D208 PREDICTIVE MODELING WEBINAR

Dr. William Sewell,
College of IT

Episode 1

WELCOME TO THE PREDICTIVE MODELING WEBINAR

When it's time, join your Webex meeting here.

Join meeting

More ways to join:

Join from the meeting link
https://wgu.webex.com/wgu/j.php?MTID=me73470c2eac9e863c6f47a3d5b6d2f26

Join by meeting number
Meeting number (access code): 120 848 5390
Meeting password: w7WPDApU

Tap to join from a mobile device (attendees only)
+1-415-655-0003,,1208485390## United States TOLL

Join by phone
+1-415-655-0003 United States TOLL
Global call-in numbers

WELCOME TO D207 EDA WEBINAR

A PILOT PRESENTATION

Mission

- To present as many types of questions as possible from the course topics
- Ask questions that cover the main concepts
- Aid you in distinguishing among the types of statistical methods
- To limit webinar to ~ 30 minutes and cover the three Statistical methods
- Ask questions that will aid in your understanding of Statistics and EDA
- By attending this Webinar, you will BE FEARLESS in your studies!

Episode 1 Slides

Do you remember the equation of a straight line?

https://www.mathsisfun.com/equation_of_line.html

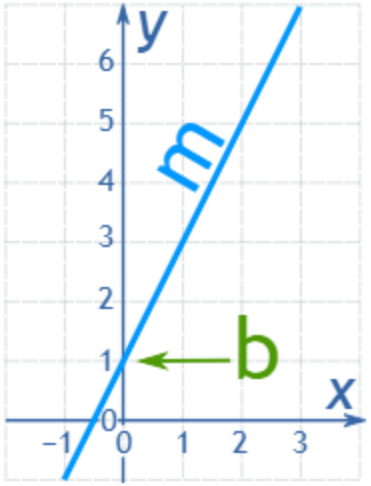
Equation of a Straight Line

The equation of a straight line is usually written this way:

$$y = mx + b$$

(or "y = mx + c" in the UK see below)

What does it stand for?



$$y = mx + b$$

Slope or Gradient

y when x=0 (see *Y Intercept*)

y = how far up

x = how far along

m = Slope or Gradient (how steep the line is)

b = value of **y** when **x=0**

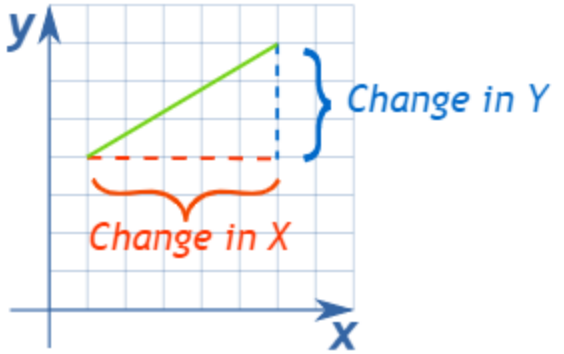
Episode 1 Slides

New!

Do you remember the equation of a straight line?

What is y? Response, dependent, criterion variable
What is b? Constant, y-intercept
What is m? The slope, coefficient, gradient
What is x? The independent, explanatory, predictor variable

$$m = \frac{\text{Change in Y}}{\text{Change in X}}$$

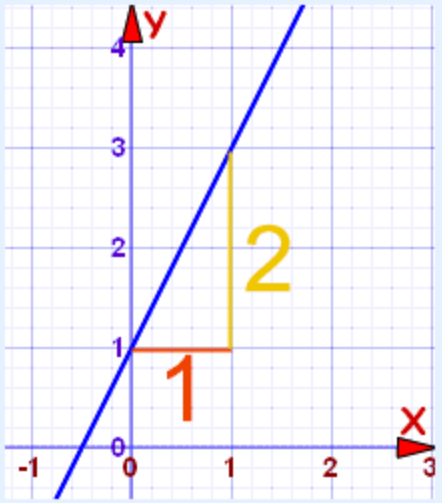


https://www.mathsisfun.com/equation_of_line.html

Episode 1 Slides

New!

How do you state the equation for a line?



$$m = \frac{2}{1} = 2$$

b = 1 *(value of **y** when x=0)*

So: y = 2x + 1

https://www.mathsisfun.com/equation_of_line.html

Episode 1 Slides

New!

1. The equation $Y' = a + b_1X_1 + b_2X_2$

   A. is simple linear regression.
   B. Is a multiple regression with three independent variables.
   C. Is a multiple regression with unknown number of independent variables.
   D. All of the above.
   E. None of the above.

https://www.questionpro.com/blog/nominal-ordinal-interval-ratio/

Episode 1 Slides

New!

2. Matching. Consider the Regression Model $Y=\beta0+\beta1X1+\beta2X2+\beta3X3+\varepsilon$

A. What is Y?
B. What is X?
C. What are Population regression parameters?
D. What is E(Y)?
E. What is Y - E(Y)?

1. regression error term ε
2. scatter plot
3. independent or predictor variables
4. dependent or response variable
5. expected value of Y
6. β0, β1 . . .

Consider the multiple linear regression model $Y = \mathrm{E}(Y) + \epsilon = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon.$

Source: ZyBooks 3.6

Episode 1 Slides

New!

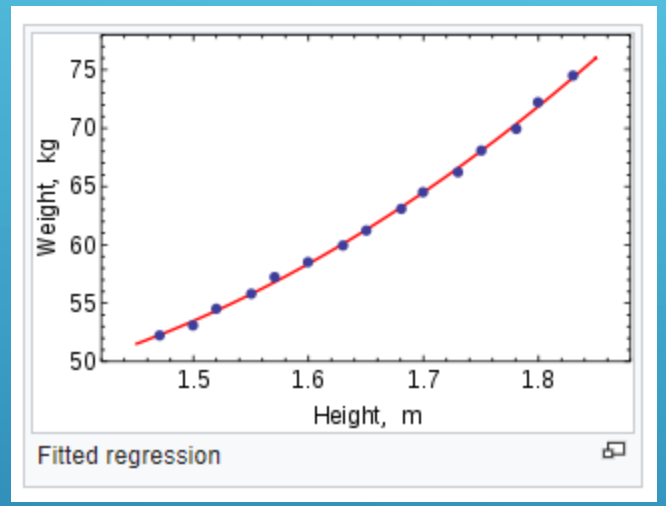
Ordinary Least Squares Method: OLS

The output from most popular statistical packages will look similar to this:

Method	Least squares
Dependent variable	WEIGHT
Observations	15

Parameter	Value	Std error	t-statistic	p-value
β_1	128.8128	16.3083	7.8986	0.0000
β_2	–143.1620	19.8332	–7.2183	0.0000
β_3	61.9603	6.0084	10.3122	0.0000

R^2	0.9989	S.E. of regression	0.2516
Adjusted R^2	0.9987	Model sum-of-sq.	692.61
Log-likelihood	1.0890	Residual sum-of-sq.	0.7595
Durbin–Watson stat.	2.1013	Total sum-of-sq.	693.37
Akaike criterion	0.2548	F-statistic	5471.2
Schwarz criterion	0.3964	p-value (F-stat)	0.0000



Fitted regression

Episode 1 Slides

New!

3. Let's calculate a Simple Linear Regression equation, given the following data.

Subject	Age x	Glucose (y)
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

New!

3. **Step 1. Calculate the product, squares, and sum of the squares.**

Subject	Age x	Glucose (y)	xy	x^2	y^2
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
∑	247	486	20485	11409	40022

Use the following equations to find a and b.

New!

3. Step 2. Use the formulas to calculate a and b.

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

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

Subject	Age x	Glucose (y)	xy	x^2	y^2
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
∑	247	486	20485	11409	40022

a = ((486 * 11409) - (247 * 20485)) / 6 (11409) - 247^2)

484979 / 7445

= 65.14

b = (6(20485) - (247 * 486)) / (6 (11409) - 247^2)

2868 / 7445

= .385225

New!

3. Step 3. Construct the Simple Linear Regression Equation.

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

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

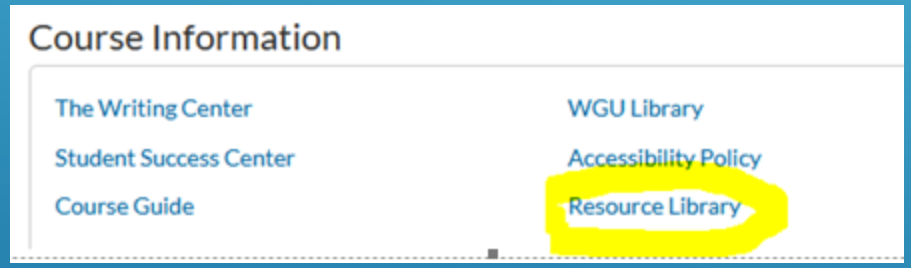
a = ((486 * 11409) - (247 * 20485)) / 6 (11409) - 247^2)
484979 / 7445
= 65.14
b = (6(20485) - (247 * 486)) / (6 (11409) - 247^2)
2868 / 7445
= .385225

y = a + bx
y = 65.14 + .385225x

Resource Tip:

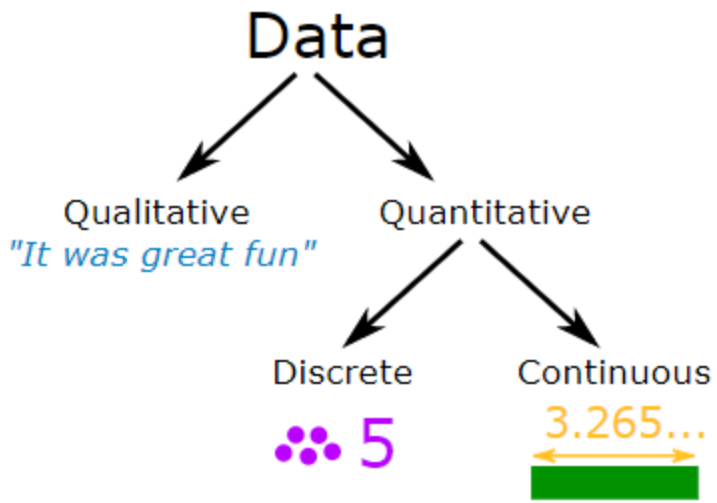
The following links are the textbooks for the course.

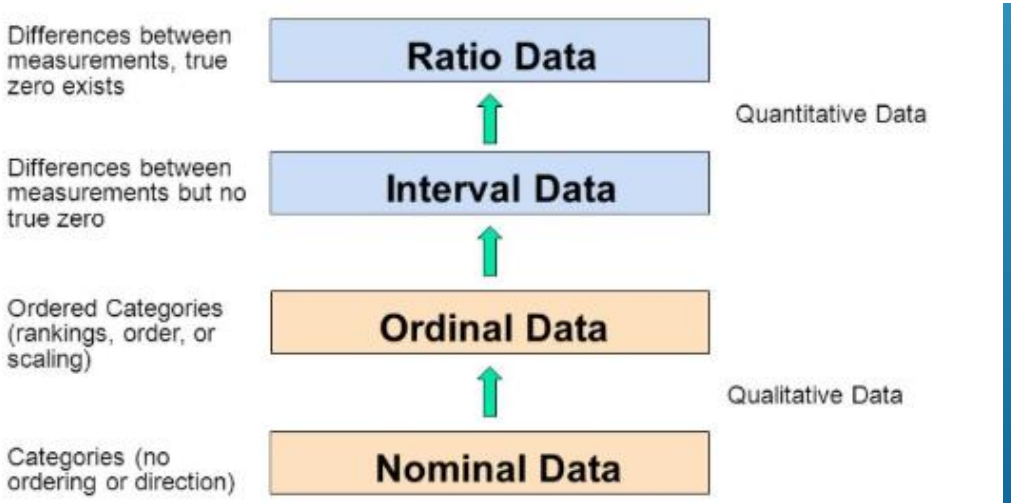
Regression Analysis with R
Regression Analysis with Python



New!

Continuous Data is not Discrete Data.





Types of data on the basis of measurement

Scale	True Zero	Equal Intervals	Order	Category	Example
Nominal	No	No	No	Yes	Marital Status, Sex, Gender, Ethnicity
Ordinal	No	No	Yes	Yes	Student Letter Grade, NFL Team Rankings
Interval	No	Yes	Yes	Yes	Temperature in Fahrenheit, SAT Scores, IQ, Year
Ratio	Yes	Yes	Yes	Yes	Age, Height, Weight

New!

1. What are the datatypes that we test in statistics?

   A. Nominal. (Nomen (Latin for “name”).

   B. Ordinal. (Labelled classes).

   C. Interval. (Think Likert-type scale).

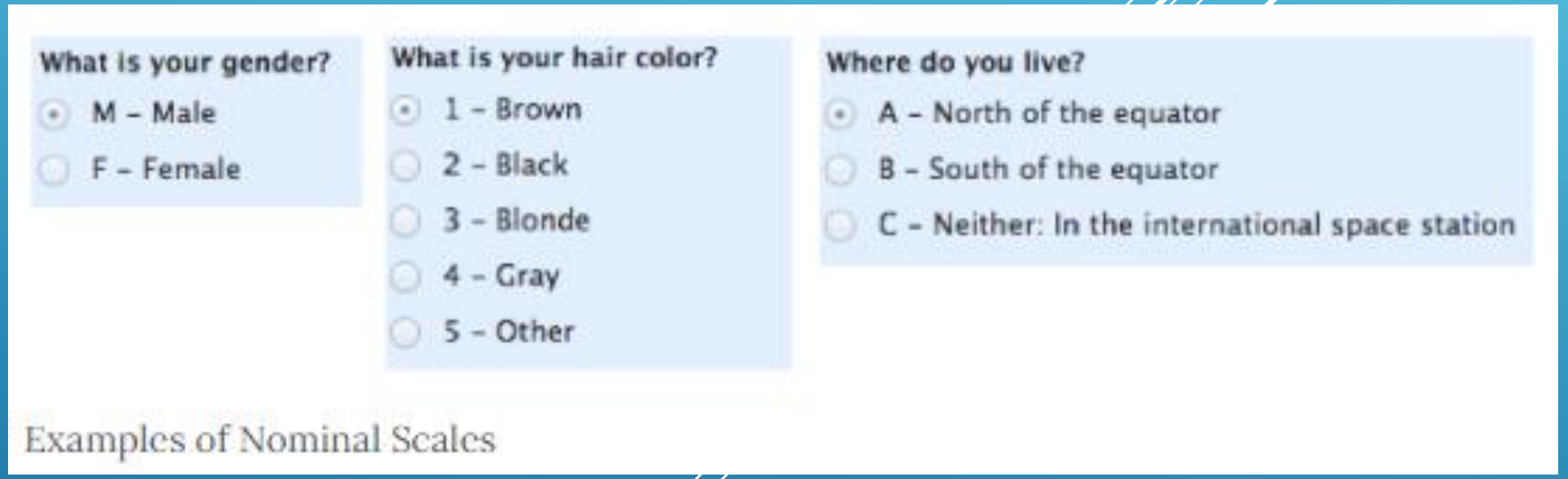
   D. Ratio. (Numbers, calculable).

   E. All of the above.

Offers:	Nominal	Ordinal	Interval	Ratio
The sequence of variables is established	–	Yes	Yes	Yes
Mode	Yes	Yes	Yes	Yes
Median	–	Yes	Yes	Yes
Mean	–	–	Yes	Yes
Difference between variables can be evaluated	–	–	Yes	Yes
Addition and Subtraction of variables	–	–	Yes	Yes
Multiplication and Division of variables	–	–	–	Yes
Absolute zero	–	–	–	Yes

3. What is *nominal* scale data and how does it work?

A. Used for labelling variables into distinct classes
B. Does not involve actual measurement
C. Does not contain numerical value
D. Also known as the continuous variable scale
E. All of the above



Examples of Nominal Scales

Examples:

Population Distribution:
1. **Urban**
2. **Suburban**
3. **Rural**

Which Brand of Computer is Most Popular?
1. **HP**
2. **Apple**
3. **IBM**

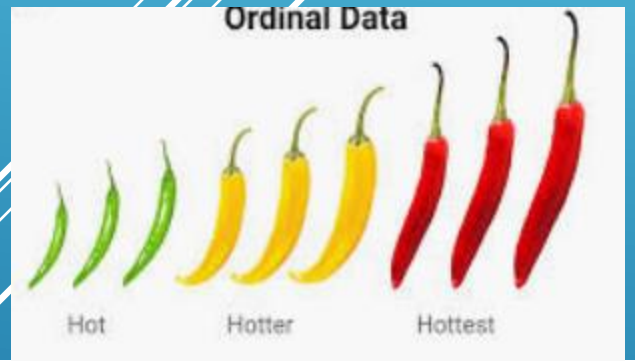
4. What is the *ordinal* scale data and how does it work?

