COMP 7990

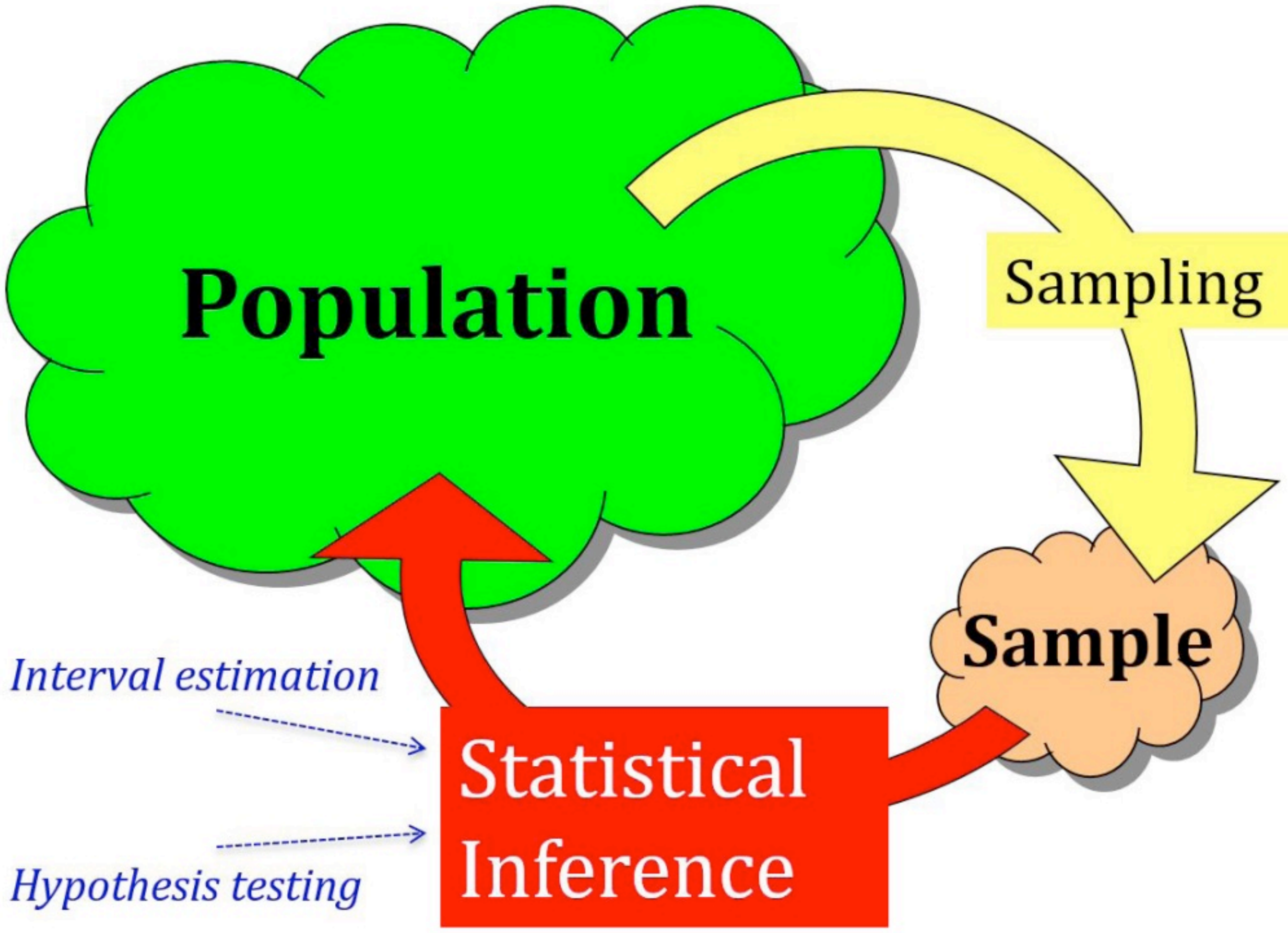
Principles and Practices of data analytics

Lecture 6：Inferential Statistics

Statistical inference

- Statistical inference is the process of drawing conclusions about the population based on information in a sample

- Example: use the sample in one election poll to draw inferences about all voters

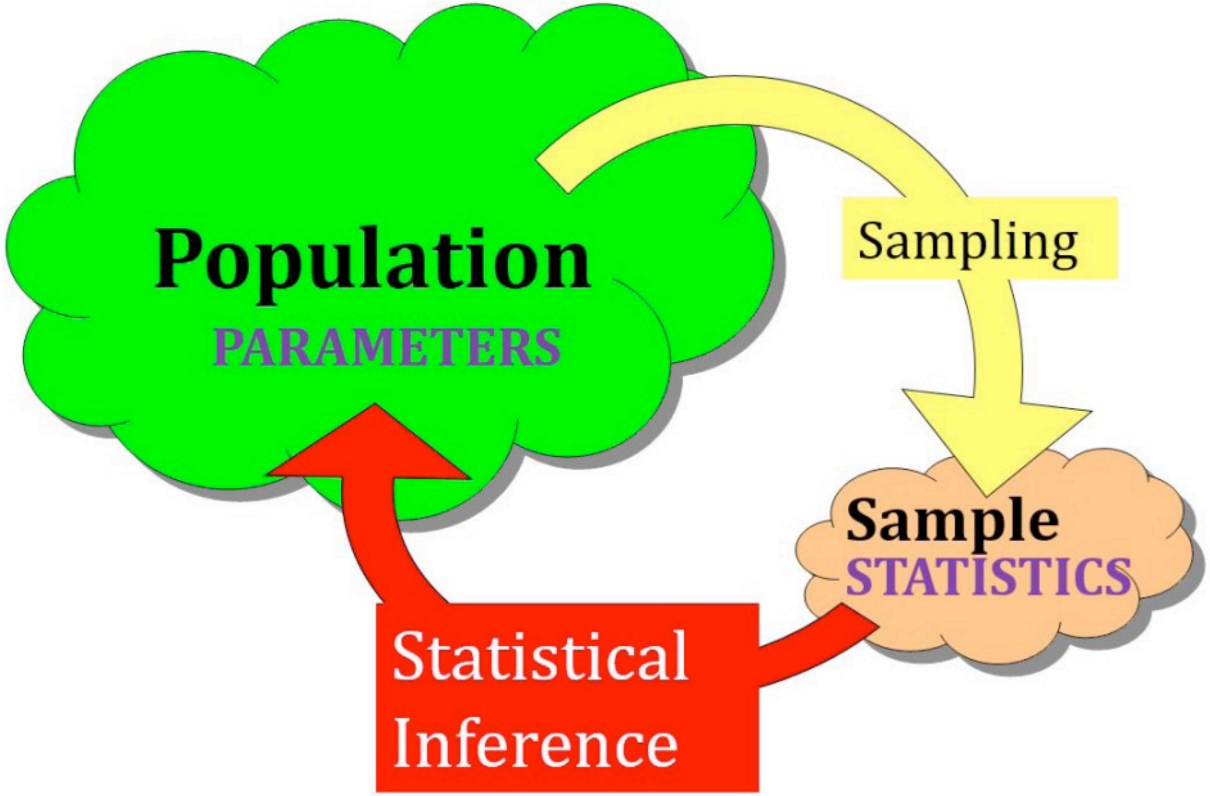


Outline

- Analyzing Differences Among Groups
- Hypothesis Testing
    - Parametric tests
    - Non-parametric tests
    - ANOVA
- Correlation

Population Parameters and Sample Statistics

- Parameters vs. Statistics
  - A **Parameter** is a number that describes some aspect of a population
  - A **Statistic** is a number that is computed from the data in a sample
    - We usually have a sample statistics and want to use it to make inferences about the population parameter



Example: parameters

- There are 2,500 managers (population) in the electronic associates. The average annual salary of 2,500 managers is \$51,800. Let y_i be the i th manager's annual salary, then

$$\mu = \frac{\sum_{i=1}^{2500} y_i}{2500} = 51800 = \text{population mean}$$

- Assume the population variance of the salary data is

$$\sigma^2 = \frac{\sum_{i=1}^{2500} (y_i - \mu)^2}{2500} = \frac{\sum_{i=1}^{2500} (y_i - 51800)^2}{2500} = 4000^2$$

- Assume that 1500 of the 2500 managers have completed the training program. Then the population proportion of completing the program is

$$p = \frac{1500}{2500} = 0.6$$

Example: parameters

Suppose we randomly select 30 managers as a sample.

- Let $x_1,\ldots, x_{30}$ be the annual salaries of the 30 managers and their average annual salary is \$51,814. Assume the sample standard deviation of the salary data is 4,002. Suppose 19 of them have completed the training program.

$$\bar{x} = \frac{\sum_{i=1}^{30} x_i}{30} = 51814 = \text{sample mean}$$

$$s^2 = \frac{\sum_{i=1}^{30} (x_i - \bar{x})^2}{30 - 1} = 4002^2 = \text{sample variance}$$

$$\hat{p} = \frac{19}{30} = 0.63 = \text{sample proportion}$$

Note: $\bar{x}$, s^2 and $\hat{p}$ are sensible estimates of $\mu = 51800$, $\sigma^2 = 4000^2$ and $p = 0.6$

Exercise

- We have total 152 students in COMP7990 section1&2. In the final exam, the average score is 51.5 and standard deviation is 15. The university randomly choose five students to check if there are calculation mistakes.

- Five students, 65, 54, 31, 22, 80
- What is the population and sample mean?
- What is the population and sample standard deviation?

Exercise

- We have total 152 students in COMP7990 section1&2. In the final exam, the average score is 51.5 and standard deviation is 15. The university randomly choose five students to check if there are calculation mistakes.
- Five students, 65, 54, 31, 22, 80
- $\bar{x} = \frac{65+54+31+22+80}{5} = 50.4$ (sample mean)
- μ=51.5 (population mean)

Standard Deviation

Standard Deviation, s: **23.901882771029** (Sample standard deviation)

Count, N: 5
Sum, Σx: 252
Mean, $\bar{x}$: 50.4
Variance, s^2: 571.3

$\sigma = 15$ (Population standard deviation)

Steps

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2},$$

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{N - 1}$$

$$= \frac{(65 - 50.4)^2 + \ldots + (80 - 50.4)^2}{5 - 1}$$

$$= \frac{2285.2}{4}$$

$$= 571.3$$

$$s = \sqrt{571.3}$$

$$= 23.901882771029$$

Point estimate

We use the statistic from a sample as a point estimate for a population parameter.

➢ Suppose x_1, x_2,..., x_n is the sample. Then

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \text{sample mean}$$ is the point estimation of μ

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \text{sample variance}$$ is the point estimation of σ^2

Population and Sample

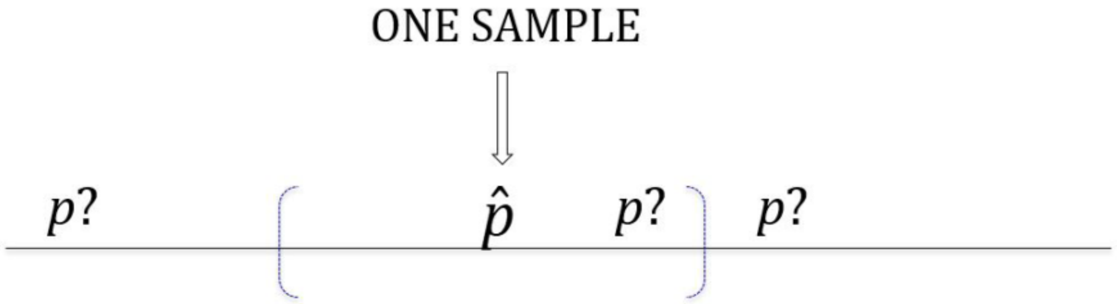
- An education official wants to estimate the proportion of adults aged 18 or older who had *read at least one book* during the previous year.
- A random sample of *1,006* adults aged 18 or older is obtained, and *835* of those adults had read at least one book during the previous year.
- The parameter is the proportion of adults 18 or older who read a book in the previous year. p is unknown The statistic is 835/1006 = 0.830, the proportion who read a book in the sample.
- $\hat{p} = 0.830$ is the **point estimate** of p

Population and Sample

- The International Dairy Foods Association (IDFA) wants to estimate the average amount of calcium male teenagers consume.
- From a random sample of *50* male teenagers, the IDFA obtained a sample mean of *1,081* milligrams of calcium consumed.
  - The parameter is the average amount of calcium that male teenagers consume. μ is unknown
- The statistic is the mean of 1,081 milligrams of calcium from the sample of 50 teenagers.
- $\bar{x} = 1081$ is the **point estimate** of μ

Problem

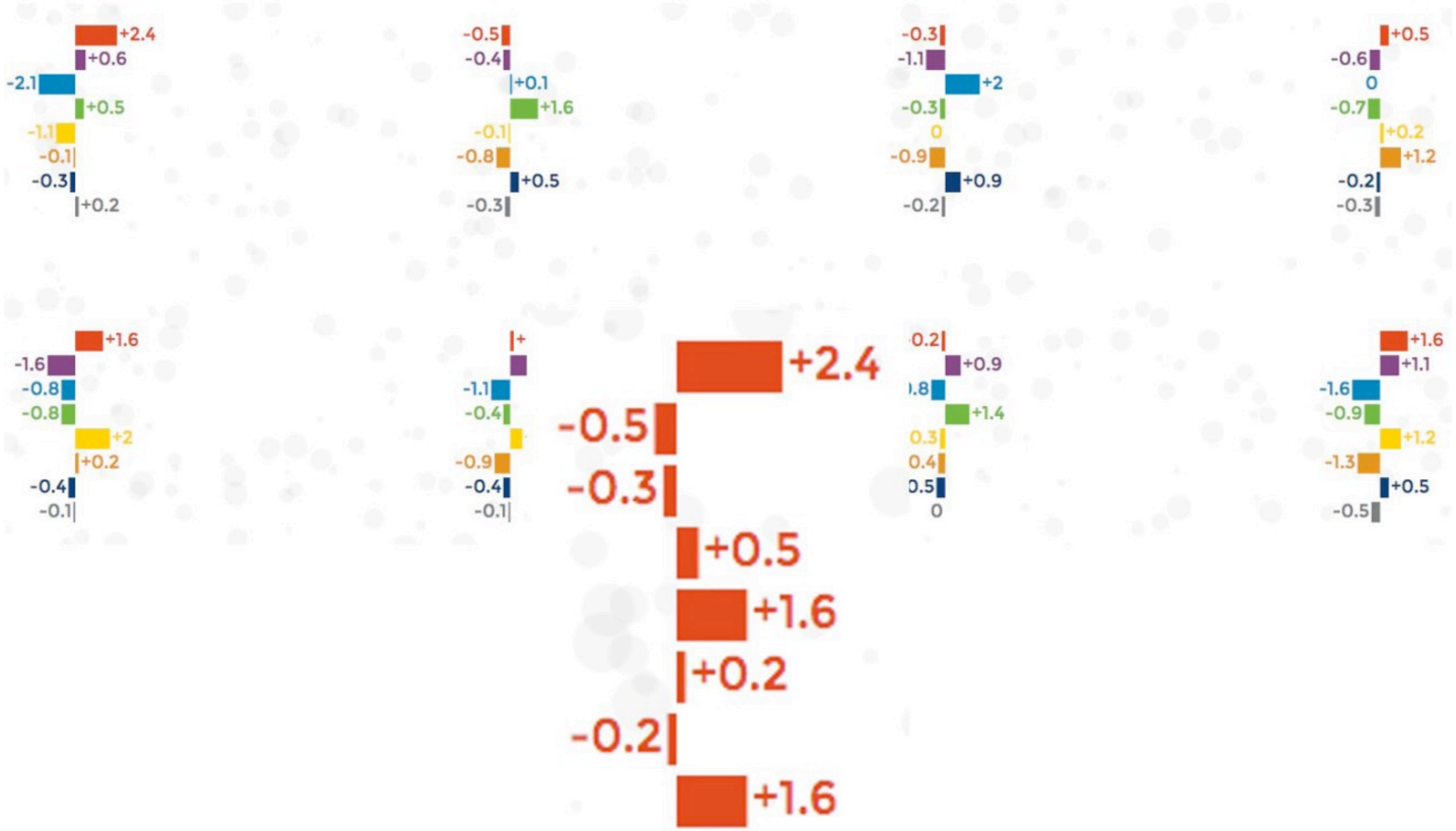
How far might the population parameter fall from the sample statistic?



- GOAL: Identify an interval of plausible values
- Key Question: For a given sample statistic, what are plausible values for the population parameter?

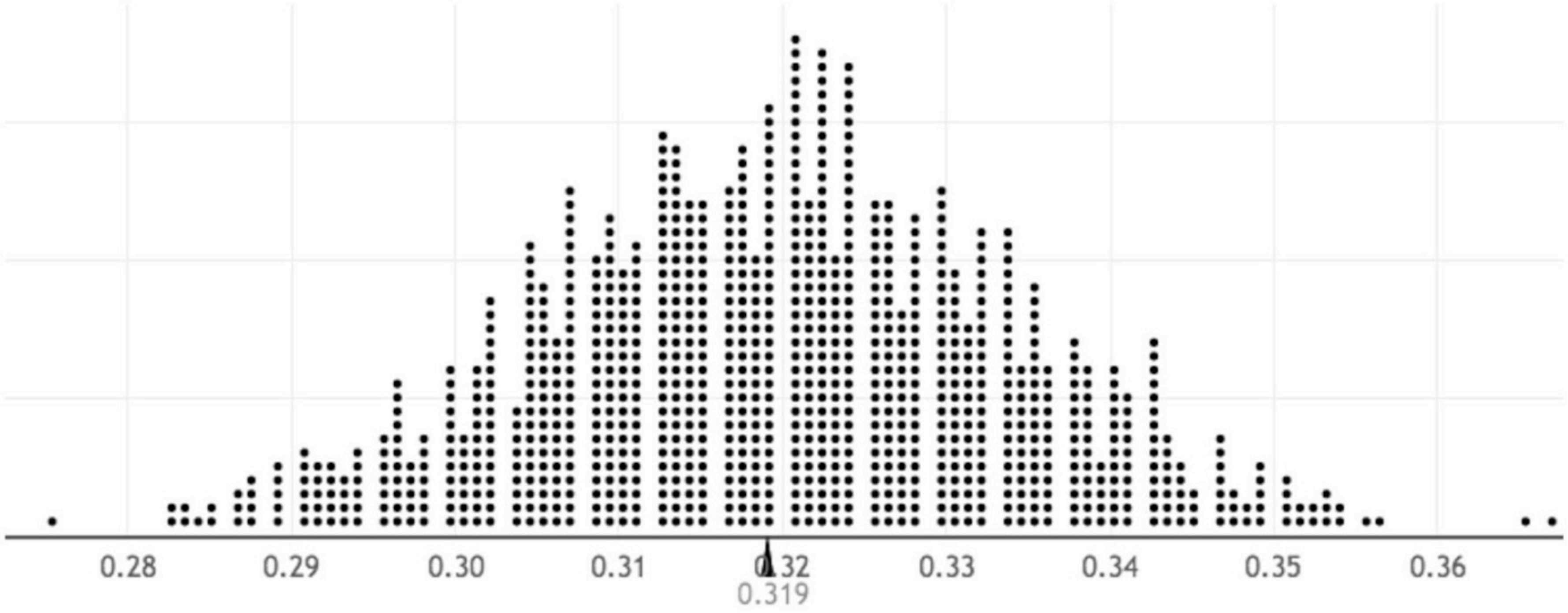
Looking at different polls

Key answer: It depends on how much the statistics varies from sample to sample!



