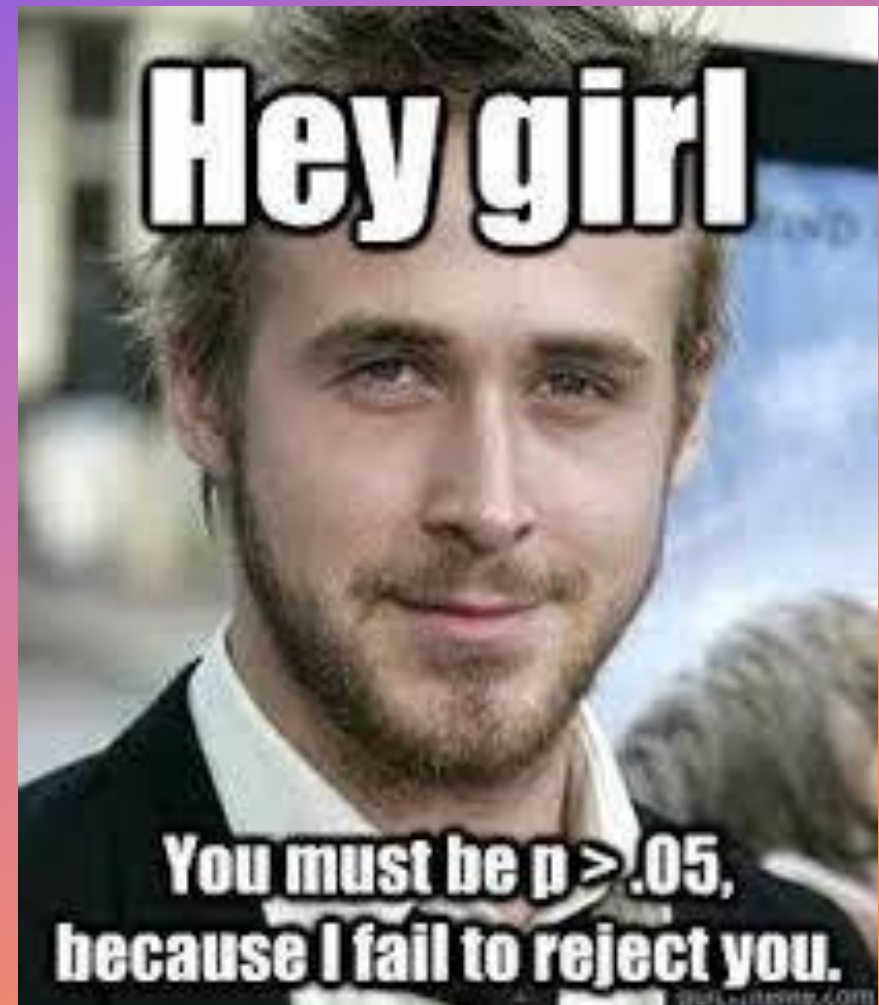




# T-TESTS, ANOVAS AND REGRESSION

Alexandros Abdel Massih  
and Priskila Sophiana



# Table of contents

- **T-Test** (function, computation, how to read result)
- **ANOVAs** (function, limitation, examples, result)
- **Regression** (when to use, linear regression, result)
- **Conclusion**



# T-TEST



# What is the main function of T-Test?

To identify whether there is a **significant difference** between our data and a known mean or between two groups.



Score
650
730
510
670
480
800
690
530
590
620
710
670
640
780
650
490
800
600
510
700



Score
630
720
462
631
440
783
673
519
543
579
677
649
632
768
615
463
781
563
488
650

# When do we use T-Test?

- The **standard deviation** of the population is unknown.
- Our data are **normally distributed, or else:**
  - Mann-Whitney U Test
  - Wilcoxon Signed Ranked Test
- The variable must be **continuous**.
- **Independence of observation** assumed.
- **Homogeneity of variance** assumed (Unpaired sample t-test).
- **No outliers**.

# Step-by-step of doing (manual computation) T-Test?

- State our **hypothesis** (null and alternative hypothesis).
- State **decision rule** (choose significance level, number of tails and degree of freedom).
- Conduct the **t-test computation**.
- **Compare the test result** to the 'decision rule'.
- **Conclusion** (null hypothesis rejected or accepted).



# Type of T-Tests

**One-Sample T-Test**

**Paired-Sample T-Test**

**Unpaired-Sample T-Test**

# One-sample T-Test



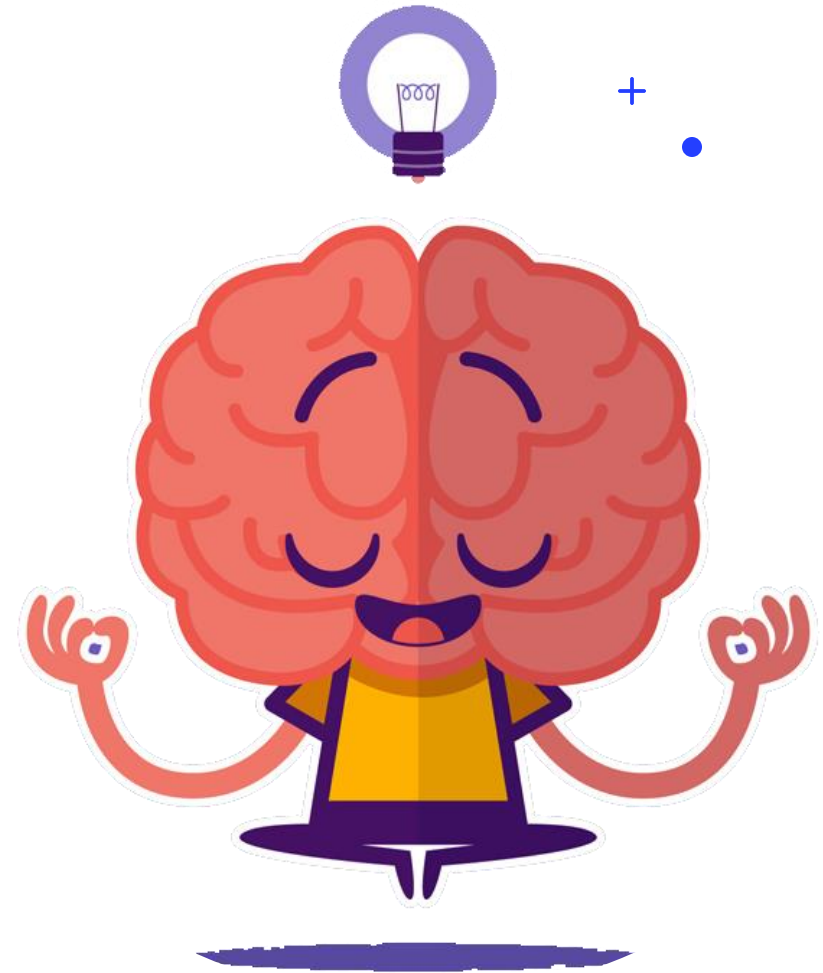
- The '**simplest**' t-test.
- Used to **compare sample mean and another known means** (e.g: compare sample mean to population mean).
- There is only **one group of sample/data** in one sample t-test.