A. Used for labelling variables into distinct classes
B. Does not involve actual measurement
C. Provides a rank order to the variable classes
D. Not a numerical variable
E. All of the above

Examples

Men's Shirt Sizes:
1 - Small
2 - Medium
3 - Large
4 - Extra Large
5 - XXL

Children's Clothing Sizes
1 - Toddler 1
2 - Toddler 2
3 - Toddler 3
4 - Toddler 4
5 - Child Small (4/5)



Example of Ordinal Scales

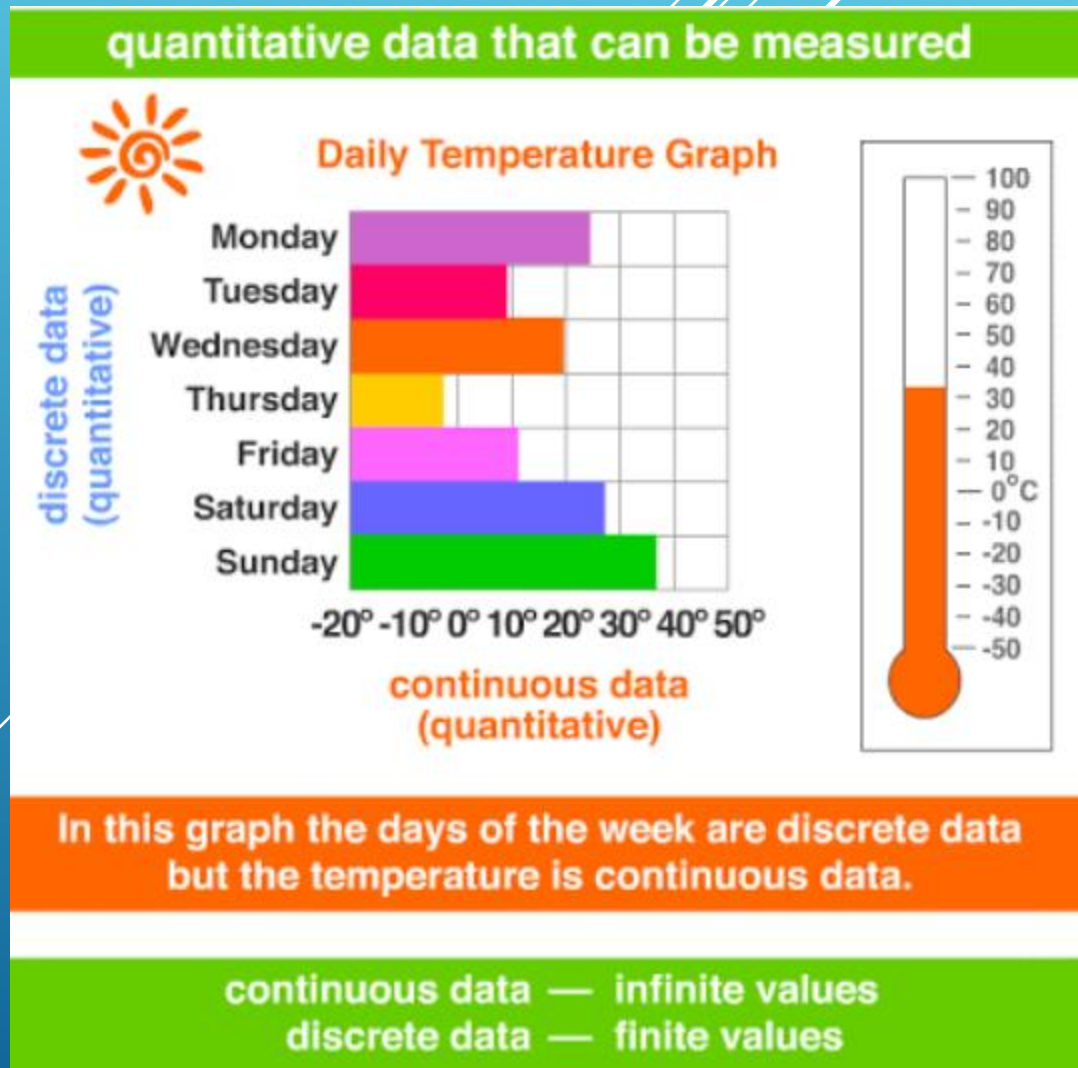
5. What is the *interval* scale data and how does it work?

A. A numerical scale
B. Ordered scale
C. Calculations permitted.
D. There can be negative values.
E. All of the above

Examples:

Temperature scale, Measurement scale, Likert scale

Example of Interval Scale



6. What is the ***ratio*** scale data and how does it work?

A. A numerical scale
B. Ordered scale
C. Calculations permitted
D. Has a true zero
E. All of the above

Examples:

Weight, Height, mean, median, mode.

Football Players Weight
1. Less than 200 pounds
2. Less than 250 pounds
3. Less than 300 pounds
4. Greater than 300 pounds

New!

1. What are the four fundamental levels (types) of measurement scales (variables)?

A. Radical, integral, invertebral, logical

B. Expository, inquisitive, rational, logical

C. Logical, radical, interval, ordinal

D. Nominal, ordinal, interval, ratio

E. All of the above.

https://www.questionpro.com/blog/nominal-ordinal-interval-ratio/

Nominal, ordinal, interval, ratio

DATA CLEANING and PREPARATION

The dataset is fairly clean that you are given. So, what more can you do to satisfy the data Cleaning or preparation task?

Tip # 1. Drop missing values and/or fill

Python **df.dropna(how='all') for missing values row-wise or column-wise single, multiple, all**
https://towardsdatascience.com/how-to-drop-rows-in-pandas-dataframes-with-nan-values-in-certain-columns-7613ad1a7f2

R drop null and missing values like NA and NaN: na.omit() and is.na()
https://www.datasciencemadesimple.com/drop-rows-with-missing-values-in-r-drop-null-values-nanan-2/

Python tip to convert all of the NA values to the mean of the remaining values
df.fillna(df.mean(), inplace=True)

R tip to convert all of the NA values to the mean of the remaining values

```
dataset <- mutate_all(dataset, ~if_else(is.na(.),mean(.,na.rm=T),.))
```

DATA CLEANING and PREPARATION

The dataset is fairly clean that you are given. So, what more can you do to satisfy the data Cleaning or preparation task?

Tip # 2. Convert your categoricals.

Python **using Pandas get_dummies() or .astype('category') to convert from object to category or OneHotEncoder**
https://pbpython.com/categorical-encoding.html

df['Cylinders'].replace(['four', 'six', 'eight'], [4, 6, 8], inplace=True)
Df['location'] = pd.factorize(df['location'])[0] ← last character is a zero

R using as.factor and unclass()
https://www.statology.org/convert-categorical-variable-to-numeric-r/

R using sapply(), as.factor and as.numeric
https://stackoverflow.com/questions/47922184/convert-categorical-variables-to-numeric-in-r

DATA CLEANING and PREPARATION

Tip # 3. Check your categoricals and find the number of unique values. Cardinality can determine If you want to keep or exclude the categorical variable from your analysis.

https://www.askpython.com/python/built-in-methods/unique-values-from-a-dataframe

Get number of unique values in Python (Pandas .value_counts())
https://thispointer.com/pandas-get-unique-values-in-single-or-multiple-columns-of-a-dataframe-in-python/

Using nunique in Python

https://datascienceparichay.com/article/pandas-count-of-unique-values-in-each-column/

Lengths(lapply(df, unique))
Can use sapply or dplyr in R per comments:

https://stackoverflow.com/questions/38492832/count-number-of-unique-levels-of-a-variable

DATA CLEANING and PREPARATION

Tip # 4. How many levels do you want to keep? K-1 groups using drop_first

Pandas get_dummies()

https://pandas.pydata.org/docs/reference/api/pandas.get_dummies.html

Using droplevels in R
https://www.geeksforgeeks.org/removing-levels-from-a-factor-in-r-programming-droplevels-function/

Code Samples

Checking for VIF is one way to reduce the number of independent variables. You can do this in your data preparation step.

Here, in R, we are using the VIF (Variance Inflation Factor) to check for multicollinearity.

In this case, all variables in the regression model have low VIF scores.

```
> library(car)
> head(mtcars)
                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
> model <- lm(mpg ~ disp + hp + wt + drat + qsec, data = mtcars)
> vif(model)
    disp       hp       wt     drat     qsec
9.110869 5.201833 7.012686 2.322343 3.191939
>
```

Here, in Python, we are using the VIF (Variance Inflation Factor) to check for multicollinearity.

```
import os
os.chdir(r'C:\Users\william.sewell\Documents\D208')
print('Get current working directory : ', os.getcwd())
```

```
Get current working directory :  C:\Users\william.sewell
Get current working directory :   C:\Users\william.sewell\Documents\D208
```

```
import pandas as pd
data = pd.read_csv('BMI.csv')

print(data.head())
```

```
   Gender  Height  Weight  Index
0    Male     174      96      4
1    Male     189      87      2
2  Female     185     110      4
3  Female     195     104      3
4    Male     149      61      3
```

You can use the map command to create Dummy (indicator) variables or use pd.get_dummies() →

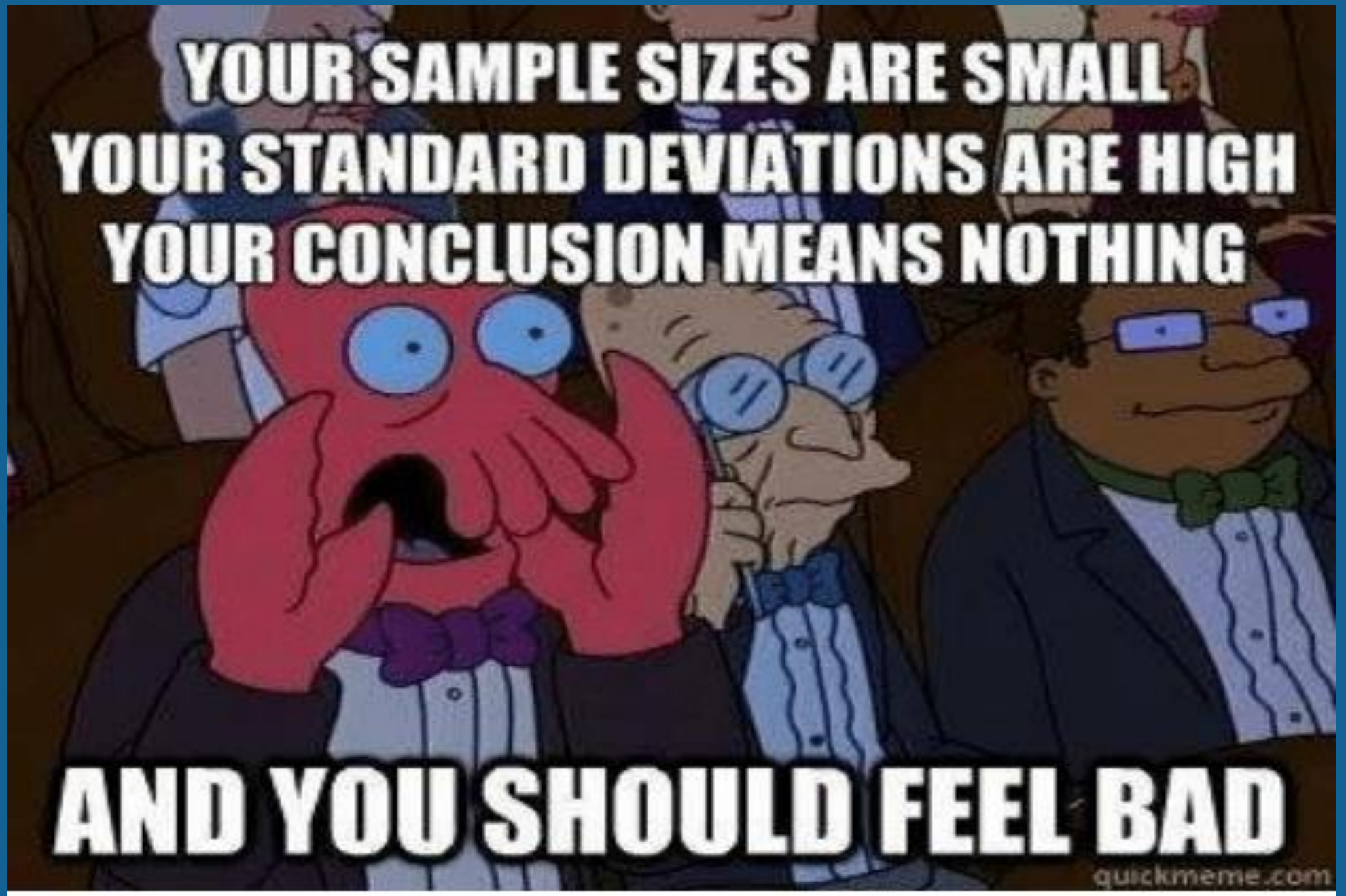
Creating dummy variables (0,1) out of categoricals is another data preparation step.

In R, you can use as.factor, as.numeric, as.integer to convert chr strings into categoricals then into values.

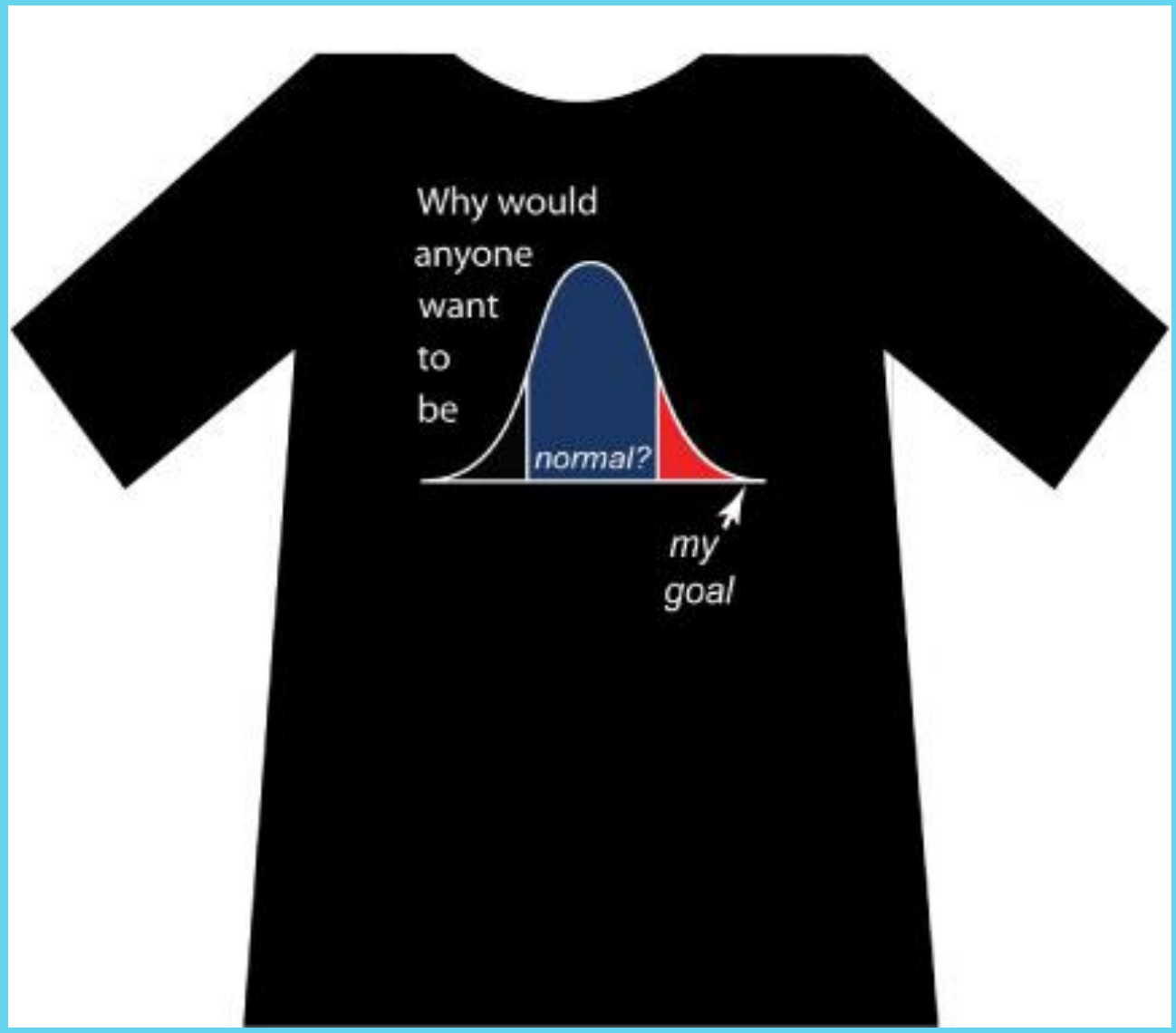
```
from statsmodels.stats.outliers_influence import variance_inflation_factor

# creating dummies for gender
data['Gender'] = data['Gender'].map({'Male':0, 'Female':1})

# the independent variables set
X = data[['Gender', 'Height', 'Weight']]

# VIF dataframe
vif_data = pd.DataFrame()
vif_data["feature"] = X.columns

# calculating VIF for each feature
vif_data["VIF"] = [variance_inflation_factor(X.values, i)
for i in range(len(X.columns))]

print(vif_data)
```

```
  feature        VIF
0  Gender   2.028864
1  Height  11.623103
2  Weight  10.688377
```