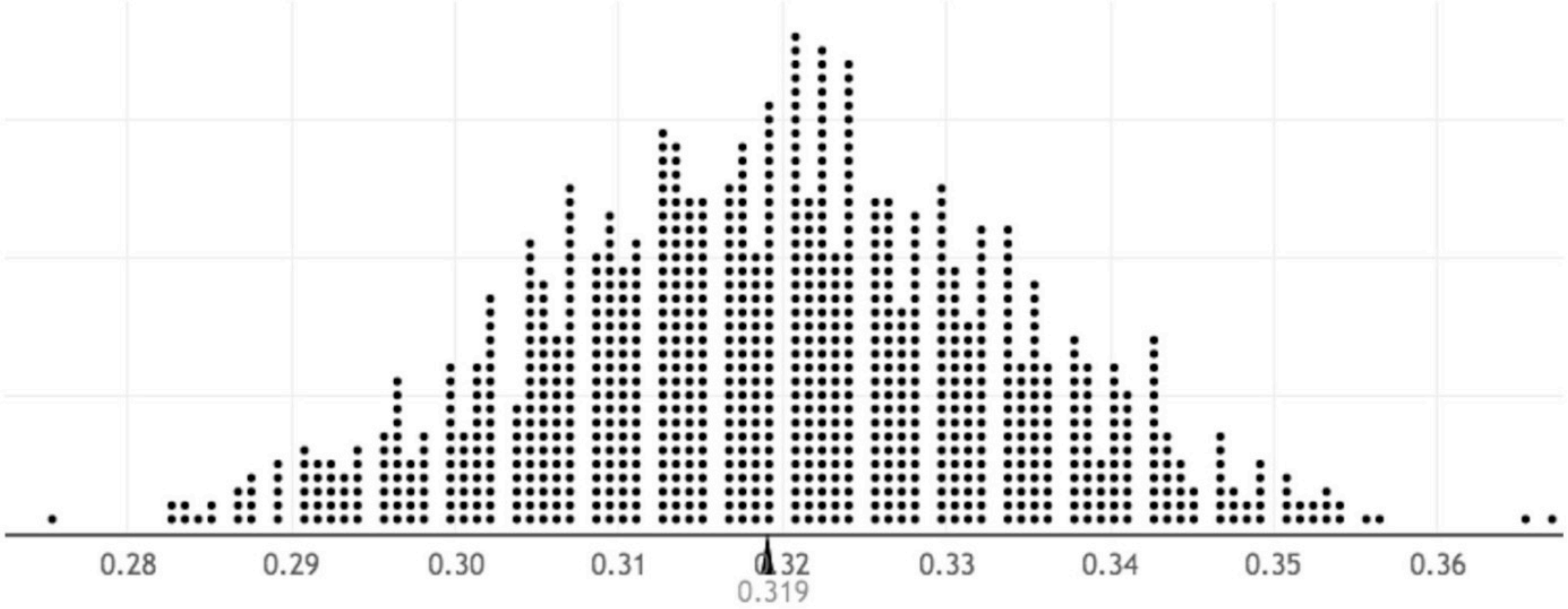
Sampling distribution

- A sampling distribution is the distribution of sample statistic computed for different samples of the same size from the same population
- A sampling distribution shows us how the sample statistic varies from sample to sample



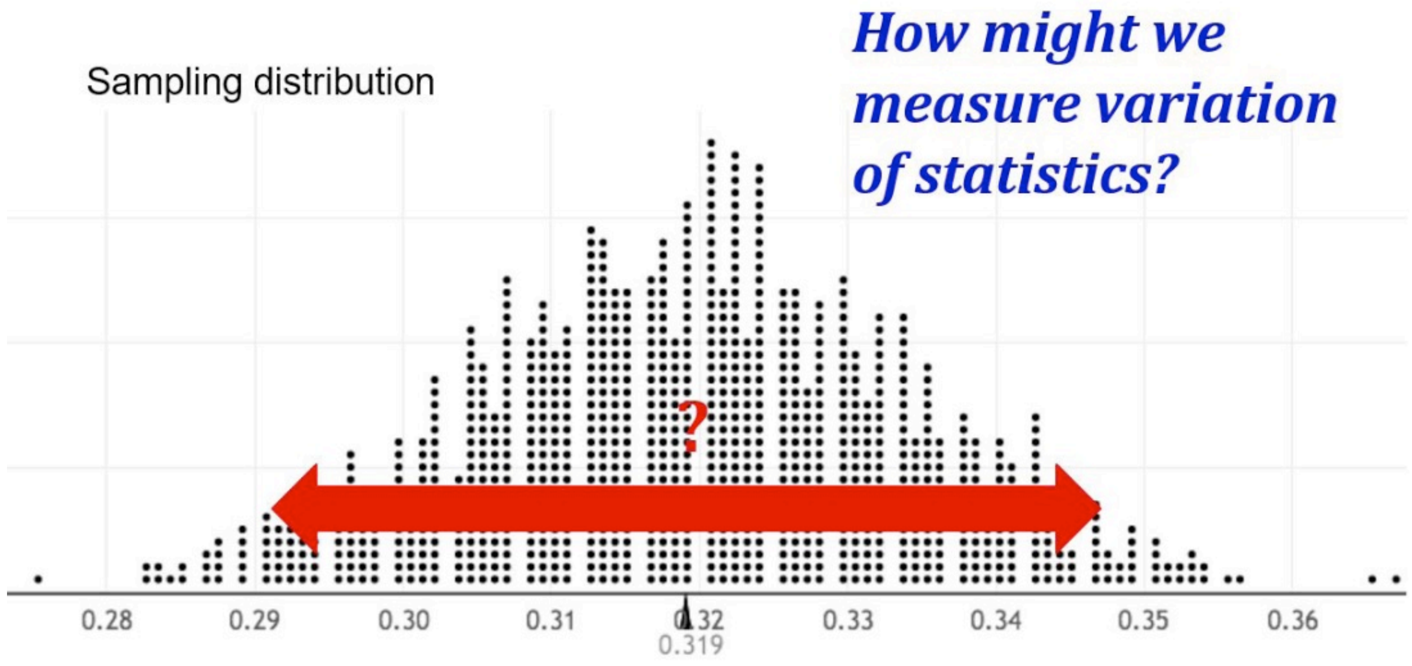
Sampling distribution

- For most of the parameters we consider, samples should be randomly selected, and the sample size is large enough
- Center: the sampling distribution will be centered at the population parameter.
- Shape: the sampling distribution will be symmetric and bell-shaped, i.e., Normal Distribution



Sampling distribution

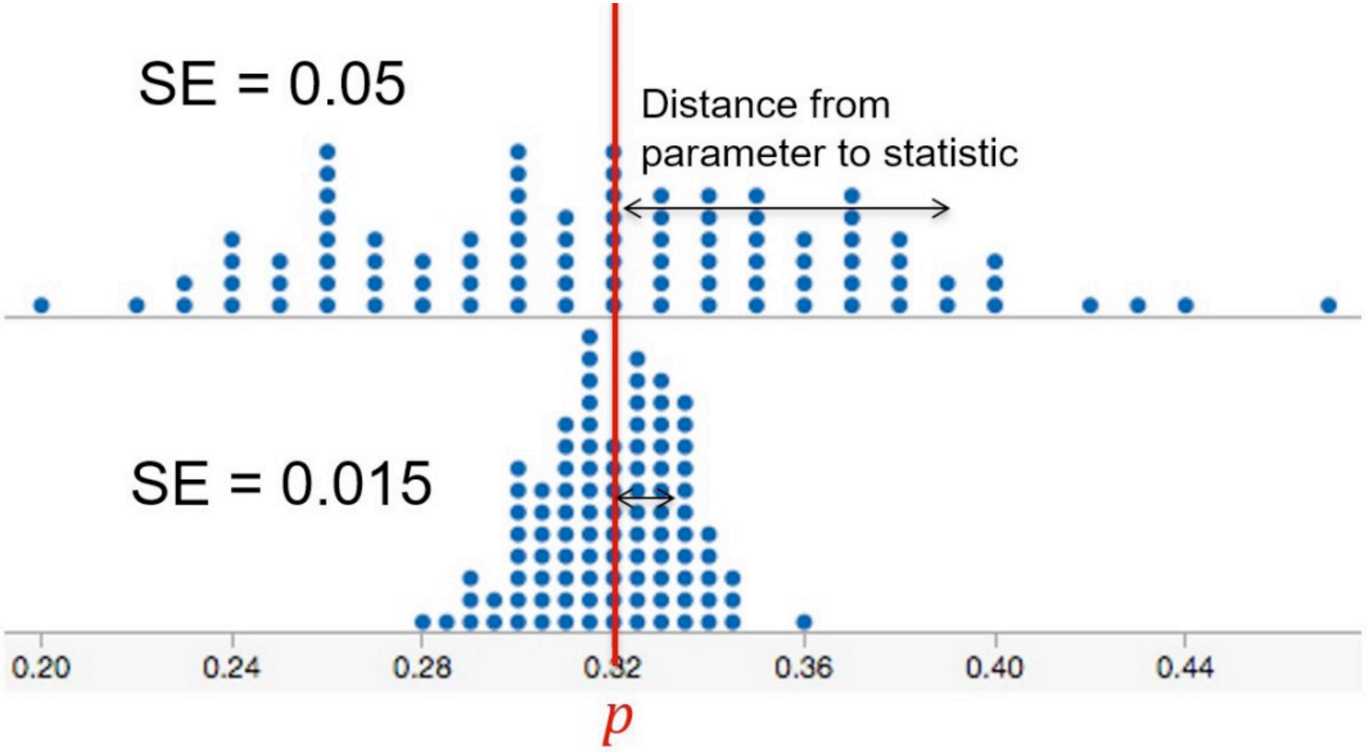
➢We really care about the spread of the statistic.



How much do statistics vary from sample to sample?

Standard Error

- ➢The standard error of a statistics, denoted SE, is the standard deviation of the sample statistic
- ➢The standard error can be calculated as the standard deviation of the sampling distribution
  - ❖ The standard error measures how much the statistic varies from sample to sample
    - o The more the statistic varies from sample to sample, the higher the standard error
    - o Lower SE means statistic closer to true parameter value
  - ❖ SE measures 'typical' distance between parameter and statistic



Confidence Intervals

- Confidence Interval
  - A Confidence Interval is a range of values we are sure our true value lies in.
  - The success rate (proportion of all samples falls into the confidence interval) is known as the confidence level.
- Calculation of 95% Confidence Interval Using the Standard Error
  - If we can estimate the standard error and the sampling distribution is relatively symmetric and bell-shaped, a 95% confidence interval can be estimated using

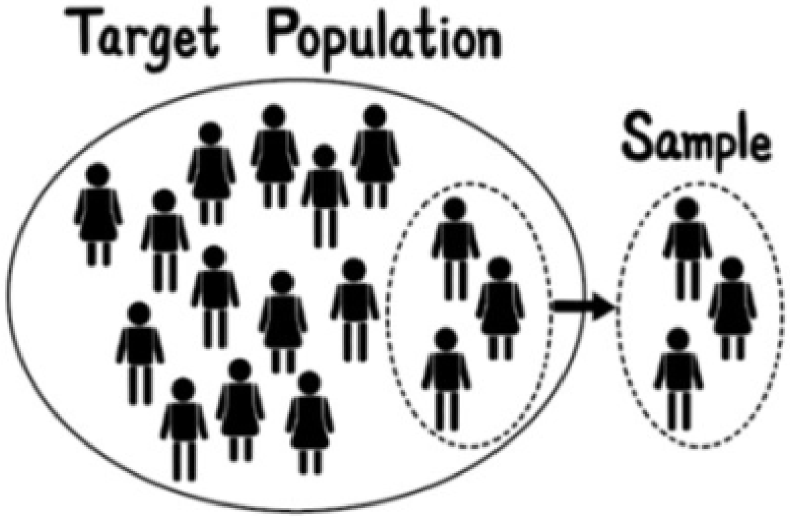
$$\text{Statistic} \pm 2 * \text{SE}$$

Estimation of population mean

- Recall that the population mean can be estimated by drawing a random sample and compute the sample mean
  - Naturally, this kind of point estimate is often not the same as the actual population mean (unless we are very lucky with our sample selection or we are working with a very special population).
- Problem: population with unknown mean, μ
- Solution: Estimate μ with $\bar{x}$
  - But $\bar{x}$ does not exactly equal to μ
  - How accurately does $\bar{x}$ estimate μ?

Estimation of population mean

- If we want to know the average monthly salary of people living in Tin Shui Wai, we can randomly select, say, 80 residents of the district and compute the average salary of these 80 people.



- Suppose the sample mean is \$9,714. Do you think that the population mean is exactly \$9,714?

Estimation of population mean

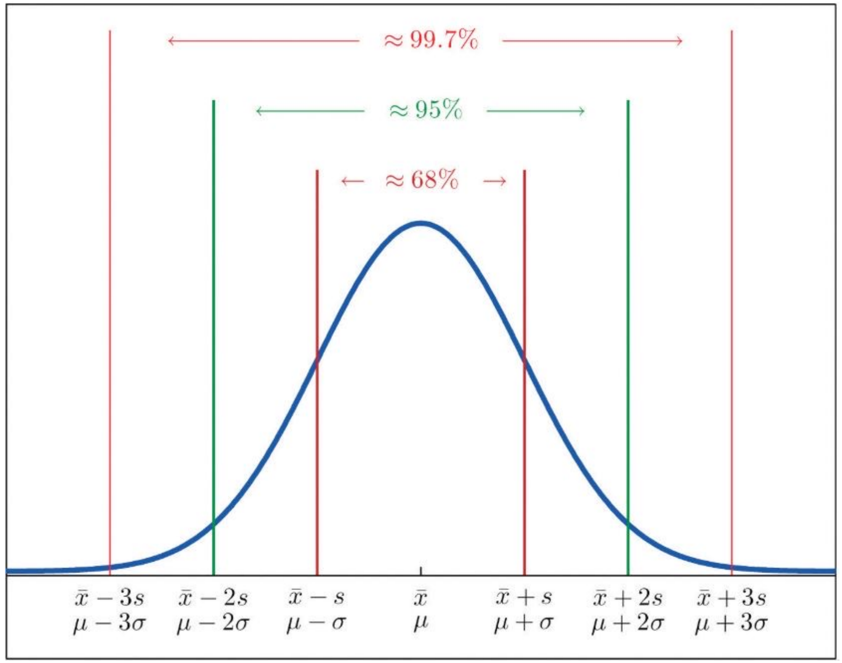
- If we repeat the procedure (i.e. drawing another random sample), do you think we will get the same estimate?
- If we repeat the procedure many times, how are the sample means (corresponding to the many different samples) going to be different?
- To estimate the average income, compare two procedures
  - Compute sample mean for a random sample of size 80
  - Compute sample mean for a random sample of size 1000
- If the procedures are repeated, which procedure do you think will produce more “stable” sample means (i.e. more reliable)?

Estimation of population mean

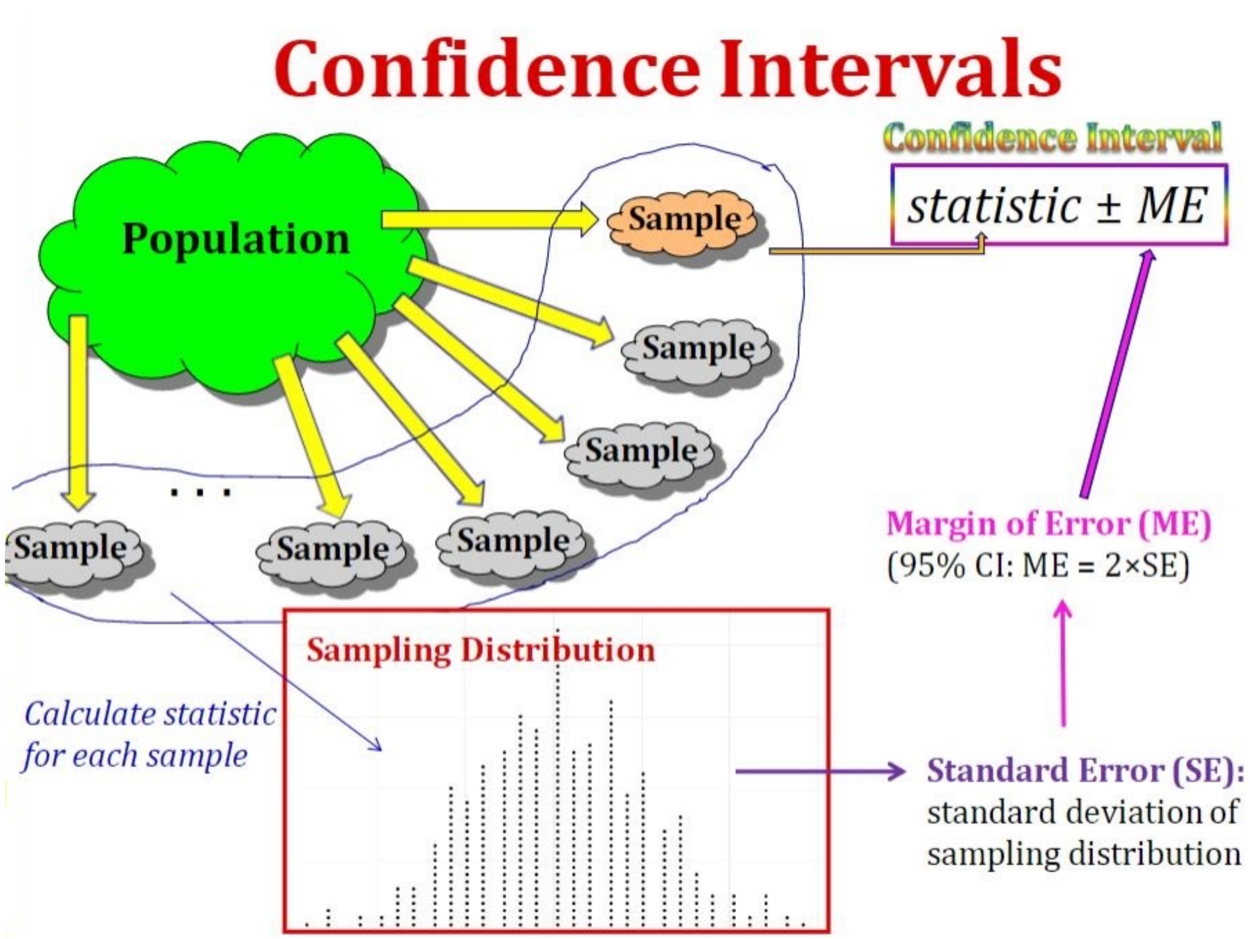
- A larger sample size leads to more stable sample means.
- With sufficient sample size, the sample mean will follow approximately a **bell-shaped distribution** around the population mean regardless of the actual distribution of the population (which is good news as we don't always know too much about the population distribution).
- Such conclusions can help us create a procedure that produces interval estimates (in contrast of a single-valued point estimate).

The empirical rule

- If the data distribution is bell-shaped, about 95% of data values will fall within 2 standard deviations of the mean
  - $\mu \pm 2\sigma$ contains 95% of the data values in the population
  - $\bar{x} \pm 2s$ contains 95% of the data values in the sample



95% confidence interval



Construction of confidence interval (an interval estimate) for a sample of size at least 30:

- Step1:Compute the sample mean $\bar{x}$ of the random sample. Suppose we know the population standard deviation σ. Compute the standard error: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

  where n is the sample size
- Step2: Construct the 95% confidence interval

$$[\bar{x} - 2\sigma_{\bar{x}}, \bar{x} + 2\sigma_{\bar{x}}]$$

Here 2 is chosen according to the confidence level 95%

- If σ is unknown, then we use s to estimate σ

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

- Then the 95% confidence interval is

$$[\bar{x} - 2s_{\bar{x}}, \bar{x} + 2s_{\bar{x}}]$$

Example

- 64 randomly selected university students were asked how much they usually spend on textbooks per year. This sample produced a mean of $1450 and a standard deviation of $300 for such annual expenses.
- Construct a 95% confidence interval for the sample mean.

