# Using Technology in Teaching Introductory Statistics at SGU

*Dishon Edwards[1,2], Stefan Hypolite[1], Sally-Ann Clement[1], Aleksandr Mylläri[1]*

[1] *School of Arts and Sciences, St. George's University, Grenada, West Indies*

[2] *TA Marryshow Community College, Grenada, West Indies*

Polynomial Computer Algebra 2024
April 15–20, 2024

Modern technology (computers, gadgets, computer algebra systems) changes the way **to do** mathematics and the way **to teach** mathematics and natural sciences.

We discuss the use of technology in teaching introductory statistics. Using computers in the course saves time by doing most of the work on the computer and leaves more time in class for explanations and discussions. It also prepares the student for the real life – no one does statistical analysis by hand today.

Computer Algebra Systems (CAS), such as Mathematica or Maple besides being easy to program basic formulas, provide a set of ready statistical tools ranging from basic descriptive statistics to fitting the models, cluster analysis, hypothesis testing and time-series analysis.

## Statistical Quantities »

Mean • Variance • StandardDeviation • Median • Quantile • Covariance • ...

## Data Smoothing »

MovingAverage • MovingMedian • ListCorrelate • ...

## Statistical Visualization »

Histogram • SmoothHistogram • QuantilePlot • BoxWhiskerChart • ...

## Hypothesis Tests »

DistributionFitTest • LocationTest • VarianceTest • ...

## Cluster Analysis »

Dendrogram • FindClusters • ClusteringTree • ...

## Probability & Statistics »

EstimatedDistribution • EmpiricalDistribution • Probability • ...

## Statistical Model Analysis »

LinearModelFit • NonlinearModelFit • GeneralizedLinearModelFit • ...

## Random Sampling »

RandomReal • RandomInteger • RandomVariate • RandomChoice • ...

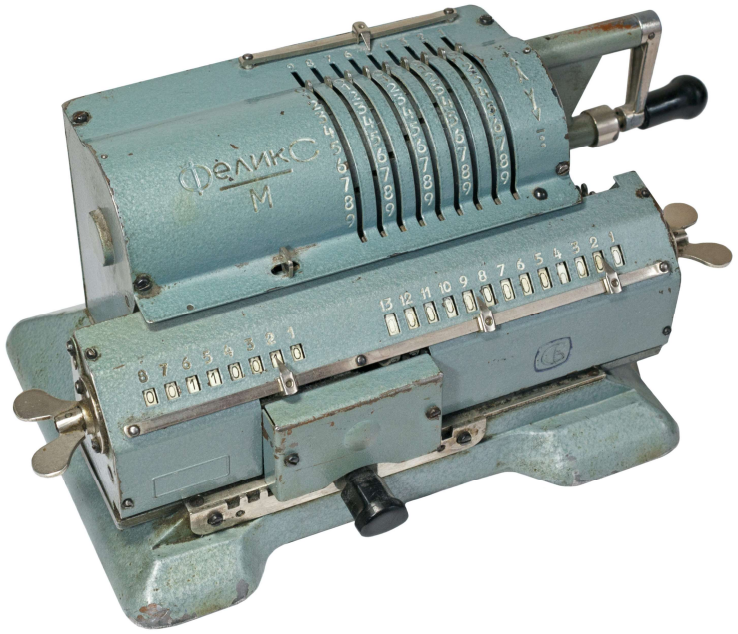
---

## Data Fitting »

FindFit — find a general nonlinear fit

Fit • Interpolation • LeastSquares • FindFormula • ...

Remembering our university studies in the 1970s...



***Example***: Find the variance and standard deviation for the following data:

| No. of order | *f* |
|---|---|
| 10 – 12 | 4 |
| 13 – 15 | 12 |
| 16 – 18 | 20 |
| 19 – 21 | 14 |
| Total | n = 50 |

**Solution:**

| No. of order | *f* | *x* | *fx* | $fx^2$ |
|---|---|---|---|---|
| 10 – 12 | 4 | 11 | 44 | 484 |
| 13 – 15 | 12 | 14 | 168 | 2352 |
| 16 – 18 | 20 | 17 | 340 | 5780 |
| 19 – 21 | 14 | 20 | 280 | 5600 |
| Total | n = 50 | | 832 | 14216 |

$$\bar{x} = \frac{\sum fx}{n} = \frac{832}{50} = 16.64$$

$$s^2 = \frac{\sum fx^2 - \frac{\left(\sum fx\right)^2}{n}}{n-1}$$

$$= \frac{14216 - \frac{(832)^2}{50}}{50-1}$$

$$= 7.5820$$

$$s = \sqrt{s^2} = \sqrt{7.5820} = 2.75$$

## Quartiles

Using the same method of calculation as in the Median, we can get $Q_1$ and $Q_3$ equation as follows:

$$Q_1 = L_{Q_1} + \left( \frac{\frac{n}{4} - F}{f_{Q_1}} \right) i \qquad Q_3 = L_{Q_3} + \left( \frac{\frac{3n}{4} - F}{f_{Q_3}} \right) i$$