Our findings indicate that height and weight are collinear > 10. Using them together results in multicollinearity →


YOUR SAMPLE SIZES ARE SMALL
YOUR STANDARD DEVIATIONS ARE HIGH
YOUR CONCLUSION MEANS NOTHING
AND YOU SHOULD FEEL BAD
quickmeme.com

YOUR GOAL SETTING SHOULD BE:



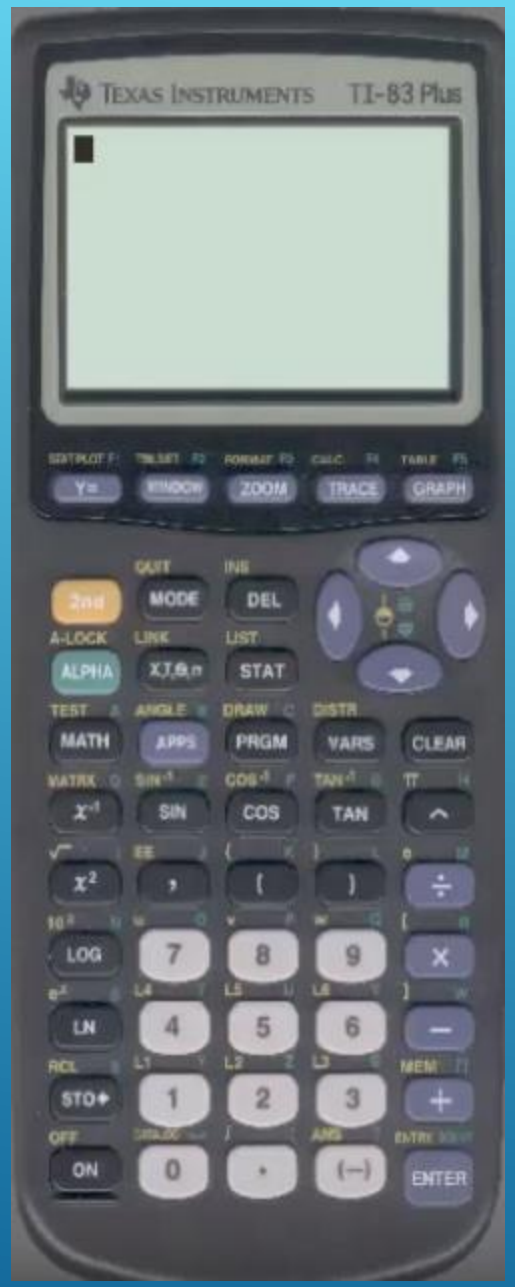
D208 PREDICTIVE MODELING WEBINAR

THANK YOU FOR JOINING US TODAY!

Presented by: Dr. William Sewell, IT

CALCULATOR CORNER

1. One Sample *t*-test
2. Chi-Square
3. Linear Regression (Includes Correlation)
4. Confidence Intervals
5. ANOVA
6. Two-Sample *t*-test
7. Logistic Regression



- **PERFORMING A SIMPLE ONE-SAMPLE T-TEST**

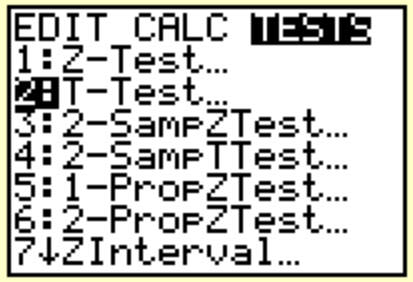
CALCULATOR CORNER

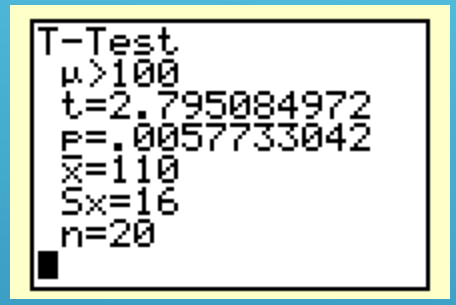
Calculate the p-value for a **T-Test** and accept/reject the null hypothesis.

Sample size: n = 20, mean = 110 and std dev = 16.

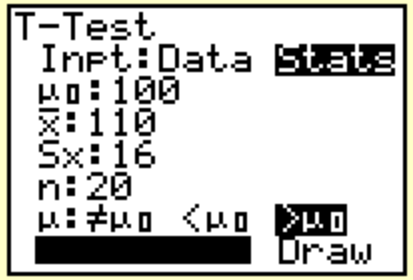
Press STAT and the right arrow twice to select TESTS.

To select the highlighted 2:T-Test…
Press ENTER.





Use right arrow to select Stats (summary values rather than raw data) and Press ENTER.
Use the down arrow to Enter the hypothesized mean, sample mean, standard deviation, and sample size.
Select alternate hypothesis.
Press down arrow to select Calculate and press ENTER.



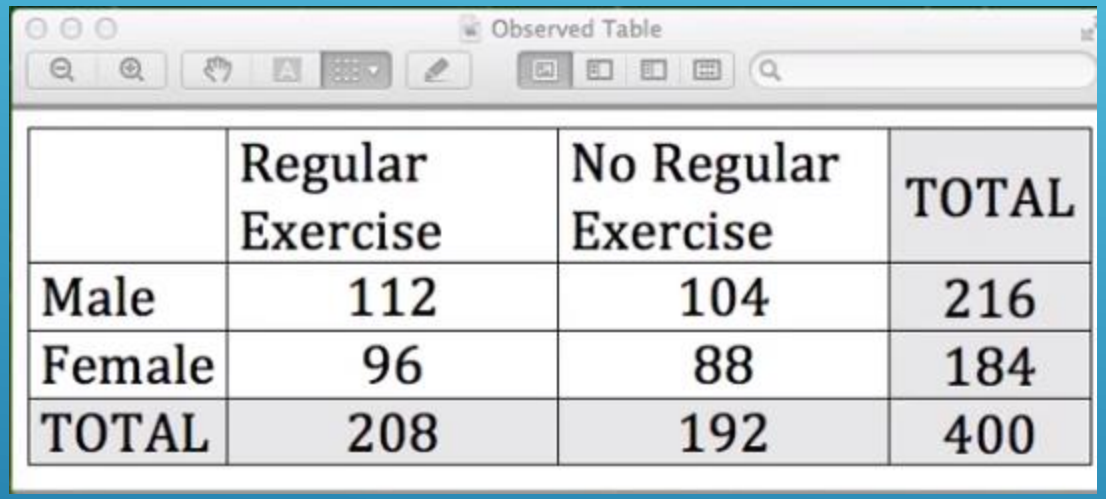
http://cfcc.edu/faculty/cmoore/TI83TTest.htm

- **CALCULATING CHI-SQUARE**

CALCULATOR CORNER

finding Chi-Square critical values

Using the following Contingency table, we want to calculate Chi-Square

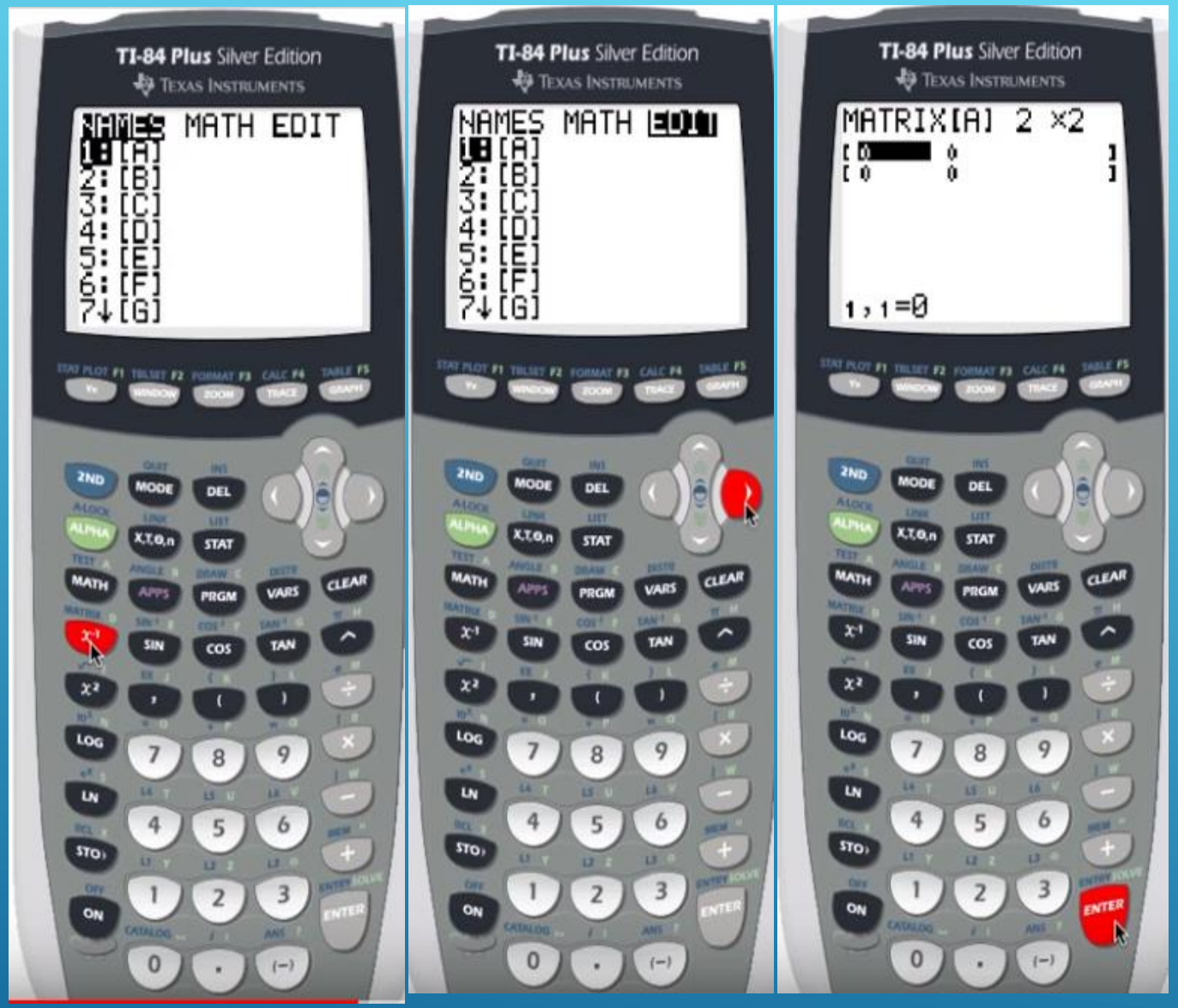


	Regular Exercise	No Regular Exercise	TOTAL
Male	112	104	216
Female	96	88	184
TOTAL	208	192	400

https://www.youtube.com/watch?v=wSszHVNWEEs

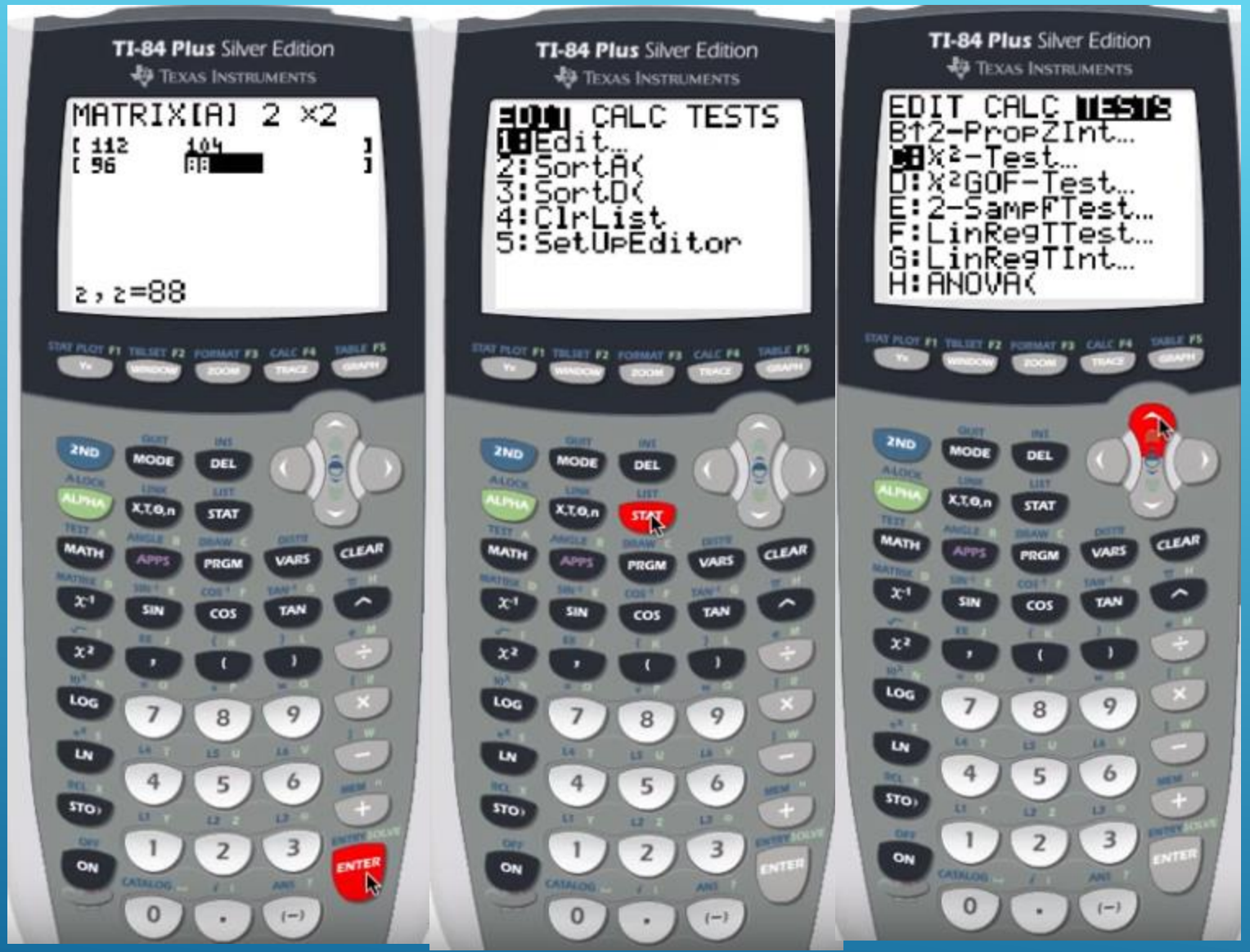
finding Chi-Square critical values

Step 1. 2^{nd} x^{-1} keys
Step 2. Right arrow to Edit, Enter
Step 3. Type in rows and columns
Step 4. Enter each value, Enter



finding Chi-Square critical values

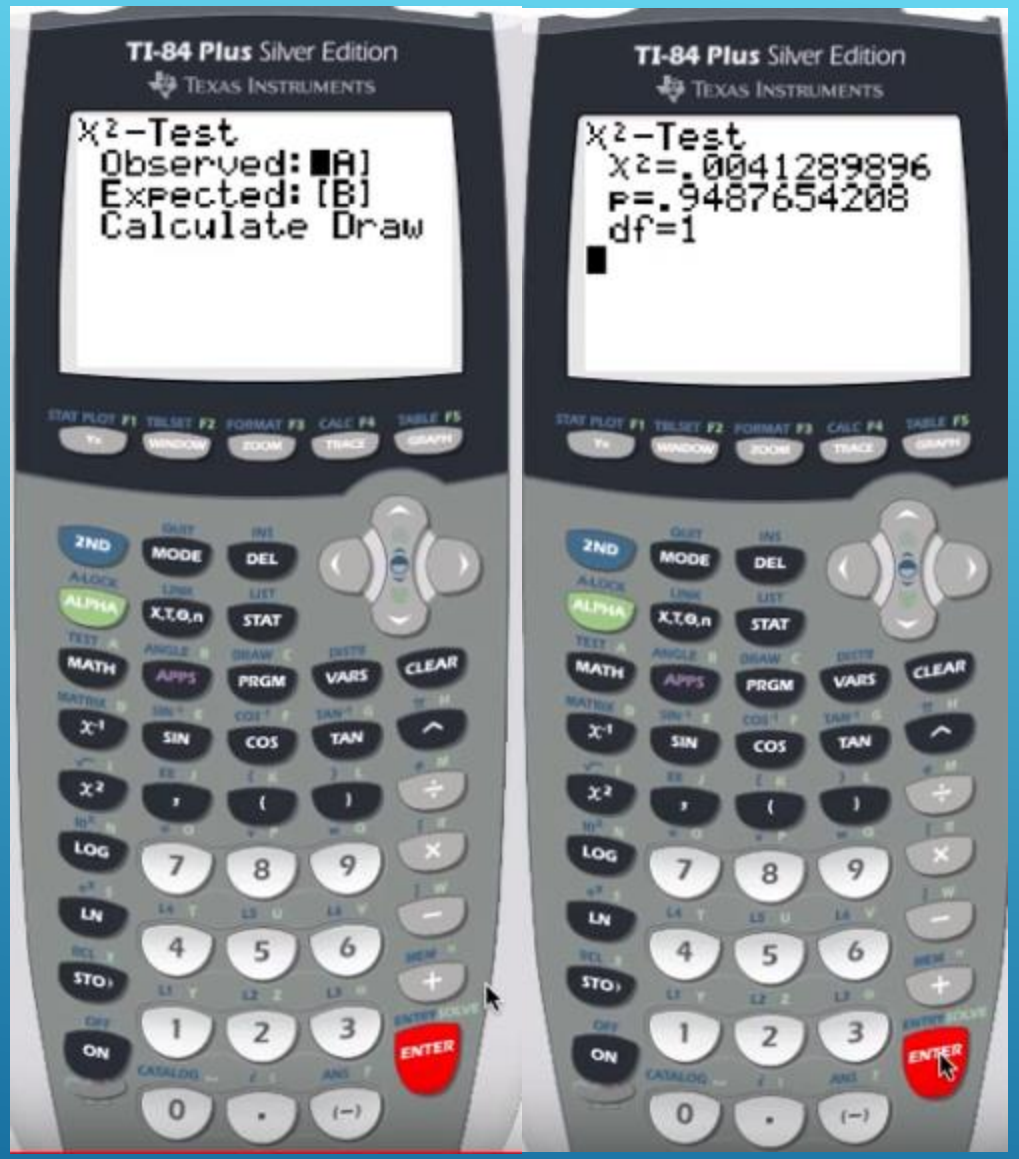
Step 5. Press Stat button
Step 6. Right arrow to Tests,
Scroll Up to Chi-Square, Enter
Step 7. Type in rows and columns
Step 8. Enter each value, Enter



finding Chi-Square critical values

Step 9. Press Stat button
Step 10. Right arrow to Tests,
Scroll Up to Chi-Square, Enter
Step 11. Press Enter, Enter, Enter

Your results are displayed as follows:
Chi-Square value
P-value
Degrees of freedom



https://www.youtube.com/watch?v=wSszHVNWEEs

- **PERFORMING AND CALCULATING LINEAR REGRESSIONS**

CALCULATOR CORNER

New!