N=64, $\bar{x} = \$1450, s = \$300$

$s_{\bar{x}} \approx \frac{300}{\sqrt{64}} = 37.5 \qquad \bar{x} \pm 2 * s_{\bar{x}} = \$1450 \pm 2 * \$37.5$

Recap the *Symbols*

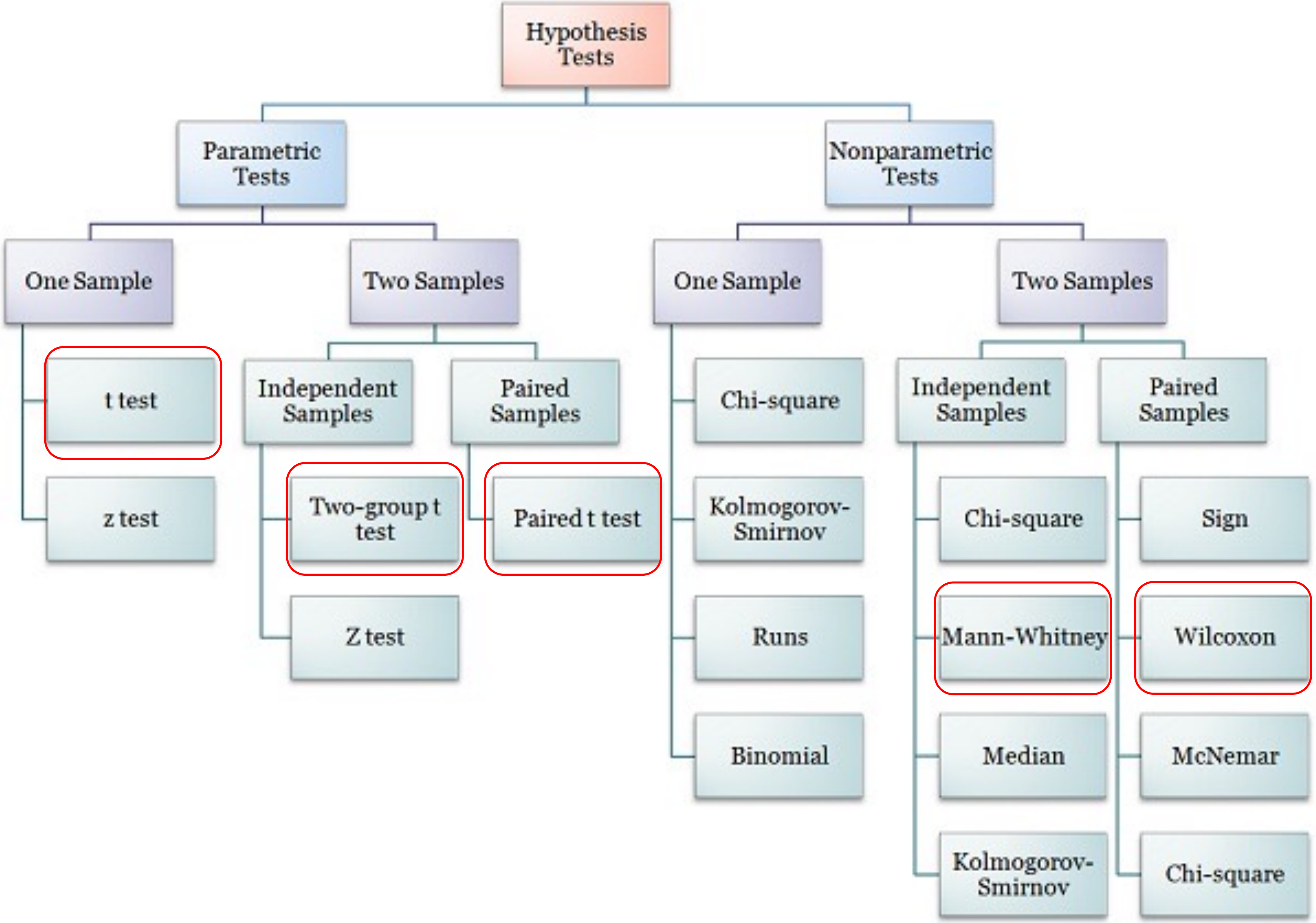
- population mean: μ
- sample mean: $\bar{x}$
- population standard deviation: σ
- sample standard deviation: *s*
- population standard error: $\sigma_{\bar{x}}$
- sample standard error: $s_{\bar{x}}$
- sample proportion: $\hat{p}$
- population proportion: *p*

Outline

- Analyzing Differences Among Groups
- Hypothesis Testing
  - Parametric tests
  - Non-parametric tests
  - ANOVA
- Correlation

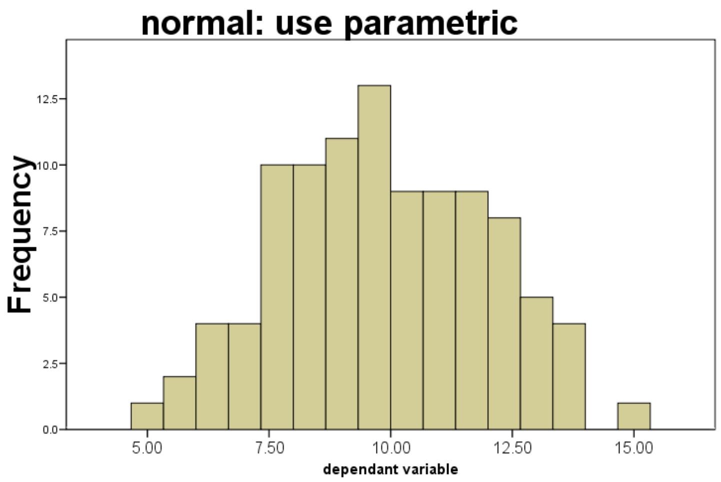
Parametric and Non-Parametric Tests

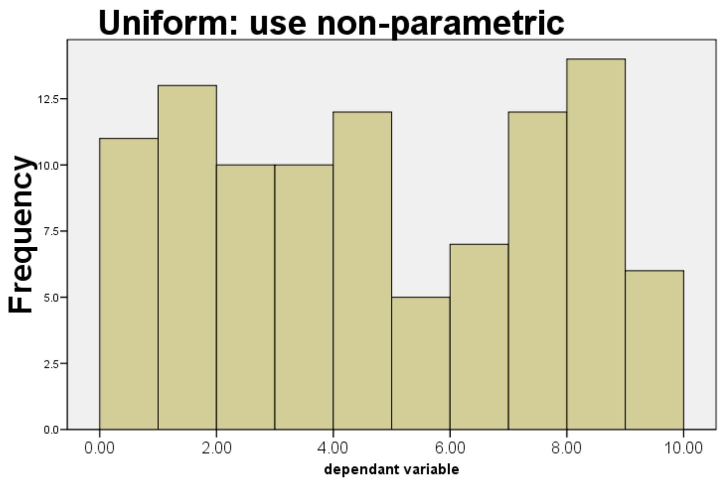
- **Parametric Tests:** assume the distribution of sample data (i.e. normality)
- **Non-Parametric Tests:** do not assume data are drawn from any particular distribution

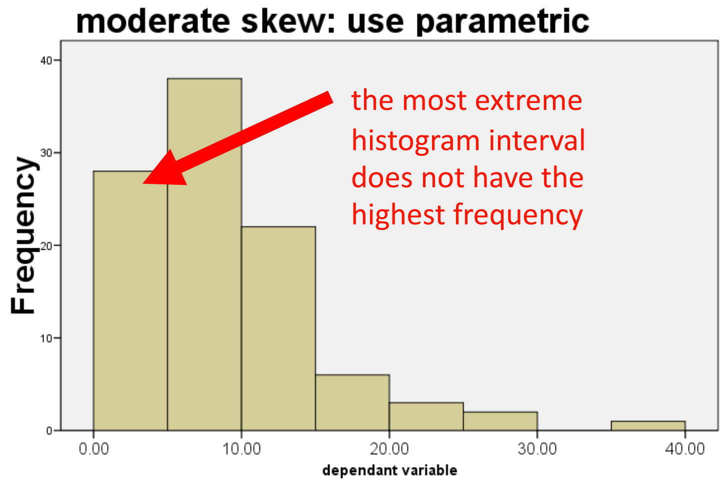


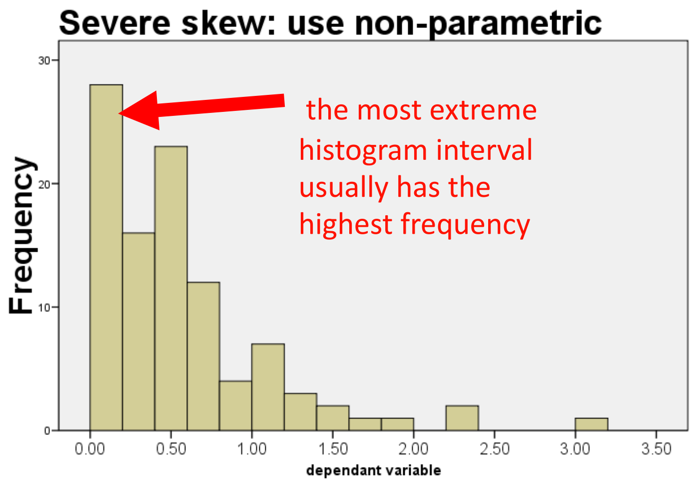
T-test assumption 1 - normality

Assumption1: The sampling distribution is normally distributed

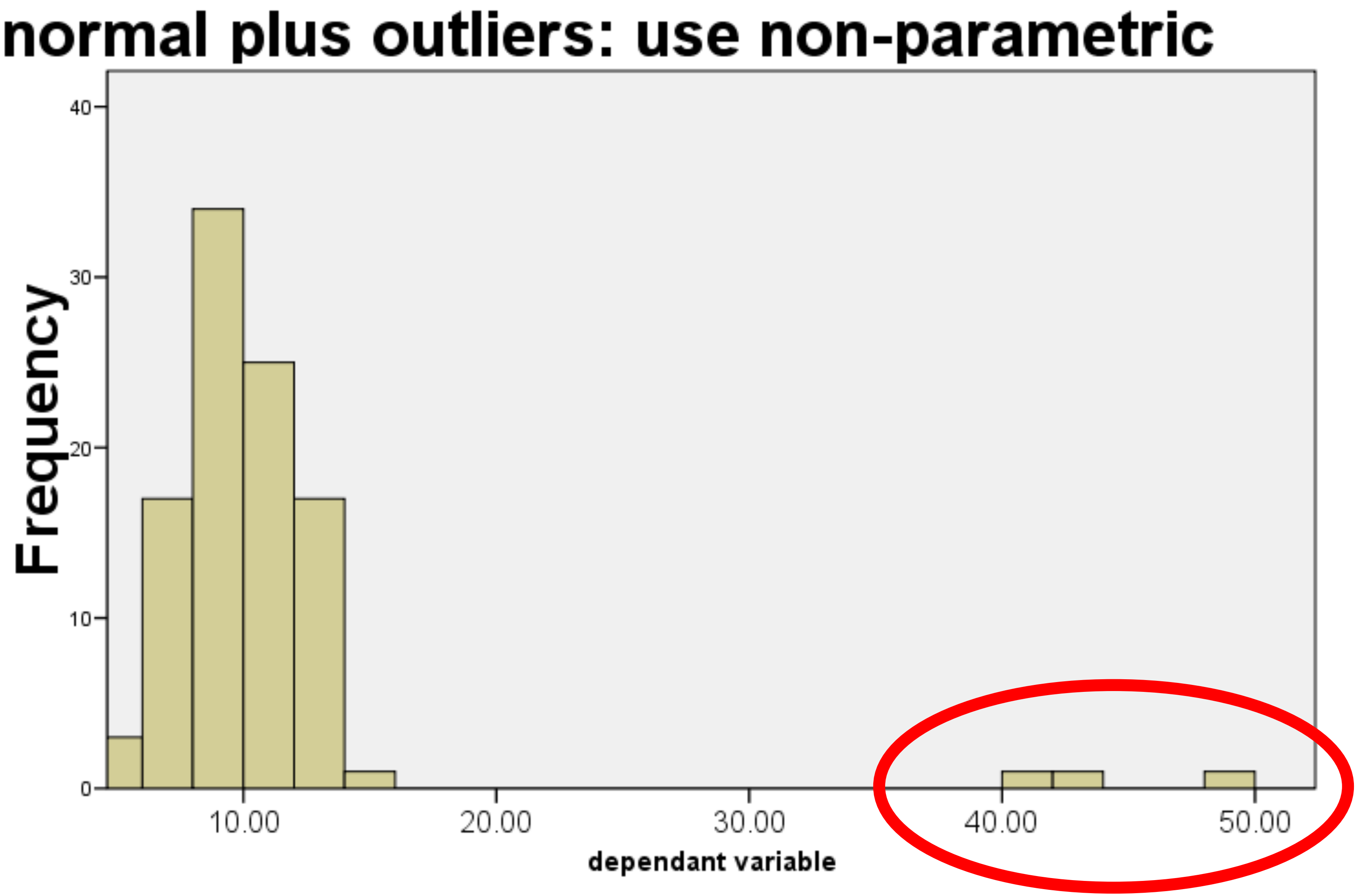








T-test assumption 2-No extreme scores



There should not include ***extreme scores*** or ***outliers***, because these have a disproportionate influence on the mean and the variance

Types of *T-test*

- Single sample t
  - One sample, compared with ***known*** population mean
  - Goal: Is current sample different from population?

- Independent samples t
  - Different (independent) samples of participants
  - Are our samples from different populations?

- Paired/Dependent Samples t
  - Same or related (dependent) samples
  - Are our samples from different populations?

Degrees of Freedom

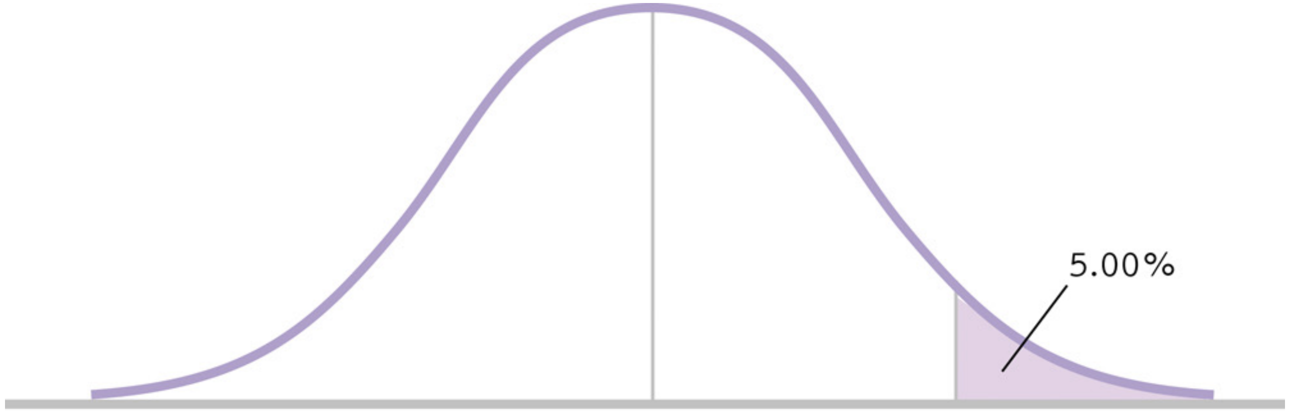
- Necessary when making estimates…
- The number of scores that are free to vary when estimating a population parameter from a sample
  - df = n – 1 (for a Single-Sample t Test)
  - Example: I decide to ask 6 people how often they floss their teeth and record their answers: average = 2 (times per week)
    - Eventual goal: Estimate population parameters (population variability).
    - How many scores are free to vary and can still produce an average of 2?

3	Free	2
5	Free	1
1	Free	0
0	Free	0
2	Free	0
1	**LOCKED**	**9**
Average = 2		**Average = 2**

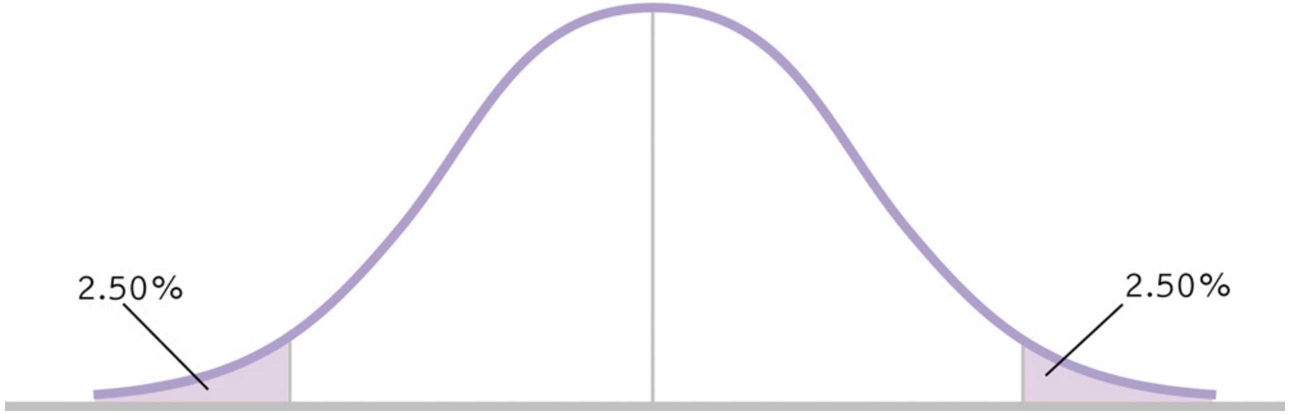
General Form of t-test

	Single sample, Paired/Dependent Samples t	Independent Samples t (same variance)
Statistic	$t = \dfrac{\overline{x} - \mu}{s_{\overline{x}}}$	$t = \dfrac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{s_{Difference}} = \dfrac{\bar{x} - \bar{y}}{\sqrt{s_{Pooled}^2 (\frac{1}{m} + \frac{1}{n})}}$ $s_{Pooled}^2 = \dfrac{df_x s_x^2 + df_y s_y^2}{df_{Total}} = (\dfrac{df_x}{df_{Total}}) s_x^2 + (\dfrac{df_y}{df_{Total}}) s_y^2$
Critical value	$t_{\alpha,n-1}$	$t_{\alpha,m+n-2}$

One Tailed vs. Two Tailed Tests



H0: A=B
Ha: A>B
Ha: A<B



H0: A=B
Ha: A≠B

From two tailed test, you can only draw the conclusion that **A≠B**, instead of A>B or A<B due to significant level.

Class Activity: Attendance in Therapy Sessions

- Our Counseling center on campus is concerned that most students requiring therapy do not take advantage of their services. Right now students attend only **4.6 sessions** in a given year! Administrators are considering having patients sign a contract stating they will attend **at least 10 sessions** in an academic year.
- Question: Does signing the contract actually influence participation/attendance?
- We had 5 patients sign the contract and we counted the number of times they attended therapy sessions

Number of Attended Therapy Sessions
6
6
12
7
8

Class Activity: Six Steps for Hypothesis Testing

1. Identify
2. State the hypotheses
3. Characteristics of the comparison distribution
4. Critical values
5. Calculate
6. Decide

Single-Sample *t* Test: Attendance in Therapy Sessions

1. Identify
   - Populations:
     - Pop 1: All clients who **sign contract**
     - Pop 2: All clients who **do not sign contract**
   - Distribution:
     - One Sample mean: **Distribution of means**
   - Test & Assumptions: Population mean is ***known*** but not standard deviation → single-sample *t* test
2. State the null and research hypotheses
   - H_0: Clients who sign the contract will attend the **same number** of sessions as those who do not sign the contract.
   - H_a: Clients who sign the contract will attend a **different number** of sessions than those who do not sign the contract.

