University of Alberta

STAT 151

Introduction to Applied Statistics I

Fall 2017

Final Exam

Exam Guide

Part 1 of 2

Table of Contents:

Chapter 16 – Confidence

This chapter covers inferential statistics, its objective is to use sample data to obtain results about the whole population.

In a first step, the goal is to describe an underlying population. Since the populations are described in form of models that are characterized by parameters (mean μ and standard deviation σ or probability p for the event of interest). At this time we will estimate those characteristics. There are two different approaches for estimating: Point Estimation and Interval Estimation. For Point Estimation, you give one value for a characteristic, which is hopefully close to the true unknown value. For Interval Estimation, you give an interval of likely values, where the width of the interval will depend on the confidence you require to have in this interval. Since we base our statement just on a sample, we see later how to give a measure of accuracy or confidence for the estimate.

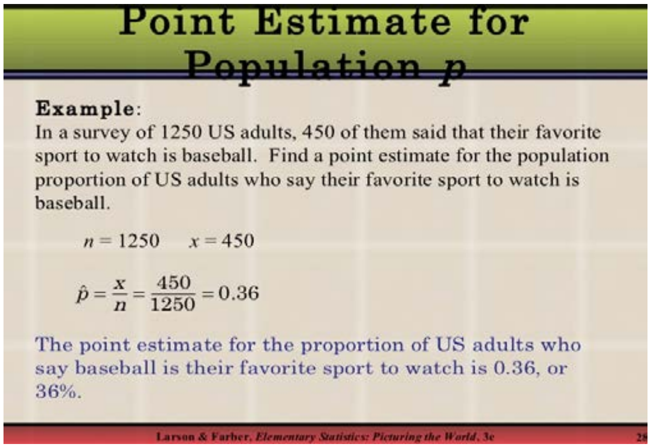
Point Estimation:

A point estimate of a possible characteristic is a single number that is based on sample data and represents the population parameter.

- The sample mean y is a point estimate for the population mean μ.
- The sample proportion p hat is a point estimate of p the population probability for Success.

A point estimate gives a single value that is supposed to be close to the true value of the characteristic but it does NOT tell how close the estimate is.

Considering we know that we would observe different values of a point estimate from sample to sample, point estimates are not enough to describe a parameter. Thus, we introduce the second type of estimate – interval estimate.



16.1 Confidence Interval
As an alternative to point estimation we can report not just a single value for the population characteristic, but an entire interval of reasonable values based on sample data. These intervals take into account of error and uncertainty. We often associate interval estimate with some level of confidence and the result is called a confidence interval.

When we don't know p or σ, we're stuck, right?
No, we will use sample statistics to estimate these population parameters. Whenever we estimate the standard deviation of a sampling distribution, we call it a standard error. The standard error for a sample proportion:

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

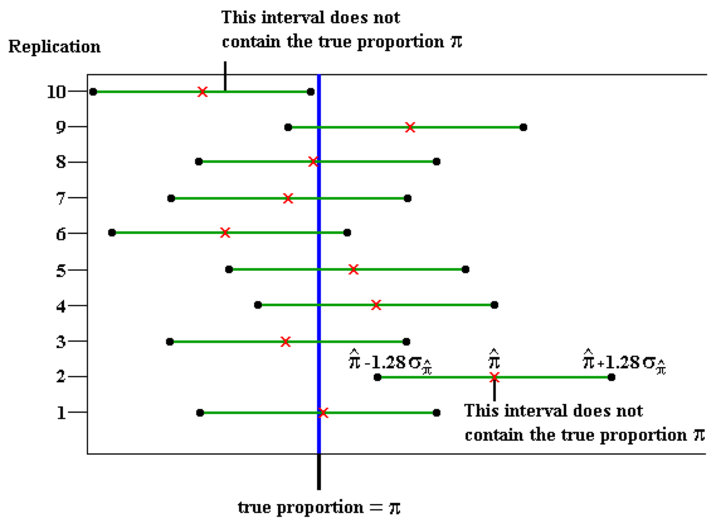
The standard error for the sample mean:

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

By the 68-95-99.7% Rule, when sampling distribution p hat is approximately Normal, we know
- about 68% of all samples will have p hat's within 1 SE of p
- about 95% of all samples will have p hats 's within 2 SEs of p
- about 99.7% of all samples will have p hats 's within 3 SEs of p

16.2 What does "95% Confidence" really mean?
Being "95% confident" means, if you were to construct 100 95% confidence intervals from 100 different samples. Of the 100 intervals, you expect 95 to capture the true parameter, and 5 not to capture the parameter. How can this happen?

- Each confidence interval uses a sample statistic to estimate a population parameter, but since samples vary, the statistics we use, and thus the confidence intervals we construct, vary as well.
- In conclusion, you cannot be sure that a specific confidence interval captures the true proportion p. Our confidence is in the process of constructing the interval, not in any one interval itself.



16.3 Margin of Error: Certainty vs. Precision

Most confidence intervals are of the form: Point estimate ± margin of error = point estimate ± critical value × SE(estimate)

- The more confident we want to be, the larger our margin of error needs to be (makes the interval wider).
- We need more values in our confidence interval to be more certain. - Because of this, every confidence interval is a balance between certainty and precision.
- The tension between certainty and precision is always there.
- Fortunately, in most cases we can be both sufficiently certain and sufficiently precise to make useful statements.
- The most commonly chosen confidence levels (C) are 90%, 95%, and 99% (but any percentage can be used).

Critical Values:

- The critical value is how far we need to deviate from the estimate to capture the central 100C% of the values on the sample distribution.
- The ‘2’ in p hat 2SE(p hat) (our 95% confidence interval) came from the 68-95-99.7% Rule.
- Using a table or technology, we find that a more exact value for our 95% confidence interval is 1.96 instead of 2.
- We call 1.96 the critical value and denote it z*

Example: To find the central 95% region on a standard normal curve, you need to cut off 2.5% at each end. The z* value for C = 0.95 has 97.5% of the area to the left. Using z-table, we find z* = 1.96

A confidence interval for small samples:

- All is not lost if the success/failure condition doesn't work
- Adjusting the calculation lets us create a confidence interval
- We need to add four observations, two successes and two failures

So instead of p̂=y/n we use the adjusted proportion p~ = (y + 2)/(y + 4)

*11.5 A Confidence Interval for Small Samples

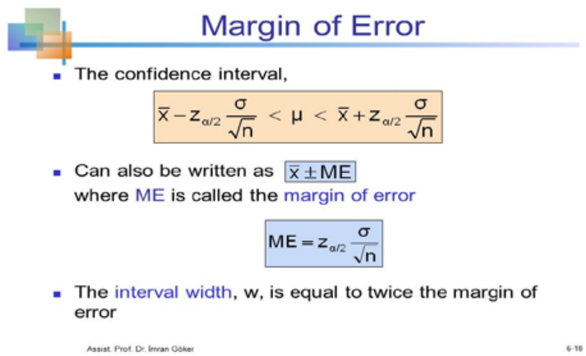
Including the synthetic observations leads to a new adjusted interval.

$$\boxed{\tilde{p} \pm z^* \sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}}$$

This form gives better performance for proportions near zero or one. It also has the advantage that we do not need to check the Success/Failure condition.

- This adjusted form gives us much better performance overall and works for proportions of 0 or 1

Margin of error:



We may like to choose the sample size n to achieve a certain margin of error, so we solve for n:
N = $(z^*/ME)^2\ \hat{p}\ (1-\hat{p})$

use $\hat{p}$ = 0.50. This is conservative as it gives a margin of error bigger than the true margin of error.

Example: If a TV executive would like to find a 95% confidence interval estimate within 0.03 for the proportion of all households that watch NYPD Blue regularly. How large a sample is needed if a prior estimate for $\hat{p}$ was 0.15?

$\hat{P}$ = 0.50
95% = C = z^* = 1.96
0.03 = ME
n≥ $(z^*/ME)^2\ \hat{P}(1-\hat{P}) = (1.96/0.03)^2\ 0.15(1-0.15) = 544.2$
N = 545, we can't have 0.2 so we round up from 544.2 to 545

Example (Revisited):
Suppose a TV executive would like to find a 95% confidence interval estimate within 0.03 for the proportion of all households that watch NYPD Blue regularly. How large a sample is needed if we have no reasonable prior estimate for $\hat{P}$?

When there's no reasonable estimate for $\hat{P}$, $\hat{P}$ = 0.50
n≥ $(z^*/ME)^2\ \hat{P}(1-\hat{P}) = (1.96/0.03)^2\ 0.5(1-0.5) = 1067.1$
N = 1068

Chapter 17: Testing Hypotheses about Proportions

Previously, population parameters were described, now we will be checking if claims about the population parameters are true, or plausible to a given degree.

Example:
Suppose p is the probability to get a HEAD when flipping a coin. Eric claims that the coin is fair, but Amy wants to test if Eric's claim is true. She took a random sample of size 1000 showed 400 HEADs.

$\hat{P}$ = (400/1000) = 0.4

Based on this result, can we state that:
i) the proportion of heads occurring is less than 0.5 (so Eric's claim is invalid)? OR
ii) the difference between 0.5 (the population proportion of heads) and 0.4 (the sample proportion of heads) may have occurred because of sampling error?

To answer these questions we have to do a hypotheses test:
A hypothesis test is a method for using sample statistics to decide between two competing claims on hypotheses about a population parameter. It follows the following procedure:
1) Define the variable, the parameter(s) of interest, and any relevant assumptions.
2) State the null hypothesis H0 and alternative hypothesis Ha
3) Gather the evidence (sample). Based on the data in the sample, we will calculate a test statistic.
4) Assess the strength of the evidence against the null hypothesis in favor of the alternative. This will be done by finding p-value.
5) Make a decision based on Step 4.
6) State the conclusion.