# CASE EXAMPLE:

In the population, the average IQ is 100. A team of researcher wants to test a new medication to see if it has either a positive or negative effect on intelligence, or no effect at all.

A sample of 30 participants who have taken the medication has a mean of 140 with a standard deviation of 20. Did the medication affect intelligence?



# CASE EXAMPLE:

In the population, the average IQ is 100. A team of researcher wants to test a new medication to see if it has either a positive or negative effect on intelligence, or no effect at all. A sample of 30 participants who have taken the medication has a mean of 140 with a standard deviation of 20. Did the medication affect intelligence?

**Let's break  
this down!**



# 1 State Hypothesis

- **Null Hypothesis:** The participant group who have taken the medicine **will have the same** mean ratio as the population with the average mean of 100.
- **Alternative Hypothesis:** The participant group who have taken the medicine will have **significantly different** mean ratio (either lower or higher) than the population with the average mean of 100.



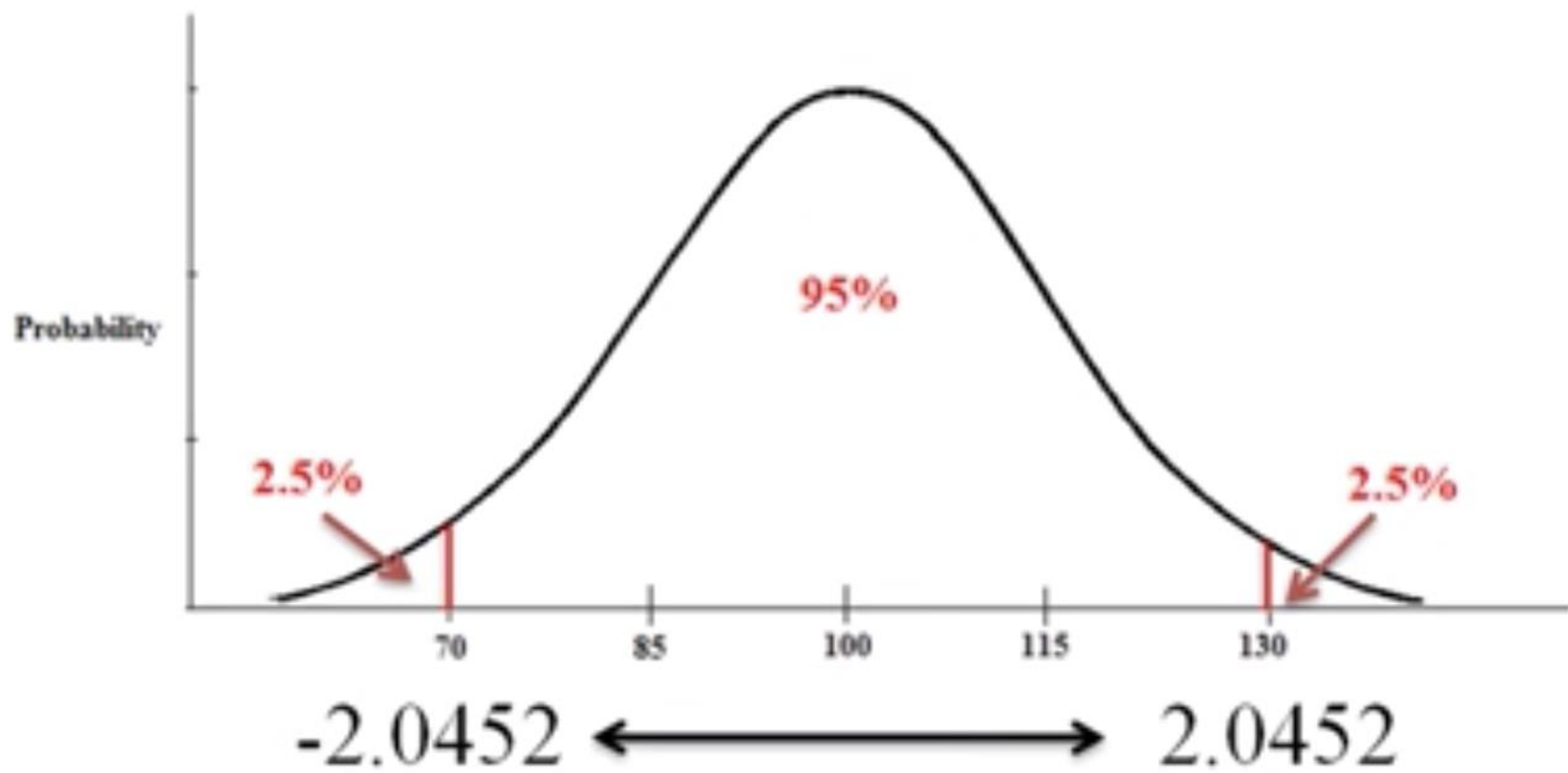
# 2 Decide Decision Rule

- **Alpha ( $\alpha$ ):** Usually **0.05**
- **Number of Tails:** **Two-tailed**
- **Degree of Freedom:**  **$n-1$  ( $30-1 = 29$ )**



	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.12	2.583	2.921	3.686	4.015
17	1.333	1.74	2.11	2.567	2.898	3.646	3.965
18	1.33	1.734	2.101	2.552	2.878	3.61	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.85
21	1.323	1.721	2.08	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.5	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.06	2.485	2.787	3.45	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.66
30	1.31	1.697	2.042	2.457	2.75	3.385	3.646
60	1.296	1.671	2	2.39	2.66	3.232	3.46
120	1.289	1.658	1.98	2.358	2.617	3.16	3.373
1000	1.282	1.646	1.962	2.33	2.581	3.098	3.3
Inf	1.282	1.645	1.96	2.326	2.576	3.091	3.291

	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
Df							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.12	2.583	2.921	3.686	4.015
17	1.333	1.74	2.11	2.567	2.898	3.646	3.965
18	1.33	1.734	2.101	2.552	2.878	3.61	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.85
21	1.323	1.721	2.08	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.5	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.06	2.485	2.787	3.45	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.66
30	1.31	1.697	2.042	2.457	2.75	3.385	3.646
60	1.296	1.671	2	2.39	2.66	3.232	3.46
120	1.289	1.658	1.98	2.358	2.617	3.16	3.373
1000	1.282	1.646	1.962	2.33	2.581	3.098	3.3
Inf	1.282	1.645	1.96	2.326	2.576	3.091	3.291