Single-Sample *t* Test: Attendance in Therapy Sessions

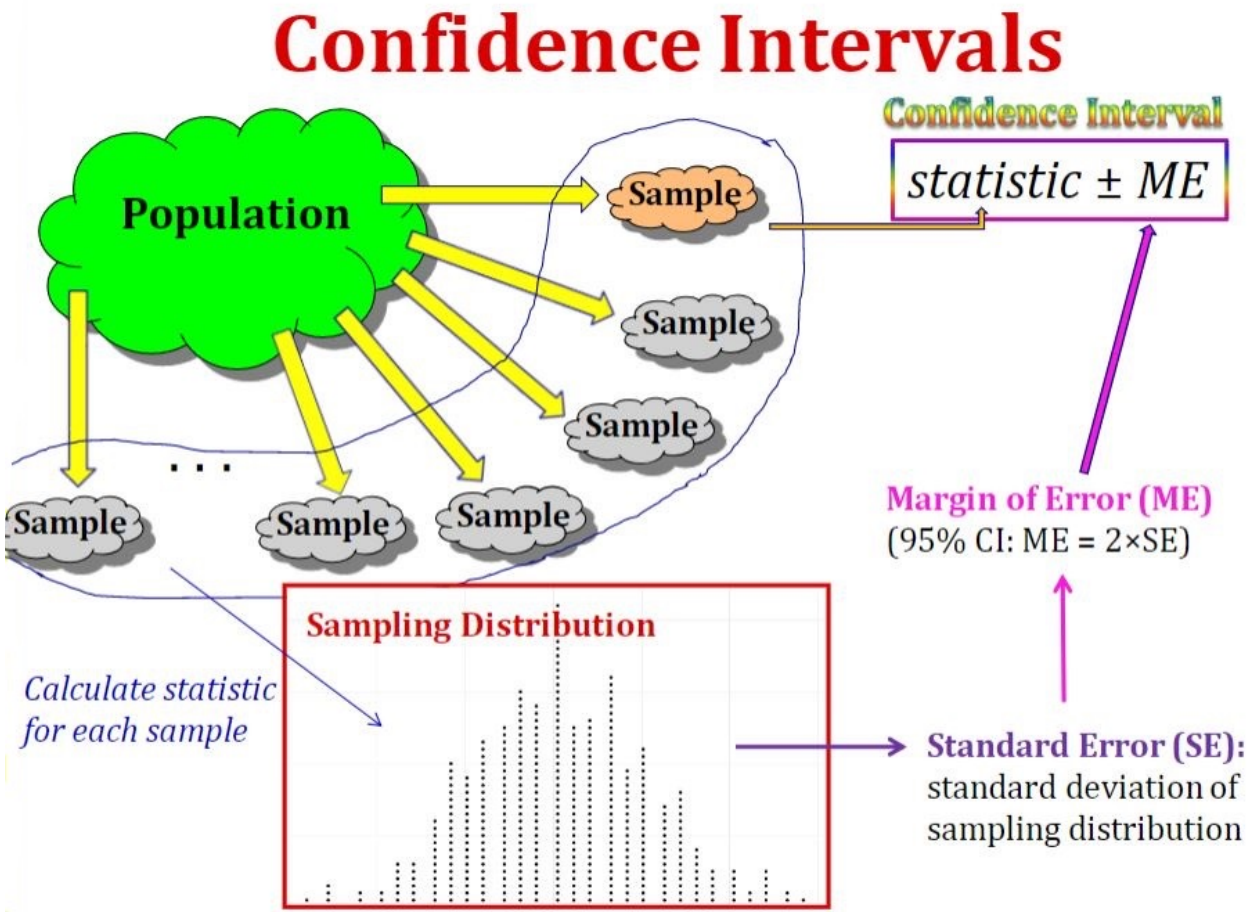
3. Determine characteristics of comparison distribution (distribution of sample means)
   - Population: μ =4.6 times
   - Sample: $\bar{x} = 7.8$ times

x	$x - \bar{x}$	$(x - \bar{x})^2$
6	-1.8	3.24
6	-1.8	3.24
12	-4.2	17.64
7	-0.8	0.64
8	0.2	0.04
$\bar{x} = 7.8$		***SS$_X$* = 24.8**

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{24.8}{5\text{-}1}} = 2.490$$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{2.490}{\sqrt{5}} = 1.114$$

95% confidence interval



Construction of confidence interval (an interval estimate) for a sample of size at least 30:

- Step1:Compute the sample mean $\bar{x}$ of the random sample. Suppose we know the population standard deviation σ. Compute the standard error: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

  where n is the sample size

- Step2: Construct the 95% confidence interval

$$[\bar{x} - 2\sigma_{\bar{x}}, \bar{x} + 2\sigma_{\bar{x}}]$$

Here 2 is chosen according to the confidence level 95%

- If σ is unknown, then we use s to estimate σ

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

- Then the 95% confidence interval is

$$[\bar{x} - 2s_{\bar{x}}, \bar{x} + 2s_{\bar{x}}]$$

Single-Sample *t* Test: Attendance in Therapy Sessions

$$\mu = 4.6,\ s_{\bar{x}} = 1.114,\ \bar{x} = 7.8,\ n = 5,\ df = 4$$

4. Determine critical value (cutoffs)
   - In Behavioral Sciences, we use *p* = .05 (5%)
   - Our hypothesis is <u>*nondirectional*</u> so our hypothesis test is <u>*two-tailed*</u>.

Significance level = α

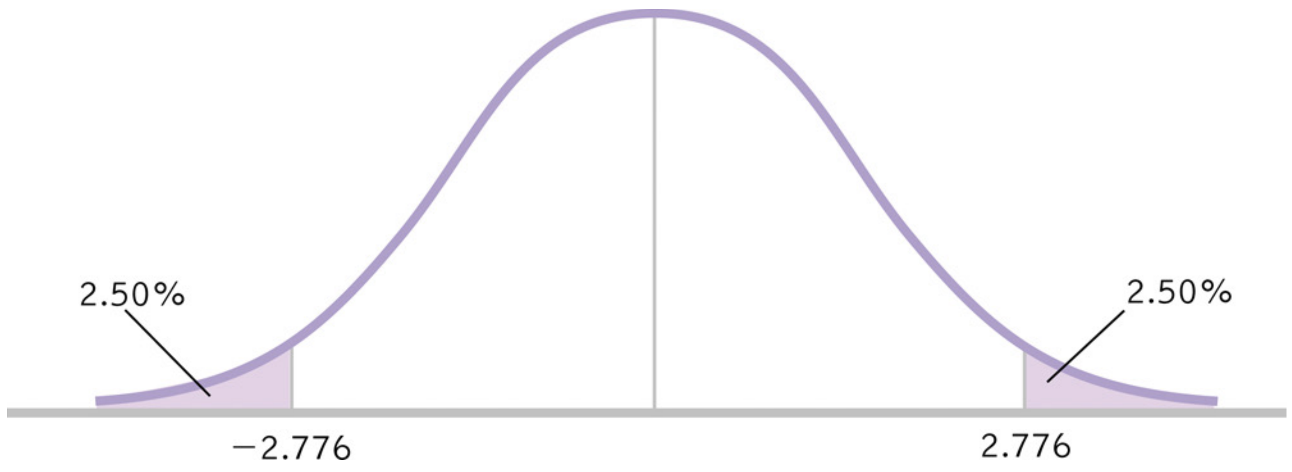
Degrees of Freedom	.005 (1-tail) .01 (2-tails)	.01 (1-tail) .02 (2-tails)	.025 (1-tail) .05 (2-tails)	.05 (1-tail) .10 (2-tails)	.10 (1-tail) .20 (2-tails)	.25 (1-tail) .50 (2-tails)
1	63.657	31.821	12.706	6.314	3.078	1.000
2	9.925	6.965	4.303	2.920	1.886	.816
3	5.841	4.541	3.182	2.353	1.638	.765
df = 4 → 4	4.604	3.747	$\boxed{2.776}$	2.132	1.533	.741
5	4.032	3.365	2.571	2.015	1.476	.727
6	3.707	3.143	2.447	1.943	1.440	.718
7	3.500	2.998	2.365	1.895	1.415	.711

Single-Sample *t* Test: Attendance in Therapy Sessions

$$\mu = 4.6,\ s_{\bar{x}} = 1.114,\ \bar{x} = 7.8,\ n = 5,\ df = 4$$

4. Determine critical value (cutoffs)

$t_{crit} = \pm\ 2.776$



Single-Sample *t* Test: Attendance in Therapy Sessions

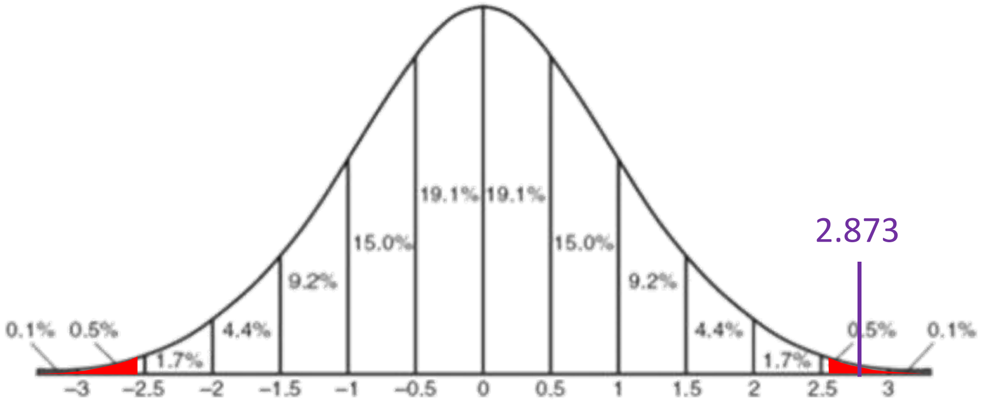
μ =4.6, $s_{\bar{x}} = 1.114,\ \bar{x} = 7.8,\ n = 5,\ df = 4$

5. Calculate the test statistic

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{7.8 - 4.6}{1.114} = 2.873$$

Single-Sample *t* Test: Attendance in Therapy Sessions

$$\mu = 4.6,\ s_{\bar{x}} = 1.114,\ \bar{x} = 7.8,\ n = 5,\ df = 4$$



6. Make a decision

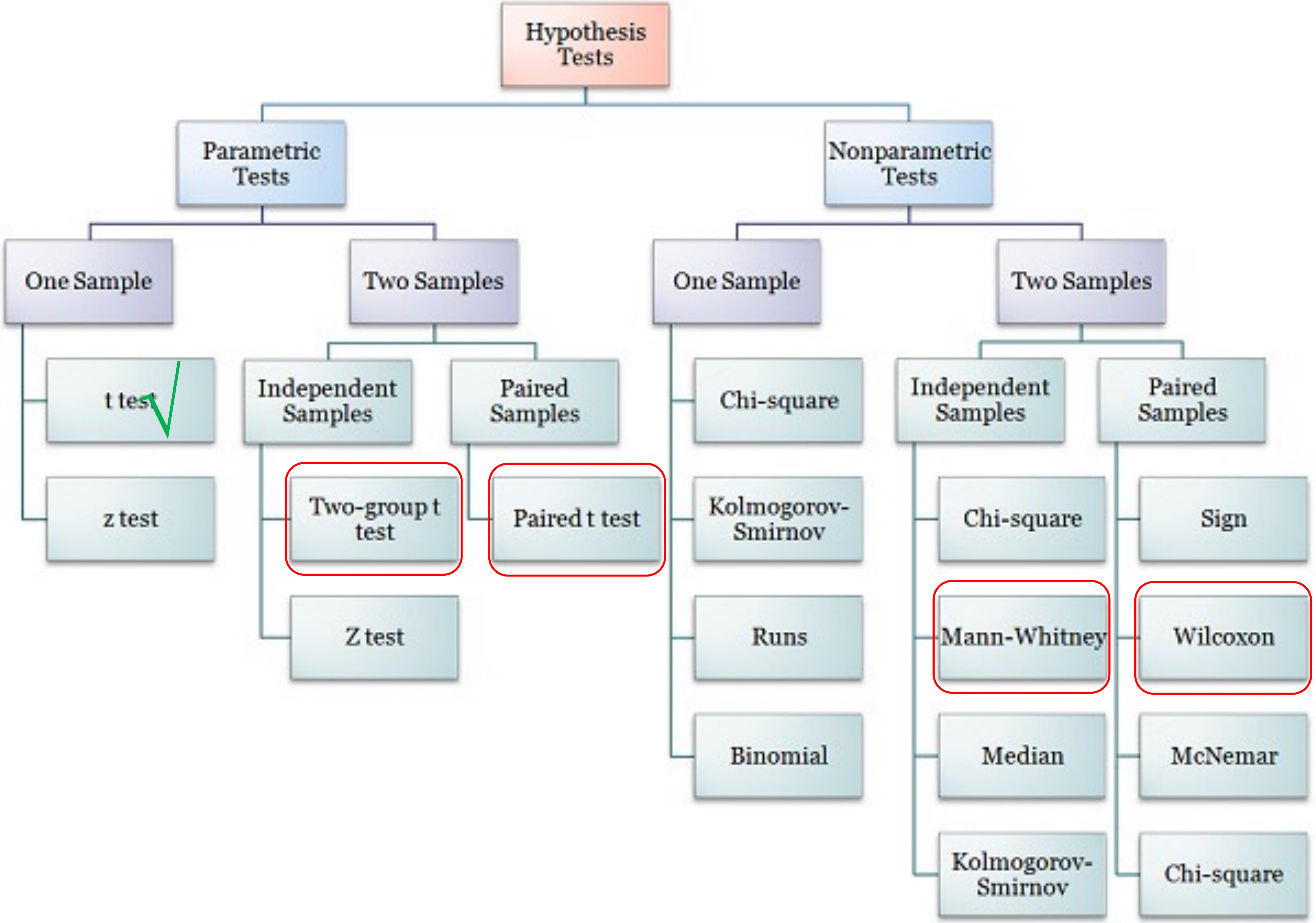
   t = 2.873 > t_{crit} = 2.776, *reject the null hypothesis*

   Clients who sign a contract will attend different number of sessions than those who do not sign a contract,

   $t(4) = 2.87$, $p < 0.05$.

Parametric and Non-Parametric Tests

- **Parametric Tests: assume** the distribution of the sample data (i.e., normality)
- **Non-Parametric Tests: do not assume** data are drawn from any particular distribution



Paired-Samples *t* Test

- *Used to compare 2 means for a within-groups design, a situation in which every participant is in both samples (paired/dependent)*

- New Terminology
  - Distribution of Mean Differences
  - Difference Scores: $X_1 – Y_1$, $X_2 – Y_2$, …

- Let's walk through an example…

Paired Samples *t* Test: Does Studying in the Exam Room Help?

I have a debate with a research assistant about context effects in studying for exams. She believes that she does far better when she studies for an exam in the same room as she later takes the exam. I told her that I could see it either hurting or helping students. We agreed to have a group of **5 participants** complete **two highly similar math exams**. For the first exam, the participants studied for and completed the exams in the **same room** while the second exam were in **different rooms**, order counterbalanced. Data for SAME and DIFFERENT rooms are below.

Student ID	SAME (*X*)	DIFFERENT (*Y*)
1	122	111
2	131	116
3	127	113
4	123	119
5	132	121

Score Difference Y-X
-11
-15
-14
-4
-11

$$\bar{x} = -11$$

Class activity: Six Steps for Hypothesis Testing

1. Identify
2. State the hypotheses
3. Characteristics of the comparison distribution
4. Critical values
5. Calculate
6. Decide

Paired Samples *t* Test: Does Studying in the Exam Room Help?

1. Identify
   - Populations:
     - Pop 1: Exam grades when studying and testing are in the **same** room.
     - Pop 2: Exam grades when studying and testing are in **different** rooms.
   - Distribution:
     - Mean of Difference Scores: **Distribution of Mean Differences**
   - Test & Assumptions: One group of participants that is studied at ***two situations***, paired-samples t test
2. State the null and research hypotheses
   - H_0: Studying and testing in the same room will result in the **same grade** as studying and testing in different rooms.
   - H_a: Studying and testing in the same room will result in **different grades** than studying and testing in different rooms.

Paired Samples *t* Test:
Does Studying in the Exam Room Help?

3. Determine characteristics of comparison distribution (distribution of mean differences)
   - Population: $\mu = 0$ (i.e., no mean difference)
   - Sample(s): $\bar{x} = -11$

Score Difference (Y-X)	Deviation Score $x - \bar{x}$	Squared Deviation $(x - \bar{x})^2$
-11	0	0
-15	-4	16
-14	-3	9
-4	7	49
-11	0	0
$\bar{x} = -11$		$SS_X = 74$

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{74}{5\text{-}1}} = 4.301$$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{4.301}{\sqrt{5}} = 1.923$$

Paired Samples *t* Test: Does Studying in the Exam Room Help?

$$\mu = 0, s_{\bar{x}} = 1.923,\ \bar{x} = -11,\ n = 5,\ df = 4$$

4. Determine critical value (cutoffs)
   - In Behavioral Sciences, we use *p* = 0.05 (5%)
   - Our hypothesis ("Studying and testing in the same room will result in **different grades** than studying and testing in different rooms.") is ***nondirectional*** so our hypothesis test is ***two-tailed***.

Significance level = α

Degrees of Freedom	.005 (1-tail) .01 (2-tails)	.01 (1-tail) .02 (2-tails)	.025 (1-tail) .05 (2-tails)	.05 (1-tail) .10 (2-tails)	.10 (1-tail) .20 (2-tails)	.25 (1-tail) .50 (2-tails)
1	63.657	31.821	12.706	6.314	3.078	1.000
2	9.925	6.965	4.303	2.920	1.886	.816
3	5.841	4.541	3.182	2.353	1.638	.765
4	4.604	3.747	$\boxed{2.776}$	2.132	1.533	.741
5	4.032	3.365	2.571	2.015	1.476	.727
6	3.707	3.143	2.447	1.943	1.440	.718
7	3.500	2.998	2.365	1.895	1.415	.711

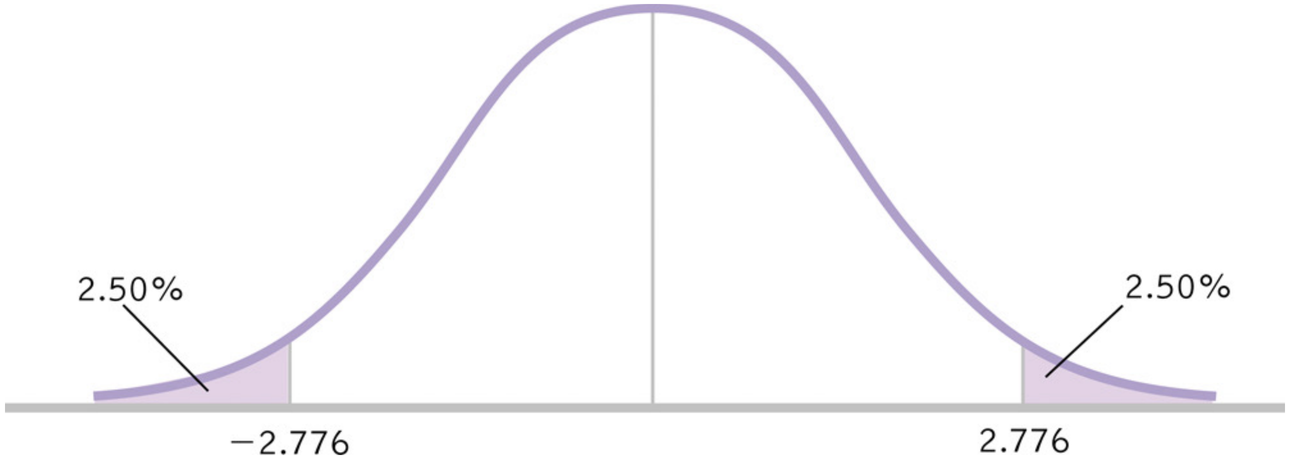
df = 4 →

Paired Samples *t* Test: Does Studying in the Exam Room Help?

$$\mu = 0, s_{\bar{x}} = 1.923,\ \bar{x} = -11,\ n = 5,\ df = 4$$

4. Determine critical value (cutoffs)

$t_{crit} = \pm\ 2.776$



Paired Samples t Test:
Does Studying in the Exam Room Help?