The null hypothesis H0 is a claim about a population parameter that is assumed to be true until it is declared false. It is generally the hypothesis of "no effect."

- We usually write down the null hypothesis in the form H0: parameter = hypothesized value.

The alternative hypothesis Ha is a claim about a population parameter that will be true ONLY when we reject the null hypothesis. In other words, this is the hypothesis that we are trying to find evidence for.

Common choices of hypotheses are:

Two-tailed Test:

- H0: population characteristic = specific value versus
- Ha: population characteristic ≠ specific value

Upper-tailed Test:

- H0: population characteristic = specific value versus
- Ha: population characteristic > specific value

Lower-tailed Test:

- H0: population characteristic = specific value versus
- Ha: population characteristic < specific value

Examples:

H0: $\mu = 100$ versus Ha: $\mu < 100$

H0: $p = 0.25$ versus Ha: $p \neq 0.25$

- We cannot test H0: $p = 0.5$ versus Ha: $\mu > 0.6$
- We cannot test H0: $\bar{y} = 100$ versus Ha: $\bar{y} \neq 100$
- We cannot test H0: $\hat{P}$ 0.65 versus Ha: $\hat{P} < 0.65$

We cannot test p hat or y bar because we are not testing sample statistics, which is what those are, we are only testing for the unknown population parameter.

Example:

According to a June 2004 Gallup poll, 28% of Americans "said there have been times in the last year when they haven't been able to afford medical care." Is this proportion higher for black Americans than for all Americans? In a random sample of 801 black Americans, 38% reported that there had been times in the last year when they had not been able to afford medical care. Which type of hypothesis test would you use?

A. One-tail upper tail

B. One-tail lower tail

C. Two-tail

D. Both A and B

H0 $p = 0.28$ Ha: $P > 0.28$ so it would be one-tail upper tail

Example: A statistics professor wants to see if more than 80% of her students enjoyed taking her class. At the end of the term, she takes a random sample of students from her large class and asks, in an anonymous survey, if the students enjoyed taking her class. Which set of hypotheses should she test?

A. H0: $p < 0.80$ HA: $p > 0.80$
B. H0: $p = 0.80$ HA: $p > 0.80$
C. H0: $p > 0.80$ HA: $p = 0.80$
D. H0: $p = 0.80$ HA: $p < 0.80$

Example:
An online catalog company wants on-time delivery for 90% of the orders they ship. They have been shipping orders via UPS and FedEx but will switch to a new, cheaper delivery service (ShipFast) unless there is evidence that this service cannot meet the 90% on-time goal. As a test the company sends a random sample of orders via ShipFast, and then makes follow-up phone calls to see if these orders arrived on time. Which hypotheses should they test?
A. H0: $p < 0.90$ HA: $p > 0.90$
B. H0: $p = 0.90$ HA: $p > 0.90$
C. H0: $p > 0.90$ HA: $p = 0.90$
D. H0: $p = 0.90$ HA: $p < 0.90$

Testing H0 vs. Ha:

- H0 will be rejected only if the sample evidence strongly suggests that H0 is false.
- Otherwise H0 will not be rejected.

So there are two possible conclusions:

- reject H0 (accept Ha)
- do not reject H0 (When H0 is not being rejected, it doesn't mean strong support for H0, but lack of strong evidence for Ha.)

Note: these decisions are not symmetric, **there is NO way you can say you <u>accept H0.</u>**

Example (con't): Suppose p is the probability to get a HEAD when flipping a coin. Eric claims that the coin is fair, but Amy wants to test if Eric's claim is true. She took a random sample of size 1000 and it showed 400 HEADs. State the hypothesis. Interpret rejection and nonrejection of H0 for this example. H0: versus Ha:

Rejection of H0:
There is enough evidence to conclude the proportion of heads is not equal to 0.5 (i.e. it is not a fair coin)

Non Rejection of H0:
There is <u>not enough</u> evidence to conclude that the proportion of heads is not equal to 0.5 (i.e. it does not mean we accept the coin is fair)

Chapter 18: Alpha Levels and Hypothesis Testing

Test statistics = a number that measures how many standard deviations away the estimate in the sample is from the hypothesized value of parameter in Ho

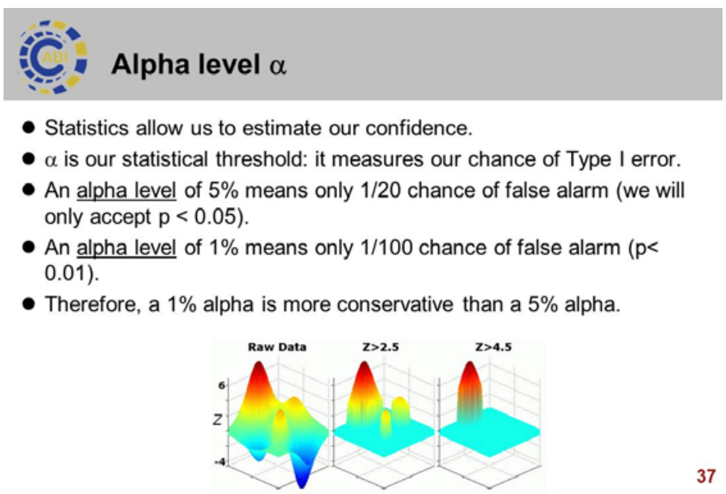
Alpha Levels

Sometimes we need to make a firm decision about whether or not to reject Ho

- When the p-value is small, it tells that our data are rare given the Ho
- How rare is "rare"?

If our P-value falls below a threshold point, we'll reject H0. We call such results statistically significant.

- The threshold is called an alpha level, denoted by a
- Common alpha levels are 0.10, 0.05, and 0.01.
- You have the option—almost the obligation—to consider your alpha level carefully and choose an appropriate one for the situation.
- The alpha level is also called the significance level.



To say a test was statistically significantly means the test statistic has a lower P-value than our alpha level.

18.5 Decision Errors

There are two types of errors we can make: Type 1 and Type 2

H0: the person is innocent

Ha: the person is guilty

Example of errors: Jury's Decision

	Did Not Commit Crime	Committed Crime
Guilty	Type I Error Convict Innocent Person	Correct Verdict Convict Guilty Person *No Error*
Not Guilty	Correct Acquittal Fail to Convict Innocent Person *No Error*	Type II Error Fail to Convict Guilty Person

	Do not reject Ho	Reject Ho
Ho is true	Correct Decision	Incorrect Decision: Type I error α
Ho is false	Incorrect Decision: Type II error β	Correct Decision

Definition:

a) type I error – the error of rejecting H0 even though H0 is true

b) type II error – the error of failing to reject H0 even though H0 is false

1) Probability of a type I error (α) is the probability of rejecting H0 even though H0 is true. α = P(H0 is rejected | Ho is true)

2) Probability of a type II error (β) is the probability of failing to reject H0 even though H0 is false. β = P(H0 is not rejected | Ho is false)

3) Power of the test ($1 - \beta$) is the probability of correctly rejecting a false H0. $1 - \beta$ = P(H0 rejected | Ho is false)

Example:

The proportion of people surviving is 30% after a specific cancer treatment. A new treatment has been proposed to be effective. To show the efficacy of the new treatment, a study has to be designed to test the following hypotheses for p (the proportion of people surviving under the new treatment).

H0: p = 30% (not effective)

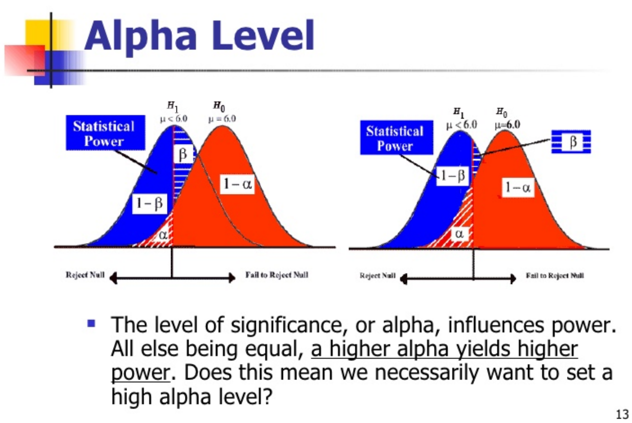
Versus:
Ha: p > 30% (effective)

Describe the type I and type II error

Type I Error (error of rejecting Ho even though it is true):
Conclude the new treatment is effective even though the treatment is not effective

Type II Error (error of failing to reject H0 even though H0 is false):
Conclude the new treatment if not effective even though it is effective
b) If the scientist wants to make sure that this new treatment is only used if it is really improving the survival rate as it may cause some side effects, then would he choose α = 0.01 or α = 0.05? He has to limit the probability for the type I error, so he chooses α = 0.01

He chooses 0.01 because he wants to plan a lower probability of a type 1 error, the error he wants to eliminate



*we choose difference α levels, because beta levels increase as alpha levels decrease so we don't always want a lower one.

a) The two types of errors that occur in the tests of hypotheses depend on each other. Lowering the value of α will increase the value of β, and vice versa.
- It makes sense that the more we're willing to accept a Type I error, the less likely we will be to make a Type II error.

b) The value of α is controlled by the experimenter.
c) The value of β is difficult, if not impossible to calculate, because we don't know what the value of the parameter really is.

d) The power increases with the effect size, where effect size = | p - p0 |

- The larger the effect size, the easier it should be to see it.
- Obtaining a larger sample size decreases the probability of a Type II error, so it increases the power.

e) The only way to reduce both types of errors is to collect more data. Otherwise, we just wind up trading off one kind of error against the other.