CALCULATING LINEAR REGRESSION

For Linear Regression

Step 1: Press the Stat key and the display shows as seen in the first picture on the left.

Step 2: Then press 1 to enter your List 1 values.

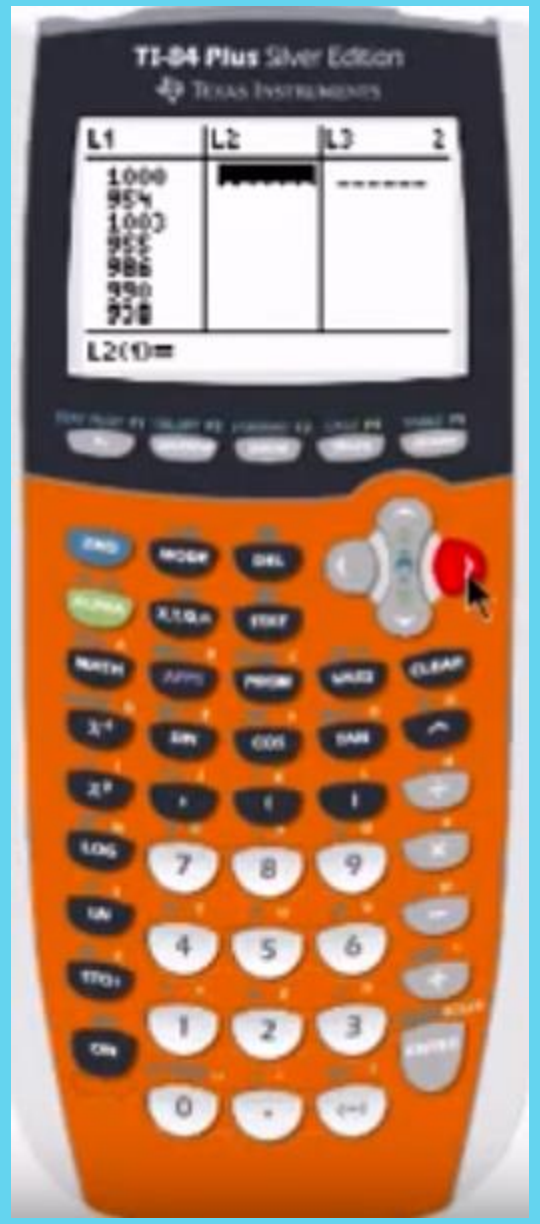
Step 3: Type the first number and press the Enter key as shown.

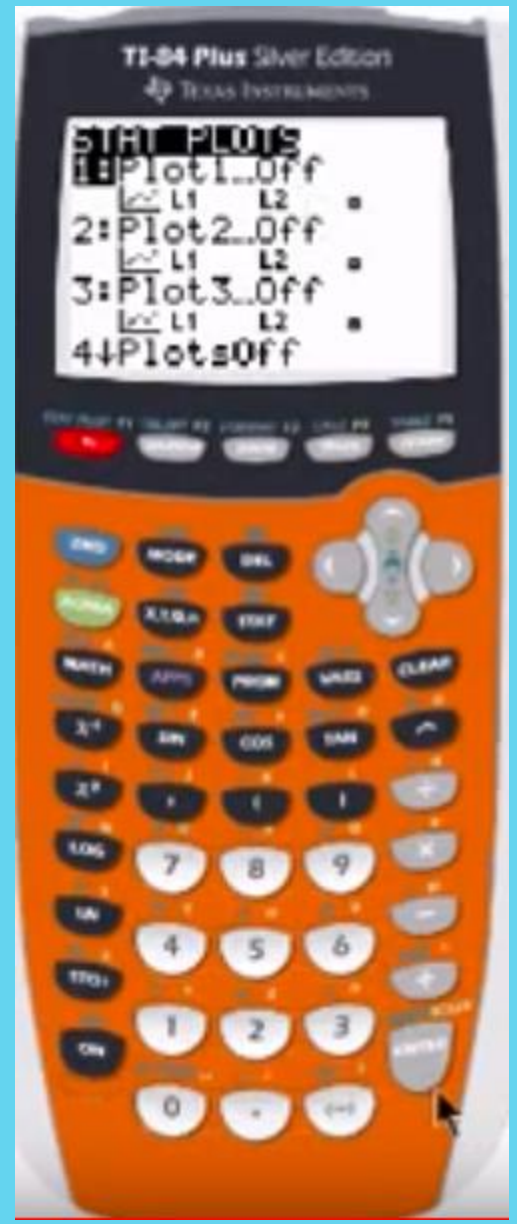
Step 4: Repeat Step 3 for each value

Step 5: Right Arrow and enter each value for List 2.

Step 6: Press 2nd then Y= key just above it.

Step 7: That brings up STAT PLOTS, choose 1 and Press Enter





New!

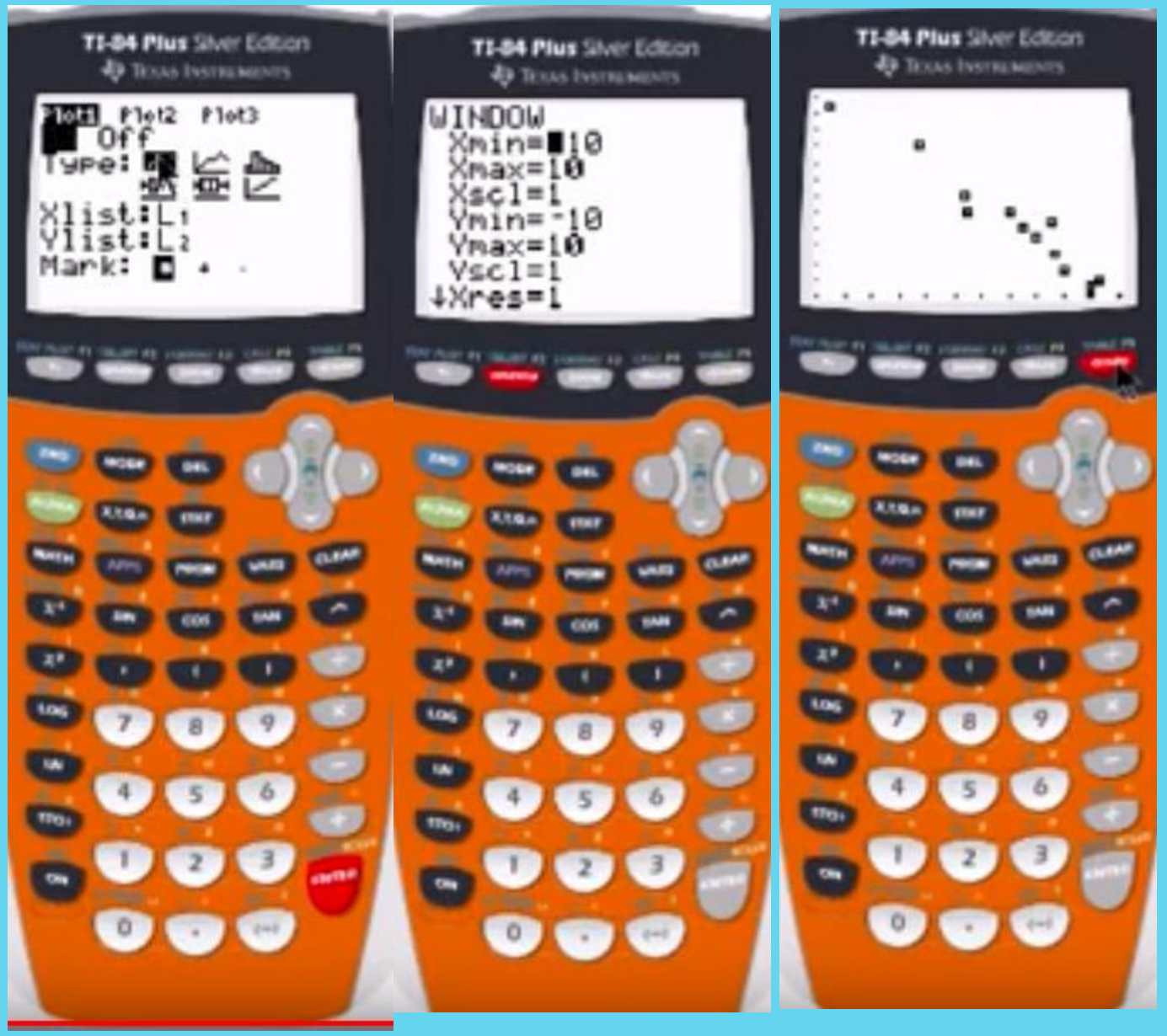
For Linear Regression

Step 8: Plot1 appears, Press Enter so that it is ON

Step 9: By default scatter plot and L1 and L2 are set. X-values are in L1, Y-values are in L2. Press F2 Table Set (shown in Red) to set up your X-axis and Y-axis values.

Step 10: Press the Graph button (Red) to view the data points.

Step 8 | Step 9 | Step 10



New!

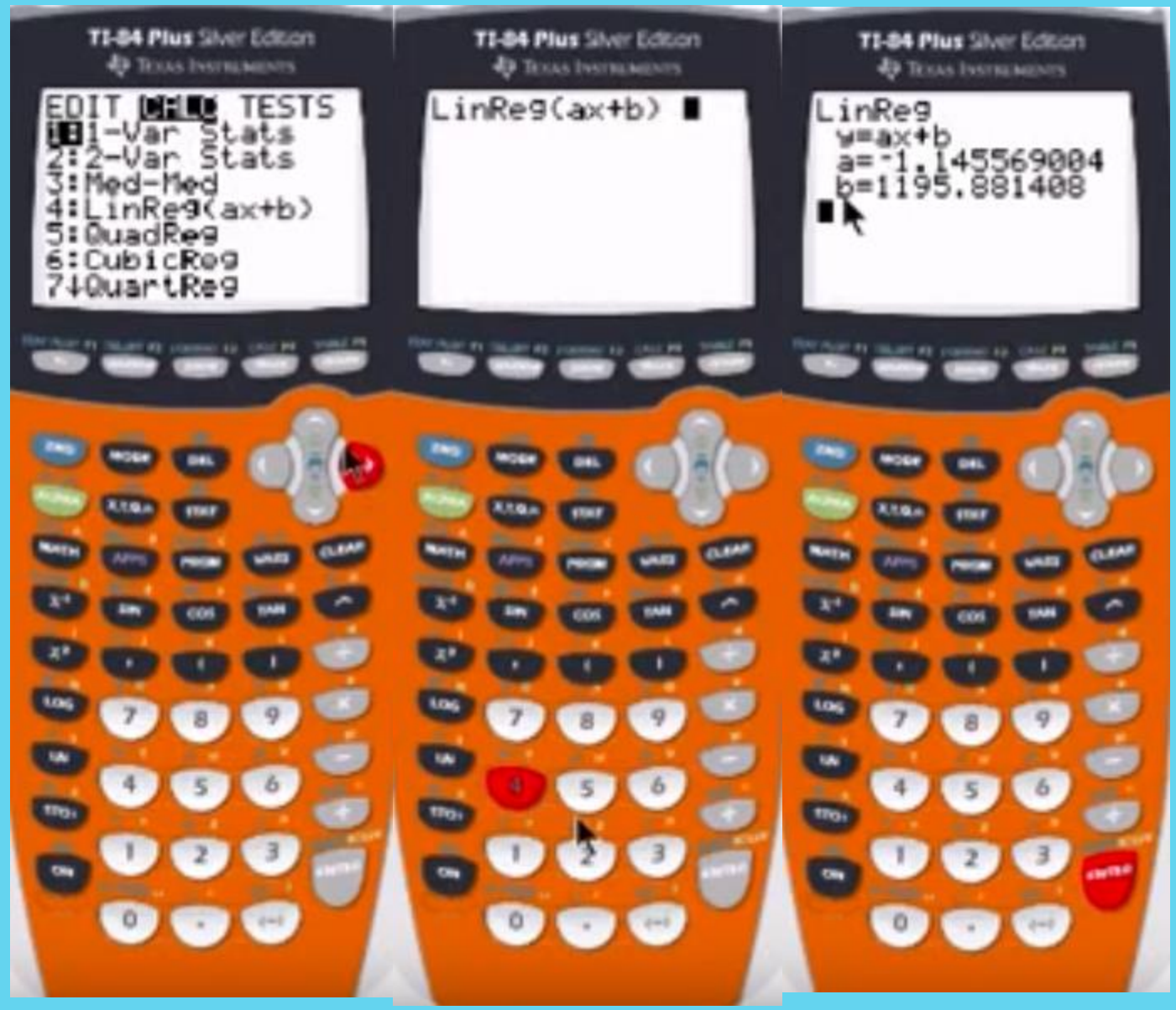
For Linear Regression

Step 11: Press STAT button right arrow to Calc and Press 4 for Lin(ear) Reg(ression) and Press Enter.

Step 12: LinReg (ax + b) displays. Press Enter

Step 13: a = Slope , b = y-intercept you will use these values for plotting your regression.

Step 11 | Step 12 | Step 13



New!