# 3 Compute T-Test

One-sample t-test

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

df = n-1

$$\bar{x} = 140$$

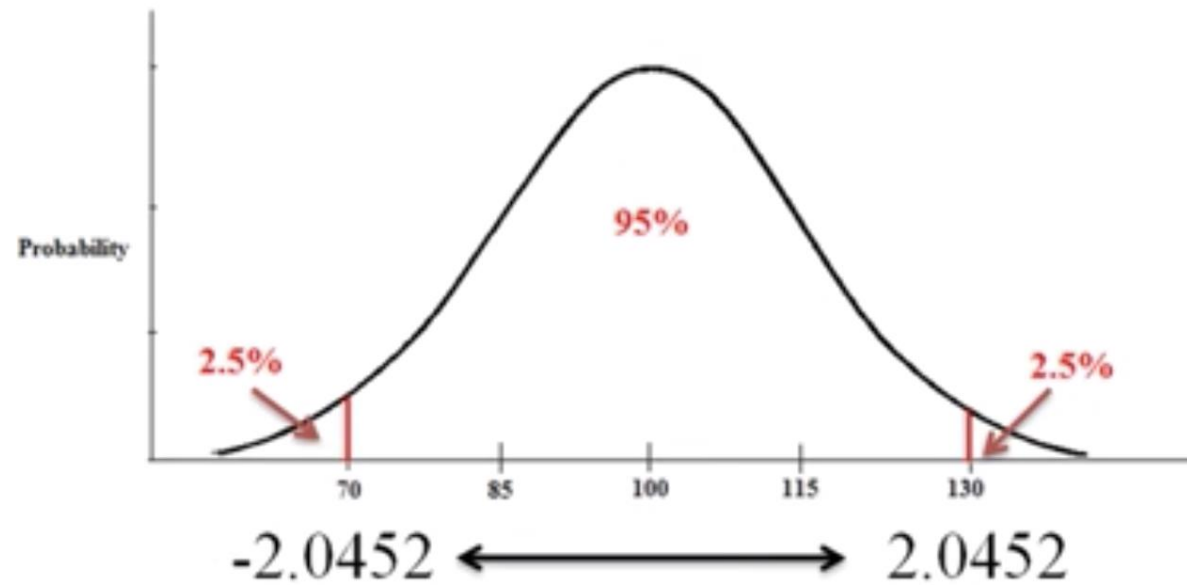
$$\mu = 100$$

$$s = 20$$

$$n = 30$$

$$t = \frac{140 - 100}{20/\sqrt{30}} = \frac{40}{3.65} = 10.96$$

# 4 Compare



-10.96 and 10.96 are both less and greater than -2.0452 and 2.0452

**Less OR Greater = Significant.**

# 5 Conclusion

**Medication significantly affected Intelligence,  
 $T = 10.96, p < 0.05$**

**Reject the Null Hypothesis!**

# SPSS EXAMPLE

According to CDC, the mean height of U.S. adults ages 20 and older is about 66.5 inches (69.3 inches for males, 63.8 inches for females).

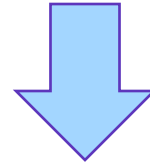
In our sample data, we have a sample of 408 college students from a single college. Let's test if the mean height of students at this college is significantly different than 66.5 inches using a one-sample  $t$  test.

The null and alternative hypotheses of this test will be:

$H_0: \mu_{\text{Height}} = 66.5$  ("the mean height is equal to 66.5")

$H_1: \mu_{\text{Height}} \neq 66.5$  ("the mean height is not equal to 66.5")

# Analyze > Compare Means > One-Sample T Test



## One-Sample Test

	(B) t	(C) df	(A) Test Value = 66.5 Significance (D)		(E) Mean Difference	(F) 95% Confidence Interval of the Difference	
			One-Sided p	Two-Sided p		Lower	Upper
Height (inches)	5.810	407	<.001	<.001	1.53176	1.0135	2.0501

- There is a **significant difference** in the mean height of the students at this college and the overall adult population in the U.S. ( $p < .001$ ).
- The average height of students at this college is about 1.5 inches taller than the U.S. adult population average (95% CI [1.013, 2.050]).

# Unpaired-sample T-Test

- Also known as **independent sample t-test**.
- Comparing **2 independent groups**.



# EXAMPLE

A biology teacher wants to compare her two classes to see if they performed any differently on the test he gave that semester.

Class A had 25 students with an average score of 70, standard deviation 15. Class B had 20 students with an average score of 74, standard deviation 25. Did these two classes perform differently on the tests?



Type

T - statistic

Degrees of freedom

One-sample t-test

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

df = n-1

Two-sample t-test

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{X_1 X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

df = n<sub>1</sub> + n<sub>2</sub> - 2

Equal variance assumed

Paired t-test

$$t = \frac{\bar{X}_D - \mu_0}{s_D/\sqrt{n}}$$

df = n-1



Type

T - statistic

Degrees of freedom

One-sample t-test

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

df = n-1

Two-sample t-test

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

df = n<sub>1</sub> + n<sub>2</sub> - 2

Equal variance  
not assumed

Paired t-test

$$t = \frac{\bar{X}_D - \mu_0}{s_D/\sqrt{n}}$$

df = n-1

# SPSS EXAMPLE

You want to know whether there is any significant difference of mile times between athlete and non-athlete group.

So you do simple research by gathering 166 athletes and 226 non-athletes to run miles while you calculate each of their mile time. Is there any significant difference in mile time between the athlete and non-athlete group?

$H_0: \mu_{\text{non-athlete}} - \mu_{\text{athlete}} = 0$  ("the difference of the means is equal to zero")

$H_1: \mu_{\text{non-athlete}} - \mu_{\text{athlete}} \neq 0$  ("the difference of the means is not equal to zero")