NOTE: increase alpha -> decrease beta -> increase the power of the test

Example:
Suppose that a manufacturer is testing one of its machines to make sure that the machine is producing more than 97% good parts (H0: p = 0.97 and HA: p >0.97). a) If the test results in a P-value of 0.102. In reality, the machine is producing 99% good parts. What probably happens as a result of our testing?
A. We correctly fail to reject H0.
B. We correctly reject H0.
C. We reject H0, making a Type I error.
D. We fail to reject H0, making a Type I error.
E. We fail to reject H0, making a Type II error
*****We choose E because it's a type 2 error, we failed to reject Ho

Example:
We conclude that it is producing more than 97% good parts when it is not. What probably happens as a result of our testing?
A.We correctly fail to reject H0
B. We correctly reject H0
C. We reject H0, making a Type I error.
D. We reject H0, making a Type II error.
E. We fail to reject H0, committing Type II error.

One Proportion z-test steps:
1) Assumption & Conditions:

- Randomization
- Independence
- The sample size n is large, that is: $np0 \geq 10$ and $n(1 – p0) \geq 10$. where p0 comes from the hypotheses.

2) Determine the type of test
a) two tailed: a. H0 : p = p0 versus Ha : p ≠ p0 35 of 51

b) lower tailed: a. H0 : p = p0 versus Ha : p < p0

c) upper tailed: a. H0 : p = p0 versus Ha : p > p0

3) Test statistic: Let p0 be a value between zero and one, and define the test statistic

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

4) P-value

Test Type:	P-value:
Upper Tail	P(z>z0)
Lower Tail	P(z < z0)
Two Tails	2 P(z > |z0|) = 2 P(z < -|z0|)

5) Decision: Reject H0, if and only if p-value ≤ α.

Example (continued): Applying those steps

Suppose p is the probability to get a HEAD when flipping a coin. Eric claims that the coin is fair, but Amy wants to test if Eric's claim is true. She took a random sample of size 1000 showed 400 HEADs.

1) H0: P = 0.5 Ha: P≠0.5
2) A: independent, sample size is large enough

N = 1000

P0 = 0.5

P hat = 400/1000 = 0.4

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

= (0.4 - 0.5)/ √((0.5x(1-0.5))/1000) = -6.32

4) P-value = 2x $P(z<-6.32)$
2x0 = 0,
Therefore there is no chance to observe this value of the sample proportion, this extreme or more extreme as a result of sampling variability alone when H0 is true.

5) P-value = 0 ≤ alpha = 0.05, reject H0

6) There is strong evidence to suggest that population proportion ≠ 0.5 at alpha 0.05

Hypothesis Test

Example:

A company claims to have 40% of the market for some product. You suspect this number, so you conduct a survey and find 38 out of 112 buyers (ie. 34%) purchased this brand. Are these data consistent with the company's claim at α=0.05 level? Let p be the true market share of this product.

1) H0 = P = 0.4

Ha = P ≠ 0.4

2) A: random (?) n = 112 Po = 0.4

np ≥10

n(1-p) ≥ 10

3) $Z0 = (\hat{p} - p0)/(\sqrt{po(1-po)/n})$

$= (38/112)-0.4/\sqrt{0.4(1-0.4)/112}) = -1.31$

4) p-value = 2 x $p(z<-1.31)$

2 x 0.0951 = 0.1902

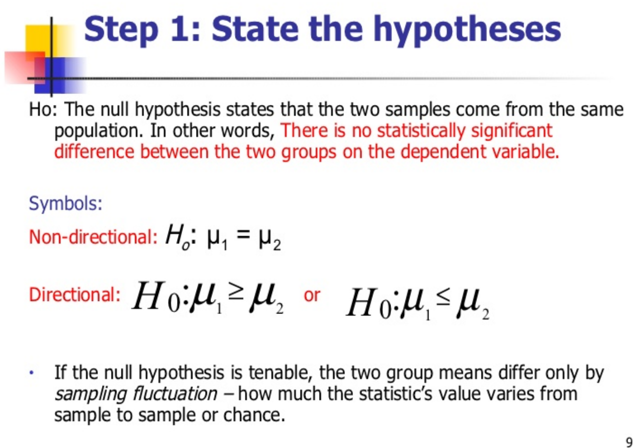
5) p-value> α = 0.05 do not reject H0

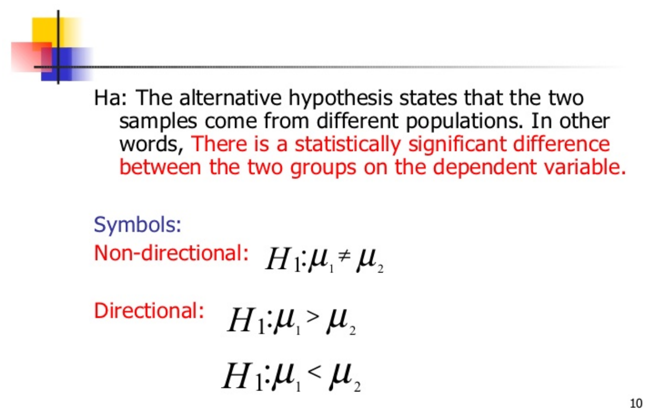
6) there is not enough evidence to conclude ≠ 0.4 to suggest the data is not consistent with the company's claim at α = 0.05

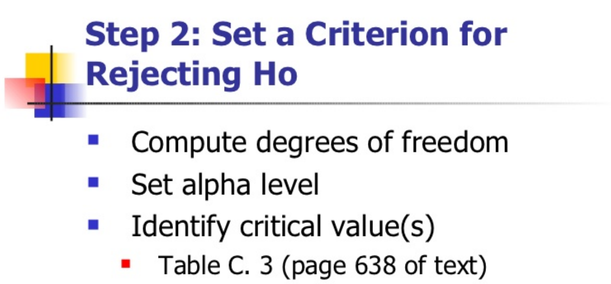
Chapter 19: Comparing Two Proportions

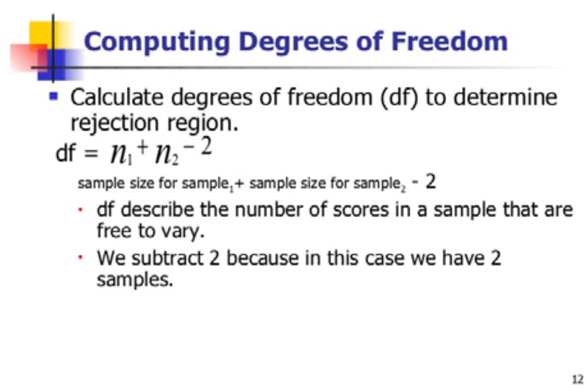
Estimating the Difference between Two Proportions You may want to compare:

- the proportion of people who play computer games in the age groups of 20 to 30 and 30 to 40.
- the proportion of defective items manufactured in two production lines The statistic for estimating the difference in two population proportions (p1 p2) that comes to mind is the difference in the sample proportions (p̂1 p̂2).









More on Degrees of Freedom

- In an Independent samples *t*-test, each sample mean places a restriction on the value of one score in the sample, hence the sample lost one degree of freedom and there are n-1 degrees of freedom for the sample.

Set alpha level

- Set at .001, .01 , .05, or .10, etc.

Identify critical value(s)

- Directional or non-directional?
- Look at page 638 Table C.3.
- To determine your CV(s) you need to know:
  - df – if df are not in the table, use the next lowest number to be conservative
  - directionality of the test
  - alpha level

Step 3: Collect data and Calculate *t* statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

variance

df

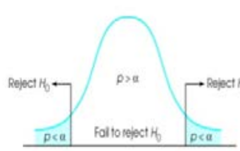
Whereby:

n: Sample size — s^2 = variance

$\bar{x}$: Sample mean — subscript$_1$ = sample 1 or group 1

subscript$_2$ = sample 2 or group 2

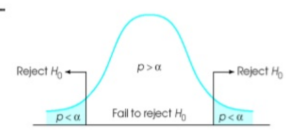
Step 4: Compare test statistic to criterion



df = 18 α = .05 , two-tailed test in this example

- critical values are ± 2.101 in this example

Step 5: Make Decision

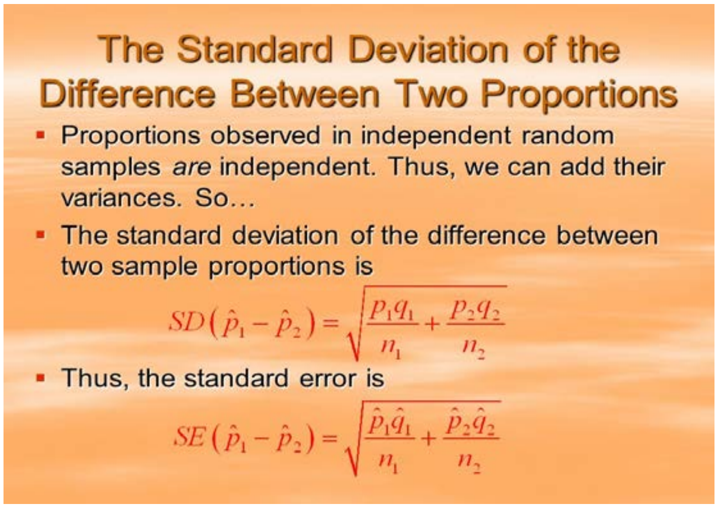


Fail to reject the null hypothesis and conclude that there is no statistically significant difference between the two groups on the dependent variable, t = , $p > \alpha$.

OR

Reject the null hypothesis and conclude that there is a statistically significant difference between the two groups on the dependent variable, t = , $p < \alpha$.