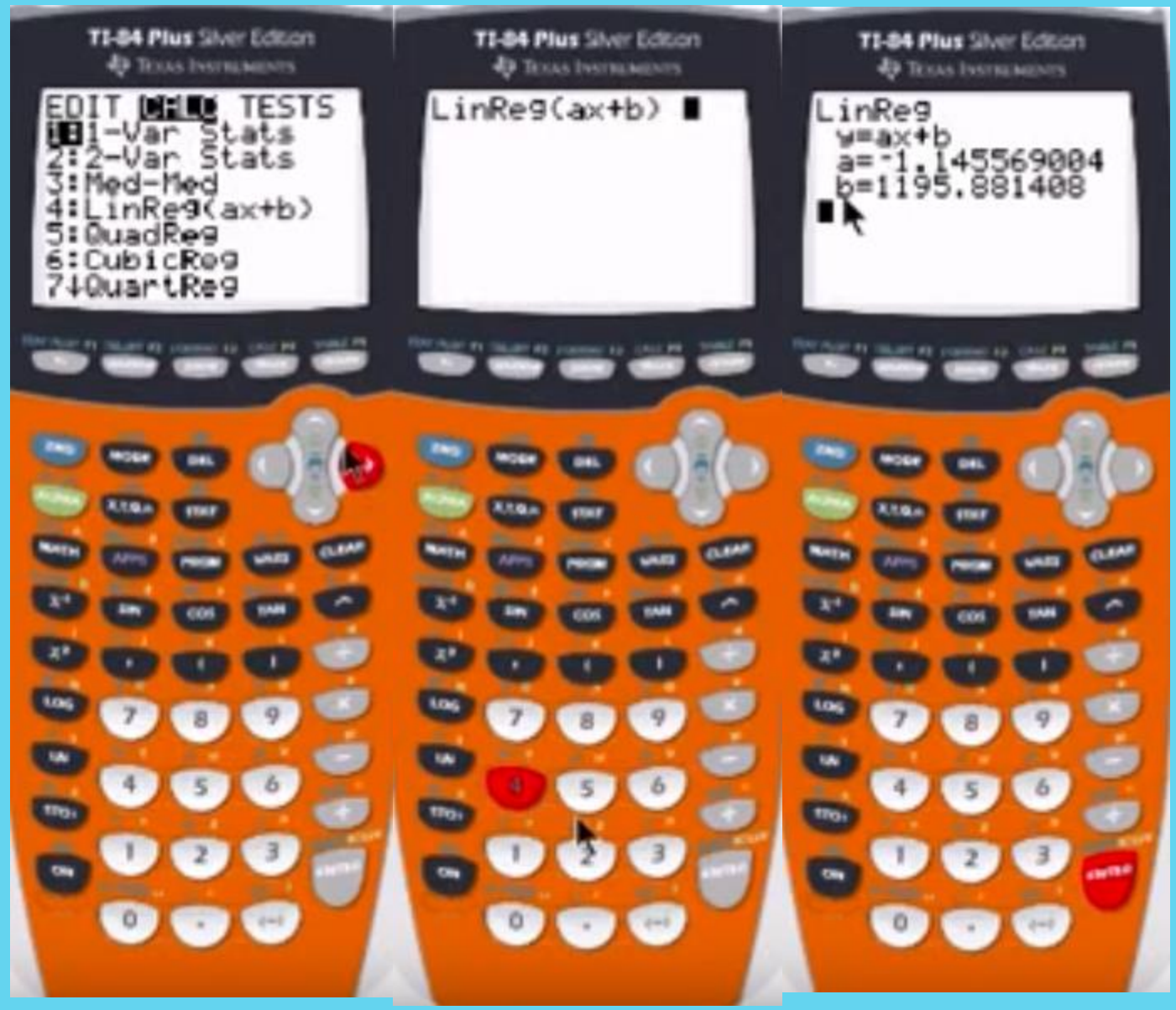
For Linear Regression

Step 14: Press STAT button right arrow to Calc and Press 4 for Lin(ear) Reg(ression) and Press Enter.

Step 15: LinReg (ax + b) displays. Press Enter

Step 16: a = Slope , b = y-intercept you will use these values for plotting your regression. To get the correlation coefficient Press 2nd (blue button) and catalog (0 button).

Step 14 | Step 15 | Step 16



New!

For Linear Regression

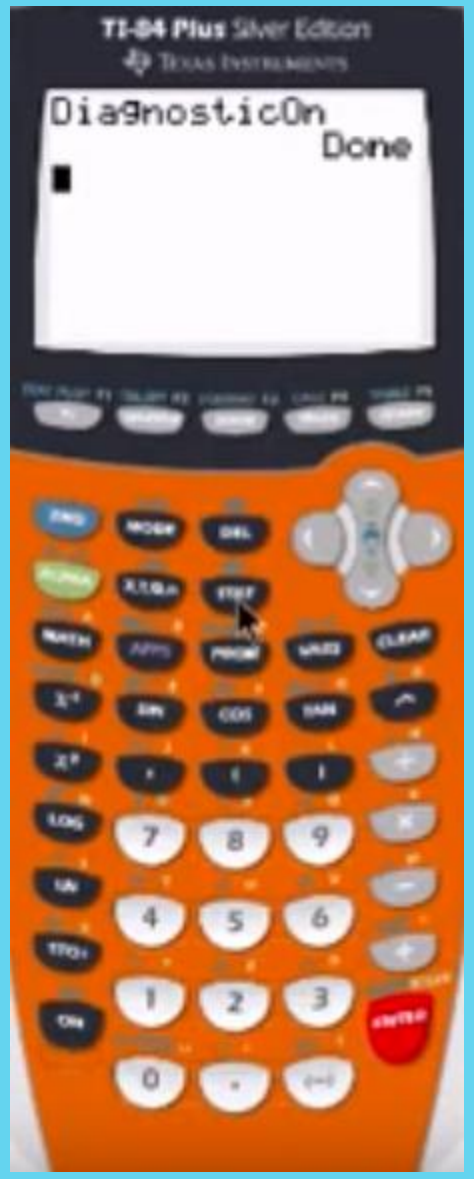
Step 17: CATALOG menu appears and arrow down to Diagnostic On and press Enter.

Step 18: The display should look like this. Now Press STAT, CALC, 4, Enter

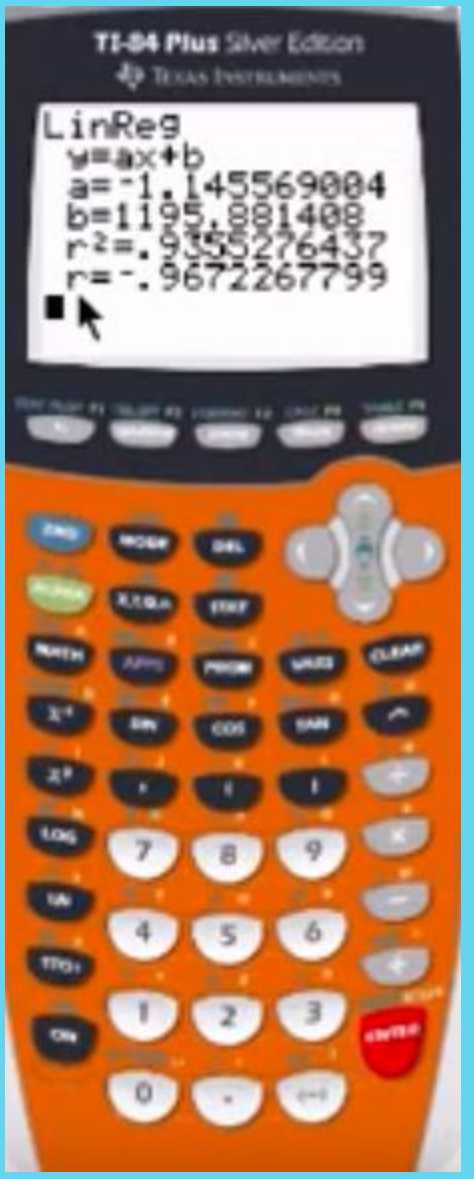
Step 19: Now you have the r^2 and r values for the correlation. The r value tells how well the line fits the data. The typical distance a point is from the regression line.

Step 17

Step 18



Step 19



New!

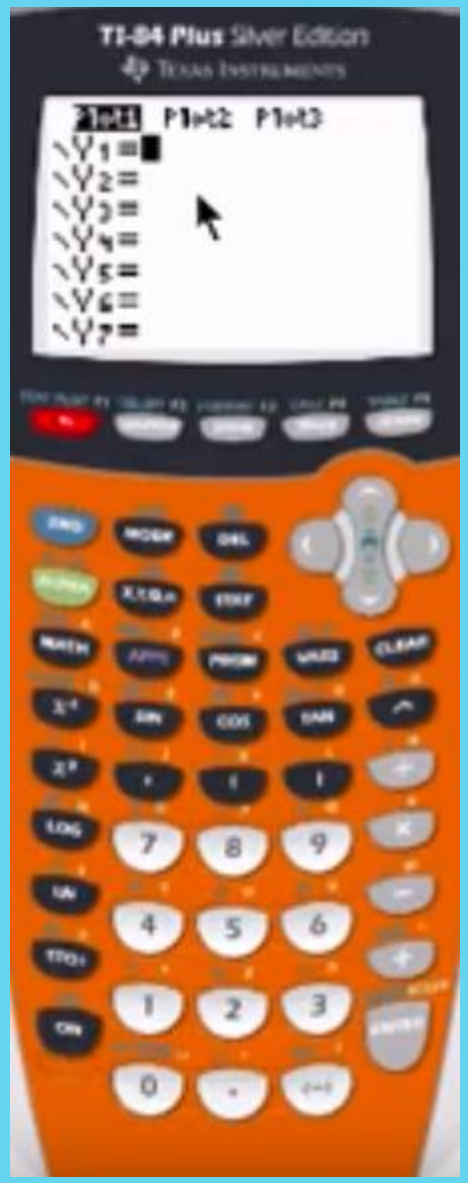
For Linear Regression

Step 20: Press y = button

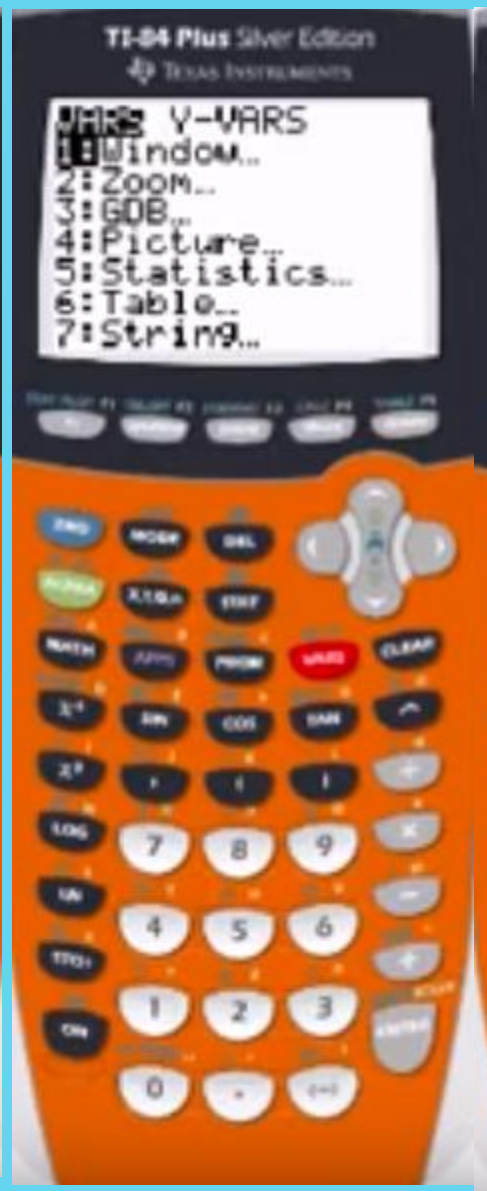
Step 21: Press VARS button, 5 (Statistics),

Step 22: Right arrow to EQ, 1 (Reg), Enter

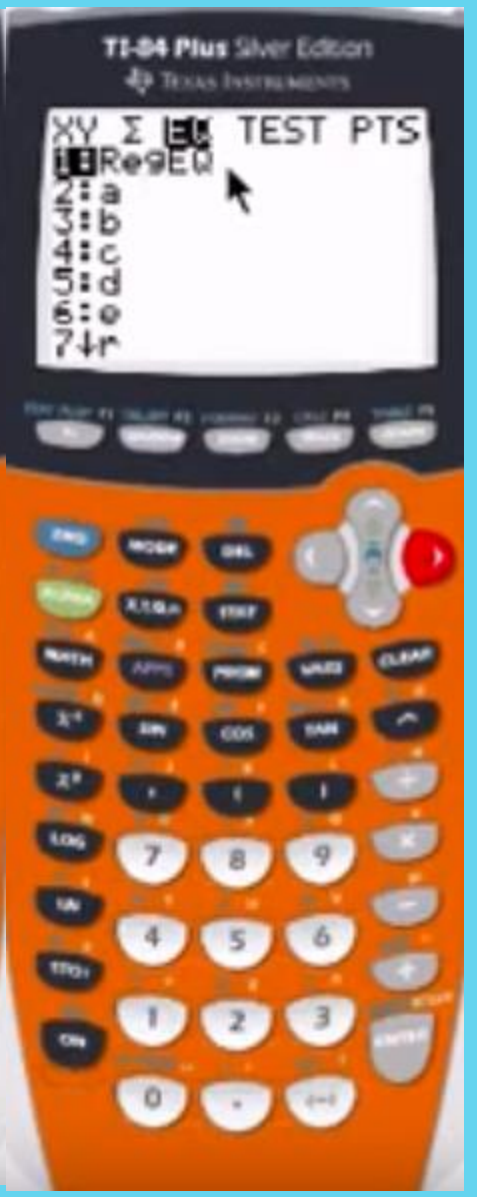
Step 20



Step 21



Step 22



New!

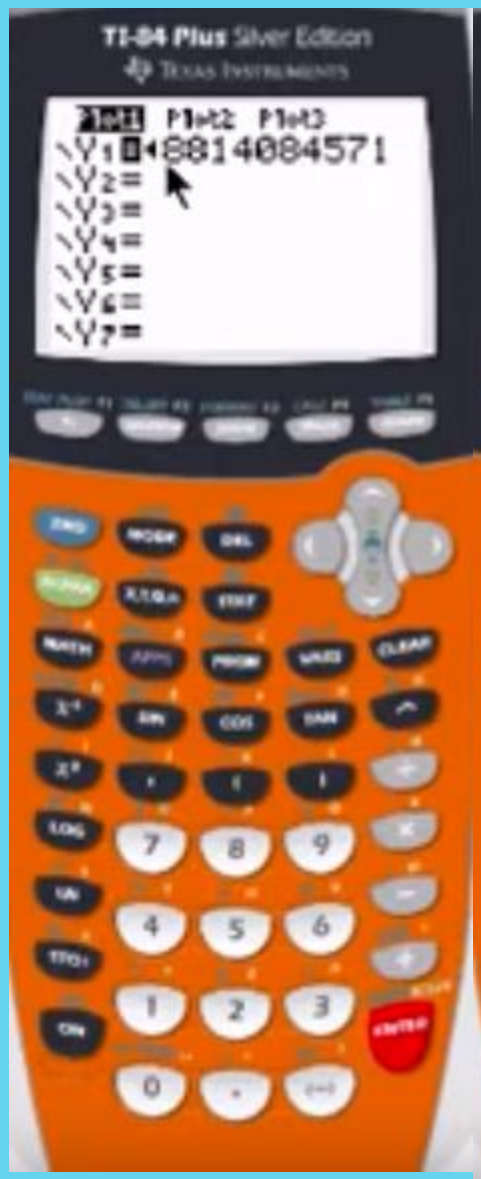
For Linear Regression

Step 23: The data is automatically entered for the regression equation. To view it, just use the left and right arrows.

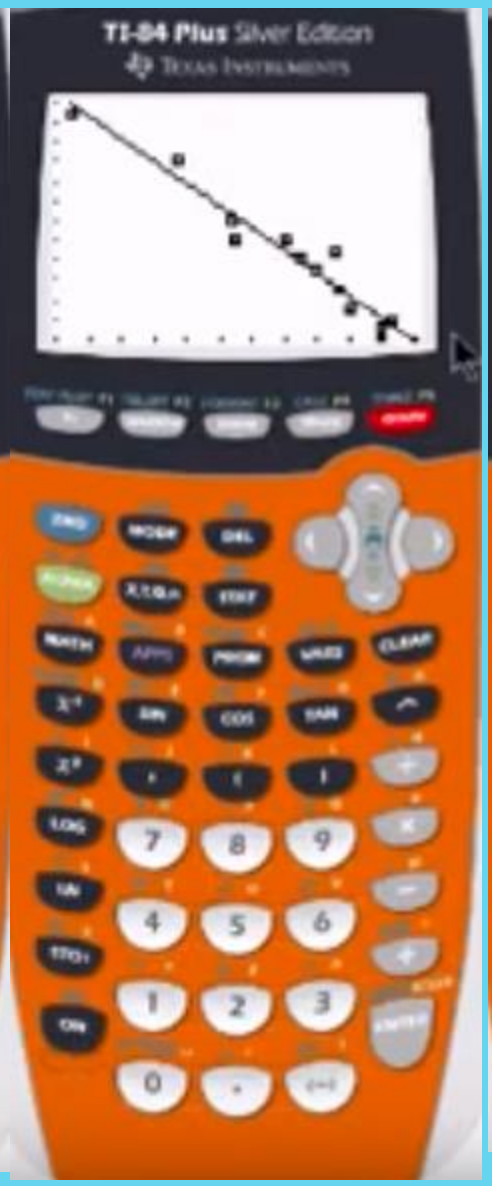
Step 24: Now Press Graph (red button) and the regression line is drawn for you.

Step 25: If you want to enter a new value to see where it would be on the graph, just press 2nd Calculate, 1, Enter.