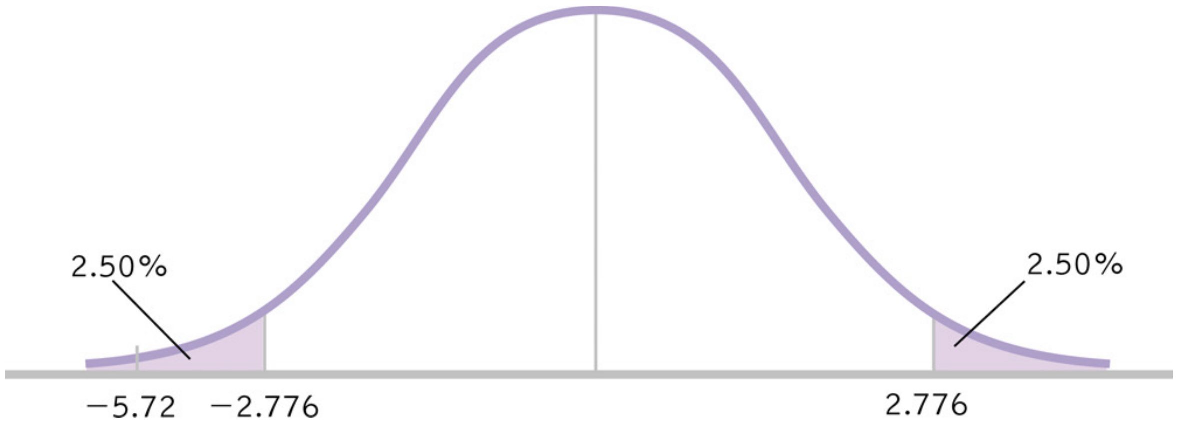
$$\mu = 0, s_{\bar{x}} = 1.923,\ \bar{x} = -11,\ n = 5,\ df = 4$$

5. Calculate the test statistic

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{-11 - 0}{1.923} = -5.720$$

Paired Samples *t* Test:
Does Studying in the Exam Room Help?

$$\mu = 0, s_{\bar{x}} = 1.923,\ \bar{x} = -11,\ n = 5,\ df = 4$$



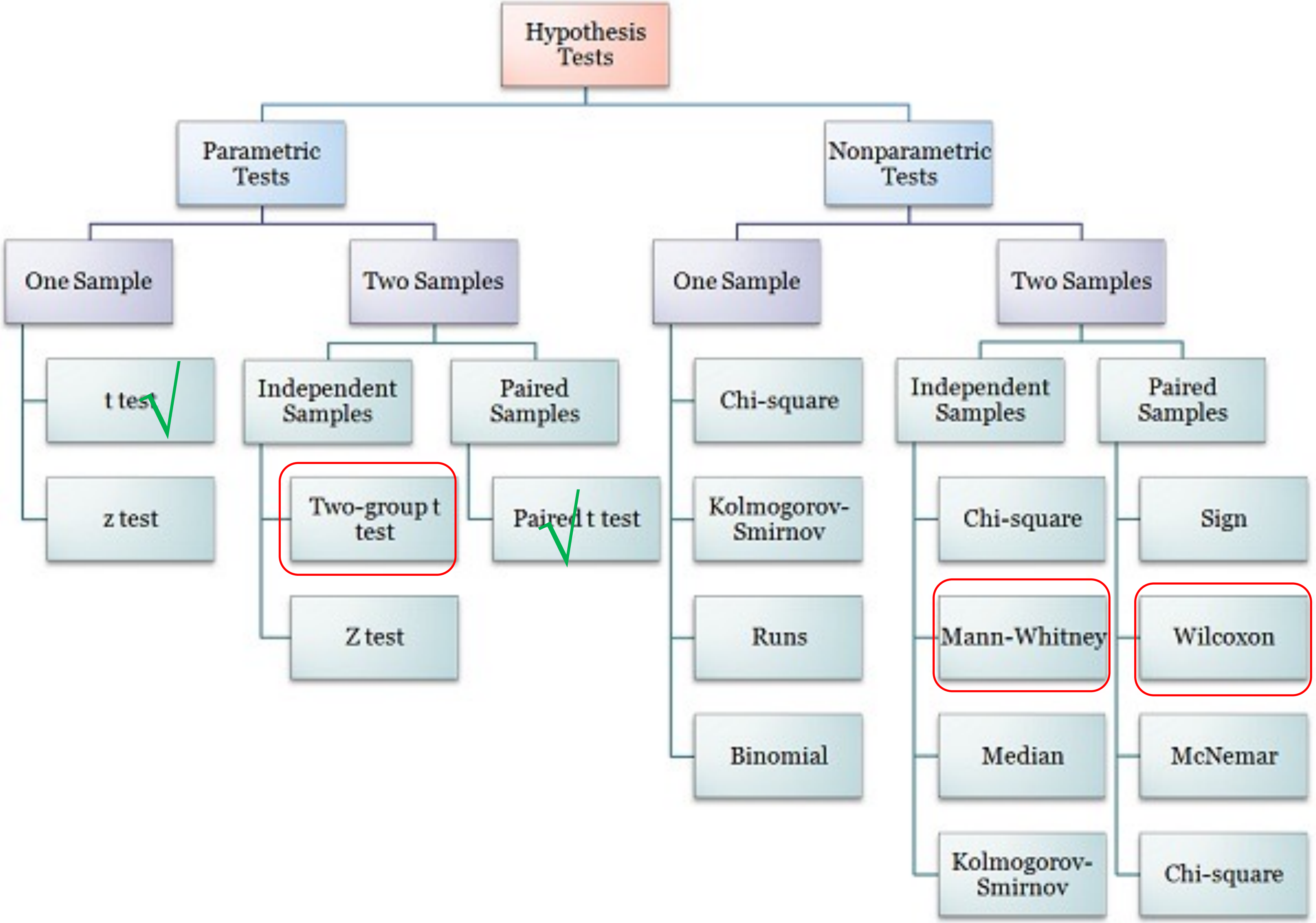
6. Make a decision

   t = -5.720< t_{crit}= -2.776, *reject the null hypothesis*

   People studying and testing in different rooms performed <u>**differently**</u> than in the same rooms,
   $t(4) = -5.720$, $p < .05$.

Parametric and Non-Parametric Tests

- **Parametric Tests:** assume the distribution of the sample data (i.e., normality)
- **Non-Parametric Tests: do not assume** data are drawn from any particular distribution



Independent Samples *t* Test

- Used to **compare two means** for a between-groups design, a situation in which each participant is assigned to only one condition.

$$t = \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{s_{Difference}} = \frac{\bar{x} - \bar{y}}{s_{Difference}}$$

- New Statistics & Terminology:
  - ❖Distribution of Differences Between Means
  - ❖df_x , df_y , df_{Total}
  - ❖Pooled Variance
  - ❖Standard Error of the Mean Difference

Independent Samples *t* Test: Gender Differences in Humor Appreciation

- I tend to believe that very few differences exist between males and females in cognitive abilities but there is some evidence that there are gender differences in, for example, humor appreciation.

- In this hypothetical study we ask: **what percentage of cartoons do men and woman consider funny?** We recruited **9** people from the psychology subject pool and asked them to view a cartoon. After the cartoon, each participant gave us a humor rating of the cartoon, from 0-100 (100 being the funniest possible). Here are those data.

Women (X)	Men (Y)
84	88
97	90
58	52
90	97
	86

Class activity: Six Steps for Hypothesis Testing

1. Identify
2. State the hypotheses
3. Characteristics of the comparison distribution
4. Critical values
5. Calculate
6. Decide

Independent Samples *t* Test: Gender Differences in Humor Appreciation

1. Identify
   - Populations:
     - Pop 1: **Women** exposed to the cartoon
     - Pop 2: **Men** exposed to the cartoon
   - Distribution:
     - Difference Between Means: Distribution of Differences Between Means
       - Not Distribution of <u>Mean Differences</u>
   - Test & Assumptions: Two groups of participants that are studied at a same time point, independent-samples t test!

2. State the null and research hypotheses

H_0: Women will categorize the **same** number of cartoons as funny as will men.

H_a: Women will categorize a **different** number of cartoons funny than will men.

Independent Samples *t* Test: Gender Differences in Humor Appreciation

3. Determine characteristics of comparison distribution (distribution of differences between means)
   - Population: $\mu_x = \mu_y$ (i.e., no difference between means)

$$t = \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{s_{Difference}} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2_{Pooled}(\frac{1}{m} + \frac{1}{n})}}$$

$$s^2_{Pooled} = \frac{df_x s^2_x + df_y s^2_y}{df_{Total}} = (\frac{df_x}{df_{Total}}) s^2_x + (\frac{df_y}{df_{Total}}) s^2_y$$

Independent Samples *t* Test: Gender Differences in Humor Appreciation

3. Determine characteristics of comparison distribution

$$s_{Difference}$$

- Standard Error of the Difference:
  a) Calculate **<u>variance</u>** for each sample
  b) Pool variances, accounting for sample size
  c) Convert from squared standard deviation to squared standard error
  d) Add the two variances
  e) Take square root to get estimated standard error for distribution of differences between means.

$$t = \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{s_{Difference}} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2_{Pooled}\left(\frac{1}{m} + \frac{1}{n}\right)}}$$

$$s^2_{Pooled} = \frac{df_x s_x^2 + df_y s_y^2}{df_{Total}} = \left(\frac{df_x}{df_{Total}}\right)s_x^2 + \left(\frac{df_y}{df_{Total}}\right)s_y^2$$

Calculating $s_{Difference}$

a) Calculate variance for each sample

Women (x)	$x-\bar{x}$	$(x-\bar{x})^2$
84	1.75	3.063
97	14.75	217.563
58	-24.25	588.063
90	7.75	60.063

$\bar{x} = 82.25$

SS_X = 868.752

Men (y)	$y-\bar{y}$	$(y-\bar{y})^2$
88	5.4	29.16
90	11.4	129.96
52	-30.6	936.36
97	14.4	207.36
86	3.4	11.56

$\bar{y} = 82.6$

SS_Y = 1314.4

$$s_x{}^2 = \frac{\sum (x-\bar{x})^2}{m-1} = \frac{868.752}{4-1} = 289.584$$

$$s_y{}^2 = \frac{\sum (y-\bar{y})^2}{n-1} = \frac{1314.4}{5-1} = 328.6$$

Independent Samples *t* Test

- Degrees of Freedom

Women (X)
84
97
58
90
m = 4

$$df_x = m - 1 = 4 - 1 = 3$$

$$df_x = 3$$

$$df_y = n - 1 = 5 - 1 = 4$$

$$df_y = 4$$

$$df_{Total} = df_x + df_y = 3 + 4 = 7$$

$$df_{Total} = 7$$

Men (Y)
88
90
52
97
86
n = 5

Pooled Variance

b) Pool variances, accounting for sample size

- *Weighted average of the two estimates of variance – one from each sample – that are calculated when conducting* ***an independent samples t test****.*

$$s_{Pooled}^2 = \left(\frac{df_x}{df_{Total}}\right)s_x^2 + \left(\frac{df_y}{df_{Total}}\right)s_y^2$$

$$= \left(\frac{3}{7}\right)\times 289.584 + \left(\frac{4}{7}\right)\times 328.6$$

$$= 124.107 + 187.771 = 311.878$$

Independent Samples *t* Test:
Gender Differences in Humor Appreciation

c) Convert from squared standard deviation to squared standard error

$$s^2_{Pooled} = 311.878$$

$$S^2_{\bar{x}} = \frac{s^2_{Pooled}}{m} = 77.970 \qquad S^2_{\bar{y}} = \frac{s^2_{Pooled}}{n} = \frac{311.878}{5} = 62.376$$

d) Add the two variances

$$S^2_{Difference} = S^2_{\bar{x}} + S^2_{\bar{y}} = 77.970 + 62.376 = 140.346$$

e) Take square root to get estimated standard error for distribution of differences between means.

$$S_{Difference} = \sqrt{S^2_{Difference}} = \sqrt{140.346} = 11.847$$

Independent Samples *t* Test:
Gender Differences in Humor Appreciation

4. Determine critical value (cutoffs)
   - ➢ In Behavioral Sciences, we use p = .05 (5%)
   - ➢ Our hypothesis ("Women will categorize a different number of cartoons funny than will men.") is ***nondirectional*** so our hypothesis test is ***two-tailed***.

$t = \pm 2.365$

$df_{Total} = 7$

Significance level = α

Degrees of Freedom	.005 (1-tail) .01 (2-tails)	.01 (1-tail) .02 (2-tails)	.025 (1-tail) .05 (2-tails)	.05 (1-tail) .10 (2-tails)	.10 (1-tail) .20 (2-tails)	.25 (1-tail) .50 (2-tails)
1	63.657	31.821	12.706	6.314	3.078	1.000
2	9.925	6.965	4.303	2.920	1.886	.816
3	5.841	4.541	3.182	2.353	1.638	.765
4	4.604	3.747	2.776	2.132	1.533	.741
5	4.032	3.365	2.571	2.015	1.476	.727
6	3.707	3.143	2.447	1.943	1.440	.718
7	3.500	2.998	$\boxed{2.365}$	1.895	1.415	.711
8	3.355	2.896	2.306	1.860	1.397	.706
9	3.250	2.821	2.262	1.833	1.383	.703
10	3.169	2.764	2.228	1.812	1.372	.700
11	3.106	2.718	2.201	1.796	1.363	.697

Independent Samples *t* Test: Gender Differences in Humor Appreciation

4. Determine critical values
   - $df_{Total} = 7$ $p = .05$ $t = \pm 2.365$

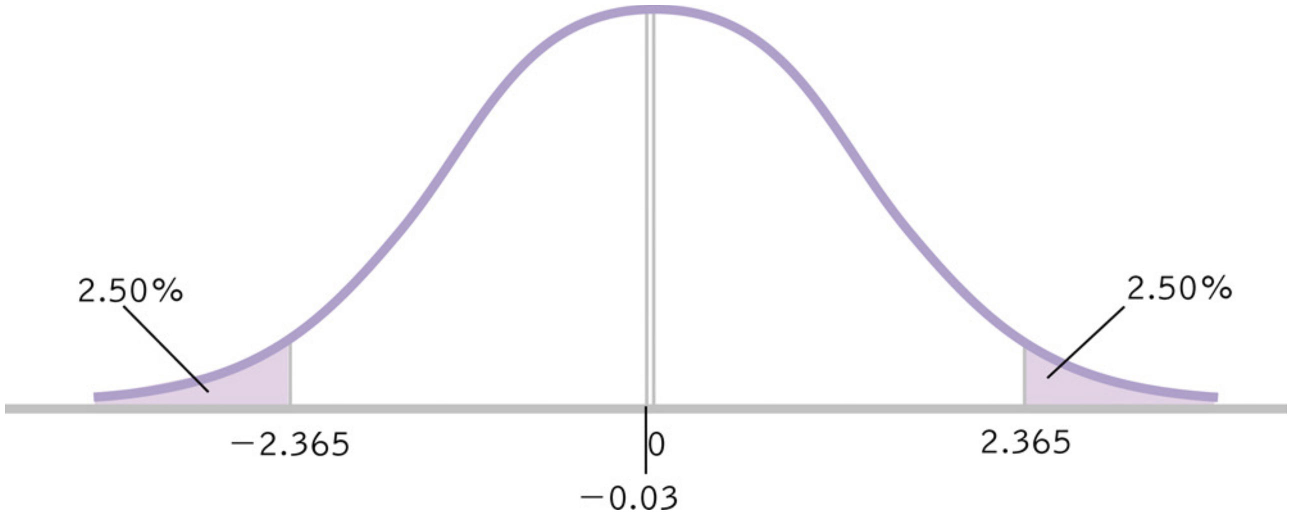
5. Calculate a test statistic

$$t = \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{s_{Difference}} = \frac{\bar{x} - \bar{y}}{s_{Difference}}$$

$$t = \frac{(82.25 - 82.6)}{11.847} = -.03$$

Independent Samples *t* Test:
Gender Differences in Humor Appreciation

6. Make a decision



➢ Fail to reject null hypothesis

❖ Men and women find cartoons equally humorous, $t(7) = -0.03$, $p > 0.05$

Exercise for t-test

Exercise: Are different diet gain the same weight?

Suppose we put people on 2 diets: the **pizza diet** and the **beer diet**. Participants are randomly assigned to either 1-week of eating exclusively pizza or 1-week of exclusively drinking beer. At the end of the week, we measure weight gain by each participant. Data for pizza and beer diets are listed below.

Do pizza and beer cause the same weight gain?

$$t = \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{S_{Difference}} = \frac{\bar{x} - \bar{y}}{\sqrt{\sigma^2_{Pooled}(\frac{1}{m} + \frac{1}{n})}}$$

$$\sigma^2_{Pooled} = \frac{df_x \sigma_x^2 + df_y \sigma_y^2}{df_{Total}} = (\frac{df_x}{df_{Total}})\sigma_x^2 + (\frac{df_y}{df_{Total}})\sigma_y^2$$

Pizza (*X*)	Beer (*Y*)
1	3
2	4
2	4
2	4
3	5

Exercise: Are different diet gain the same weight?

1. Identify
2. State the hypotheses
3. Characteristics of the comparison distribution
4. Critical values
5. Calculate
6. Decide

Exercise: Are different diet gain the same weight?

1. Identify
   - Populations:
     - Pop 1: Participants who **eat pizza** exclusively for a week
     - Pop 2: Participants who **drink beer** exclusively for a week
   - Distribution:
     - Difference Between Means: Distribution of Differences Between Means
       - Not Distribution of <u>Mean Differences</u>
   - Test & Assumptions: Two group of participants have no weight gain difference, Independent Samples t

Exercise: Are different diet gain the same weight?

2. State the null and research hypotheses

- H_0: Participants eat pizza exclusively will **gain the same weight** as those drink beer.
- H_a: Participants eat pizza exclusively will **not gain the same weight** as those drink beer.

Exercise: Are different diet gain the same weight?

3. Determine characteristics of comparison distribution (distribution of differences between means)

➢ Population: $\mu_x = \mu_y$ (i.e., no difference between means)

$$t = \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{S_{Difference}} = \frac{\bar{x} - \bar{y}}{\sqrt{\sigma^2_{Pooled}(\frac{1}{m} + \frac{1}{n})}}$$

$$\sigma^2_{Pooled} = \frac{df_x\sigma_x^2 + df_y\sigma_y^2}{df_{Total}} = (\frac{df_x}{df_{Total}})\sigma_x^2 + (\frac{df_y}{df_{Total}})\sigma_y^2$$

Calculating $s_{Difference}$

a) Calculate variance for each sample

Pizza (X)	$x-\bar{x}$	$(x-\bar{x})^2$
1	-1	1
2	0	0
2	0	0
2	0	0
3	1	1

$\bar{x} = 2$

$$\sigma_x^{\;2} = \frac{\sum (x-\bar{x})^2}{m-1} = \frac{2}{5-1} = 0.5$$

Beer (Y)	$y-\bar{y}$	$(y-\bar{y})^2$
3	-1	1
4	0	0
4	0	0
4	0	0
5	1	1

$\bar{y} = 4$

$$\sigma_y^{\;2} = \frac{\sum (y-\bar{y})^2}{n-1} = \frac{2}{5-1} = 0.5$$

Independent Samples *t* Test

- Degrees of Freedom

Pizza (X)
1
2
2
2
3

$m = 5$

$$df_x = m - 1 = 5 - 1 = 4$$

$$df_x = 4$$

$$df_y = n - 1 = 5 - 1 = 4$$

$$df_y = 4$$

$$df_{Total} = df_x + df_y = 4 + 4 = 8$$

$$df_{Total} = 8$$

Beer (Y)
3
4
4
4
5

$n = 5$

Pooled Variance

b) Pool variances, accounting for sample size

- *Weighted average of the two estimates of variance – one from each sample – that are calculated when conducting an independent samples t test.*

$$\sigma^2_{Pooled} = \left(\frac{df_x}{df_{Total}}\right)\sigma^2_x + \left(\frac{df_y}{df_{Total}}\right)\sigma^2_y$$

$$= \left(\frac{4}{8}\right) \times 0.5 + \left(\frac{4}{8}\right) \times 0.5$$

$$= 0.5$$

Independent Samples *t* Test: Are different diet gain the same weight?

c) Convert from squared standard deviation to squared standard error

$$\sigma^2_{Pooled} = 0.5$$

$$\sigma^2_{\bar{x}} = \frac{\sigma^2_{Pooled}}{m} = \frac{0.5}{5} = 0.1 \qquad \sigma^2_{\bar{y}} = \frac{\sigma^2_{Pooled}}{n} = \frac{0.5}{5} = 0.1$$

d) Add the two variances

$$\sigma^2_{Difference} = \sigma^2_{\bar{x}} + \sigma^2_{\bar{y}} = 0.1 + 0.1 = 0.2$$

e) Take square root to get estimated standard error for distribution of differences between means.

$$\sigma_{Difference} = \sqrt{\sigma^2_{Difference}} = \sqrt{0.2} = 0.4472136$$

Independent Samples *t* Test: Are different diet gain the same weight?

4. Determine critical value (cutoffs)

- In Behavioral Sciences, we use α = .05 (5%)
- Our hypothesis is ***nondirectional***, so our hypothesis test is ***two-tailed***.

$t = \pm 2.306$

$df_{Total} = 8$ →

Significance level = α

Degrees of Freedom	.005 (1-tail) .01 (2-tails)	.01 (1-tail) .02 (2-tails)	.025 (1-tail) .05 (2-tails)	.05 (1-tail) .10 (2-tails)	.10 (1-tail) .20 (2-tails)	.25 (1-tail) .50 (2-tails)
1	63.657	31.821	12.706	6.314	3.078	1.000
2	9.925	6.965	4.303	2.920	1.886	.816
3	5.841	4.541	3.182	2.353	1.638	.765
4	4.604	3.747	2.776	2.132	1.533	.741
5	4.032	3.365	2.571	2.015	1.476	.727
6	3.707	3.143	2.447	1.943	1.440	.718
7	3.500	2.998	2.365	1.895	1.415	.711
8	3.355	2.896	$\boxed{2.306}$	1.860	1.397	.706
9	3.250	2.821	2.262	1.833	1.383	.703
10	3.169	2.764	2.228	1.812	1.372	.700
11	3.106	2.718	2.201	1.796	1.363	.697

Independent Samples *t* Test: Which diet gain more weight?