- If directional, indicate which group is higher or lower (greater, or less than, etc.).



The variance is that formula without the square root ^

For this we must have the following assumptions and conditions meet: Independence Assumptions:

Randomization Condition: The data in each group should be drawn independently and at random from a homogeneous population or generated by a randomized comparative experiment.

The 10% Condition: If the data are sampled without replacement, the sample should not exceed 10% of the population.

Independent Groups Assumption: The two groups we're comparing must be independent of each other.

Sample Size Condition: Each of the groups must be big enough

Success/Failure Condition: Both groups are big enough that at least 10 successes and at least 10 failures have been observed in each

$n1p1 \geq 10$ and $n1(1 – p1) \geq 10$ and

$n2p2 \geq 10$ and $n2(1 – p2) \geq 10$.

Example:
Suppose we want to compare therapies. The criterion for the comparison is the probability to survive at least 5 years after therapy. The study produced the following data:

	Therapy 1	Therapy 2
# of people sampled	100	80
# of people survive at least 5 years after therapy	90	70

$\hat{P}1 = 0.90$ $\hat{p}2 = 0.875$

A = random (?)

2 independent groups (believe yes)
$n1p1 \geq 10$ and $n1(1 – p1) \geq 10$ and
$n2p2 \geq 10$ and $n2(1 – p2) \geq 10$. Both work

$(\hat{P}1-\hat{P}2) \pm z \times \sqrt{}(\hat{P}1(1-\hat{P})/n) + (\hat{P}2(1-\hat{P})/n2))$
$= (0.9 - 0.875)\pm 1.96\sqrt{}(0.9(1-0.9)/100) + (0.875(1-0.875)/80)$
$= (-0.068, 0.118)$

Not enough evidence to suggest $\hat{P}1-\hat{P}2 \neq 0$

Two Proportion Z-Test
Example:
a) Construct a 95% CI for the proportion of low APGAR scores for preterm births (less than 37 weeks gestation)

$\hat{P}1 = 25/100$
$N1 = 11.8$
$C = 95\%$
$Z^* = 1.96$

$\hat{P}1 \pm z^* = \sqrt{}(\hat{P}1(1-\hat{P})/n1) = (0.138, 0.286)$

b) Construct a 95% CI for the proportion of low APGAR scores for full term births.

$\hat{P}2 = 19/482$

$N2 = 282$

$C = 95\%$

$Z^* = 1.96$

$\hat{P}2 \pm z^* = \sqrt{(}\hat{P}2(1-\hat{P}2)/n2) = (0.024, 0.075)$

c) Construct a CI for the difference in proportions of low APGAR. Interpret.

We are 95% confident that the proportion of low APGAR score for preterm births is below 8.4% to 2.4% higher than the corresponding proportion for full-term births

d) Does the interval contain zero?

No, there is difference in the low low APGAR score

Below preterm and full-term births

e) Are the results in parts c and d contradictory?

No, they aren't

<u>Two-proportion z - test</u>

1. Determine the type of test

a) lower tailed: o $H0 : p1 – p2 = p0$ versus $Ha : p1 – p2 < p0$

b) upper tailed: o $H0 : p1 – p2 = p0$ versus $Ha : p1 – p2 > p0$

c) two tailed: o $H0 : p1 – p2 = p0$ versus $Ha : p1 – p2 \neq p0$

2. Assumption & Condition:

a) The sample is randomly chosen and the sampled values are independent

b) Assume the two samples are independent of each other

c) Check the condition: Both sample sizes are large, that is:

$n1p1 \geq 10$ and $n1(1 – p1) \geq 10$

and $n2p2 \geq 10$ and $n2(1 – p2) \geq 10$.

3. Test statistics

$$Z = \frac{p_1 - p_2}{\sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$z = \frac{(p_1 - p_2) \pm \left[\frac{1}{2n_1} + \frac{1}{2n_2}\right]}{\sqrt{\hat{p}(1-\hat{p})\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

Where:

$$\hat{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

If p1 – p2 = 0, then we are hypothesizing that there is no difference between the two proportions, which implies the standard deviations for each proportion are the same. Since this is the case, we combine (or pool) the counts to get an overall

Two-Proportion z-Test

Two-proportion z-test

The conditions for the two-proportion *z*-test are the same as for the two-proportion *z*-interval. We are testing the hypothesis

$$H_0: p_1 - p_2 = 0.$$

Because we hypothesize that the proportions are equal, we pool the groups to find

$$\hat{p}_{\text{pooled}} = \frac{Success_1 + Success_2}{n_1 + n_2}$$

and use that pooled value to estimate the standard error.

$$SE_{\text{pooled}}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_{\text{pooled}}\,\hat{q}_{\text{pooled}}}{n_1} + \frac{\hat{p}_{\text{pooled}}\,\hat{q}_{\text{pooled}}}{n_2}}$$

Now we find the test statistic using the statistic,

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{SE_{\text{pooled}}(\hat{p}_1 - \hat{p}_2)}.$$

When the conditions are met and the null hypothesis is true, this statistic follows the standard Normal model, so we can use that model to obtain a P-value.

4. P-Value

Test-Type	P-value
Upper Tail	P(z > zo)
Lower Tail	P(z < zo)
Two Tails	2 P(z >|zo|)

5. Decision

- p-value ≤ α -> H0 is rejected.
- p-value > α H0 is not rejected.

Example:

Find if the proportions of red M&M's in the plain and peanut variety differ at a significance level of 0.05

The sample:

	Plain (1)	Plain (2)
Sample Size	56	64
Number of Red M&M's	12	16

1) H0: P1 - P2 = 0 (P1 = P2)

Ha = P1 - P2 ≠ 0 (p1 ≠ p2)

2) A: - random sample (?)
- 2 groups are independent
- $n1\hat{P}1 \geq 10$ & $n1(1-\hat{P}1) \geq 10$
- $n2\hat{P}2 \geq 10$ & $n2(1-\hat{P}2) \geq 10$

$$Z = \frac{p_1 - p_2}{\sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})}}$$

3)

= ((12/16)-(16/64))/ **√(**(28/120)(1-28/120)(1/56+1/64)) = -0.

4) P-value

(2 x<-0.96)

= 2 x 0.3228 = 0.6459

5) P-value > α = 0.05, do not reject Ho

6) There is not enough evidence to conclude

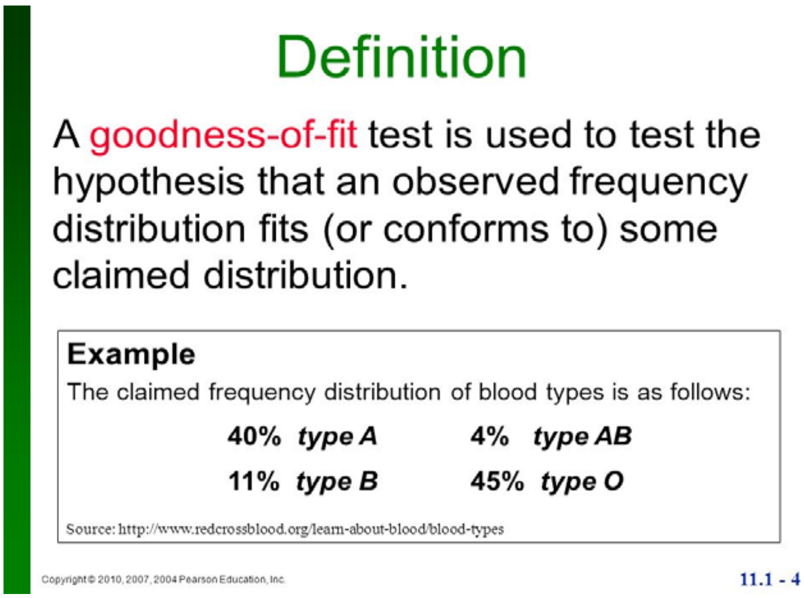
Chapter 23: Comparing Counts and Tests and Confidence Intervals

This chapter will cover 2 types of tests:

1. Tests of hypotheses for experiments with more than 2 categories, called goodness-of-fit tests.
2. Tests of hypotheses about contingency tables, called independence and homogeneity tests.

Goodness-of-Fit Test or Univariate x^2 Test

- a test of whether the frequency distribution of a categorical variable with more than 2 categories from a sample matches the probability distribution predicted by a model.



Tests

Carry out a hypothesis test and see whether the advertised label is giving a true description of the contents of the seed bag.

1) State hypothesis

H0: p01 = 0.5; p02 = 0.25; p03 = 0.15; p04 = 0.05; p05 = 0.05

Ha: at least 1 kind of seeds is not in the claiming proportions

Kinds of Seeds	Advertised Proportion	Observed Count	Expected Count
K1	0.5	480	500 (1000 x 0.5)
K2	0.25	233	250 (1000 x 0.25)
K3	0.15	160	150
K4	0.05	63	50
K5	0.05	64	50

N = 1000

2) Assumptions

Random ✓

E ≥ 5 ✓

3) Test Statistic

$\chi^2{}_O$

$(480-500)^2/500 + (233-250)^2/250 + (160-150)^2/150 + (65-50)^2/50 + (64-50)^2/50 = 9.92$

4) P-value $P(\chi^2 > 9.92)$

$\sim\chi^2$ with df = (c-1)

(5 - 1 = 4)

0.025 < P-value < 0.05

5) P-value < α = 0.05

Reject Ho

6) Conclusion: There is enough evidence to conclude there is at least one kind of seeds is not in the claiming proportions

Example:
On a bag of candies with 6 different colors, it states that (proportion column)
You randomly take one bag and started counting and find the following summarized data (number column)

Colour	Number	Proportion	E(np)
Brown	0.3	50	(167 x 0.3) = 50.1
Red	0.2	32	33.4
Yellow	0.2	20	33.4
Orange	0.1	18	16.7
Green	0.1	22	16.7
Blue	0.1	25	16.7

3) $\chi^2{}_O = \sum(O-E)^2/E$
$(50-50.1)^2/50.1 + (32-33.4)^2/33.4 + (18-16.7)^2/16.7 + (22-16.7)^2/16.7 + (25-16.7)^2/16.7 = 11.34$
4) P-Value = $P(X^2>11.34)$ ->
$0.025 <$ P-value < 0.05

5) p-value $< \alpha$ -0.05 reject Ho

6) There is enough evidence to conclude at least 1 colour of candies is not in the claiming proportions

χ^2 Test for Homogeneity and Independence in a 2-Way-Table

A Test of Homogeneity

- A test comparing the distribution of counts for two or more groups on the same categorical variable is called a chi-square test of homogeneity.
- A test of homogeneity is actually the generalization of the two proportion z-test.
- The statistic that we calculate for this test is identical to the chi square statistic for goodness-of-fit.