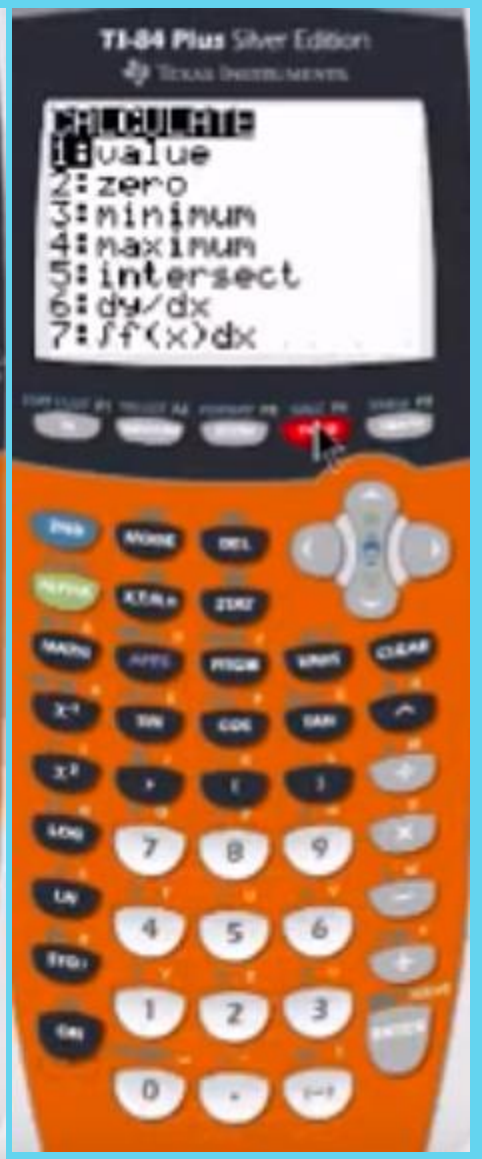
Step 23



Step 24



Step 25



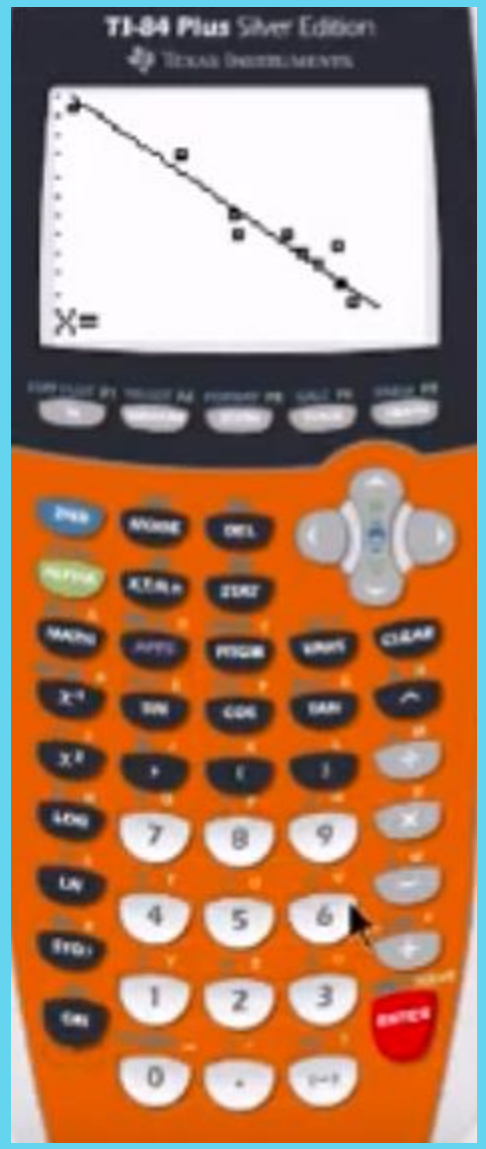
New!

For Linear Regression

Step 26: Enter the X value and Press Enter. The corresponding Y value is displayed.

Step 27: The new data point is displayed.

Step 26



Step 27

- **CALCULATING CONFIDENCE INTERVALS**

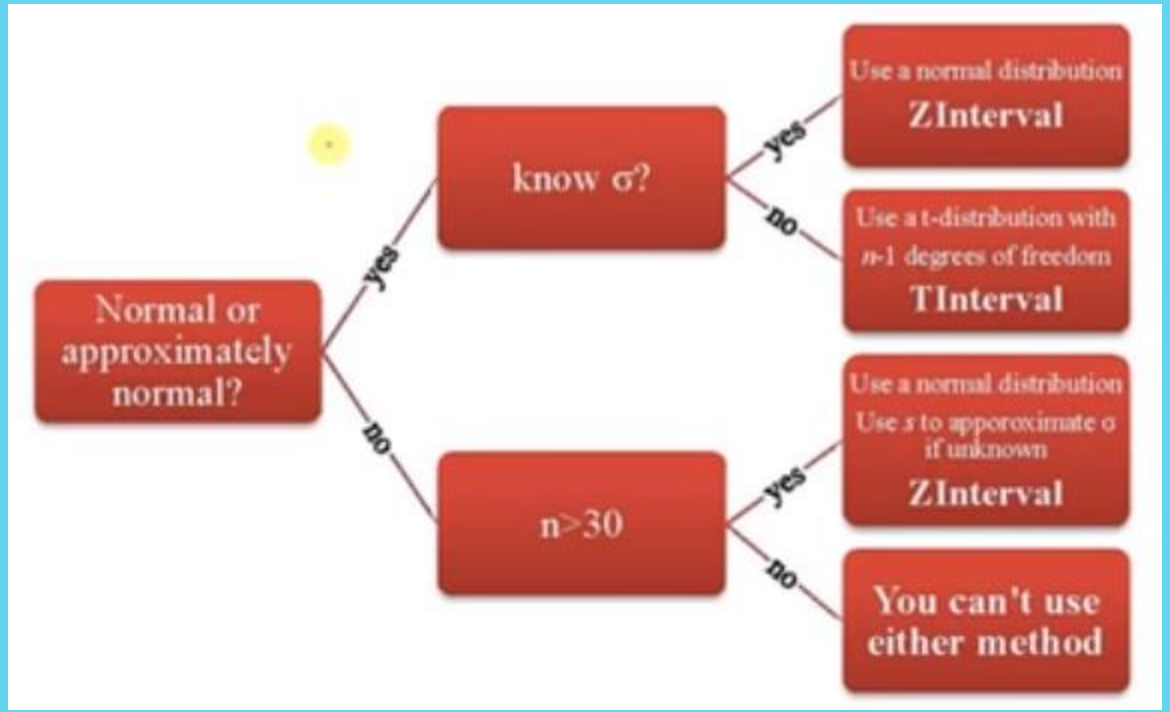
CALCULATOR CORNER

New!

CALCULATING CONFIDENCE INTERVALS

For Confidence Intervals

Here is a handy chart to know whether to Use a Z Interval or a T Interval to find a Confidence Interval or not.



New!

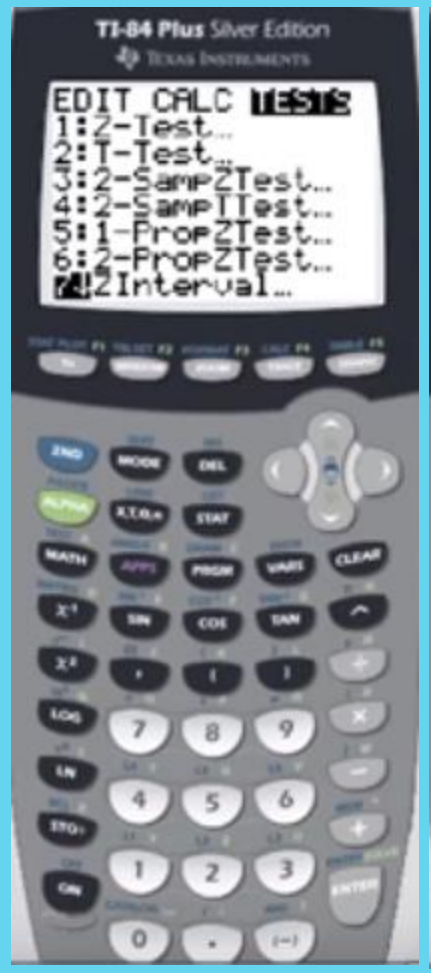
Finding Confidence Intervals

Step 1: Press STAT button, arrow to TESTS, arrow down to 7 (Z Interval), Press Enter.

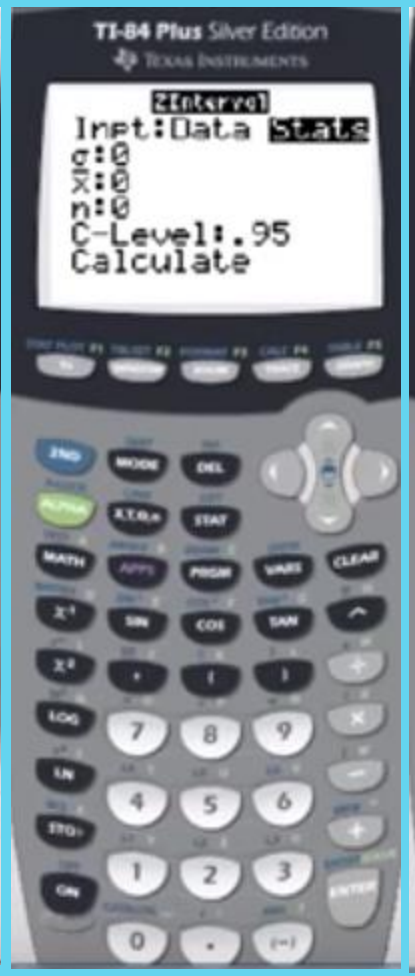
Step 2: Arrow to Stats, Press Enter
Now you can enter σ (std dev) X-bar (mean), n (sample pop), and the Confidence level. Arrow down and enter each value, then arrow down to calculate and Press Enter.

Step 3: As an example, we entered std dev = 12, mean = 120, population = 150

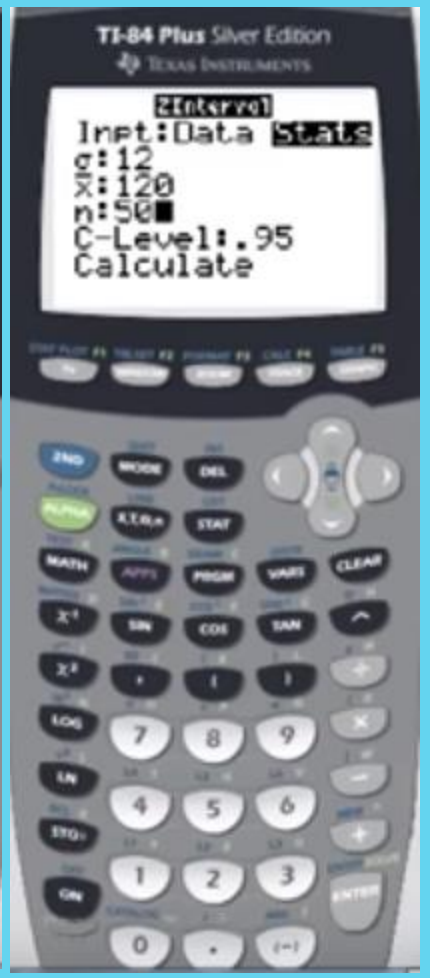
Step 1



Step 2



Step 3



New!

Finding Confidence Intervals

Step 4: The Confidence Interval is displayed with upper (123.33) and lower (116.67) bounds.

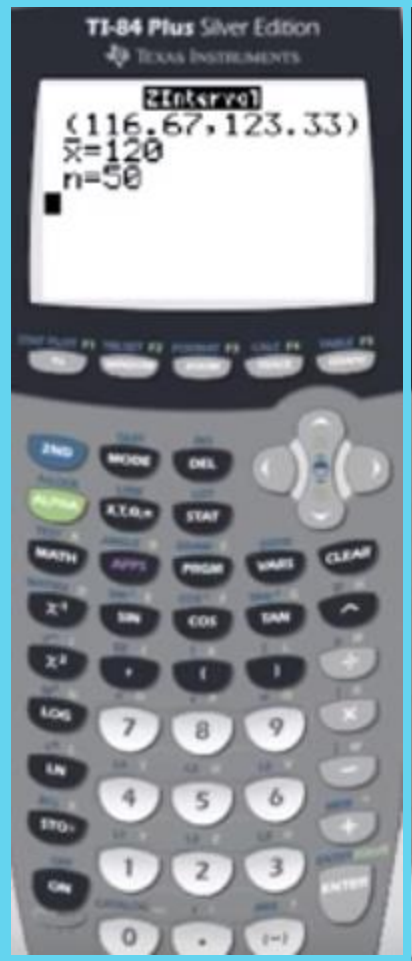
Step 5: if you don't have the population std. dev. You can run a T Interval Press STAT, TESTS, 8 (T Interval) for the T distribution.

Step 6: Notice now, we enter a mean, a Sample std dev, and a sample population. Press Calculate.

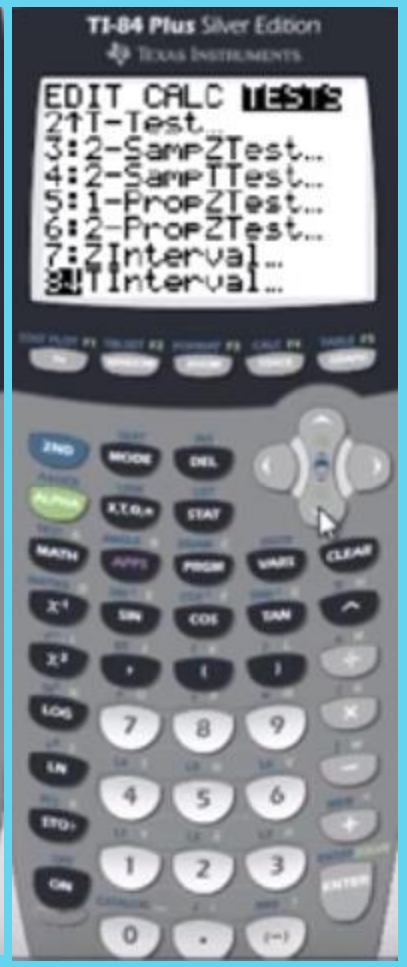
Step 7: Again, we see the Confidence Interval.

Finally, you can Enter raw data into a List and use the List to run your Conf. Interval.

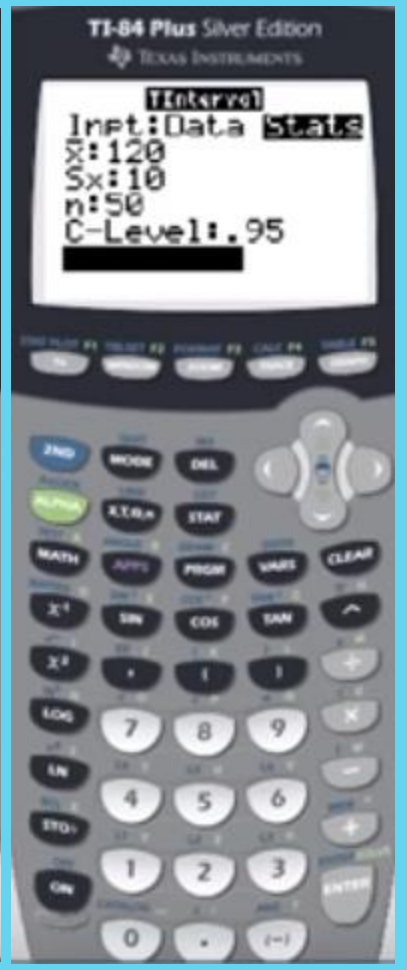
Step 4



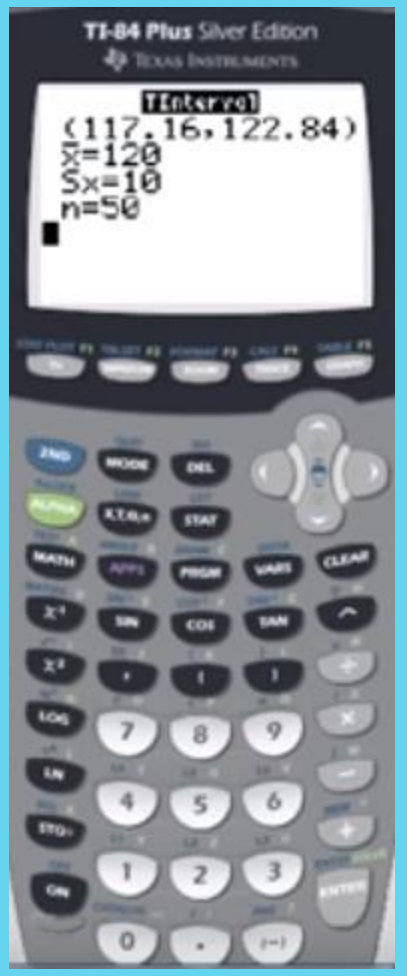
Step 5



Step 6



Step 7



- **CALCULATING AND UNDERSTANDING ANOVA**

CALCULATOR CORNER

New!

CALCULATING ANOVA

ANOVA Step-by-Step

In a Rifle company, we took six soldiers from each platoon and Measured their scores on the firing range to determine training Effectiveness.

