	
Women	Observed	348	152	433	592	511	2036
	Expected	**281.99**	**128.12**	**383.81**	**618.24**	**623.84**	
Column total		504	229	686	1105	1115	3639

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Chi-square Statistic χ 2

- There are (2 − 1)(5 − 1) = 4 degrees of freedom.

- $$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 108.35.$$

- From the χ 2 table, $P < 0.005$. A computer package gives $P \approx 0$.

- We can conclude that the proportions in the various income categories differ between men and women.

Chi-Square Test for Independence

Chi-Square Test for Independence

- The test is applied when you have **two categorical variables** from a single population.
- It is used to determine whether there is a significant association between the two variables.
- For a chi-square Independence test, the hypotheses take the following form:
  - H0: Variable A and Variable B are independent.
  - Ha: Variable A and Variable B are not independent.

Problem 1

- A Study of 11, 160 alcohol drinkers on university campuses revealed:

	Never consumed alcohol	Occasional alcohol consumers	Frequent alcohol consumers	Row total
Trouble with police	71	154	398	623
No trouble with police	4992	2808	2737	10537
Col. total	5063	2962	3135	11160

- Does this sample provide enough evidence for a relationship between Alcohol drinking and getting into trouble with police?

State the Hypothesis

- Null hypothesis

  H0 : Alcohol drinking and trouble with police are independent variables

- Alternate Hypothesis

  Ha : Alcohol drinking and trouble with police are not independent variables

Expected Frequency Count

Expected = (row total * col total) / overall total

		Never consumed alcohol	Occasional alcohol consumers	Frequent alcohol consumers	Row total
Trouble with police	observed	71	154	398	623
	expected	**282.2**	**165.4**	**175.0**	
No trouble with police	observed	4992	2808	2737	10537
	expected	**4780.4**	**2796.0**	**2960.0**	
Col. total		5063	2962	3135	11160

Chi-square Statistic χ 2

- There are (2 − 1)(3 − 1) = 2 degrees of freedom.
- $$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 496.6$$
- From the χ 2 table, $P < 0.0005$.
- We can conclude that Alcohol drinking and getting into trouble with police are not independent variables, however chi-square test does not provide any information about the nature of the relationship shared by these variables.

Problem 2

- Administrators in a large urban district take a random sample of 50 seventh graders and compare the pre-algebra achievement levels of those who attended pre-school and those who did not.
- If achievement is independent of attending pre-school then the proportions at each level should be equal.
- Use the counts in the frequency table to determine if there is an association between attending pre-school and pre-algebra achievement.

	Below grade level	At grade Level	Advanced
Pre-school	8	6	6
No Pre-school	6	15	9

Problem 2 : Solution

Ho: Pre-algebra achievement is independent of pre-school attendance.
Ha: There is a relationship between Pre-algebra achievement and pre-school attendance

Expected counts	Below grade level	At grade level	Advanced
Pre-school	5.6	8.4	6
No Pre-school	8.4	12.6	9

Chi-squared = 2.85 with p-value = .239, much larger than alpha = .05.

- Since our p-value is so high, our sample does not provide significant evidence that pre-algebra achievement is related to pre-school attendance.

Practice Problem

- Suppose you take a random sample of 20 students who are using a new algebra text which features group work and unit summaries and a second sample of 30 students who are using a more traditional text. You compare student achievement on the state test given to all students at the end of the course. Use the frequency table to determine if the proportions from each group are equal at each performance level.

Identify the chi-square test to be used!

	Below grade level	At grade level	Advanced
New text	8	6	6
Old text	6	15	9

Distribution-Free Tests

Introduction

- Does not require Sample to come from a normal population or any partcular population.

  (Nonparametric test)

- No restriction on the sample size.
- Can contain outliers.

Restriction:

- The sample should come from a population which is approximately symmetric(=> mean = median)

Two types of Distribution Free Tests

1) Wilcoxon Signed-Rank test

- Analogous to one sample t test for population mean.