# Analyze > Compare Means > Independent-Sample T Test

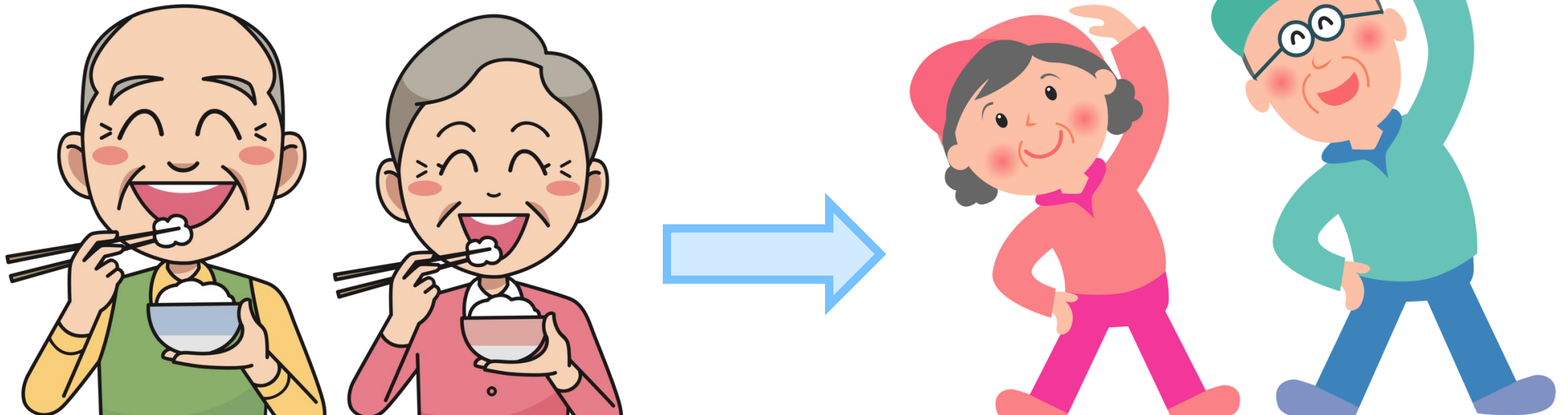
Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means <span style="float: right;">C</span>						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Mile time	Equal variances assumed	102.98	.000	13.475	390	.000	0:02:14	0:00:10	0:01:55	0:02:34
	Equal variances not assumed			15.047	315.846	.000	0:02:14	0:00:08	0:01:57	0:02:32

- There was a **significant difference** in mean mile time between non-athletes and athletes ( $t_{315.846} = 15.047, p < .001$ ).
- The average mile time for athletes was 2 minutes and 14 seconds faster than the average mile time for non-athletes.

# Paired-sample T-Test

- Also known as dependent-sample T-Test.
- We use this t-test to compare 2 different data set from the same group of sample.
- Focusing on the '**difference**'.



# EXAMPLE

A group of researchers wants to test a new anti-hunger weight loss pill. They have 10 people rate their hunger both before and after taking the pill. Does the pill do anything?



Type

T - statistic

Degrees of freedom

One-sample t-test

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

df = n-1

Two-sample t-test

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{X_1 X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

df =  $n_1 + n_2 - 2$

Paired t-test

$$t = \frac{\bar{X}_D - \mu_0}{s_D/\sqrt{n}}$$

df = n-1

# SPSS EXAMPLE

The sample dataset has placement test scores (out of 100 points) for four subject areas: English, Reading, Math, and Writing. Students in the sample completed all 4 placement tests when they enrolled in the university. Suppose we are particularly interested in the English and Math sections, and want to determine whether students tended to score higher on their English or Math test, on average. We could use a paired  $t$  test to test if there was a significant difference in the average of the two tests.

$H_0: \mu_1 - \mu_2 = 0$  ("the difference between the paired population means is equal to 0")

$H_1: \mu_1 - \mu_2 \neq 0$  ("the difference between the paired population means is not 0")

## Analyze > Compare Means > Paired-Sample T Test

Paired Samples Test

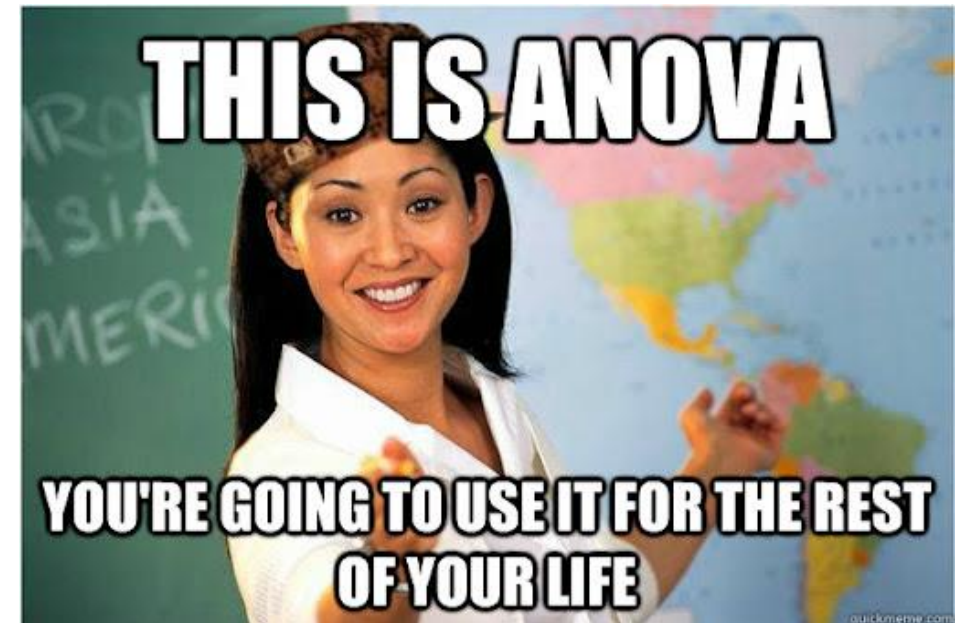
	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
English - Math	17.30	9.50303	.4763	16.3608	18.2337	36.313	397	.000

- There was a **significant difference** between English and Math scores ( $t_{397} = 36.313, p < 0.001$ ).
- On average, English scores were 17.3 points higher than Math scores (95% CI [16.36, 18.23]).