In this test, however, we ask whether choices have changed (i.e., there is no model). The expected counts are found directly from the data and we have different degrees of freedom

Assumptions and Conditions

The assumptions and conditions are the same as for the chi square goodness-of-fit test:

Counted Data Condition: The data must be counts.

Randomization Condition.

Expected Cell Frequency Condition: The expected count in each cell must be at least 5.

Test statistics:

To find the expected counts, we multiply the row total by the column total and divide by the grand total. We calculated the chi-square statistic as we did in the goodness-of-fit test:

Chi-square Example

$$X^2 = \frac{(O_i - E_i)^2}{E_i}$$

$$X^2 = \frac{(56-43.25)^2}{43.25} + \frac{(29-41.75)^2}{41.75} + \frac{(31-43.75)^2}{43.75} + \frac{(55-42.25)^2}{42.25} = 15.22$$

In this situation we have (R – 1)(C – 1) degrees of freedom, where R is the number of rows and C is the number of columns. We'll need the degrees of freedom to find a P-value for the chi-square statistic.

Example: Suppose we define 3 income strata: high income group (with income > $100,000), medium income group (with income of $50,000 to $100,000), and low income group (with an income of less than $50,000). Furthermore, assume that we take one sample of 250 households from California and another sample of 150 households from Wisconsin, and the collected the info on the incomes of these households are shown in the following table:

	California	Wisconsin	Total
High Income	70	34	104
Medium Income	80	40	120
Low Income	100	76	176
Total	250	150	400

Use appropriate test to test the null hypothesis that the proportions of households in California and Wisconsin who belong to various income groups are the same at the level of significance 0.025.

1) Ho: the proportions of households that belong to different income groups are the same in both states

Ha: these proportions of households that belong to different income groups are not the same in both states

2) A: random (?)

$E \geq 5$

E = (row total x column total/grand total)

(104 x 250)/400 = 65
(104 x 150)/400 = 39
(120 x 250)/400 = 75
Wisconsin x Medium income = 45
California x low income = 110
Wisconsin x Low income = 66

3)

$(70 - 65)/65^2 + (34 - 39)^2/39 + (80 - 75)^2/75 + (180 - 110)^2/110 + (40 - 45)^2/45 + (76 - 66)^2/66 =$

$4.34 \sim \underline{\chi^2}$

4) P - value ($\underline{\chi^2} > 4.34$)

5) P-value > 0.1 > α = 0.025

Do NOT reject Ho

6) There is not enough evidence to conclude the proportions of households in California and Wisconsin who belong to various income groups are not the same at α = 0.025

The χ^2 Test for Independence

- Contingency tables categorize counts on two (or more) variables so that we can see whether the distribution of counts on one variable is contingent on the other.

- Tests of independence examine counts from a single group for evidence of an association between two categorical variables.
- A chi-square test of independence uses the same calculation as a test of homogeneity; the only difference is what you think.

Assumptions and Conditions:

We still need counts and enough data so that the expected values are at least 5 in each cell. If we're interested in the independence of variables, we usually want to generalize from the data to some population.

In that case, we'll need to check that the data are a representative random sample from that population.

Example: A survey was conducted to evaluate the effectiveness of a new flu vaccine that had been administered in a small community. It consists of a two-shot sequence of two weeks. A survey of 1000 residents the following spring provided the following information:

	No Vaccine	One shot	Two shot	Total
Flu	24	9	13	46
No Flu	289	100	565	954
Total	313	109	578	1000

1) Ho: The number of flu shots and the incident of the flu are independent (always assume independent in Ho)

Ha: The number of flu shots and the incident of the flu are not independent/ associated/ related

2) a random (?)

$E \geq 5$

If Ho is true (indep) P(flu w/ no vaccine)

= (46/100) x (313/1000)

If a sample of 1000 residents, if Ho is true, what is the expected number of people getting the flu without a vaccine?

1000 x (46/1000) x (313/1000) = (46 x 313/1000)

3)

$(24-14.398)^2/14398)$ +

$(19-5.01)^2/5.01$ +

$(13-26.54)^2/26.54$ +

$(289 - 298.6)^2/298.6$ +

$(100-103.99)^2/103.99$ +

$(565 - 551.41)^2/551.41$ =

17.313

Following X^2 distribution with $\alpha f = (r - 1)(c - 1) =$

$(2 - 1)(3 - 1) = 2$

4) P-value = $P(x^2 > 17.313) < 0.005$

5) P-value < $\alpha = 0.05$ reject Ho

6) There is enough evidence to suggest the number of flu shots and incidence of flu are not independent at $\alpha = 0.05$

Chi-Square Test (χ^2)

When is the Chi-Square Test used?

The chi-square test is used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories.

Also, the chi-square test is used to test for independence of two or more different categories.

If there is a significant difference, it basically implies that $\chi^2 > \sigma$, where σ is the stated significance level with usual values of 1%, 5% or 10%.

Take note that the significance level (σ) is always given in a problem.

Example: A study has been conducted and resulted in the following data:

Smoker
Heart disease?

	Yes	No	Total
Yes	23 (38x92/36.6)	15 (28.45)	38
No	69 (82.45)	259 (245.55)	328
Total	92	274	366

1) Ho: There is no association between H.D. and the factor smoker

Ha: There is an association between H.D. and smoker

2) random (?)

E $\geq$ 5

3)

$(23 - 9.55)^2/9.55 +$
$(15 - 28.95)^2/28.95 +$
$(69-87.45)^2/87.45 +$
$(259-245.55)^2/245.55 =$
28.23
$\sim X^2$ with df $(r-1)(c-1) = (2-1)(2-1) = 1$

4) P-value $P(x^2>28.23) < 0.05$

5) P-value $<\alpha$ 0.05 reject Ho

6) There is enough evidence to conclude there is a association between heart disease and the factor smoker at $\alpha = 0.05$

Note: The two tests are equivalent, more precisely, the z-test about the difference in proportion is equivalent to the chi-sq. test for independence when each variable has only 2 categories

- $X^2 = Z^2$
- P-values are the same

Ex) Income groups (H M L)
Whether data fits model -> goodness of fit
Goodness of fit test
Df = 3-1 = 2

H	10%
M	60%
L	30%

Ex) Do people from different income groups have the same proportion of smokers to non-smokers?

Income Groups

smoker

	H	M	L
Y			
N			

X^2 test for homogeneity df (2-1)(3-1)

Ex) Are income groups of smoker independent?
X^2 test for independent
Important: need to know which hypothesis to take

Confidence Interval for a Population Mean

Finding it for μ when σ is unknown
Assumptions for using the t-statistic:

Independence Assumption:

- Independence Assumption. The data values should be independent.
- Randomization Condition: The data arise from a random sample or suitably randomized experiment. Randomly sampled data (particularly from an SRS) are ideal.

10% Condition: When a sample is drawn without replacement, the sample should be no more than 10% of the population.
Normal Population Assumption:

- We can never be certain that the data are from a population that follows a Normal model, but we can check the Nearly Normal Condition: The data come from a distribution that is unimodal and symmetric.
  - Check by making a histogram or Normal probability plot.
- You can also ensure normality by checking sample size is large enough

A confidence interval for the population mean μ (when σ is unknown) is given by

$$\overline{X} \pm t \frac{s}{\sqrt{n}}$$

can also be Y bar instead of X bar
where t* is the critical value for the t distribution with df = n – 1 confidence level C. In other words, t* is the upper (1– C/2) critical value for the t(n – 1) distribution

Note: When Gosset corrected the model for the extra uncertainty, the margin of error got bigger.
Your confidence intervals will be just a bit wider and your P-values just a bit larger than they were with the Normal model.
By using the t-model, you've compensated for the extra variability in precisely the right way

Example: (Using the battery lifetime example from Ch3) We have a random sample of n = 4 observations on y = battery lifetime (hrs): 5.9, 7.3, 6.6, 5.7 NOTE: y = 6.375, s = 0.7274 (calculated in Ch3) Find the 95% confidence interval for the mean battery lifetime.