2) Wilcoxon Rank-sum test or Mann-Whitney test

- Analogous to two sample t test (comparing two population means)

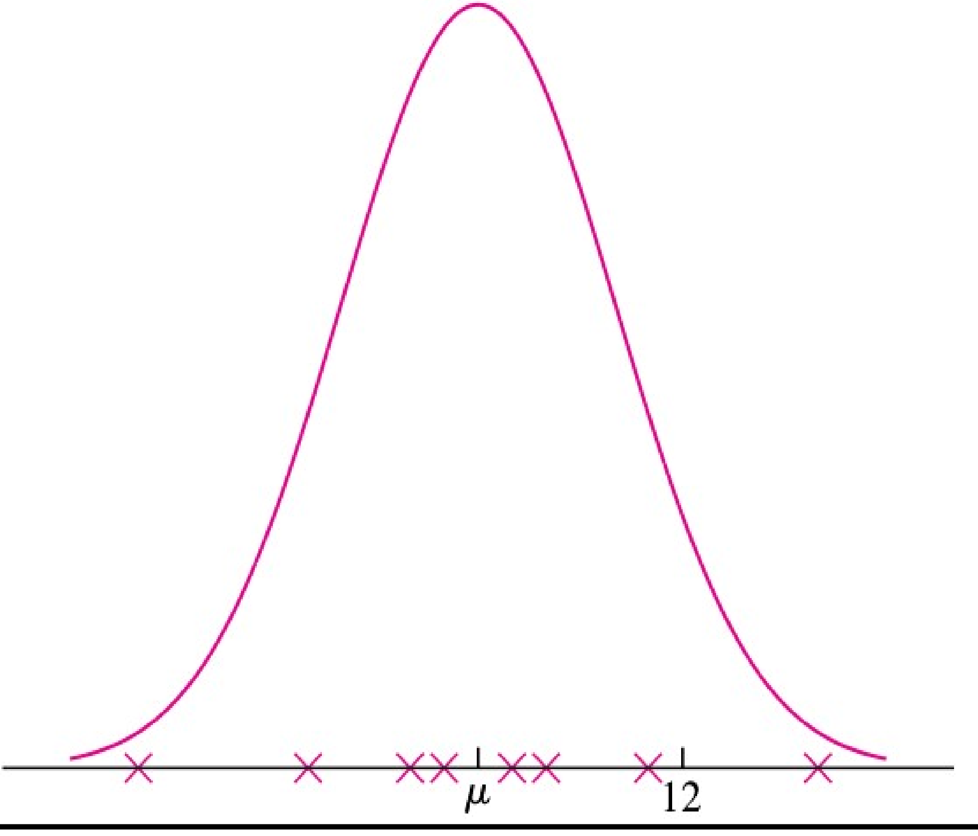
1) Wilcoxon Signed-Rank test (Small Samples)