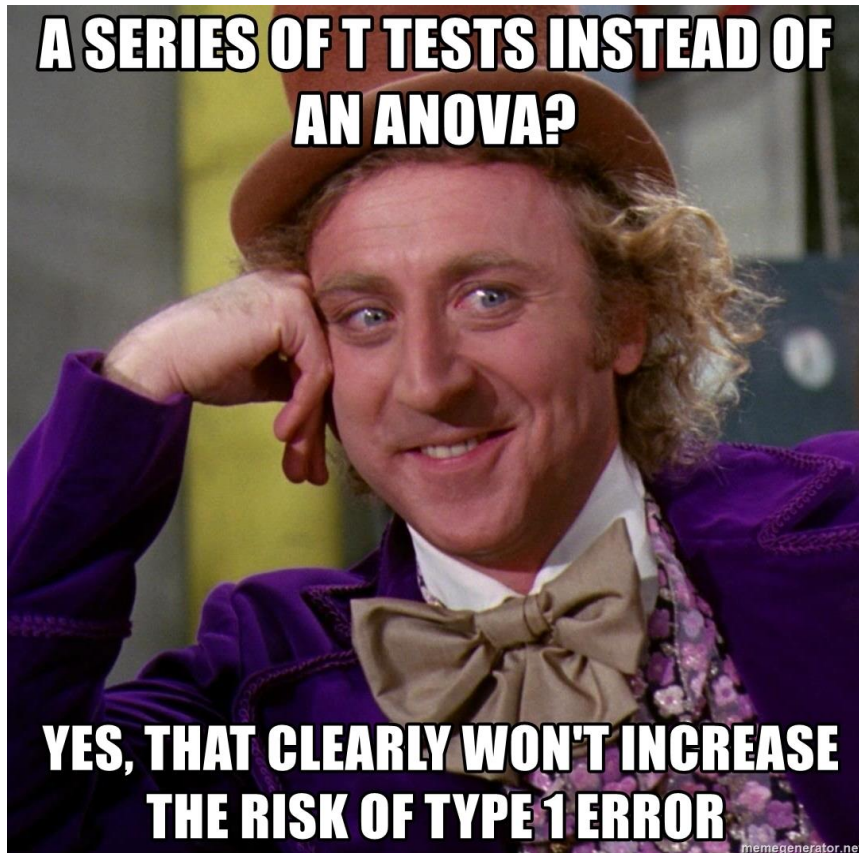




# ANOVAS



# Why use ANOVAS?

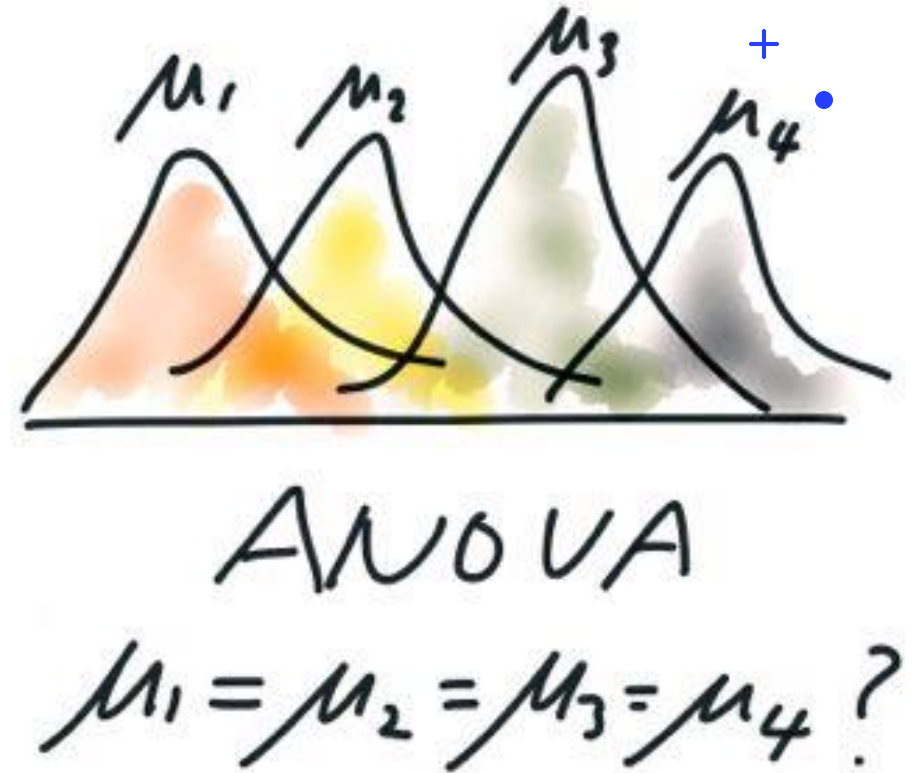


- + • A T-Test can only be used when comparing two groups. (Effects of green or peppermint tea on weight loss).
- o • But what if we want to measure the effects of green tea, peppermint tea AND lemon tea?)
- Instead of doing multiple T-Tests between groups, **we can run an ANOVA.**
- ANOVAS are powerful because they can **measure variation between groups and within groups.**

# Situations when we can use ANOVAS

We can use ANOVAs to compare:

- Mean of 1 continuous variable between 3+ levels of a categorical variable.
- Mean of 1 continuous variable between 3+ levels of 2+ categorical variables.



# But what are ANOVAS really?

When we run an ANOVA, we are actually testing an **hypothesis**:

- **The null hypothesis ( $H_0$ )**: The means between the groups are equal.
- **The alternative hypothesis ( $H_1$ )**: The means between the groups are significantly different.



# But what are ANOVAS really?

After running an ANOVA on the software, we should be looking at **3 different values**:

- **The F value:** Variation between groups / Variation within groups
- **P-value:** The probability of obtaining test results at least as extreme as the results actually observed, under the assumption that the null hypothesis is correct.
- **The Degree of Freedom (df):** The number of values in the final calculation of a statistic that are free to vary.

# ATTENTION!!!!



**Before running an ANOVA, always check for:**

- **Independence of observations** – this is an assumption of the model that simplifies the statistical analysis.
- **Normality** – the distributions of the residuals are normal.
- **Equality (or "homogeneity") of variances**, called homoscedasticity — the variance of data in groups should be the same.

# One-way ANOVA

- The one-way analysis of variance (ANOVA) is used to **determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.**

- Where:

- $H_0$  = the null hypothesis
- $\mu$  = group mean
- $k$  = number of groups

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



# Example

- We are interested in determining whether there are differences in leg strength between amateur, semi-professional and professional football players. The force/strength measured on an isokinetic machine is the dependent variable.
- We can run a one-way ANOVA to determine if there is any difference in leg strength between those three groups.





# Software Example

ANOVA

Time

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	91.467	2	45.733	4.467	.021
Within Groups	276.400	27	10.237		
Total	367.867	29			



# When to use a one-way ANOVA

- Use a one-way ANOVA when you have collected data about **one categorical independent variable** and **one quantitative dependent variable**.
- The **independent variable** is the **cause**. Its value is *independent* of other variables in your study.
- The **dependent variable** is the **effect**. Its value *depends* on changes in the independent variable.  
(Optional slide!)

# Two ways ANOVA

- A two-way ANOVA is used to determine **how the mean of a quantitative variable changes according to the levels of two categorical variables.**
- We use a two-way ANOVA when we want to know how two independent variables, in combination, affect a dependent variable.



# Example

We would want to know if there is an interaction Socioeconomic background (working class, middle class or upper class) and age (below 30 or above 30) for anxiety level at a certain firm. The anxiety level is the outcome, or the variable that can be measured.

Socioeconomic background and age would be the two categorical variables. These categorical variables are also the independent variables, which are called **factors** in a Two Way ANOVA.