Example: Based on the grouped data below, find the Interquartile Range

| Time to travel to work | Frequency |
|---|---|
| 1 – 10 | 8 |
| 11 – 20 | 14 |
| 21 – 30 | 12 |
| 31 – 40 | 9 |
| 41 – 50 | 7 |

***Solution:***

***1st Step:*** *Construct the cumulative frequency distribution*

| ***Time to travel to work*** | ***Frequency*** | ***Cumulative Frequency*** |
|---|---|---|
| ***1 – 10*** | ***8*** | ***8*** |
| ***11 – 20*** | ***14*** | ***22*** |
| ***21 – 30*** | ***12*** | ***34*** |
| ***31 – 40*** | ***9*** | ***43*** |
| *41 – 50* | *7* | *50* |

**2nd Step: Determine the $Q_1$ and $Q_3$**

$$\text{Class } Q_1 = \frac{n}{4} = \frac{50}{4} = 12.5$$

Class $Q_1$ is the 2nd class

Therefore,

$$Q_1 = L_{Q_1} + \left( \frac{\frac{n}{4} - F}{f_{Q_1}} \right) i$$

$$= 10.5 + \left( \frac{12.5 - 8}{14} \right) 10$$

$$= 13.7143$$

$$\text{Class } Q_3 = \frac{3n}{4} = \frac{3(50)}{4} = 37.5$$

Class $Q_3$ is the 4th class
Therefore,

$$Q_3 = L_{Q_3} + \left( \frac{\frac{n}{4} - F}{f_{Q_3}} \right) i$$

$$= 30.5 + \left( \frac{37.5 - 34}{9} \right) 10$$

$$= 34.3889$$

## Interquartile Range

$$IQR = Q_3 - Q_1$$

$IQR = Q_3 - Q_1$

calculate the IQ

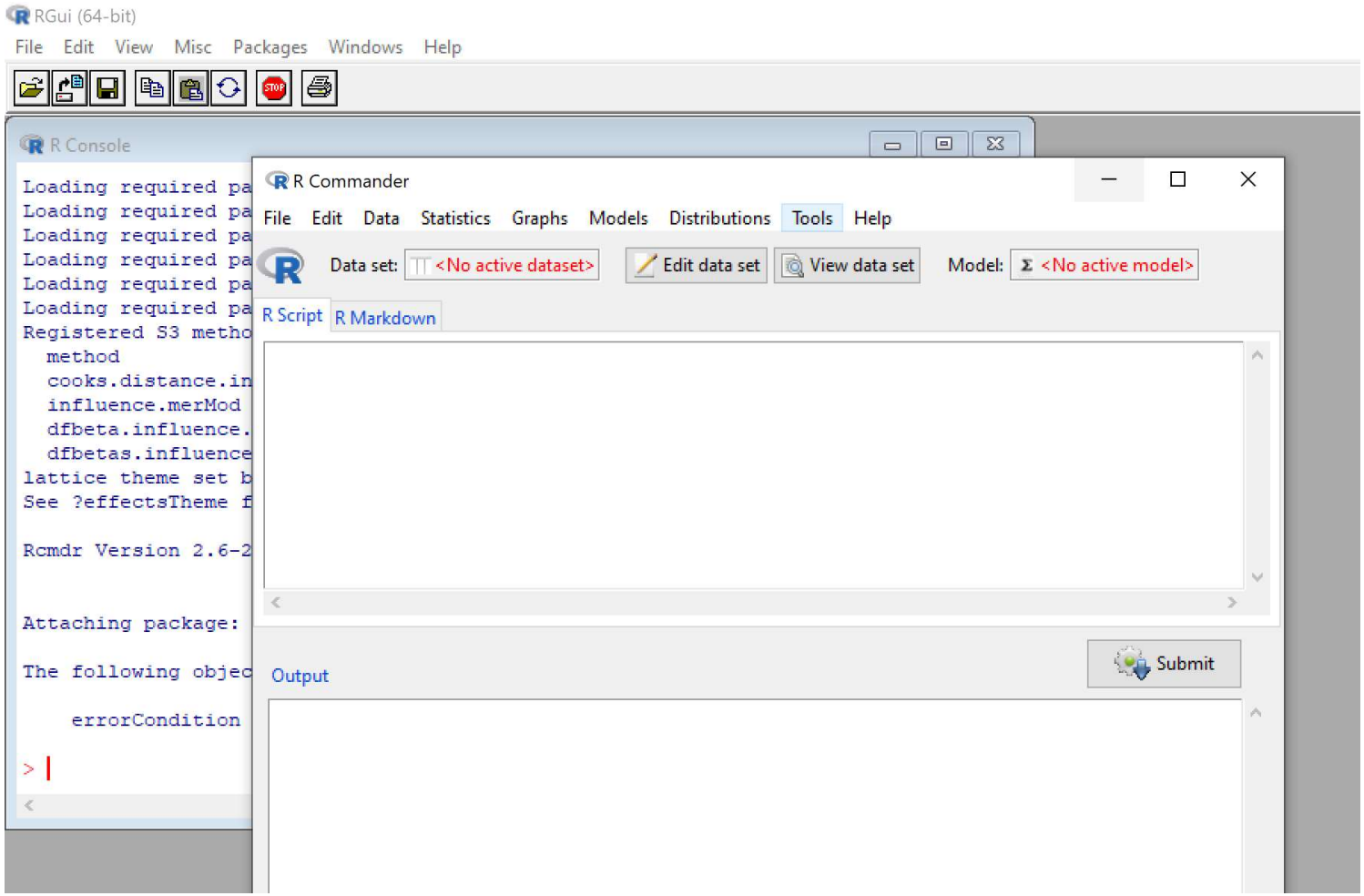
$IQR = Q_3 - Q_1 = 34.3889 - 13.7143 = 20.6746$