Procedure

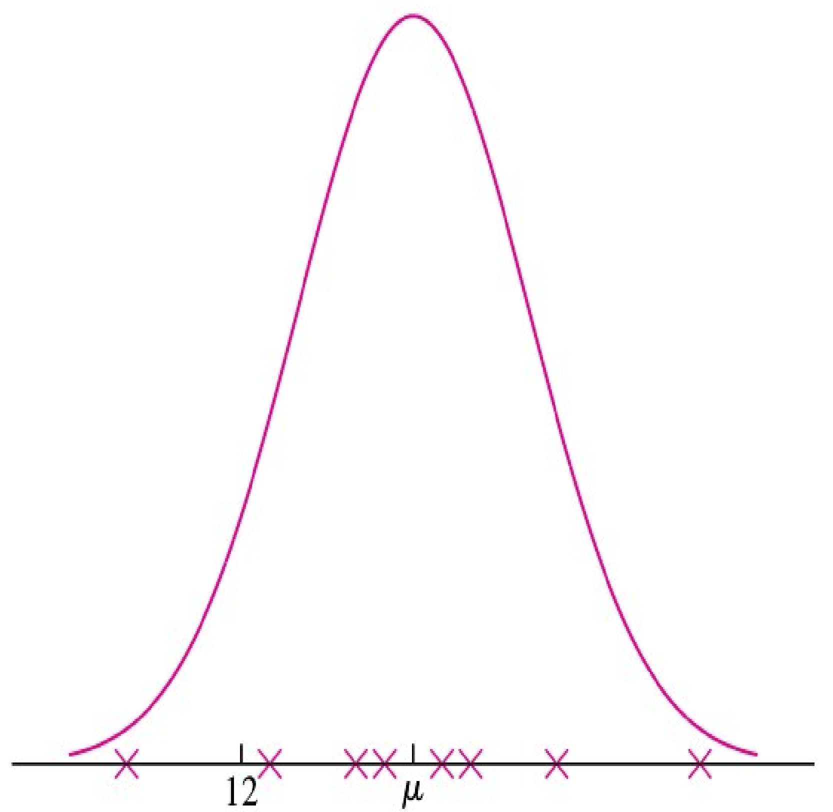
- State appropriate null and Alternate Hypothesis.
- Subtract μ_0 from each item in the sample.
- Assign ranks to the items according to closeness of the absolute values of their differences to 0(zero).
- Obtain:
  - S+ : Sum of all positive ranks (Used as the test - statistic)
  - S- : Sum of all negative ranks.
- Consult Wilcoxon Signed rank distribution to get the P-value.

  Df = sample size.
- If P-value < alpha [reject H0]

Stating a conclusion

<table>
<tr><th>Null Hypothesis</th><th>Alternate Hypothesis</th><th>S+ values</th></tr>
<tr><td>$\mu \geq \mu_0$</td><td>$\mu < \mu_0$

</td><td>Small Values of S+ provides evidence against H0</td></tr>
</table>

Stating a conclusion

Null Hypothesis	Alternate Hypothesis	S+ values
$\mu \le \mu_0$	$\mu > \mu_0$ 	Large Values of S+ provides evidence against H0

Problem 1

- The nickel content, in parts per thousand by weight, is measured for six welds.
- The results are 9.3, 0.9, 9.0, 21.7, 11.5, and 13.9.
- Let μ represent the mean nickel content for this type of weld.
- It is desired to test:

$H0 : \mu \geq 12$ versus $H1 : \mu < 12$

Problem 1 : Solution

- There are two outliers: 0.9, 21.7, hence t test cannot be used
- Wilcoxon Signed-rank test is appropriate.

x	x - 12	Signed Rank
0.9	-11.1	-6
9.0	-3.0	-4
9.3	-2.7	-3
11.5	0.5	-1
13.9	1.9	2
21.7	9.7	5

- S+ = 2 + 5 = 7, S- = 6 + 4 + 3 + 1= 14 , df = 6 (Sample Size)
- P-value : > 0.1094 [as Prob of observing a value 7 or more is 0.1094]
- Hence we cannot reject H_0

Handling Ties

- Difference : 3, 6, 6, 9
- Rank : 1, 2.5, 2.5, 4 (Avg of 2, 3)

- Difference : 3, 6, 6, 6, 9
- Rank : 1, 3, 3, 3, 5 (Avg of 2, 3, 4)

Handling values equal to μ_0

- No rank is assigned to that item (denoted by -)
- Sample size is reduced by the number of items which equal μ_0 .

Problem 2

- An article presents measurements of total solids, in g/L, for seven sludge specimens.
- The results (rounded to the nearest gram) are 20, 5, 25, 43, 24, 21, and 32.
- Assume the distribution of total solids is approximately symmetric.

An environmental engineer claims that the mean concentration of total solids is equal to 18 g/L. Can you conclude that the claim is false?