1) Independent ✓
2) Normality (assume) ✓

$\bar{y} \pm t^* (s/\sqrt{n}) = 6.375 \pm 3.182 \times (0.7274/\sqrt{4})$

(5.218, 7.532)
We are 95% confident that the mean battery lifetime is between 5.218 and 7.532 hours

Example:
A scientist interested in monitoring chemical contaminants in food, and thereby the accumulation of contaminants in human diets, selected a random sample of n = 50 male adults. It was found that the average daily intake of dairy products was y = 756 grams with a standard deviation of s = 35 grams. Find a 95% confidence interval for the mean daily intake of dairy products for men

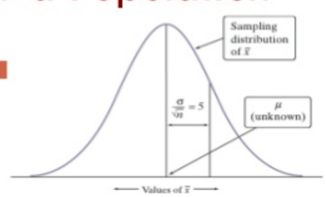
1) Independent ✓
2) Normality $n \geq 30 \rightarrow \bar{y} \sim N$

$\bar{y} \pm t^* (s/\sqrt{n}) = 756 \pm 2.014 \times (35/\sqrt{50}) = (746, 766)$
2.014 from t* -> C = 95%
Df = n-1, 50-1 = 49 -> 45

Confidence Interval for a Population Mean (Cont…)

13

Population distribution : $N(\mu, \sigma = 20)$;
SRS of size $n = 16$.
Sample mean $\bar{x} = 240.79$



Calculate a 95% confidence interval for μ.

z^*	1.645	1.960	2.576
C	90%	95%	99%

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}} = 240.79 \pm 1.96 \cdot \frac{20}{\sqrt{16}}$$
$$= 240.79 \pm 9.8 = (230.99, 250.59)$$

Example: IQ test scores
The SRS IQ test scores of 31 girls in Region A as follows:

113 102 105 … 95

This has a sample mean y =105.84 and a sample standard deviation of s = 15. The shape of the population distribution is unimodel and relatively symmetric.

a) Give a 99% confidence interval for the true mean IQ μ of all girls in the district.

$\bar{y} \pm t^* (s/\sqrt{n}) = 105.84 \pm 2.750 \times (15/\sqrt{31})$
t* -> C = 99%
Df = n-1 = 31-1 = 30
= (98.43, 113.25)

b) Give a 90% confidence interval for the true mean IQ μ of all girls in the district.
$\bar{y} \pm t^* (s/\sqrt{n}) = 105.84 \pm 1.697 \times (15/\sqrt{31})$
T* -> C=90%
Df = 30
= (101.26, 110.42)

Higher confidence = the wider the confidence interval

c) If the sample mean of IQ test scores of 20 girls in Region A is 105.84, give a 90% confidence interval for the true mean IQ μ of all girls in the district
20 girls
S = 15
$\bar{y} \pm t^* (s/\sqrt{n}) = 105.84 \pm 1.729 (15/\sqrt{20})$
t* -> C= 90%
Df = n-1, 20-1 = 19
= (100.04, 111.64)

Smaller size needs wider bound

Note:
Margin of error $m = t^*(s/\sqrt{n})$ gets smaller when
t* gets smaller, which is the same as smaller $(1 - \alpha)$.
To obtain a smaller margin of error, you must accept lower confidence.

- n gets larger. Increasing the sample size gives more accuracy.
- σ gets smaller. The less inherent variation in the population you are studying, the more accurate your estimate will be.

we can control t and n, but we cannot control σ .*

Example: A researcher found that a 98% confidence interval for the mean hours per week spent studying by college students was (13, 17). Which is true?
a) There is a 98% chance that the mean hours per week spent studying by college students is between 13 and 17 hours
b)We are 98% confident that the mean hours per week spent studying by college students is between 13 and 17 hours
c) Students average between 13 and 17 hours per week studying on 98% of the weeks
d) 98% of all students spend between 13 and 17 hours studying per week.

Previously, population parameters were described, now we will be checking if claims about the population parameters are true, or plausible to a given degree.

Example: A company is advertising that the mean lifetime of their light bulbs is 1000 hours. A person suspects the mean lifetime of the light bulbs is less than 1000 hours (company is lying in their advertisement), so he picks a sample of 100 light bulbs and find the average lifetime of these 100 light bulbs is $y = 998$. Based on this result, can we state that:

i) the mean lifetime of this company's light bulb, on average, is less than 1000 hours (so this company is lying in their advertisement)? OR

ii) the difference between 1000 hours (the average lifetime for the population) and 998 hours (the average lifetime for the sample) may have occurred because of sampling variability?

Chapter 20: Inference about Mean

20.1 The CLT Revisited

Now that we know how to create confidence intervals and test hypotheses about proportions, it'd be nice to be able to do the same for means.

Just as we did before, we will base both our confidence interval and our hypothesis test on the sampling distribution model.

Recall: If we use the statistic $\bar{y}$ for estimating the population mean μ, we can use the following information from the CLT in order to obtain a confidence interval for μ.

- $\mu_{\bar{y}} = \mu$
- $\sigma_{\bar{y}} = \bar{y} / \sqrt{n}$ standard deviation of $\bar{y}$
- The standard error of $\bar{y}$ is $SE(\bar{y}) = s/\sqrt{n}$ AND
- If the population distribution is originally normal, then the sampling distribution is also normal OR
- If the population distribution is non normal, but it has $n \geq 30$, then we can assume that the sampling distribution of $\bar{y}$ is approximately normal.

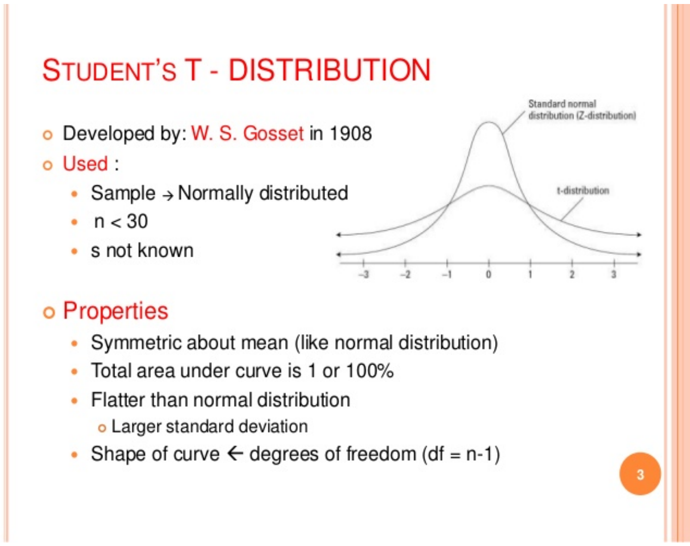
20.2 Gossets t

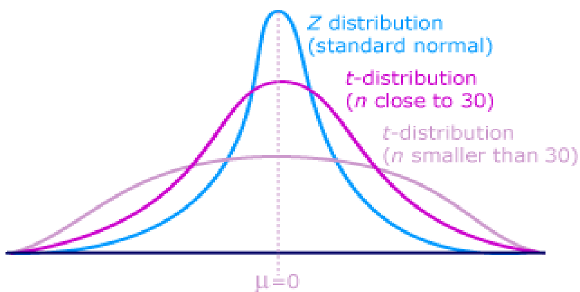
Until now, all statistical tools that were introduced were based on the assumption that population standard deviation σ is known. In practice, this assumption is very artificial and is never fulfilled in any real live situation. All procedures introduced until now are based on the normal distribution, which requires the population standard deviation σ. In most situations, σ is unknown and has to be replaced by the sample standard deviation s, it causes variability in the result. In order to calculate a confidence interval, we need to fix the problem of variability by introducing another distribution called the Student's t-distribution.

The t-distribution only depends on one parameter, which is called the degrees of freedom (df).

Properties of the t-distribution:

- Its density curves look quite similar to the standard normal curve. They are symmetric about 0, single-peaked, and bell-shaped.
- The spread of the t-distributions is a bit larger than that of the standard normal curve. (As we are now using an estimate for the population standard deviation, we must accept slightly more error in our estimation.)
- As degrees of freedom (d.f.) gets bigger, the t-density curve gets closer to the standard normal density curve. (NOTE: Table t) In another words, as degrees of freedom increases, the spread of the corresponding t density curve decreases.
- In fact, the t-model with infinite df is exactly normal.





20.4 Hypothesis Test for the Mean

Example (continued): A company is advertising that the average lifetime of their light bulbs is 1000 hours. A random sample of size 100 showed the average lifetime of their light bulbs is 998 hours with s = 5. You want to test H0: μ = 1000 versus Ha: μ < 1000. But can the difference between μ and y be explained by the sampling variability? To find this out, we calculate the test statistic that will relate the sample value ȳ with the claimed value from the null hypothesis μ0.

Test statistic: $t = \dfrac{\overline{x} - \mu_0}{s / \sqrt{n}}$

If H0 is true, this test statistic is approximately standard normal distributed (CLT: sample size of 100 is large enough), so that the value from a random sample can be judged by the standard normal distribution. For this sample, (998-1000)/(5/√ 100) = -4

That is, if the null hypothesis H0: μ = 1000 is true, y = 998 is 4 standard deviations less than what we would expect it to be. We know that this t test statistic is very low and is unlikely to occur. Let's calculate the probability to observe such a small or even smaller value, if H0 is in fact true (i.e. we calculate the p-value): p-value = P(t <- 4) ≈ 0 There is virtually no chance of observing this value of the test statistic t this extreme as a result of chance variation alone when H0 is true. Hence, there is almost no chance of seeing a sample mean y value as extreme as that observed. The evidence by the sample is compelling for H0 not to be true. We will reject H0 in favor of Ha.

One Sample t-Test for a Population Mean μ

1. The assumptions and conditions for the one-sample t-test for the mean are the same as for the one-sample t-interval.

2. Hypothesis

Test Type	
Upper Tail	Ho: $\mu = \mu o$ Ha: $\mu > \mu o$
Lower Tail	Ho: $\mu = \mu o$ Ha: $\mu < \mu o$
Two Tails	Ho: $\mu = \mu o$ Ha: $\mu \neq \mu o$

3. Test Statistic

Test statistic: $t = \frac{\overline{x} - \mu_0}{s / \sqrt{n}}$

4. P-Value

Test Type	P-Value
Upper Tail Ho: $\mu = \mu o$ Ha: $\mu > \mu o$	$P(t > to)$
Lower Tail Ho: $\mu = \mu o$ Ha: $\mu < \mu o$	$P(t < to)$
Two Tails Ho: $\mu = \mu o$ Ha: $\mu \neq \mu o$	$2 \times P(t > \lvert to \rvert)$

5. Decision: Reject H0, if and only if p-value $\leq \alpha$.

Remark:

Confidence intervals and hypothesis tests are built from the same calculations.