4. Determine critical values
   - $df_{Total} = 8$ $\alpha = .05$ $t = \pm 2.306$

5. Calculate a test statistic

$$t = \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{S_{Difference}} = \frac{\bar{x} - \bar{y}}{S_{Difference}}$$

$$t = \frac{2-4}{0.4472136} = -4.472136$$

Independent Samples *t* Test: Are different diet gain the same weight?

6. Make a decision

$$t=-4.472136<-2.306$$

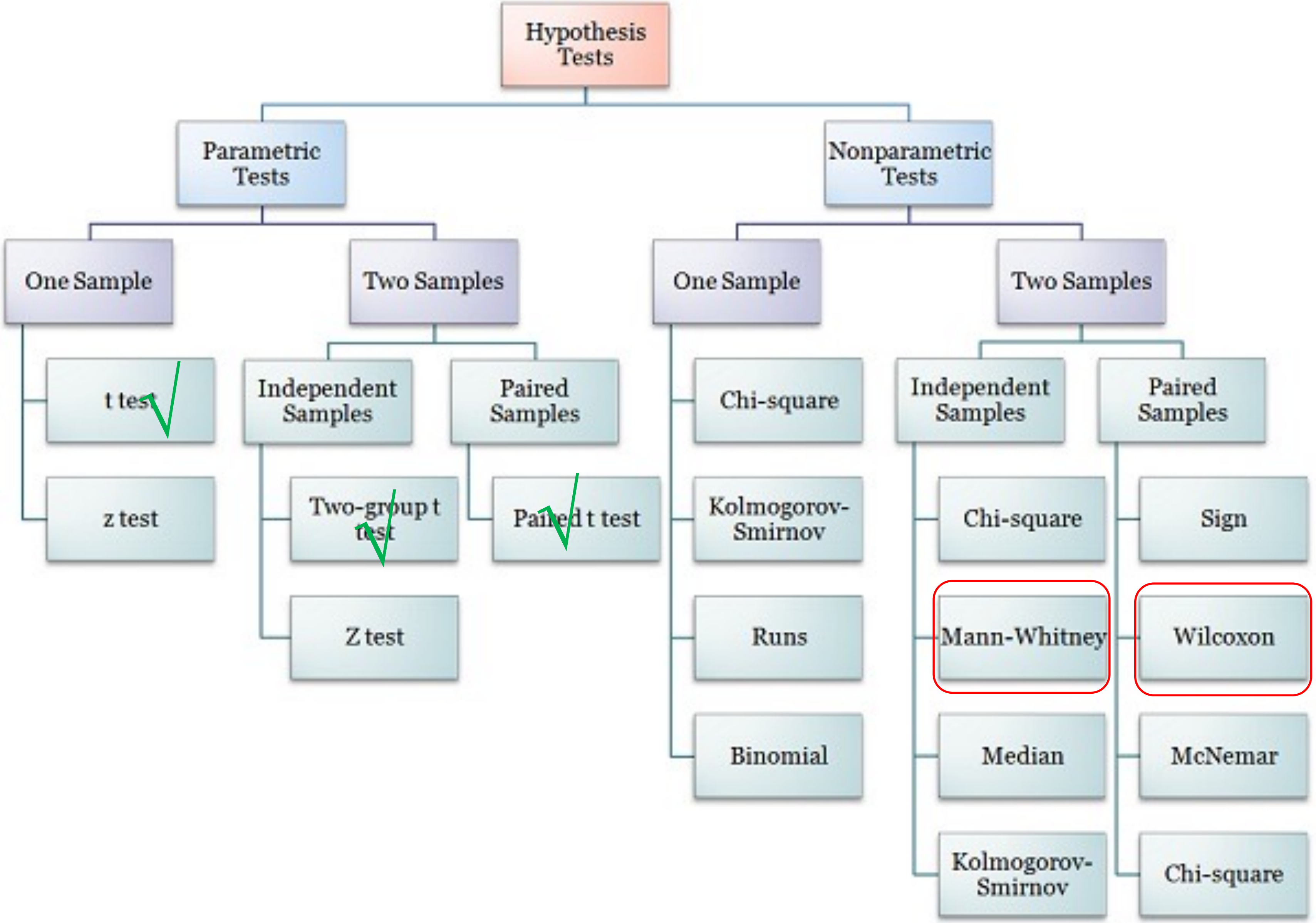
- ➢ Reject null hypothesis at α=0.05
  - ❖ Eat pizza has different weight gain with drink beer

Outline

- Analyzing Differences Among Groups
- Hypothesis Testing
  - Parametric tests
  - Non-parametric tests
  - ANOVA
- Correlation

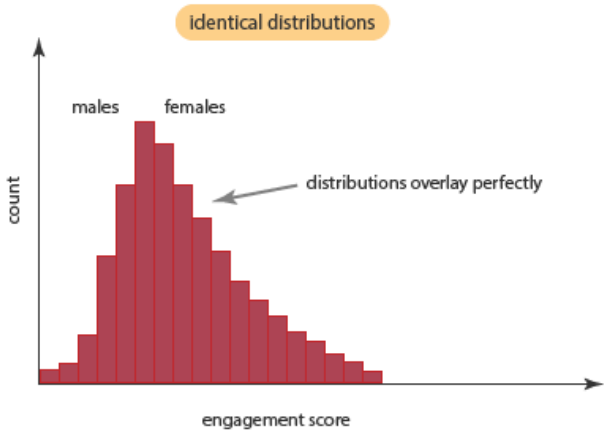
Parametric and Non-Parametric Tests

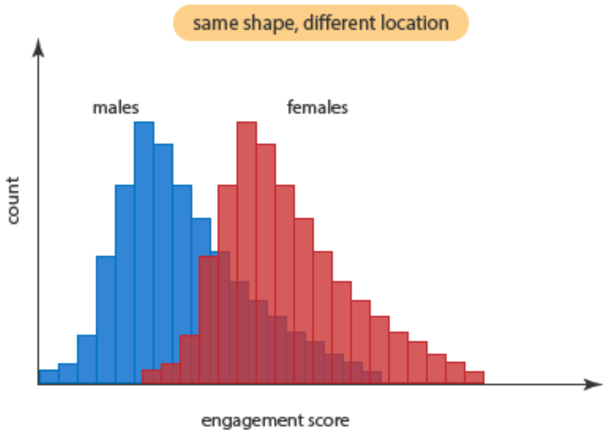
- **Parametric Tests:** assume the distribution of the sample data (i.e., normality)
- **Non-Parametric Tests: do not assume** data are drawn from any particular distribution



Rationale of Mann-Whitney U

- Imagine two samples of scores drawn at random from the **same population**
- The two samples are combined into one larger group and then ranked from lowest to highest
- In this case there should be a **similar number** of high and low ranked scores in each original group
- If however, the two samples are from **different populations** with different **medians** then most of the scores from one sample will be lower in the ranked list than most of the scores from the other sample
  - the sum of ranks in each group will differ





Mann-Whitney U test: ranking the data

- Used for 2 condition independent samples
- To avoid making the assumptions about the data that are made by parametric tests, the Mann-Whitney U test first converts the data to ranks

Sample 1		Sample 2	
Score	Rank 1	Score	Rank 2
7	3	6	2
13	8	12	7
8	4	4	1
9	5.5	9	5.5

Mann-Whitney U test: ranking the data

Step 1: Rank all scores together, ignoring which group they belong to

Sample 1		Sample 2	
Score	Rank 1	Score	Rank 2
7	3	6	2
13	8	12	7
8	4	4	1
9	5.5	9	5.5

Tied scores take the mean of the ranks they occupy. In this example, ranks 5 and 6 are shared in this way between 2 scores. (Then the next highest score is ranked 7)

Mann-Whitney U test: sum of ranks

Step2: Add up the ranks for Sample 1 and Sample 2

Sample 1		Sample 2	
Score	Rank 1	Score	Rank 2
7	3	6	2
13	8	12	7
8	4	4	1
9	5.5	9	5.5
Sum of ranks	**T_1=20.5**		**T_2=15.5**

Mann-Whitney U test: Step3-4

- Step3: Select the larger one, and call it T_X. In this case T_X is the rank total for Sample1, which is 20.5.
- Step4: Calculate N_1, N_2 and N_X.
  - N_1 is the number of scores in the group that gave you the T_1 rank total (Sample1)
    - $N_1=4$
  - N_2 is the number of scores in the group that gave you the T_2 rank total (Sample2)
    - $N_2=4$
  - N_X is the number of scores in the group that gave the larger rank total, Tx (in this case, the number of people in Sample1)
    - $N_x=4$

N_1, N_2 and N_X do not need to be equal!

Mann-Whitney U test: Step5

- Find U by working through the formula below. Remember that T_x is the larger rank total.

$$U = N_1 * N_2 + N_x * \frac{N_x + 1}{2} - T_x$$

$$= 4 * 4 + 4 * \frac{4 + 1}{2} - 20.5 = 5.5$$

Mann-Whitney U test: Step6

Step6: Use a table of critical U values for the Mann-Whitney test

For $N_1 = 4$ and $N_2 = 4$, the critical value of U is 0. To be statistically significant, our obtained U has to be **equal** to or **less** than this critical value. (Note that this is different from many statistical tests, where the obtained value has to be equal to or larger than the critical value)

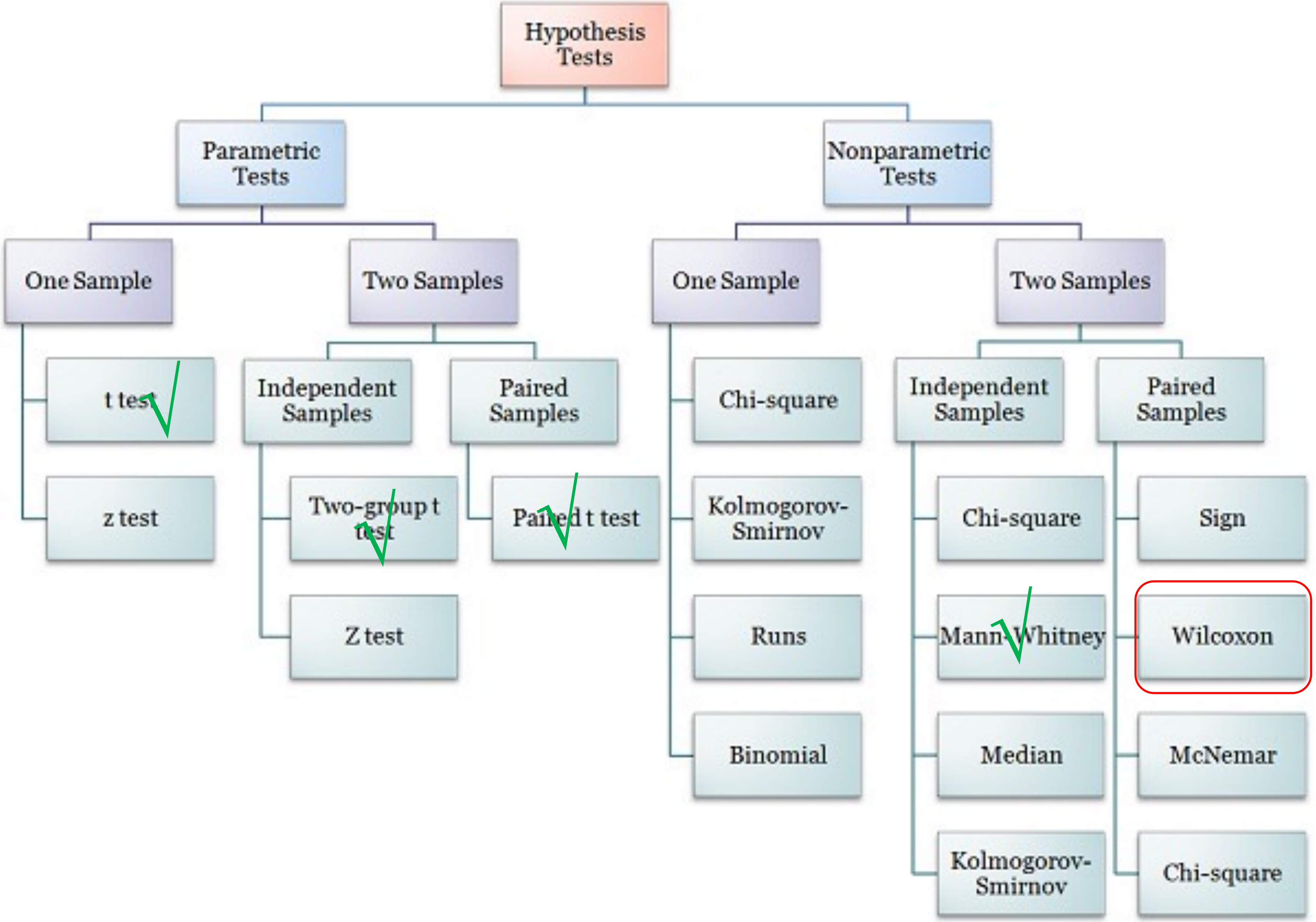
Our obtained U = 5.5, which is larger than 0. Therefore our obtained value of U is likely to occur by chance than the one in the table:
we can conclude that the difference that we have found between the ratings for the two samples **is likely to have occurred by chance**.

Table 3 Critical values of *U* (5% significance).

n_1 \ n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																				
2								0	0	0	0	1	1	1	1	1	2	2	2	2
3					0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4				0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5			0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6			1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7			1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8		0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9		0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10		0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11		0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12		1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13		1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14		1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15		1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16		1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17		2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18		2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19		2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20		2	8	13	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

Parametric and Non-Parametric Tests

- **Parametric Tests:** assume the distribution of the sample data (i.e., normality)
- **Non-Parametric Tests: do not assume** data are drawn from any particular distribution



Wilcoxon signed ranks test

- Use this when the same participants perform both conditions of your study:
  - i.e., it is appropriate for analyzing the data from a repeated-measures design with two conditions.
- Use it when the data do not meet the requirements for a parametric test
  - if the data are not **normally distributed**
  - if the variances for the two conditions are **markedly different**
  - if the data are measurements on an **ordinal scale**

Logic behind the Wilcoxon test

- The data are ranked to produce two rank totals, one for each condition.
- If there is a systematic difference between the two conditions, then most of the ***high ranks*** will belong to one condition and most of the ***low ranks*** will belong to the other one.
  - the rank totals will be quite different and one of the rank totals will be quite small.
  - if the two conditions are similar, then high and low ranks will be ***distributed evenly*** between the two conditions and the rank totals will be fairly similar and quite large.
- The Wilcoxon test statistic "W" is simply the smaller of the rank totals.
  - The SMALLER it is (taking into account how many participants you have) then the less likely it is to have occurred by chance.
  - A table of critical values of W shows you how likely it is to obtain your particular value of W purely by chance.

Wilcoxon test--example

- Suppose we want to know if people's ability to report words accurately was affected by which ear they heard them in.
- To investigate this, we performed a listening task.
  - Each participant heard a series of words, presented randomly to either their left or right ear, and reported the words if they could.
  - Each participant thus provided two scores
    - the number of words that they reported correctly from their left ear
    - the number of words that they reported correctly from their right ear
- Question: Do participants report more words from one ear than the other?

Although the data are measurements on a ratio scale ("number correct" is a measurement on a ratio scale), the data were found to be positively skewed (i.e. not normally distributed) and so we use the Wilcoxon test.

Which test should we use if the data are normally distributed?

Wilcoxon test--step 1-2

Number of words reported:

Participant	Left ear	Right ear
1	25	32
2	29	30
3	10	7
4	31	36
5	27	20
6	24	32
7	27	26
8	29	33
9	30	32
10	32	32
11	20	30
12	5	32

- ***Step1***: Find the difference between each pair of scores.
- ***Step2***: Rank these differences, ignoring any "0" differences and ignoring the sign of the difference (i.e. whether it is a positive or negative difference).

Difference (d)	Ranked difference Based on |d|
-7	7.5
-1	1.5
3	4
-5	6
7	7.5
-8	9
1	1.5
-4	5
-2	3
0	ignore
-10	10
-27	11

Wilcoxon test--step 3-5

- ***Step 3***: Add together the ranks belonging to scores with a positive sign (shaded in the table on the right):
  - 4 + 7.5 + 1.5 = 13
- ***Step 4***: Add together the ranks belonging to scores with a negative sign (unshaded in the table on the right):
  - 7.5 + 1.5 + 6 + 9 + 5 + 3 +10 + 11 = 53
- ***Step 5***: choose W, which is the smaller sums in step3 and step4
  - W = 13

Difference (d)	Ranked differnce
-7	7.5
-1	1.5
3	4
-5	6
7	7.5
-8	9
1	1.5
-4	5
-2	3
0	ignore
-10	10
-27	11

Wilcoxon test--step 6-7

- ***Step 6***: calculate N, which is the number of differences (omitting "0" differences).
  - We have 12 – 1 = 11 differences.
  - this is NOT the same as degrees of freedom. We only use N–1 here because we have one difference which equals zero. if we had two zero differences, we would use N-2, and so on.
- ***Step 7***: Use the table of critical Wilcoxon values.
  - With an N of 11, the critical value for a two-tailed test at the 0.05 significance level is 11.

Table of critical Wilcoxon values

	One Tailed Significance levels:		
	0.025	0.01	0.005
	Two Tailed significance levels:		
N	0.05	0.02	0.01
6	0	-	-
7	2	0	-
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68

Wilcoxon test--step 8

- Step 8: Compare the obtained W with the critical value
  - With the Wilcoxon test, an obtained W is significant if it is **LESS than** or **EQUAL** to the critical value.
  - Our obtained value of 13 is **larger** than 11, and so we can conclude that there is **no significant difference** between the number of words recalled from the right ear and the number of words recalled from the left ear.