There are two measures with ANOVA, df between groups and df within groups. Why? Because the dispersion of the data within the group can mitigate the difference between the groups.

With ANOVA, we want to find out if there is a significant difference Among all groups. If there is, then we can t-test individually between The groups.

Rifle Marksmanship Scores			
Plt A	Plt B	Plt C	Plt D
22	26	24	19
18	20	22	18
20	19	18	17
21	21	16	21
19	28	19	22
17	16	15	24
117	130	114	121
19.5	21.7	19.0	20.2
1.87	4.50	3.46	2.64

ANOVA	Groups		
df between	k - 1	4-1	3 numerator
df within	N - k	24 -4	20 denominator
df total			23
F crit =	3.0983		
F score =	0.754		
p-value=	0.5329		

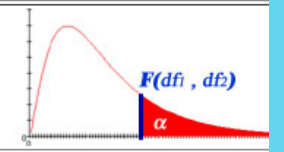
New!

ANOVA Step-by-Step

F- Table for our example.

The F_{crit}

F Table for α = 0.05



/	df_1=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
df_2=1	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433	241.8817	243.9060	245.9499	248.0131	249.0518	250.0951	251.1432	252.1957	253.2529	254.3144
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848	19.3959	19.4125	19.4291	19.4458	19.4541	19.4624	19.4707	19.4791	19.4874	19.4957
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5494	8.5264
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877	5.6581	5.6281
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.3650
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.3080	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.8780
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432

New!

Performing ANOVA on a TI-83 Calculator

Step 1: Press STAT button,
Press Enter.

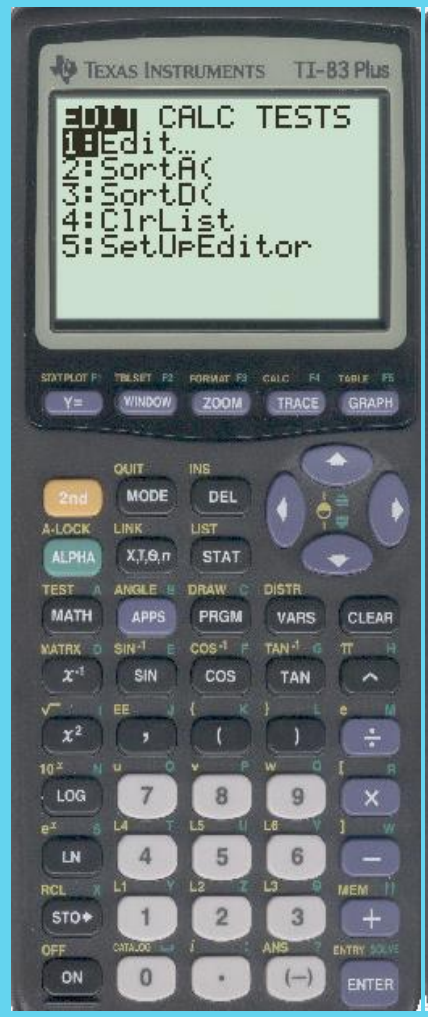
Step 2: Type the Rifle scores from Plt A
Pressing Enter after each score.

Step 3: Use the Right Arrow key to move to the next group (Platoon) to enter those scores.

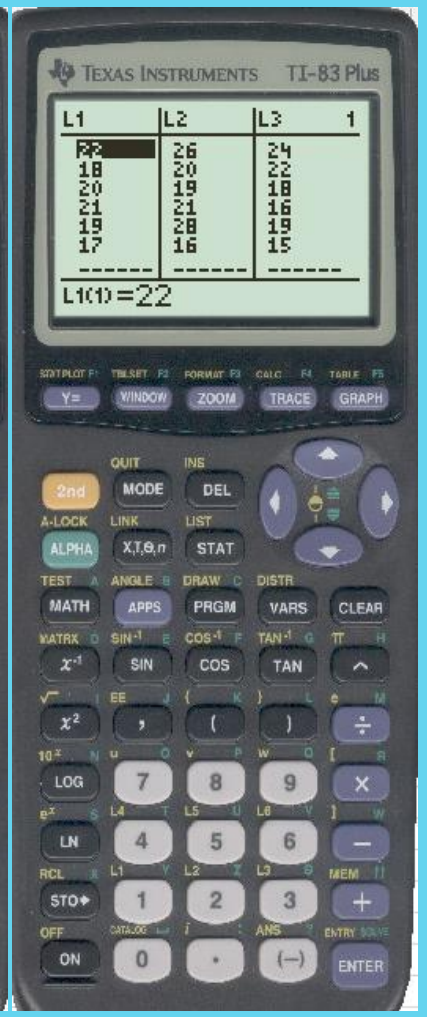
Step 4: When all the scores are entered
Press STAT

Step 5: Right Arrow to TESTS

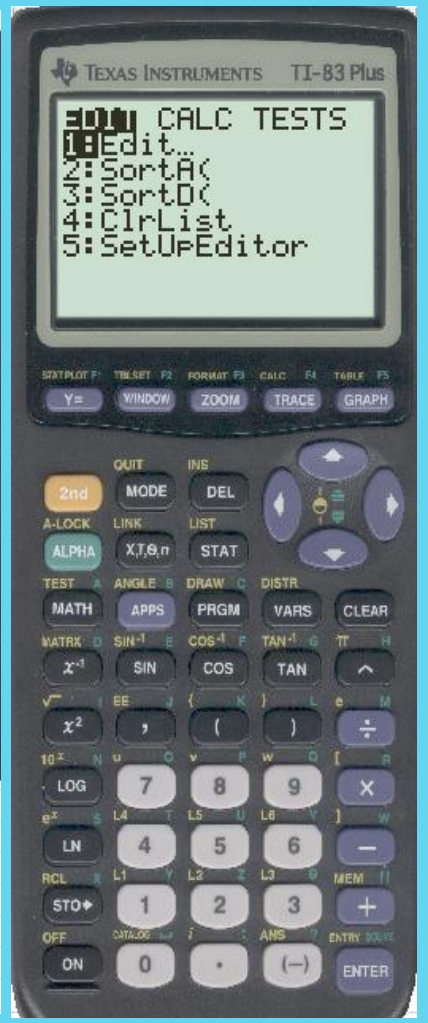
Step 1



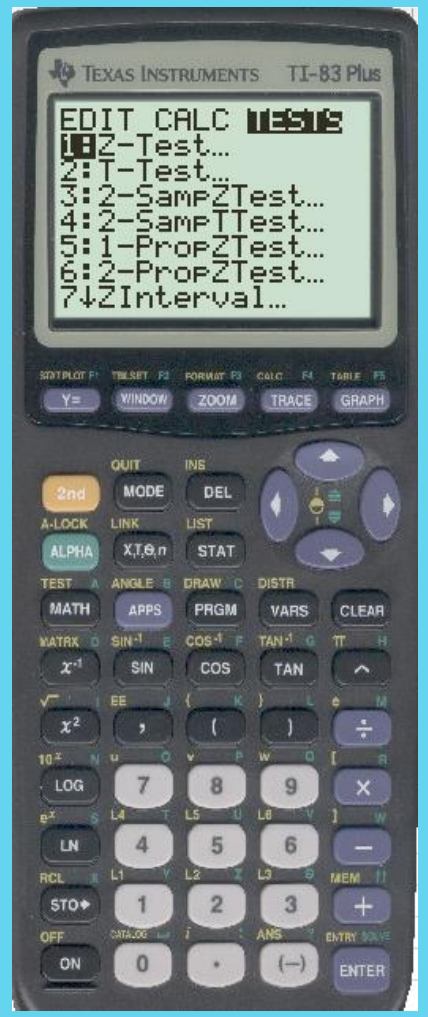
Step 2/3



Step 4



Step 5



New!

Performing ANOVA on a TI Calculator

Step 6: Arrow down to Test F: ANOVA.

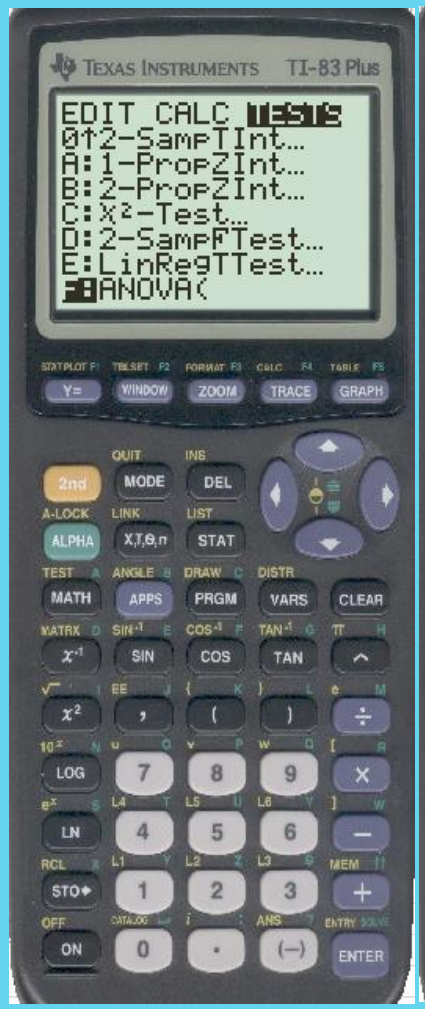
Step 7: Press Enter and ANOVA(appears.

Step 8: Do the following:
- a. Press 2nd L1.
- b. Press , key.
- c. Repeat for L1 through L4 lists.
- d. Press) .

Step 9: Press Enter.

Conclusion: Fcalc < Fcrit. P-value Confirms that there is no significant Difference among the groups.

Step 6



Step 7

Step 8

Step 9

New!

CALCULATING ANOVA

ANOVA Step-by-Step

In a Rifle company, we took six soldiers from each platoon and Measured their scores on the firing range to determine training Effectiveness.

There are two measures with ANOVA, df between groups and df within groups. Why? Because the dispersion of the data within the group can mitigate the difference between the groups.

With ANOVA, we want to find out if there is a significant difference Among all groups. If there is, then we can t-test individually between The groups.

Rifle Marksmanship Scores			
Plt A	Plt B	Plt C	Plt D
22	26	24	19
18	20	22	18
20	19	18	17
21	21	16	21
19	28	19	22
17	16	15	24
117	130	114	121
19.5	21.7	19.0	20.2
1.87	4.50	3.46	2.64

ANOVA	Groups		
df between	k - 1	4-1	3 numerator
df within	N - k	24 -4	20 denominator
df total			23
F crit =	3.0983		
F score =	0.754		
p-value=	0.5329		

- **CALCULATING A TWO-SAMPLE *t*-TEST**

CALCULATOR CORNER

New!

Two-Sample *t*-test Step-by-Step

Two brands of automobiles were tested for actual fuel mileage in comparison to each other for 30 different cars each. The mean fuel mileage for Brand A was 24.5 and for Brand B it was 29.2. The standard deviation was 2.9 and 3.6 respectively as shown below.

Is there a significant difference? Does Brand B have better overall fuel mileage as they contend?

	Auto Mileage Figures	
	Brand A	Brand B
Mean	24.5	29.2
Std dev	2.9	3.6
n	30	30

New!

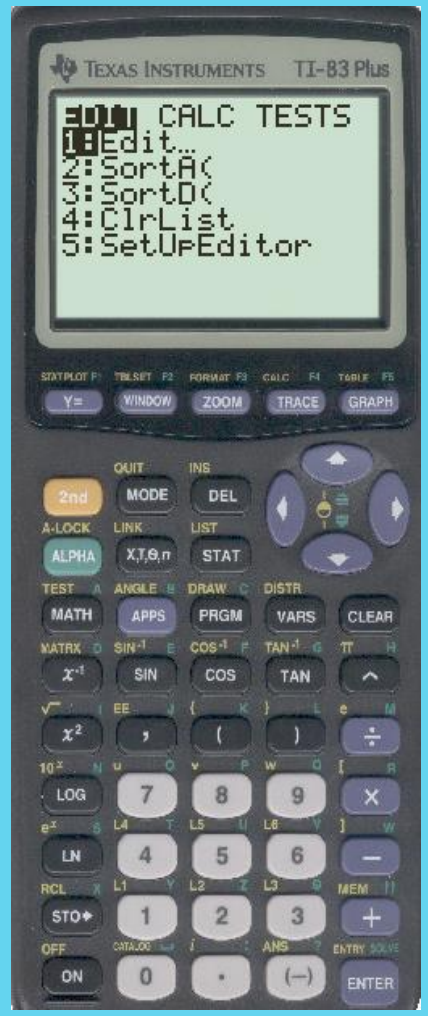
Performing a Two Sample *t*-Test on a TI-83 Calculator

Step 1: Press STAT button, Right Arrow to TESTS.

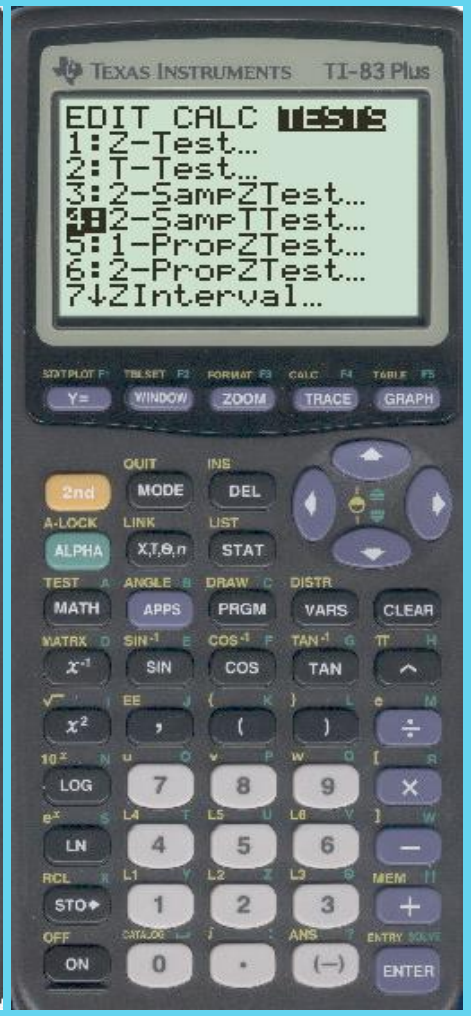
Step 2: Arrow down to 2-SampTTest and Press Enter.

Step 3: Use the Right Arrow key to move to Stats and Press Enter.

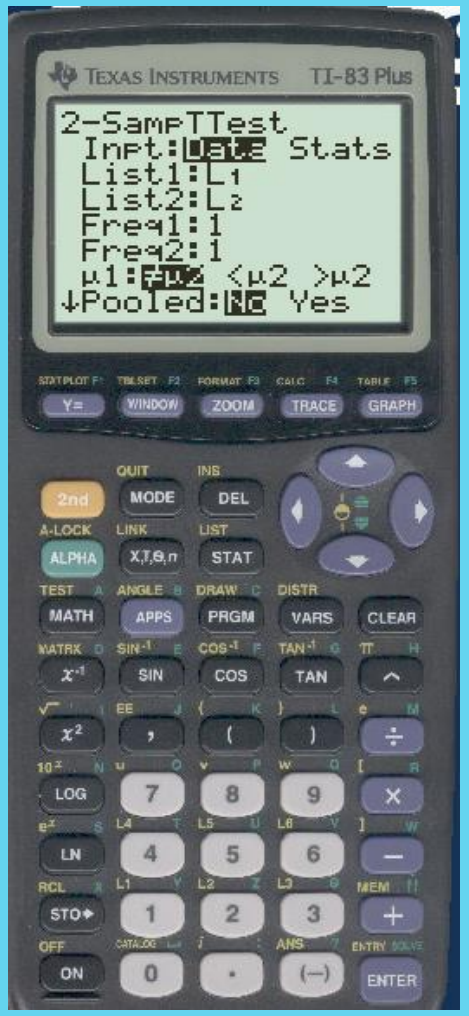
Step 1



Step 2



Step 3



New!

Performing a Two-Sample *t*-Test on a TI-83 Calculator

Step 4: The screen should look like this. Now, arrow down down x-bar 1 and begin entering the data from the scenario.

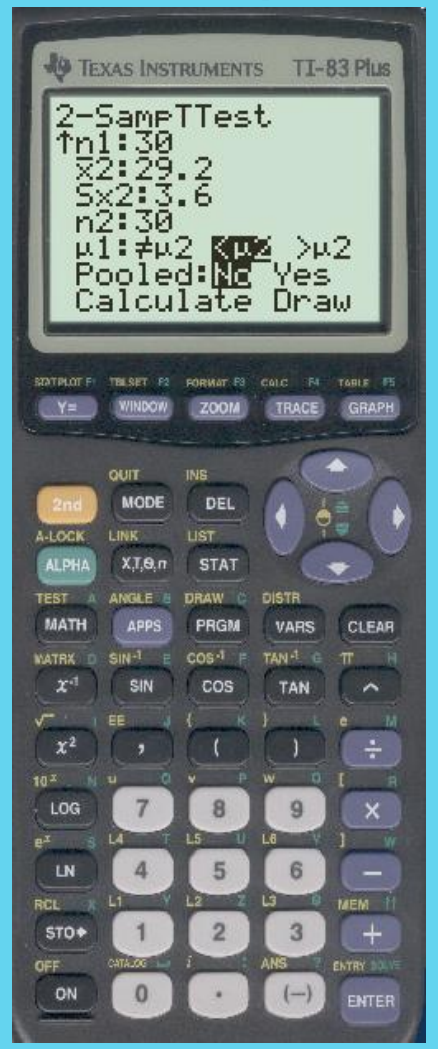
Step 5: Your screen should look like this. Now, arrow down and select <u2 since the alternative hypothesis is that Brand B is greater than Brand A.