- In fact, they are complementary ways of looking at the same question.
- The confidence interval contains all the null hypothesis values we can't reject with these data.

More precisely, a level C confidence interval contains all of the possible null hypothesis values that would not be rejected by a two-sided hypothesis test at alpha level (1 – C).

- So a 95% confidence interval matches a 0.05 level test for these data.

Confidence intervals are naturally two-sided, so they match exactly with two-sided hypothesis tests.

- When the hypothesis is one sided, the corresponding alpha level is (1 – C)/2.
- So a 95% confidence interval matches a 0.025 level test for these data

Example: Poisoning by the pesticide DDT affects the nervous system and ought to slow the "absolutely refractory period," the time required for a nerve to recover after a stimulus. This period is known to be 1.3ms in normal rats and follows a normal distribution. Measurements were taken on 4 rats exposed to DDT, and we get $\bar{y} = 1.75$ and $s = 0.13$. Do we have evidence that DDT poisoning slows nerve recovery at the $\alpha = 0.05$ level? The parameter of interest is the true mean absolutely refractory period μ in poisoned rats.

1) H0: $\mu = 1.3$ms   Ha: $\mu > 1.3$
2) A: independent (?) normal
3) $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ = (1.75 - 1.3)/(0.13/√4) = 6.92 ~ t distribution with df = 4-1 = 3
4) P-value = P(t>6.92) from table t
5) P-value < α = 0.05, reject H0
6) There is enough evidence to conclude that DDT poisoning slows nerve recovery at the α = 0.05 level

Q: Can you conclude that every rat exposure to DDT poisoning has a nerve recovery period greater than 1.3ms?

A: No, because that is an average, not every single one is above average, otherwise it wouldn't be the average

Example: A health center reports that the mean systolic blood pressure for males over 35 years of age is 128. The medical director of a large company looks at the records of 72 executives and finds the mean blood pressure in this group is y = 126.07 with s = 15. Is this evidence that the company's executives have a different mean blood pressure than the general population at α = 0.05 level?

1) H0: $\mu = 128$ Ha: $\mu \neq 128$
2) A: independent (?) normal
3) $t = \dfrac{\overline{x} - \mu_0}{s / \sqrt{n}}$ = (126.07 - 128/(15/√72) = -1.09 ~ t distribution with df = 72 - 1 = 71 (round down to 60)
4) P-value = 2 x P(t>1.09) from table t
5) P-value > α = 0.05, do not reject H0
6) There is NOT enough evidence to conclude that the company's executives have a different mean blood pressure than the general population at α = 0.05 level

20.5: Determining the sample size

One of the important decisions, before drawing a sample, is how many experimental units from the population should be sampled. That is: what is the appropriate sample size? The answer depends on the specific object of investigation and the precision or accuracy one wants to insure. A measure for the accuracy in estimation is the margin of error. In general, the researcher chooses the largest value ME that is acceptable for the margin of error. Then the researcher determines what confidence level C he wants to attain in his claims in the study. From this, the necessary sample size can be determined. To find the sample size needed for a particular confidence level with a particular margin of error (ME), solve this equation for n

$$ME = t\left(\frac{\sigma}{\sqrt{n}}\right)$$

We don't know most of the values. To overcome this: We can use s from a small pilot study. We can use z* in place of the necessary t value. Thus

$$n \geq \left(\frac{\sigma t}{ME}\right)^2$$

Example: Suppose you want to estimate the average daily yield μ of a chemical process and you want to insure with 95% confidence that the estimate is not more than 4 tons of the true mean yield μ. Assume a previous sample would have shown a sample standard deviation of s = 21 tons. Find the minimum sample size needed.

ME = not more than 5 tons
c-> Z* 95%
= 1.96

$n \geq \left(\frac{\sigma t}{ME}\right)^2 = (1.96 \times 21/4)^2 = 105.8 \rightarrow 106$

Example: The financial aid office wishes to estimate the mean cost of textbooks per quarter for students at a particular college. For the estimate to be useful, it should be used be within $20 of the true population mean. How large a sample should be used to be 95% confident of achieving this level of accuracy if the financial aid uses a standard deviation of $100
ME = 20
C = 95%
S = 100

$n \geq \left(\frac{\sigma t}{ME}\right)^2 = (1.96 \times 100/20)^2 = 96.04 \rightarrow 97$

Chapter 21: Comparing Means

21.1 Independent populations

The two populations under investigation don't have any impact on each other are independent.
Example: comparing the height of males and females in a certain company
Example: comparing the grades of university students in Canada and US.

Since we are investigating two populations, we will introduce the following notations: Notation:

Population

	Mean	Standard Deviation
Population 1	$\mu 1$	$\sigma 1$
Population 2	$\mu 2$	σ 2

Sample

	size	mean	Standard deviation
Population 1	n1	$\bar{y}1$	s1
Population 2	n2	$\bar{y}2$	s2

In the presence of two populations, it is usually the goal to compare them. To decide if the sample means are significantly different, we cannot just compare the difference of the mean. Instead, we will see that the observed difference between the two samples depends on how big the difference is compared to the inherent variability in the populations. In order to do inferential statistics using this difference we have to investigate the distribution of this statistic.

Sample Distribution of $\bar{y}1$ - $\bar{y}2$ from two independent samples assuming unequal variances

- For the mean: $\mu\ \bar{y}1\text{-}\bar{y}2 = \mu 1 - \mu 2$
- For standard deviation (assuming unequal variability)

$\sigma\ \bar{y}1\text{-}\bar{y}2 = \sqrt{}(\sigma^2 1/n1) + (\sigma^2 2/n2)$

- We still don't know the true standard deviations of the two groups, so we need to estimate and use the standard error

$$\text{t-stat} = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

- If n1 and n2 are both large or both populations are normal distributed, then the sampling distribution of ȳ1-ȳ2 is (approximately) normal.

NOTE: Since σ 1 and σ 2 are unknown, we need to use the t distribution as the sampling distribution

t-Confidence Interval for Comparing Two Population Means Assumptions

(OPTIONAL) Pooled two-sample procedures (cont.)

When both population have the *same* standard deviation, the **pooled estimator of σ^2** is:

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

The sampling distribution for $\bar{x}_1 - \bar{x}_2$ has exactly the *t* distribution with ***$(n_1 + n_2 - 2)$* degrees of freedom**.

A level *C* confidence interval for $\mu_1 - \mu_2$ is $(\bar{x}_1 - \bar{x}_2) \pm t^* s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

(with area C between $-t^*$ and t^*)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

To test the hypothesis H_0: $\mu_1 - \mu_2 = 0$ against a one-sided or a two-sided alternative, compute the pooled two-sample *t* statistic for the $t(n_1 + n_2 - 2)$ distribution.

where t * represents the the 100(1-(α-2)th percentile of the t-distribution with 2 n1 + n2 − degrees of freedom.

Example: A company wants to show that a vitamin supplement decreases the recovery time from a common cold. They selected randomly 70 adults with a cold. 35 of those were randomly selected to receive the vitamin supplement. The data on the recovery time for both samples is shown below

Population	1 No sample	2 vitamin
Sample size	35	35
Sample mean	6.9	5.8
Sample standard deviation	2	1.2

Now test the claim of the company that vitamin supplement decreases the average recovery time from a common cold at level of significance 0.05.

1) H0 = μ1 - μ2 = 0 Ha: μ1 - μ2 > 0
2)

A:

- 2 independent population
- Independent observation
- Normality n1≥ & n2 ≥ 30 -clt->

a) n1~n2 ✓
b) n1 ≥ 15 & n2 ≥ 15 ✓
c) (Slarger/Ssmaller) < 2 (2/1.2) < 2 ✓

3) $t = \dfrac{\overline{X}_1 - \overline{X}_2}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ = (6.9 - 5.8) - 0/(1.649**√**1/35 + 1/35) = 1.925

sp(**√**(25-1)² + (35-1)² 1.2²/(35 + 35 - 2) = 1.64924

4) P-value = P(t>19.25) < 0.025
5) P-value α < 0.025 reject Ho
6) There is enough evidence to conclude the claim of the company that vitamin supplement decreases the average recovery time from a common cold at level of significant 0.025

Calculate a 95% Confidence Interval for the difference in average recovery time μ1 - μ2.