# Software example?

```
. anova Int_Politics Gender##Edu_Level
```

```
Number of obs =      58      R-squared      = 0.8809  
Root MSE      = 3.84552      Adj R-squared = 0.8695
```

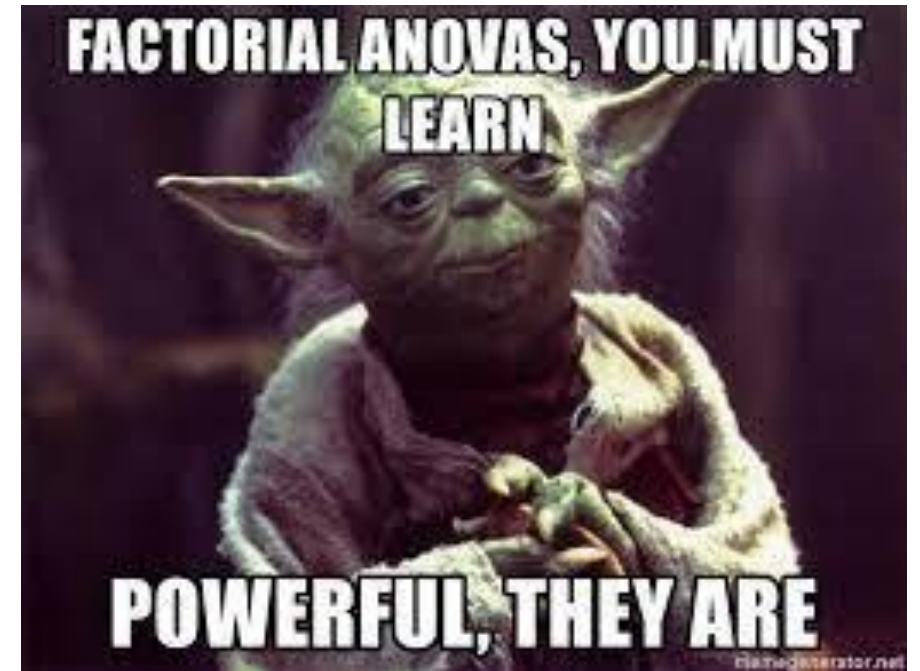
Source	Partial SS	df	MS	F	Prob > F
Model	5688.76791	5	1137.75358	76.94	0.0000
Gender	6.86428571	1	6.86428571	0.46	0.4987
Edu_Level	5486.85619	2	2743.4281	185.52	0.0000
Gender#Edu_Level	216.843562	2	108.421781	7.33	0.0016
Residual	768.977778	52	14.7880342		
Total	6457.74569	57	113.293784		

# What does a two way ANOVA tell us?

- The results from a Two Way ANOVA will calculate both a **main effect** and an **interaction effect**:
- A **main effect** is the effect of one independent variable on the dependent variable (Effect of age on anxiety levels).
- An interaction effect happens when one explanatory interacts with another explanatory variable on a response variable (Do old and socio-economically unfavored individuals experience more anxiety)? (optional slide)

# Factorial ANOVAS

- A factorial ANOVA is an Analysis of Variance test with more than one independent variable, also called factor.
- Factorial ANOVAS can be:
  - **Three way ANOVAS**
  - **Five way ANOVAS**
- Five and three will represent the number of independent factor in each ANOVAS.



# Do not always trust your result!



- + • A one way ANOVA will tell you that at least two groups were different from each other. But **it won't tell you which groups were different.**
- ANOVAS can only be used if observations are drawn from a normal distribution and though it is remarkably robust, it may not yield exact p-values when the observations come from distributions that are heavier in the tails than the normal.
- Corrections have been developed that can avoid increases in the type I error rate.



# REGRESSION



# What is Regression?

- **Predictive statistic.**
- There are 3 main function of regression:
  - Determining the **strength of predictor(s).**
  - Forecasting **an effect.**
  - Knowing the **trend** of forecasting.



# Types of regression



- **Simple Linear regression**
- **Multiple Linear regression**
- **Logistic regression**
- Other type of regression

# Simple Linear Regression

- Used when we want to make a prediction about an **outcome (dependent) variable** based on **ONE predictor (independent) variable**.
- Linear regression is a way to model a relationship between two sets of variables. **The result is a linear regression equation that can be used to make predictions about data.**



Imagine there is a new 'effective' sleeping pill to fight insomnia. The researcher conducts this clinical trial to see how effective is the sleeping pill to the improvement of sleep quality.

**So the research will be based on:**

- 1. The doses of sleeping pills.**
- 2. Hours of sleep.**

# Context:

- **Determining the strength of predictors:** What is the strength of the relationship between doses of sleeping pills and hours of sleep in patients with insomnia.
- **Forecasting an effect:** How many hours of sleep can a patient with insomnia retain per 20mg of sleeping pill doses?
- **Knowing the trend of forecasting:** How many hours can the patient sleep per night if they are about to be given 100mg doses of sleeping pill?

# Assumption of Simple Linear Regression

- **Linearity:** The relationship between  $X$  (independent) and the mean of  $Y$  (dependent) is linear.
- **Homoscedasticity:** A condition in which the variance of the residual, or error term, in a regression model is constant.
- **Independence:** Observations are independent of each other.
- **Normality:** For any fixed value of  $X$ ,  $Y$  is normally distributed.

# The equation

- Linear regression analysis will produce an equation that has the form:

$$Y = a + bx$$

- Where **y** and **x** are respectively the "**dependent**" and "**independent**" variables
- Typically, you choose a value to substitute for the independent variable and then solve for the dependent variable.



# How to find the equation

- The formula:

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

**How to interpret it:**

**a: Slope**

**b: Y-intercept**

**x: Independent variable values**

**y: Dependent variable values**

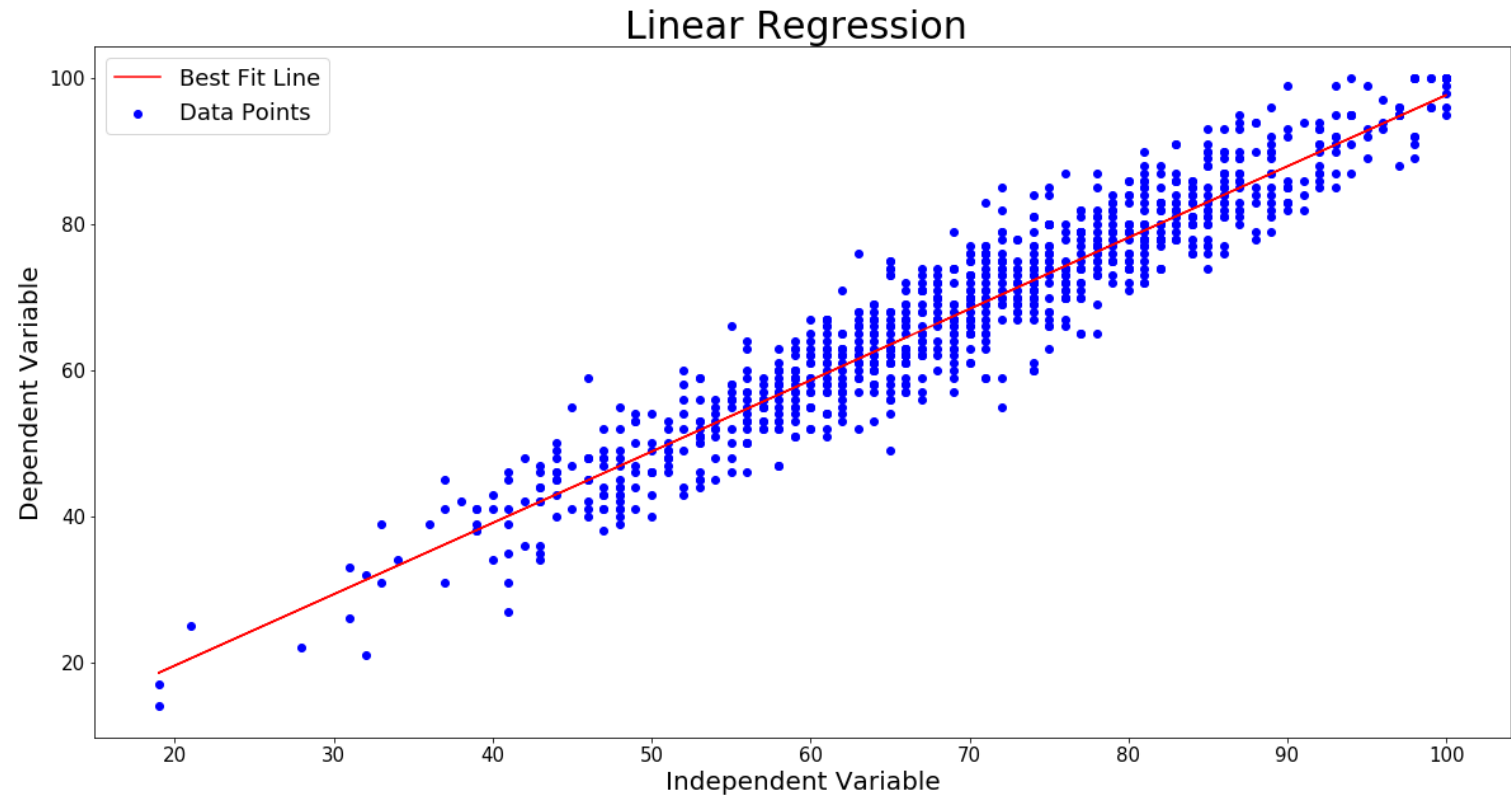
**n: Sample size**

# A visual representation of a simple linear regression

- Simple linear regression takes a set of observations and finds the line of best fit (Also called regression line).
- The distance between the line and points is called "error" and we use the methods of least squares to minimize those errors as to create the line of best fit.
- The line of best fit would in this case take the form of:

$$Y = a + bx + \epsilon$$

- **Where epsilon is the error term**



# Don't always trust your simple linear regression

**Many questions can arise after looking at a simple linear regression graph:**

- Is the dependent variable really related to the independent variable?
- IS the relationship between the two variables strong enough to claim they are related?
- Do we have enough samples to affirm or reject a relation?



# Interpreting a simple linear regression

- Instead of using the best fit line to estimate future values, it is better to use **confidence intervals**.
- **The p-value** of the regression has to be below 0.05 to claim the relationship is significant.
- We can measure how well our model fits by comparing the variance we can explain relative to the variance we cannot explain.

# Interpreting a simple linear regression

**The R-Squared computes the proportion of the variation in  $y$  that is explained by the systematic portion of the model.**

- $\text{Var}(y) = \text{Var}(bx) + \text{Var}(\varepsilon)$

**$R^2: \text{Var}(bx) / \text{Var}(y)$**

- If  $R^2=0$ , our model doesn't explain anything
- If  $R^2=1$ , our model can account for all the observed variation

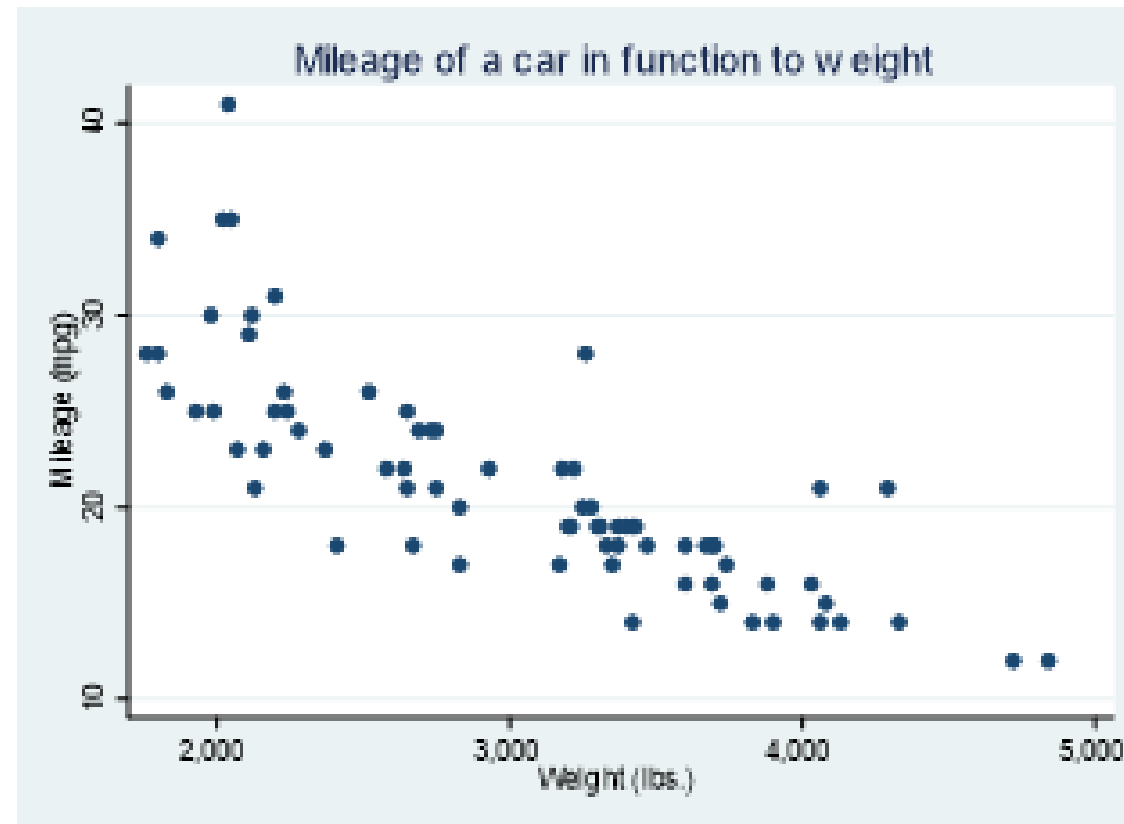
# An example to better unders

- Let's say we want to see if the weight of a car can significantly impact the mileage per gallon of the car.
- **The data set would like this:**
  - The weight is the independent variable
  - The mileage (mpg) is the dependent variable
  - N=74