Step 6: Arrow down and select Pooled No, since we do not assume that variances are equal, arrow down to Calculate and Press Enter.

Step 4

Step 5

Step 6



New!

Performing a Two-Sample *t*-test on a TI-83 Calculator

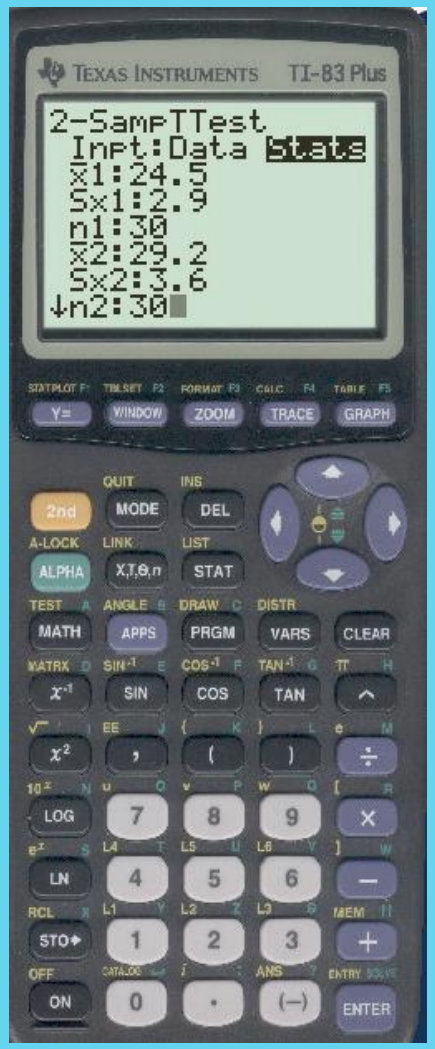
Step 4: The screen should look like this. Now, arrow down down x-bar 1 and begin entering the data from the scenario.

Step 5: Your screen should look like this. Now, arrow down and select <u2 since the alternative hypothesis is that Brand B is greater than Brand A.

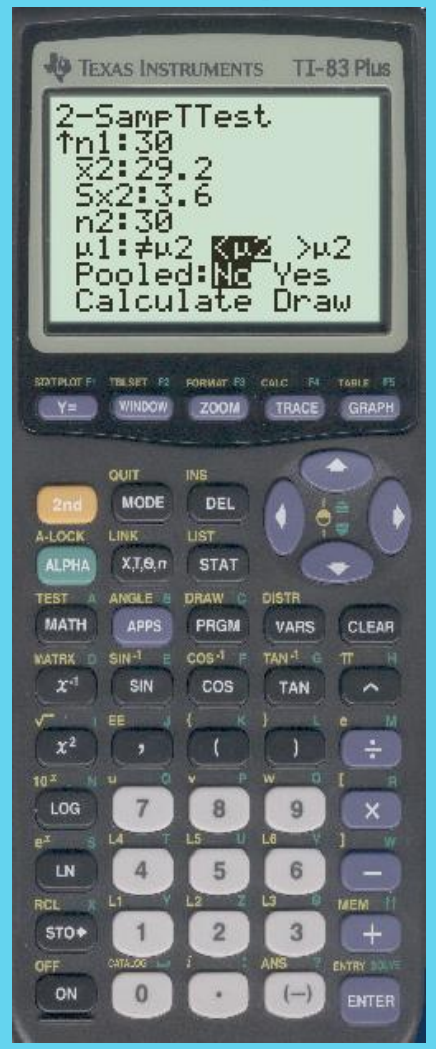
Step 6: Arrow down and select Pooled No, since we do not assume that variances are equal, arrow down to Calculate and Press Enter.

Step 4

Step 5



Step 6



New!

Performing a Two-Sample *t*-test on a TI-83 Calculator

Step 7: The T-value is negative since u1 < u2 for the fuel mileage.

Step 8: Your screen should look like this. Now, arrow down and select <u2 since the alternative hypothesis is that Brand B is greater than Brand A. The T-value is greater 5.56 than the t-critical value of 1.684 at a alpha level of 0.05.

The p-value is also significant 3.869 e -7 = 0.0000003869

Therefore, we reject the null hypothesis.

Step 7

TABLE of CRITICAL VALUES for STUDENT'S *t* DISTRIBUTIONS

Column headings denote probabilities (α) ***above*** tabulated values.

d.f.	0.40	0.25	0.10	0.05	0.04	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	7.916	12.706	15.894	31.821	63.656	127.321	318.289	636.578
2	0.289	0.816	1.886	2.920	3.320	4.303	4.849	6.965	9.925	14.089	22.328	31.600
3	0.277	0.765	1.638	2.353	2.605	3.182	3.482	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.333	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.191	2.571	2.757	3.365	4.032	4.773	5.894	6.869
6	0.265	0.718	1.440	1.943	2.104	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.046	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.004	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	1.973	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	1.948	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	1.928	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	1.912	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	1.899	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	1.887	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	1.878	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	1.869	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	1.862	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	1.855	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	1.850	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	1.844	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	1.840	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	1.835	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	1.832	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.256	0.685	1.318	1.711	1.828	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	1.825	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	1.822	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	1.819	2.052	2.158	2.473	2.771	3.057	3.421	3.689
28	0.256	0.683	1.313	1.701	1.817	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	1.814	2.045	2.150	2.462	2.756	3.038	3.396	3.660
30	0.256	0.683	1.310	1.697	1.812	2.042	2.147	2.457	2.750	3.030	3.385	3.646
31	0.256	0.682	1.309	1.696	1.810	2.040	2.144	2.453	2.744	3.022	3.375	3.633
32	0.255	0.682	1.309	1.694	1.808	2.037	2.141	2.449	2.738	3.015	3.365	3.622
33	0.255	0.682	1.308	1.692	1.806	2.035	2.138	2.445	2.733	3.008	3.356	3.611
34	0.255	0.682	1.307	1.691	1.805	2.032	2.136	2.441	2.728	3.002	3.348	3.601
35	0.255	0.682	1.306	1.690	1.803	2.030	2.133	2.438	2.724	2.996	3.340	3.591
36	0.255	0.681	1.306	1.688	1.802	2.028	2.131	2.434	2.719	2.990	3.333	3.582
37	0.255	0.681	1.305	1.687	1.800	2.026	2.129	2.431	2.715	2.985	3.326	3.574
38	0.255	0.681	1.304	1.686	1.799	2.024	2.127	2.429	2.712	2.980	3.319	3.566
39	0.255	0.681	1.304	1.685	1.798	2.023	2.125	2.426	2.708	2.976	3.313	3.558
40	0.255	0.681	1.303	1.684	1.796	2.021	2.123	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	1.781	2.000	2.099	2.390	2.660	2.915	3.232	3.460

- **CALCULATING LOGISTIC REGRESSION**

CALCULATOR CORNER

New!

Logistic Regression Step-by-Step

Florida Fish and Wildlife want to estimate future populations of fish in a large lake.

Using the data below and predict when the lake will reach maximum estimated population.

Month	Population (1,000s)
0	1000
1	3000
2	6000
3	9000
4	10000
5	10500

Performing a Logistic Regression on a TI-83/84 Calculator

Step 1: Press STAT, Enter, Input List 1, List 2

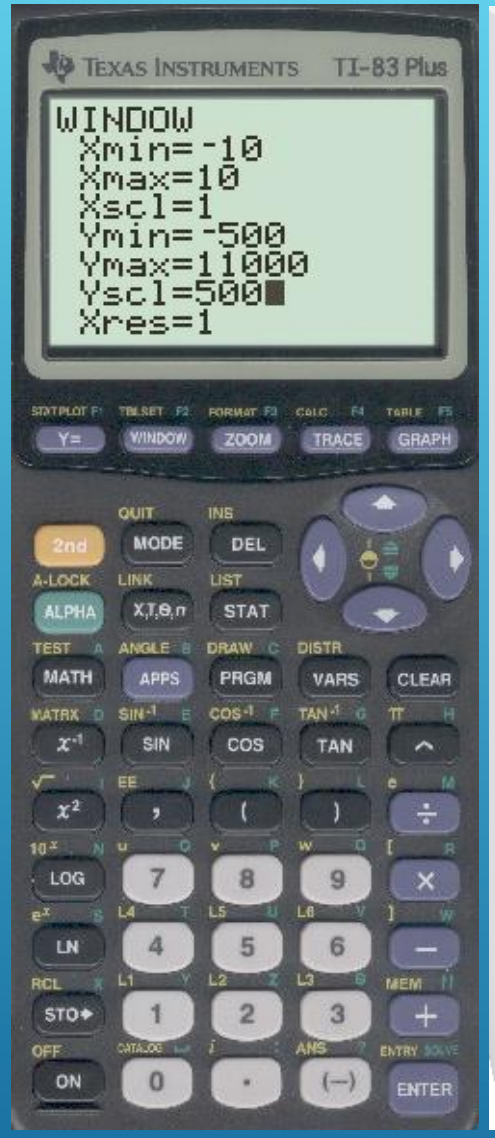
Step 2: Press Window and set size as shown

Step 3: Press Graph, to see the data points (TI-84) May have to turn Diagnostics On For TI-83. (2nd Catalog Diagnostics On). If Data points do not appear, Press 2nd Stat Plot Enter (On) , choose Type (Scatter Plot).

Step 1

Step 2

New!



Step 3

New!

Performing a Logistic Regression on a TI-83/84 Calculator

Step 4: Press Stat, right arrow, down to B Logistic regression

Step 5: Press Enter, Vars, right arrow, Enter, Enter

The regression equation is stored in Y_1, Enter

Step 4

Step 5

New!

Performing a Logistic Regression on a TI-83/84 Calculator

Step 6: The screen should look like this. The values for variables a, b, c are given.

The regression equation is:

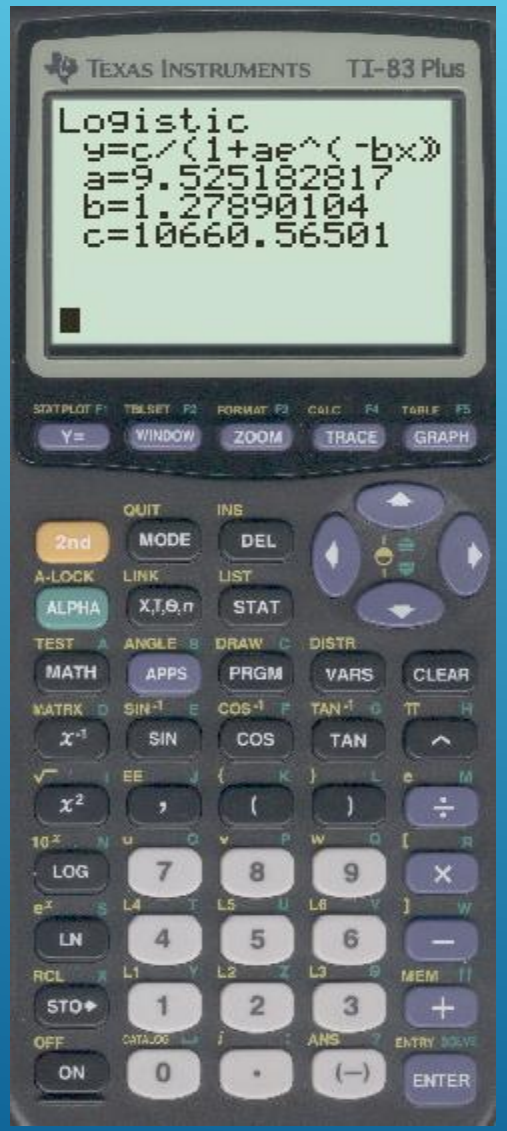
$$y = 10660.6/1 + 9.525\, e^{-1.279x}$$

Notice these contain the values for c in the numerator and b, a in the denominator.

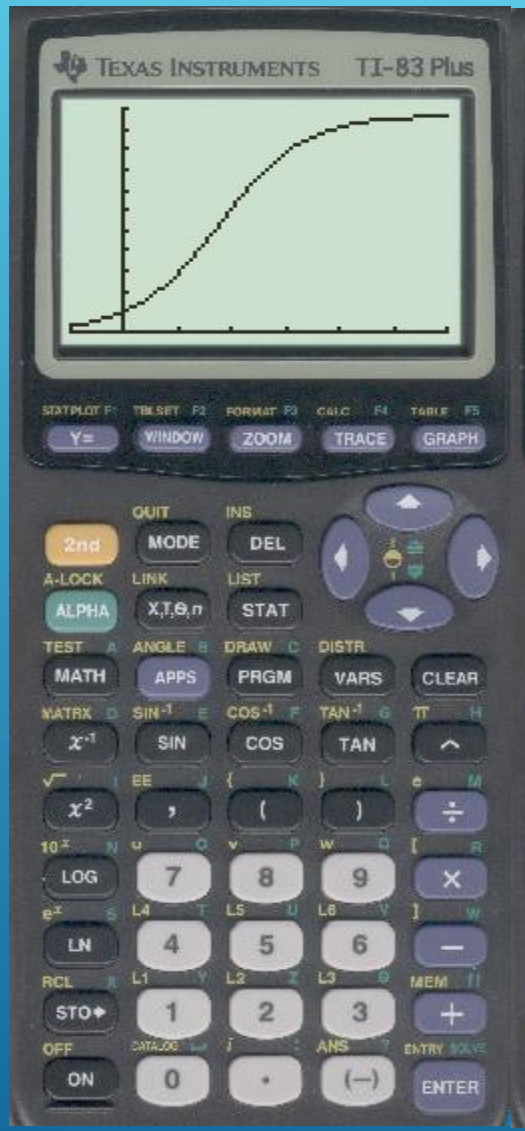
Step 7: Press Graph, S-curve appears.

Step 8: Press Vars, right arrow, Enter, Enter, (12) to predict the fish population for twelve months.

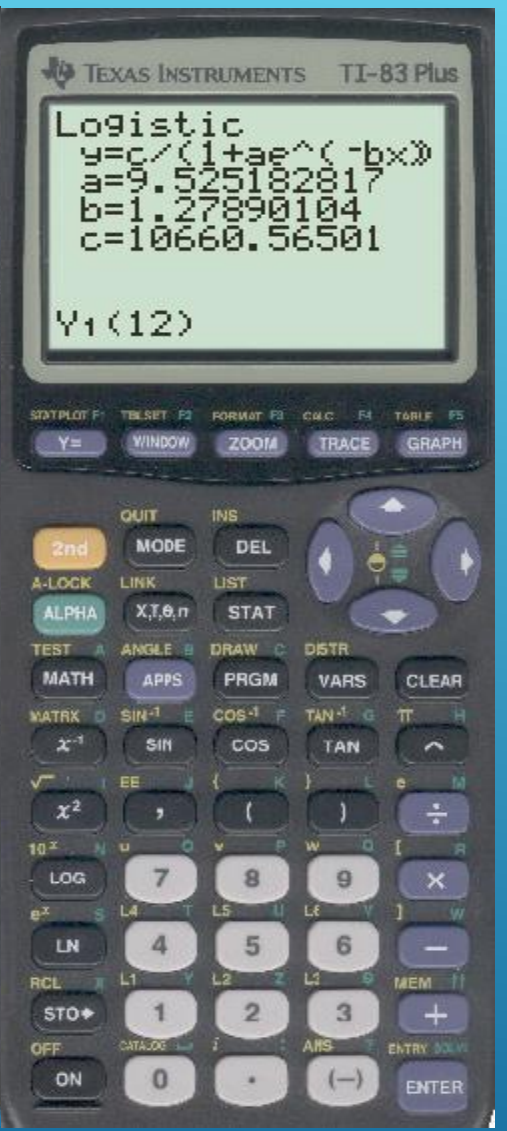
Step 6



Step 7



Step 8



New!

Performing a Logistic Regression on a TI-83/84 Calculator

Step 9: Press y= , Vars, 5 (Statistics), Enter

Step 10: Right Arrow to EQ, Press Enter

Step 11: Type 12 and Press Enter to predict the fish population for twelve months.

Step 12. Press 2nd TABLE, arrow down to 12

Notice that the fish populations are The same for months 13, 14, …

This means that the fish population Is expected to achieve maximum Size at 12 months.

Step 9 | Step 10 | Step 11 | Step 12