We would write this as follows: *“A Wilcoxon test showed that the number of words reported correctly was not significantly affected by which ear they were presented to ($W(11) = 13$, $p > .05$, two tailed test).*"

Outline

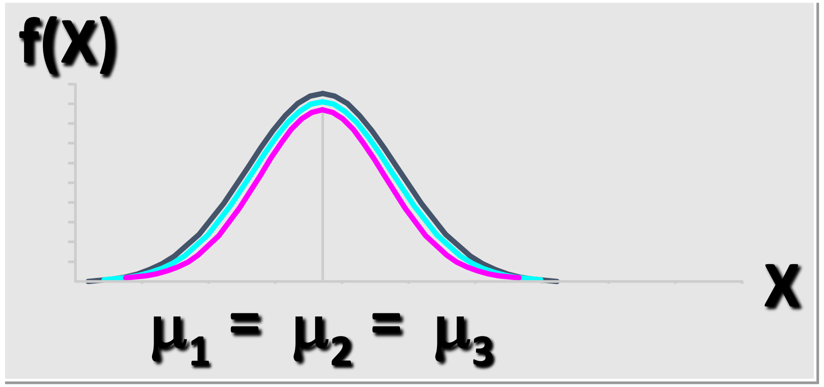
- Analyzing Differences Among Groups
- Hypothesis Testing
  - Parametric tests
  - Non-parametric tests
  - ANOVA
- Correlation

One-Way ANOVA F-Test Hypotheses

- Tests the equality of 2 or more population means (μ)

H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_p$

- All population means are equal
- No treatment effect

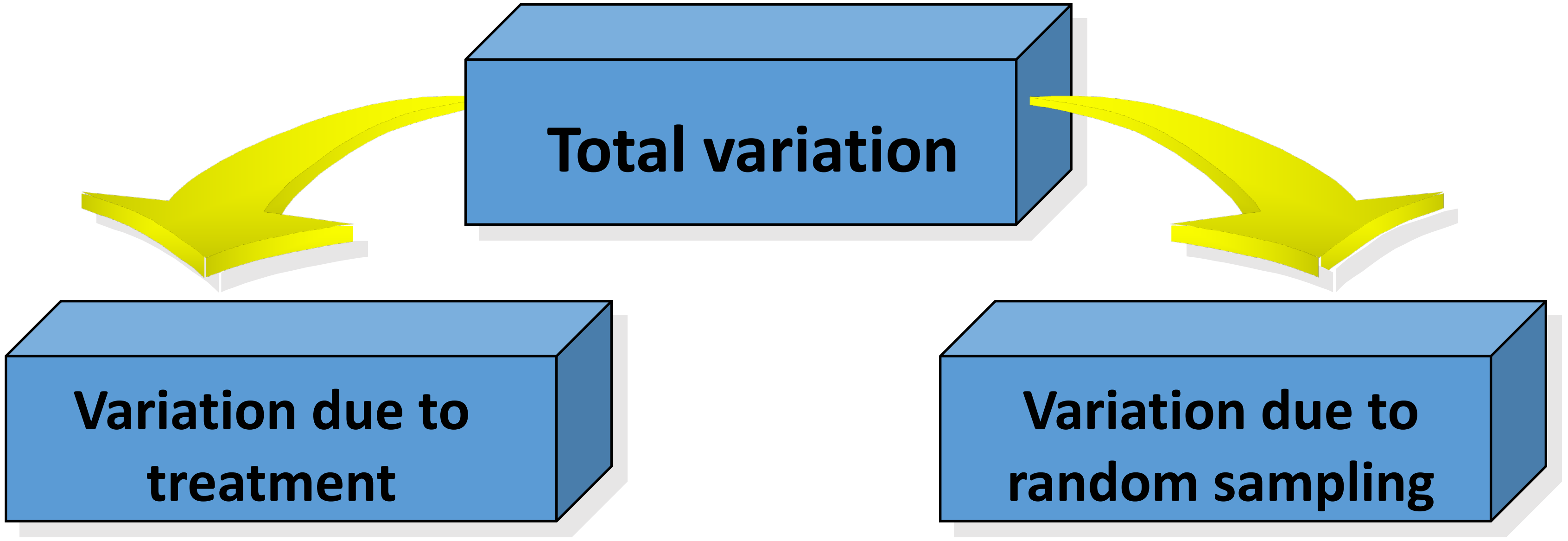


H_a: Not All μ_j Are Equal

- At least 1 population mean is Different
- Treatment Effect
- NOT $\mu_1 = \mu_2 = \ldots = \mu_p$
- **Or $\mu_i \neq \mu_j$ for some *i, j*.**

- Treatment variation between groups could be significantly greater than the in group variation
- Variation measures are obtained by ‘Partitioning’ total variation

One-Way ANOVA: Partitions Total Variation

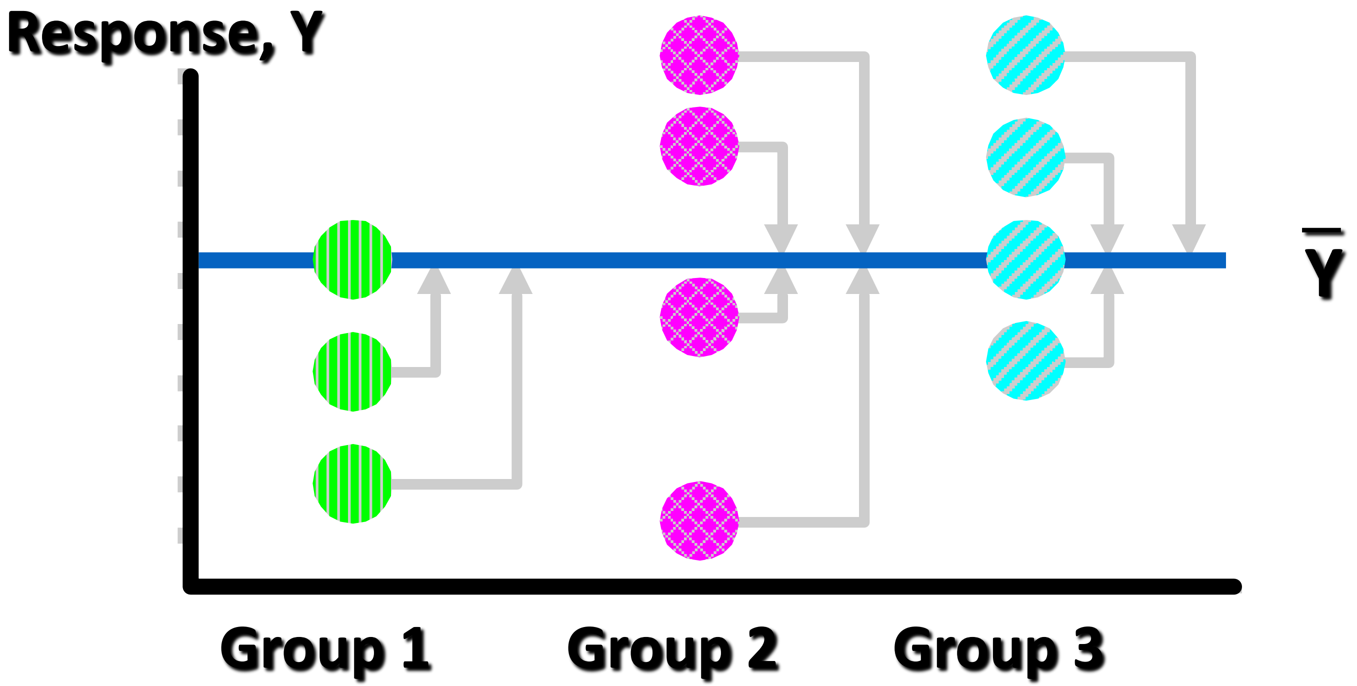


Sum of Squares Treatment (SST)
Between group variation

Sum of Squares Error (SSE)
In group variation

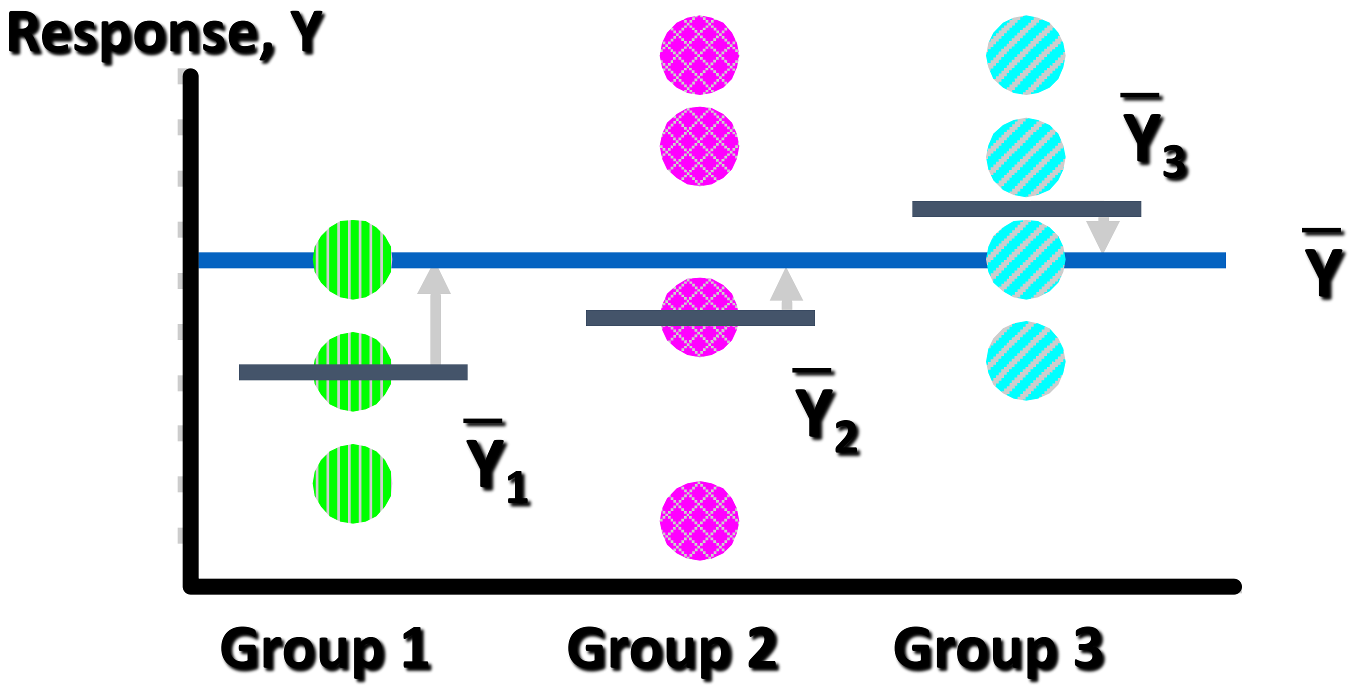
Total Variation

$$SS(Total) = \left(Y_{11} - \bar{Y}\right)^2 + \left(Y_{21} - \bar{Y}\right)^2 + \ldots + \left(Y_{ij} - \bar{Y}\right)^2$$



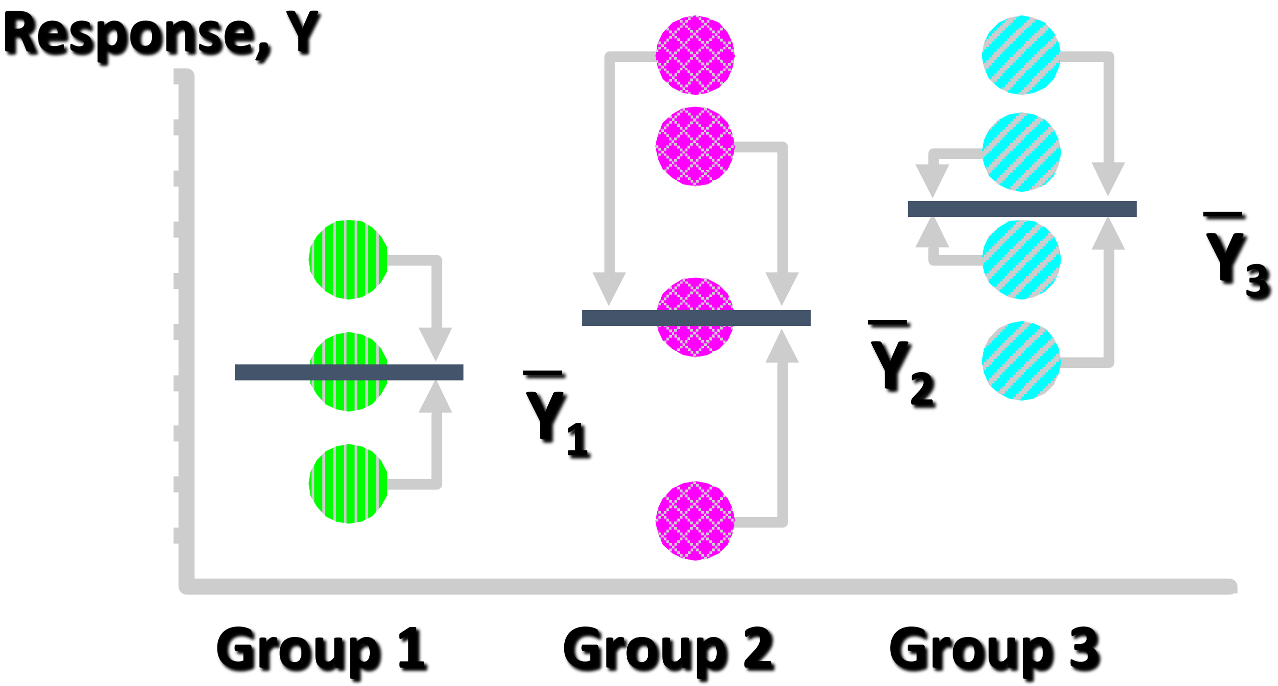
Between group variation

$$SST = n_1\left(\overline{Y}_1 - \overline{Y}\right)^2 + n_2\left(\overline{Y}_2 - \overline{Y}\right)^2 + \ldots + n_p\left(\overline{Y}_p - \overline{Y}\right)^2$$



In group variation

$$SSE = \left(Y_{11} - \bar{Y}_1\right)^2 + \left(Y_{21} - \bar{Y}_1\right)^2 + \ldots + \left(Y_{jp} - \bar{Y}_p\right)^2$$



One-Way ANOVA F-Test: Test Statistic

1. Test Statistic

➢ $F = MST / MSE = \dfrac{SST/(p-1)}{SSE/(n-p)}$

❖ *MST* : mean square for treatment

❖ *MSE* : mean square for error

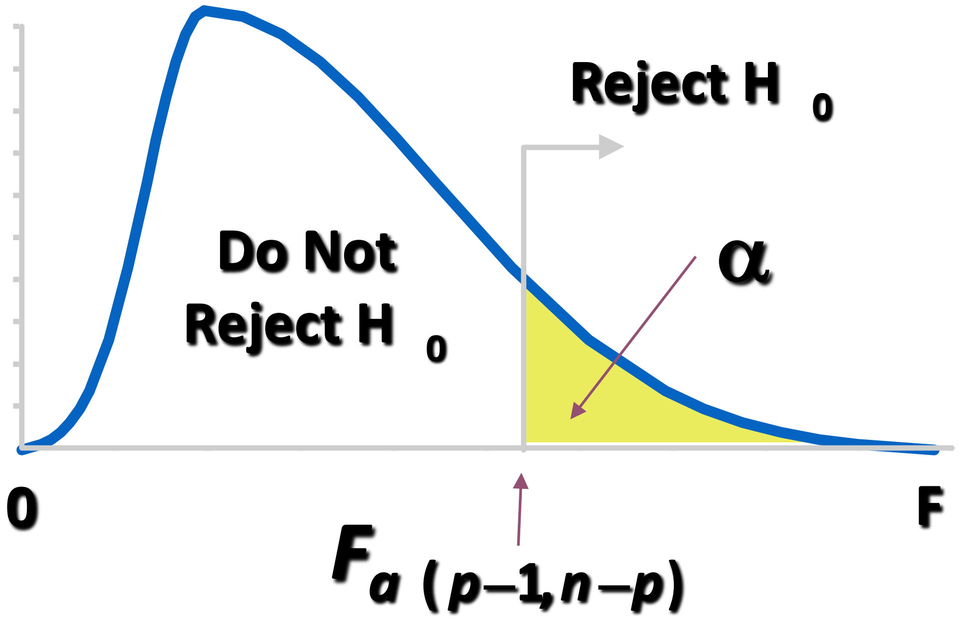
2. Degrees of Freedom

➢ $\nu_1 = p - 1$

➢ $\nu_2 = n - p$

❖ p = Populations, Groups, or Levels

❖ n = Total Sample Size



Always One-Tail!

One-Way ANOVA F-Test Example

- As a vet epidemiologist, you want to see if three food supplements have different mean milk yields.
  - You assign 15 cows, 5 per food supplement.
- Question: At the **.05** level, is there a difference in **mean** yields?

Step 1: Hypothesis

H_0: $\mu_1 = \mu_2 = \mu_3$

H_a: **Not All Equal**

Step 2: Calculate Degrees of Freedom

p=3, n=15

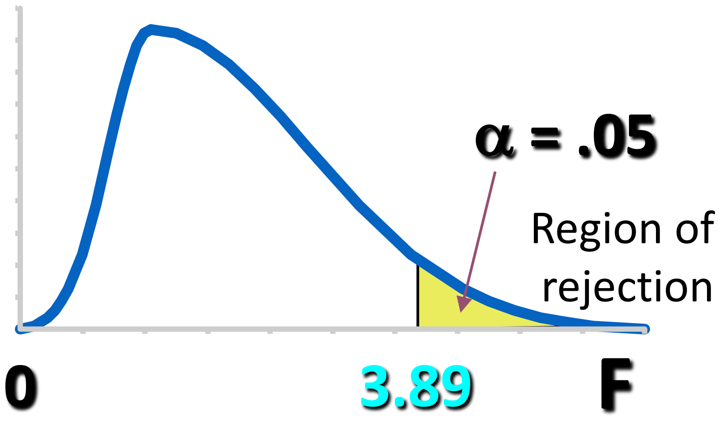
v1=p-1=3-1=2

v2=n-p=15-3=12

Food1	Food2	Food3
25.40	23.40	20.00
26.31	21.80	22.20
24.10	23.50	19.75
23.74	22.75	20.60
25.10	21.60	20.40

One-Way ANOVA F-Test: Example

Step3: Find critical value(s):



Step4: Test Statistic:

$$F = \frac{SST/(p-1)}{SSE/(n-p)}$$

p=3, n=15

$$\bar{Y}_1 = 24.93$$
$$\bar{Y}_2 = 22.61$$
$$\bar{Y}_3 = 20.59$$
$$\bar{Y} = 22.71$$

$$SST = n_1(\bar{Y}_1 - \bar{Y})^2 + n_2(\bar{Y}_2 - \bar{Y})^2 + n_3(\bar{Y}_3 - \bar{Y})^2$$
$$= 5(24.93\text{-}22.71)^2 + 5(22.61 - 22.71)^2 + 5(20.59 - 22.71)^2$$
$$= 47.164$$

$$SSE = (Y_{11} - \bar{Y}_1)^2 + (Y_{21} - \bar{Y}_1)^2 + \ldots + (Y_{51} - \bar{Y}_1)^2$$
$$+ (Y_{12} - \bar{Y}_2)^2 + (Y_{22} - \bar{Y}_2)^2 + \ldots + (Y_{52} - \bar{Y}_2)^2$$
$$+ (Y_{13} - \bar{Y}_3)^2 + (Y_{23} - \bar{Y}_3)^2 + \ldots + (Y_{53} - \bar{Y}_3)^2$$
$$= (25.40 - 24.93)^2 + (26.31 - 24.93)^2 + \ldots + (25.10 - 24.93)^2$$
$$+ (23.40 - 22.61)^2 + (21.80 - 22.61)^2 + \ldots + (21.60 - 22.61)^2$$
$$+ (20.00 - 20.59)^2 + (22.20 - 20.59)^2 + \ldots + (20.40 - 20.59)^2$$
$$= 11.0532$$

$$F = \frac{\frac{47.164}{3-1}}{\frac{11.0532}{12}} = \frac{23.5820}{0.921} = 25.6$$

One-Way ANOVA F-Test: Example

- Step 5: Make decision

  F=25.6>3.89(critical value)

  Reject the null hypothesis at α= .05

- Conclusion: At least one food supplement is different from the others

Table F The F Distribution

$\alpha = .05$

df_D \ df_N	1	2	3	4	5	6	7	8	9	10
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	$\boxed{3.89}$	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Post Hoc Tests

- If the t-test is significant, you have a difference in population means. You know where.

- If the F-test is significant, you have a difference in population means. But you don't know where.

- With 3 means, could be A=B>C or A>B>C or A>B=C.
  - We need a test to tell which means are different.

Tukey HSD(Honestly Significant Difference)
The sizes of three groups should be the same

- Step1: compute all possible absolute differences between means

$$\bar{Y}_1 = 24.93 \qquad \bar{Y}_1 - \bar{Y}_2 = 24.93 - 22.61 = 2.32$$

$$\bar{Y}_2 = 22.61 \qquad \bar{Y}_1 - \bar{Y}_3 = 24.93 - 20.59 = 4.34$$

$$\bar{Y}_3 = 20.59 \qquad \bar{Y}_2 - \bar{Y}_3 = 22.61 - 20.59 = 2.02$$

- Step2: find the critical value q
- Step3: compute HSD
- $HSD = \dfrac{\bar{Y}_i - \bar{Y}_j}{\sqrt{\dfrac{MSE}{n}}} = \dfrac{\bar{Y}_i - \bar{Y}_j}{\sqrt{\dfrac{0.9211}{5}}} = \dfrac{\bar{Y}_i - \bar{Y}_j}{0.43}$ ($\bar{Y}_i$ is always larger than $\bar{Y}_j$)
- Step4: compare absolute differences with HSD
  - HSD(G1 to G2): 5.39>3.77 **different each other**
  - HSD(G1 to G3): 10.09>3.77 **different each other**
  - HSD(G2 to G3): 4.70>3.77 **different each other**

Critical Values for the Tukey Q Test

d.f. = N – K (ANOVA Error or Within d.f.) for ANOVA: Single Factor
d.f. = N – R*C (ANOVA Within d.f.) for ANOVA: Two-Factor With Replication
d.f. = (R – 1)*(C – 1) (ANOVA Error d.f.) for ANOVA: Two-Factor Without Replication

Number of Groups (Treatments) = K

Error df	Number of Groups (Treatments)								
	2	3	4	5	6	7	8	9	10
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07
2	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47

Exercise for ANOVA

Exercise—compare 4 different table designs

- Four types of tables were designed and sold in the same shop. The numbers of order records for each month are listed below.
- Question: At the α=0.05 level, is there a difference in their order records?

Design 1	Design 2	Design 3	Design 4
11	12	23	27
17	10	20	33
16	15	18	22
14	19	17	26
15	11		28

One-Way ANOVA F-Test Example

- We want to see if 4 table designs have different mean order records.
  - You assign 19 records (5,5,4,5).