Problem 2 : Solution

- $H_0 : \mu = 18$ versus $H_1 : \mu \neq 18$
- The sum of the positive signed ranks is $S_+ = 23$.
- Df = 7
- P-value = P(Slow <= 5) + P(Sup >= 23)

  = 0.0781 + 0.0781
- Hence we cannot reject H0, hence cannot conclude that mean concentration differs from 18 g/L.

x	$x - 18$	Signed Rank
20	2	1
21	3	2
24	6	3
25	7	4
5	−13	−5
32	14	6
43	25	7

See WilcoxonSignedRankTest Document for details

1) Wilcoxon Signed-Rank test (Large Samples)

Introduction

- Whenever n (sample size) > 20, the test statistic S+ is approximately normally distributed as:

$$S+ \sim N(n(n+1)/4)\ ,\ n(n+1)(2n+1)/24)$$

- Hence Wilcoxon Signed-rank test can be performed by computing z-score for S+.

2) Wilcoxon Rank-sum test
or
Mann-Whitney test
(for small samples)

Introduction

- The Wilcoxon sign test assumes independence, meaning that the paired observations are randomly and independently drawn.
- The populations must be continuous.

- It can be used as an alternative to the paired Student's t-test, t-test for matched pairs when the population cannot be assumed to be normally distributed **to assess whether their population mean ranks differ .**

Procedure

- State appropriate null and Alternate Hypothesis.
- Combine both the samples (m + n)
- assigning ranks 1, 2, . . . , m + n to them.
- The test statistic, denoted by W , is the sum of the ranks corresponding to X 1 , . . . , X m .
- Compute P-value by consulting Wilcoxon rank-sum test distribution, where df = Sample size.
- if **H1 : μ X < μ Y** , the values in the X sample will tend to be smaller than those in the Y sample, so the rank sum W will tend to be smaller as well.
- if **H1 : μ X > μ Y** , W will tend to be larger.

Problem 1

Resistances, in m , are measured for five wires of one type and six wires of another type. The results are as follows:

X : 36 28 29 20 38

Y : 34 41 35 47 49 46

Use the Wilcoxon rank-sum test to test

$H_0 : \mu_X \geq \mu_Y$ versus $H_1 : \mu_X < \mu_Y$

Problem 1 : Solution

Value	Rank	Sample
20	1	*X*
28	2	*X*
29	3	*X*
34	4	*Y*
35	5	*Y*
36	6	*X*
38	7	*X*
41	8	*Y*
46	9	*Y*
47	10	*Y*
49	11	*Y*

- $W = 1+2+3+6+7 = 19$
- Determining P-value:
- Entering the table with $m = 5$ and $n = 6$, we find that the area to the left of $W = 19$ is 0.0260.
- Since $P < 0.05$, we reject H0.

Problem 2

A woman who has moved into a new house is trying to determine which of two routes to work has the shorter average driving time. Times in minutes for six trips on route A and five trips on route B are follows:

A:	16.0	15.7	16.4	15.9	16.2	16.3
B:	17.2	16.9	16.1	19.8	16.7	

Can you conclude that the mean time is less for route A?

Problem 2 : Solution

μX – reprsents mean time for route B.

μY – reprsents mean time for route A.

$$H0 : \mu X \leq \mu Y \text{ versus } H1 : \mu X > \mu Y$$

or

$$H0 : \mu X - \mu Y \leq 0 \text{ versus } H1 : \mu X - \mu Y > 0$$

Problem 2 : Solution

Value	Rank	Sample
15.7	1	*Y*
15.9	2	*Y*
16	3	*Y*
16.1	4	*X*
16.2	5	*Y*
16.3	6	*Y*
16.4	7	*Y*
16.7	8	*X*
16.9	9	*X*
17.2	10	*X*
19.8	11	*X*

- W = 42 (Counting all X(route B) ranks)
- Determining P-value:
- Entering the table with m = 5 and n = 6, we find that the area to the left of W = 42 is 0.0152
- Since P < 0.05, we reject H0 and conclude that mean lifetime for route A is less.

2) Wilcoxon Rank-sum test
or
Mann-Whitney test
(for Large samples)

Introduction

- When both sample sizes m and n are greater than 8, it can be shown by advanced methods that the null distribution of the test statistic W is approximately normal:

$$W \sim N\left(m(m+n+1)/2 \; , \; mn(m+n+1)/12 \right)$$

- Test is performed by computing the z-score of W , and then using the normal table to find the P-value.

Thank you !