$(\bar{y}_1 - \bar{y}_2) \pm t^* s_p\sqrt{(1/n_2)/(1/n_2)} = (6.9 - 5.8) \pm 2 \times (1.64925\sqrt{(1/35) + (1/35)}$
$= (0.3115, 1.8885)$

With 95% confidence, the mean recovery time without vitamin is between 0.3115 & 1.8885 longer than with vitamin

21.2 Two-sample t-Test for Comparing Two Population Means
(or called nonpooled t test)

Assumption:
1. Independent Groups Assumption: The two groups we are comparing must be independent of each other. (See Ch 22 if the groups are not independent of one another)
2. Independence Assumption (Each condition needs to be checked for both groups.): a. Randomization Condition: Was the data collected with suitable randomization (representative random samples or a randomized experiment)? b. 10% Condition: We don't usually check this condition for differences of means. We will check it for means only if we have a very small population or an extremely large sample.
3. Normal Population Assumption: Nearly Normal Condition: This must be checked for both groups. A violation by either one violates the condition. In other words, n1 and n2 are large or both populations are approximately normal distributed.
4. Unequal standard deviations ($\sigma_1 \neq \sigma_2$) if any of the equal standard deviation conditions fail. In other words, if any of the following condition holds:
a. the sample sizes are not approximately equal (more than 1-2 observations difference)
b. both sample sizes are not at least 15
c. the informal standard deviation ratio is greater than two or the IQRs are different on the boxplot

1) Normal proportion
$Y_1 \sim N$ & $y_2 \sim N \rightarrow \bar{y}_1 - \bar{y}_2 \sim N$

2) $n \geq 30$ & $n \geq 30$ -clt-> $\bar{y}_1 - \bar{y}_2 \sim N$

Unequal standard deviations
$n_1 \neq n_2$
$n_1 < 15$ AND/OR $n_2 < 15$

S larger / S smaller > 2 = use unequal S.D

Decision:

- p-value ≤ α H0 is rejected. You report that the results are statistically significant at level α.
- \- p-value > α H0 is not rejected. You report that the results are not significant at level α.

Example: Refer to the previous "Vitamin Supplement" Example with the following revised data:

Population	1 No sample	2 vitamin
Sample size	35	35
Sample mean	6.9	5.8
Sample standard deviation	2.9	1.2

1) H0 = μ1 - μ2 = 0   Ha: μ1 - μ2 > 0
2)

A:

- 2 independent population
- Independent observation
- Normality n1≥ & n2 ≥ 30 -clt-> ȳ1 - ȳ2 ~ N

d) n1~n2 ✓
e) n1 ≥ 15 & n2 ≥ 15 ✓
f) (Slarger/Ssmaller) < 2 (2.9/1.2) > 2 X

3)

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

~ distribution with degrees of freedom = 45.3 (computer)
No computer = df = min(n1-1, n2-1) = min(35-1, 35-1)
= min (34, 34) = 34

4) p-value = p(t>2.07) -> 0.01 < p-value < 0.025

6) There is enough evidence to conclude that the claim of the company that vitamin supplement decreases the average recovery time from a common cold at level of significant 0.05.

Two-sample t-Confidence Interval for the difference in Two Population Means
Assumption: Same as the two-sample t-test

The level 100C% Confidence Interval for μ1 − μ 2 :

Two Sample *t* Procedure: Confidence Interval

- Draw an SRS of size n_1 from a normal population with unknown mean μ_1, and draw an independent SRS of size n_2 from another normal population with unknown mean μ_2. A level C confidence interval for $\mu_1 - \mu_2$ is given by

$$(x_1 - x_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

where t^* is the critical value with area C between $-t^*$ and t^* under t^* density curve with degrees of freedom equal to the smaller of $n_1 - 1$ and $n_2 - 1$.

22

where t* is the critical value of the t-distribution with the given number of degrees of freedom.
df = min(n1 – 1, n2 – 1)

Chapter 22: Paired Samples and Blocks

- Data are paired when the observations are collected in pairs or the observations in one group are naturally related to observations in the other group.
- Paired data arise in a number of ways. Perhaps the most common is to compare subjects with themselves before and after a treatment.
- When pairs arise from an experiment, the pairing is a type of blocking. When they arise from an observational study, it is a form of matching.
- If you know the data are paired, you can (and must!) take advantage of it.
- To decide if the data are paired, consider how they were collected and what they mean. There is no test to determine whether the data are paired.
- Once we know the data are paired, we can examine the pairwise differences.
- Because it is the differences we care about, we treat them as if they were the data and ignore the original two sets of data

Example 1:
Compare the resting pulse and pulse after exercise. To control for all other influences, you take both measurements on every individual in one sample.
We are interested in the difference in the population means $\mu d = \mu 1 - \mu 2$.
For statistical inference, the difference of each of the paired observations in the sample:
Sample 1 value – Sample 2 value. This creates one sample of size n of measurement differences.

Sample 1 Observation Value	Sample 2 Observation Value	Difference
y11	y21	Yd1 = y11 - y21
y12	y22	Yd2 = y12 - y22
...	...	...
y1n	y2n	Ydn = y1n - y2n

- Now that we have only one set of data to consider, we can return to the simple one-sample t-test.

- Mechanically, a paired t-test is just a one-sample t-test for the mean of the pairwise differences. The sample size is the number of pairs.

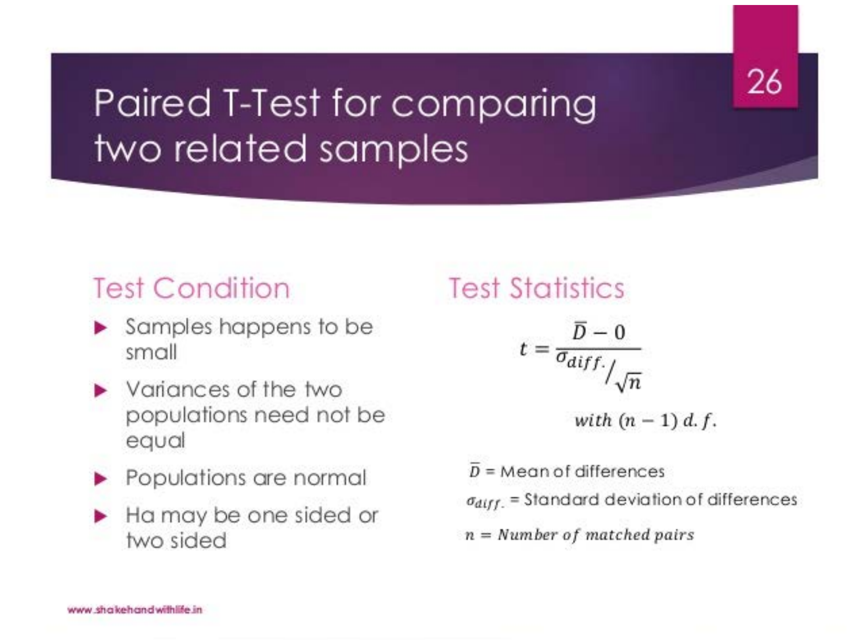
22.1 Paired t-Test for Comparing Two Population Means

Assumptions:

- Paired data Assumption: The data must be paired.
- Independence Assumption: The differences must be independent of each other.
- Randomization Condition: Randomness can arise in many ways. What we want to know usually focuses our attention on where the randomness should be.
- 10% Condition: When a sample is obviously small, we may not explicitly check this condition.
- Normal Population Assumption: We need to assume that the population of differences follows a Normal model.
- Nearly Normal Condition: Check this with a histogram or Normal probability plot of the differences

Hypothesis:

Test Type
Upper Tail Ho: µd = do vs. Ha: µd > do
Lower Tail Ho: µd = do vs. Ha: µd < do
Two Tails Ho: µd = do vs. Ha: µd ≠ do



Critical value and p-value

Test Type	p-value
Upper Tail	$P(t > t_o)$
Lower Tail	$P(t < t_o)$
Two Tails	$2 \times P(t > \lvert t_o \rvert)$

Decision:

- p-value $\leq \alpha$ H0 is rejected. You report that the results are statistically significant at level α.
- p-value $> \alpha$ H0 is not rejected. You report that the results are not significant at level α

22.2 Paired t-Confidence Interval for μd

Assumption: Same as the paired t-test

The level 100C% Confidence Interval for μd:

Confidence Interval for Mean Difference

Dependent samples

The confidence interval for difference between population means, μ_d, is

$$\bar{d} - t_{n-1,\alpha/2} \frac{S_d}{\sqrt{n}} < \mu_d < \bar{d} + t_{n-1,\alpha/2} \frac{S_d}{\sqrt{n}}$$

Where
n = the sample size
(number of matched pairs in the paired sample)

Estimating Confidence Intervals of the Difference of Generalization Performances of two Classifier Models

- If $\bar{d}$ and s_d are the mean and standard deviation of the normally distributed differences of n random pairs of errors, a $(1 - \alpha)100\%$ confidence interval for $\mu_D = \mu_1 - \mu_2$ is :

$$\bar{d} - t_{\alpha/2} \frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2} \frac{s_d}{\sqrt{n}},$$

  where $t_{\alpha/2}$ is the t-value with $v = n - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.
- Thus, if the interval contains 0.0 we can conclude on significance level α that the difference is 0.0.

Example 2:
The effect of exercise on the amount of lactic acid in the blood was examined.
Blood lactate levels were measured in eight males before and after playing three games of racquetball.

Player	Before	After	Difference
1	13	18	-5
2	20	37	-17
3	17	40	-23
4	13	35	-22
5	13	30	-17
6	16	20	-4
7	15	33	-18
8	16	19	-3

Let's test if the lactate level before exercise is lower than the lactate level after exercise at a level of significance of 0.05

NOTE1: d = ybefore – yafter

NOTE2: $\bar{d}$ = −13.63, sd = 8.28

1) Ho: µd = 0 Ha: µd < 0
2) A:
- Sample is paired
- Independent observation (assume)
- Population distribution of different is approximately normal (assume)

3) To = ȳd - do/(sd/√n) = (-13.63 - 0)/(8.28/√8) = -4.656 ~ t dist. w/ df = n - 1 = (8-1) = 7

4) p-value = p(t < -4.656) = p(t > 4.656) < 0.005

5) p-value < α = 0.05, reject Ho

Example: Random samples of 50 men and 50 women are asked to imagine buying a birthday present for their best friend. We want to estimate the difference in how much they are willing to spend. We would use a

a. Two-sample t hypothesis test

b. Two-sample t confidence interval

c. Paired t hypothesis test

d. Paired t confidence interval

End of Part 1