Step 1: Hypothesis

H_0: $\mu_1 = \mu_2 = \mu_3$

H_a: **Not All Equal**

Design 1	Design 2	Design 3	Design 4
11	12	23	27
17	10	20	33
16	15	18	22
14	19	17	26
15	11		28

Step 2: Calculate Degrees of Freedom

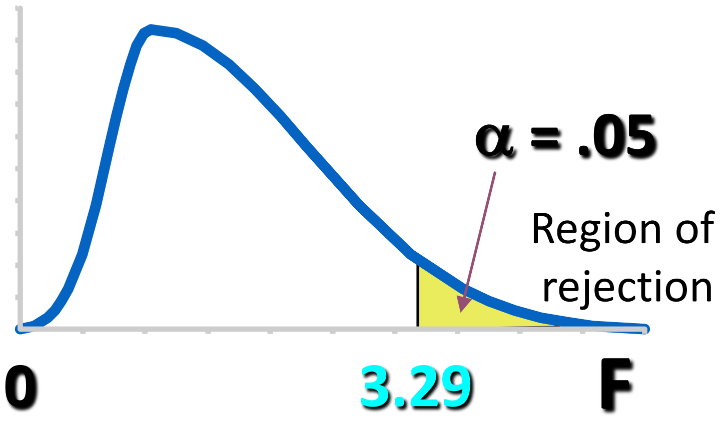
p=4, n=19

v1=p-1=4-1=3

v2=n-p=19-4=15

One-Way ANOVA F-Test: Example

Step3: Find critical value(s):



Step4: Test Statistic:

$$F = \frac{SST/(p-1)}{SSE/(n-p)}$$

p=4, n=19

$$\bar{Y}_1 = 14.6$$

$$\bar{Y}_2 = 13.4$$

$$\bar{Y}_3 = 19.5$$

$$\bar{Y}_4 = 27.2$$

$$\bar{Y} = 18.63158$$

$$SST = n_1\left(\bar{Y}_1 - \bar{Y}\right)^2 + n_2\left(\bar{Y}_2 - \bar{Y}\right)^2 + n_3\left(\bar{Y}_3 - \bar{Y}\right)^2 + n_4\left(\bar{Y}_4 - \bar{Y}\right)^2$$

$$= 5(14.6\text{-}18.63158)^2 + 5(13.4 - 18.63158)^2$$

$$+ 4(19.5 - 18.63158)^2 + 5(27.2 - 18.63158)^2$$

$$= 588$$

$$SSE = \left(Y_{11} - \bar{Y}_1\right)^2 + \left(Y_{21} - \bar{Y}_1\right)^2 + \ldots + \left(Y_{51} - \bar{Y}_1\right)^2$$

$$+ \left(Y_{12} - \bar{Y}_2\right)^2 + \left(Y_{22} - \bar{Y}_2\right)^2 + \ldots + \left(Y_{52} - \bar{Y}_2\right)^2$$

$$+ \left(Y_{13} - \bar{Y}_3\right)^2 + \left(Y_{23} - \bar{Y}_3\right)^2 + \ldots + \left(Y_{43} - \bar{Y}_3\right)^2$$

$$+ \left(Y_{14} - \bar{Y}_4\right)^2 + \left(Y_{24} - \bar{Y}_4\right)^2 + \ldots + \left(Y_{54} - \bar{Y}_4\right)^2$$

$$= 158$$

$$F = \frac{\frac{588}{4-1}}{\frac{158}{15}} = \frac{196}{10.5} = 18.7$$

One-Way ANOVA F-Test: Example

- Step 5: Make decision

  F=18.7>3.29(critical value)

  Reject the null hypothesis at α= .05

  Conclusion:

At least one order record is different from the others

Table F The *F* Distribution

α = .05

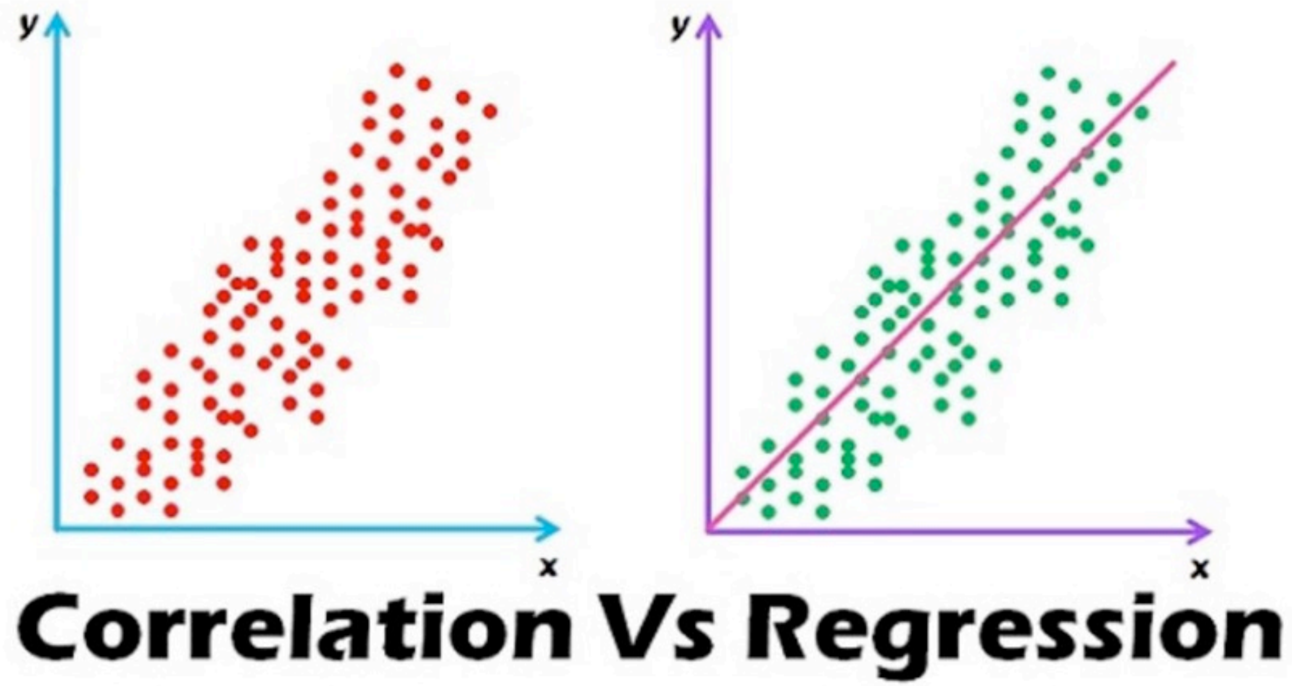
df_D \ df_N	1	2	3	4	5	6	7	8	9	10
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	$\boxed{3.29}$	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Outline

- Analyzing Differences Among Groups
- Hypothesis Testing
  - Parametric tests
  - Non-parametric tests
  - ANOVA
- Correlation

Linear regression and Correlation

- Correlation coefficient: measure of the strength and direction of linear relationship between two quantitative variables
- Linear regression fits a model predicting a quantitative response (dependent) variable (y) based on a quantitative explanatory (independent) variable (x)



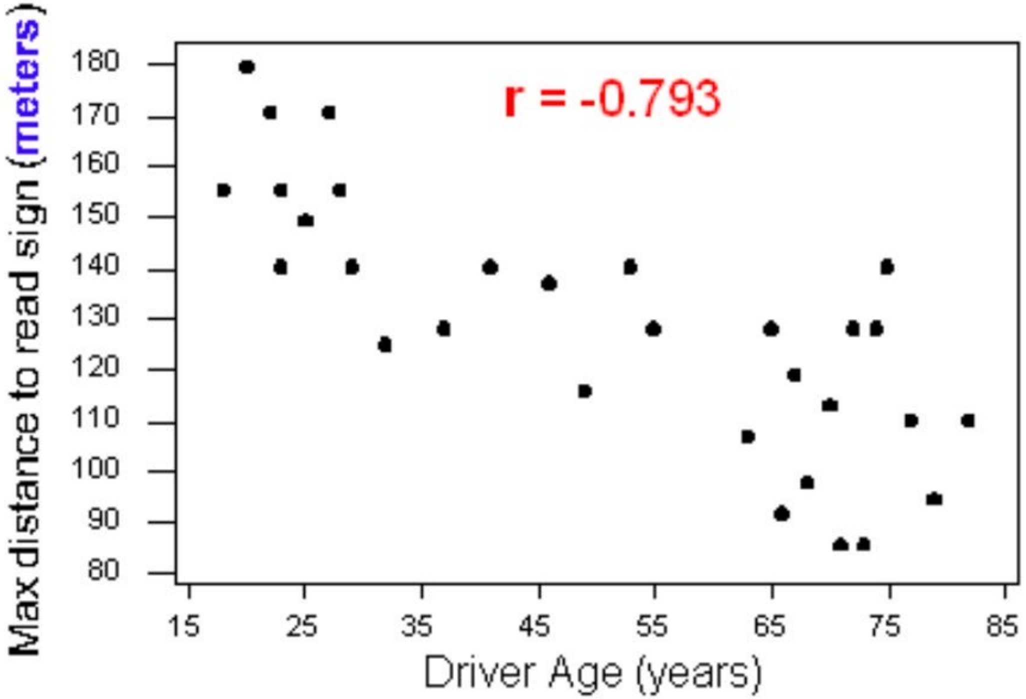
Pearson Correlation Coefficient

- The correlation coefficient (r) is a numeric measure of the strength and direction of a linear relationship between two quantitative variables

$$r_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

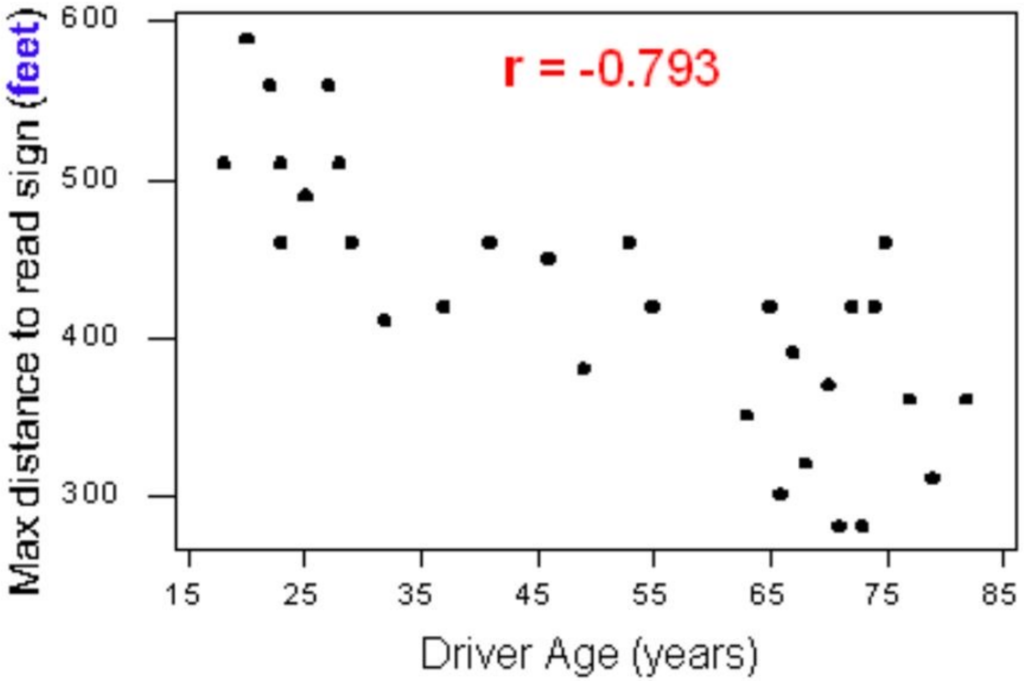
Pearson Correlation Coefficient

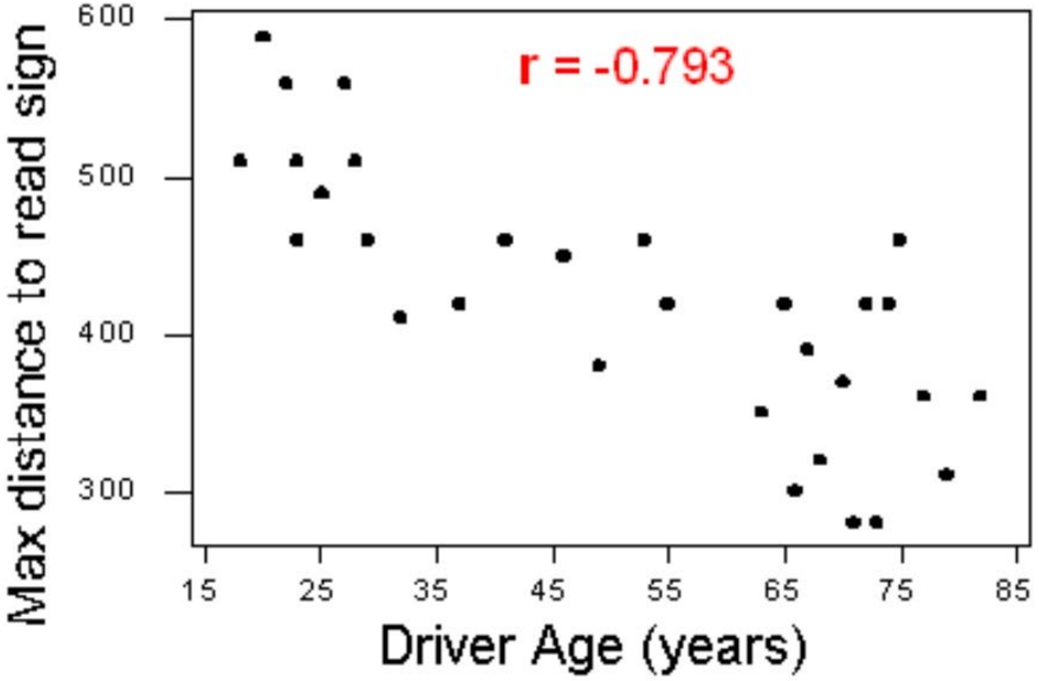
- $-1 \leq r \leq 1$
- The sign indicates the direction of relationship
  - positive relationship: $r > 0$
  - negative relationship: $r < 0$
  - no linear relationship: $r \approx 0$
- The closer r is to ± 1, the stronger the linear relationship
- The correlation between X and Y is the same as the correlation between Y and X



Properties of r

- r has no units and does not depend on the units of measurement

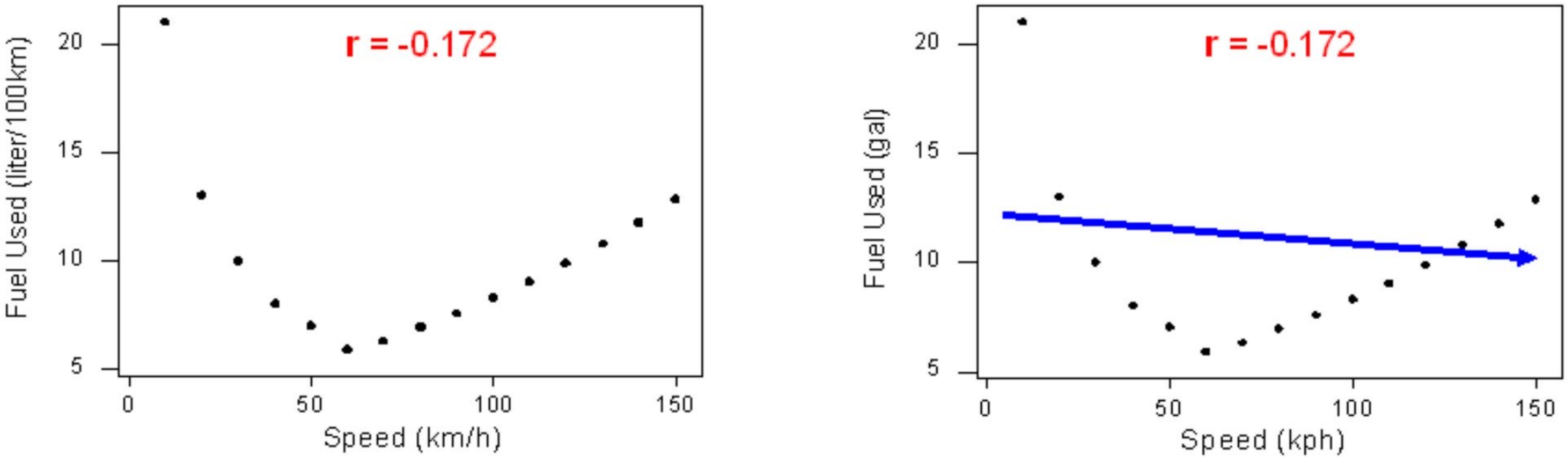




- Notice that the y-values have changed, but the correlations are the same.

Properties of r

- The correlation measures only the strength of a linear relationship between two variables.



- Notice that the correlation r = -0.172 indicates a weak linear relationship. This makes sense because the data does not closely follow a linear form.
- Always make a scatterplot of the data before calculating and interpreting the meaning of r.

Spearman's Correlation Coefficient

- Spearman's rank correlation coefficient can be defined as a special case of Pearson coefficient applied to ranked variables.
  - not restricted to linear relationships.
  - measures monotonic association (only strictly increasing or decreasing, but not mixed) between two ranked variables.
  - In other words, rather than comparing means and variances, Spearman's coefficient looks at the relative order of values for each variable.
  - can be used with both continuous and discrete data.

Steps To Calculate Spearman's Rank Correlation Coefficient

- Step1: Assign ranks 1, 2, 3, ..., n to the value of each variable.
  - Ranking can be descending in order or ascending in order.
  - However, both data sets should use the same ordering.
- Step2: For each pair of values (x, y), we will calculate

  $$d = \text{rank}(x) - \text{rank}(y)$$

  - We call the difference d.

Spearman's Rank Table

Sample size (n)	p = 0.05	p = 0.025	p = 0.01
4	1.0000	-	-
5	0.9000	1.0000	1.0000
6	0.8286	0.8857	0.9429
7	0.7143	0.7857	0.8929
8	0.6429	0.7381	0.8333
9	0.6000	0.7000	0.7833
10	0.5636	0.6485	0.7455
11	0.5364	0.6182	0.7091
12	0.5035	0.5874	0.6783
13	0.4825	0.5604	0.6484
14	0.4637	0.5385	0.6264
15	0.4464	0.5214	0.6036

- Step3:We calculate Spearman's Rank Order Correlation Coefficient as follows:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

- Step4: Compare the obtained r_s with the values in the Spearman's Rank Table. The values in this table are the minimum values of r from a sample that need to be reached for Spearman's Rank Correlation Coefficient value to be significant at the level shown.

Class Activity

- The following is a small selection of countries ranked according to the Human Development Index (HDI) and Income per capita (GDP per capita at PPP in USD). These data came from UNDP-HDR 2003. [Note: Income rank is a rank with respect to these ten countries and does not correspond to the GNP per capita ranking provided by UNDP.]

Country	Income rank (x)	HDI rank (y)	d = rank x - rank y	d^2
Norway	3	1	2	4
Iceland	2	2	0	0
Sweden	10	3	7	49
Australia	8	4	4	16
Netherlands	5	5	0	0
Belgium	7	6	1	1
United States	1	7	-6	36
Canada	6	8	-2	4
Japan	9	9	0	0
Switzerland	4	10	-6	36
				$\sum d^2 = 146$

Class Activity

- We have included column d and column d^2 in the above table.
- The Spearman's Rank Order Correlation Coefficient is:

$$r_s = 1 - \frac{6(146)}{10(10^2 - 1)}$$

$$r_s = 0.115$$

Spearman's Rank Table

Sample size (n)	p = 0.05	p = 0.025	p = 0.01
4	1.0000	-	-
5	0.9000	1.0000	1.0000
6	0.8286	0.8857	0.9429
7	0.7143	0.7857	0.8929
8	0.6429	0.7381	0.8333
9	0.6000	0.7000	0.7833
10	0.5636	0.6485	0.7455
11	0.5364	0.6182	0.7091
12	0.5035	0.5874	0.6783
13	0.4825	0.5604	0.6484
14	0.4637	0.5385	0.6264
15	0.4464	0.5214	0.6036

- In our example, we have a sample size of ten and according to the table, to achieve a significance level of 5% ($p = 0.05$), the r_s value has to be at least 0.5636. $r_s = 0.115$ is smaller than 0.5636, so we conclude that our Spearman's Rank Correlation Coefficient is not statistically different from zero. Therefore, it is not significant.