	weight	mpg
1.	2,930	22
2.	3,350	17
3.	2,640	22
4.	3,250	20
5.	4,080	15
6.	3,670	18
7.	2,230	26
8.	3,280	20
9.	3,880	16
10.	3,400	19
11.	4,330	14
12.	3,900	14
13.	4,290	21
14.	2,110	29
15.	3,690	16
16.	3,180	22
17.	3,220	22
18.	2,750	24
19.	3,430	19
20.	2,120	30
21.	3,600	18
22.	3,600	16
23.	3,740	17
24.	1,800	28
25.	2,650	21
26.	4,840	12
27.	4,720	12
28.	3,830	14
29.	2,580	22
30.	4,060	14
31.	3,720	15
32.	3,370	18
33.	4,130	14
34.	2,830	20
35.	4,060	21

# An example to better understand

- Before we perform simple linear regression, let's first create a scatterplot of weight vs. mpg so we can visualize the relationship between these two variables and check for any obvious outliers.
- Since we want to quantify this relationship, we will now perform a simple linear regression.



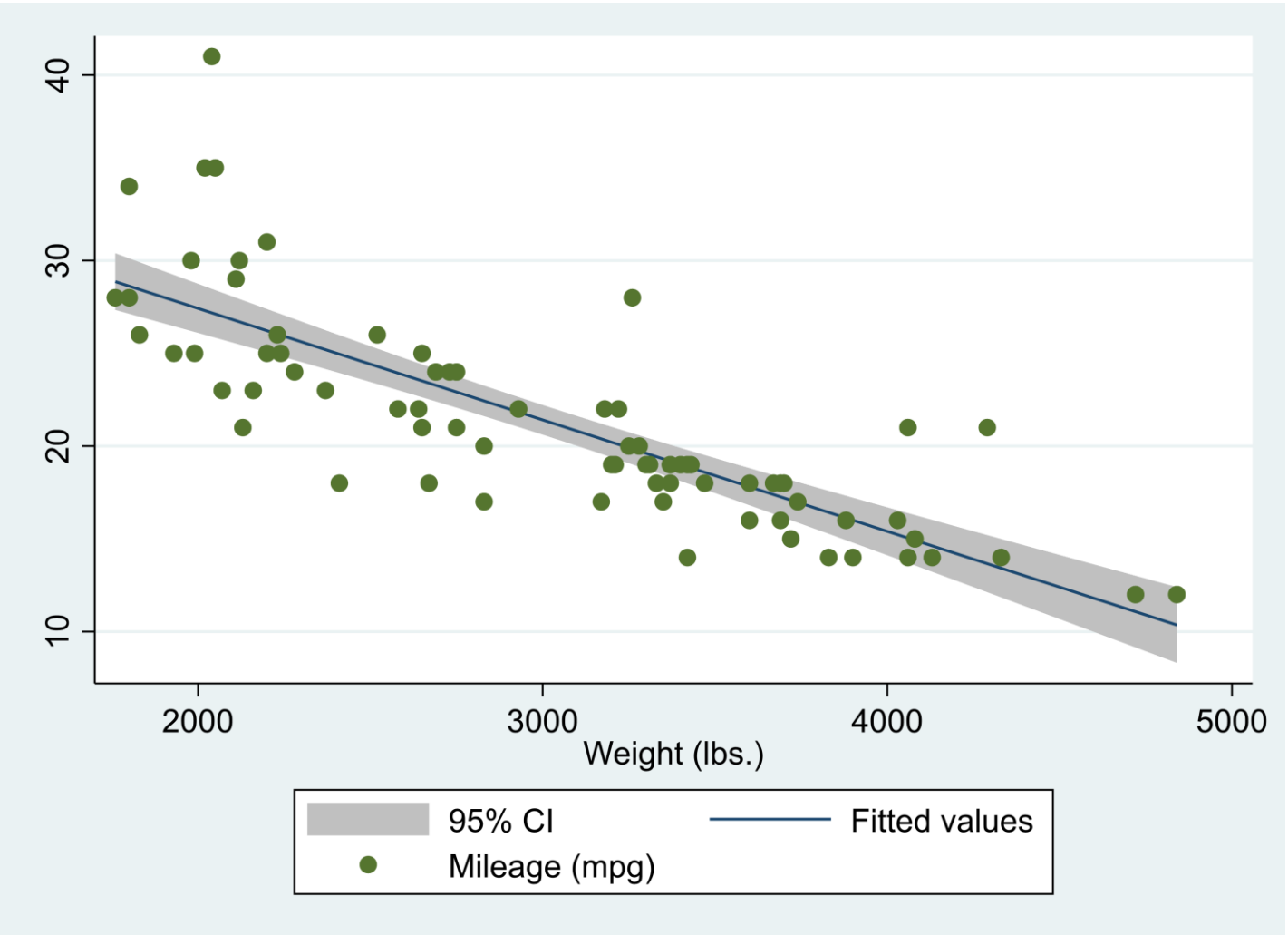
# An example to better understand

Interpreting the results:

Source	SS	df	MS	Number of obs	=	74
Model	1591.9902	1	1591.9902	F(1, 72)	=	134.62
Residual	851.469256	72	11.8259619	Prob > F	=	0.0000
Total	2443.45946	73	33.4720474	R-squared	=	0.6515
				Adj R-squared	=	0.6467
				Root MSE	=	3.4389
mpg	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
weight	-.0060087	.0005179	-11.60	0.000	-.0070411	-.0049763
_cons	39.44028	1.614003	24.44	0.000	36.22283	42.65774



# An example to better understand



# Conclusion

Purpose	Detail	Statistics
<b>Comparing Means</b>	One sample group, parametric	One-sample T-Test
	Two sample group, parametric	Unpair-sample T-Test
	Two group of data from one sample group, parametric	Paired-sample T-Test
	More than two independent group, parametric	One-way ANOVA
	More than two group of related group, parametric	Repeated ANOVA
<b>Modelling, predicting, estimating relationship</b>	One independent variable, one dependent variable.	Simple Linear Regression
	More than one independent variable, one dependent variable.	Multiple Linear Regression

**NOT SURE IF SHOULD USE T-TEST,  
REGRESSION**

**OR ANOVA**