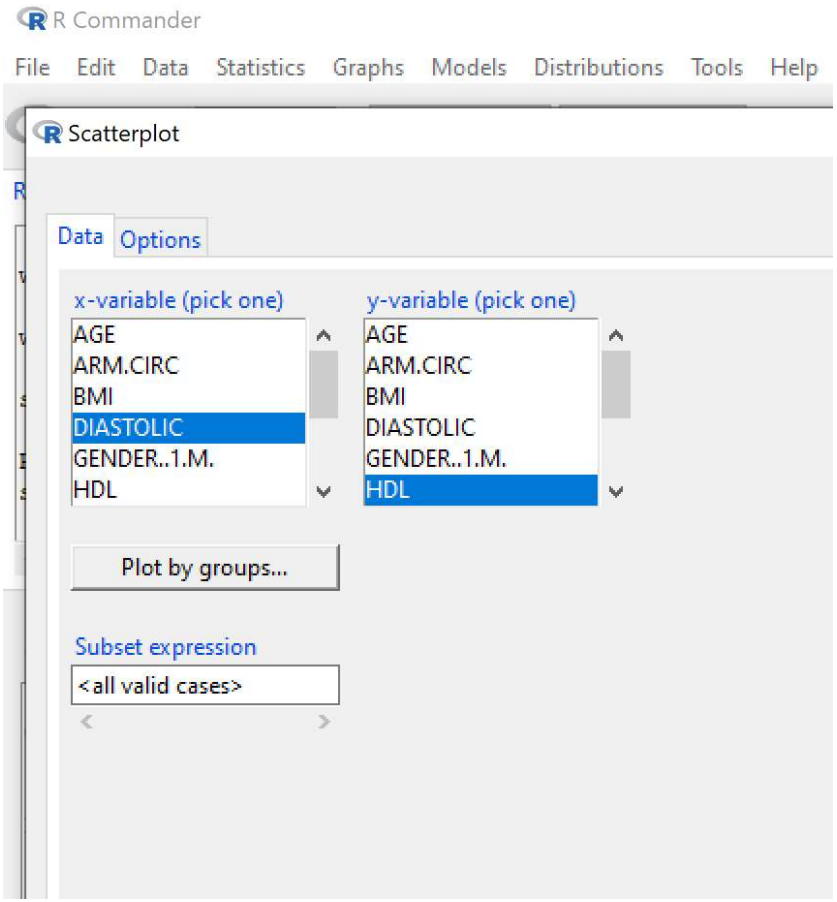
# Using technology

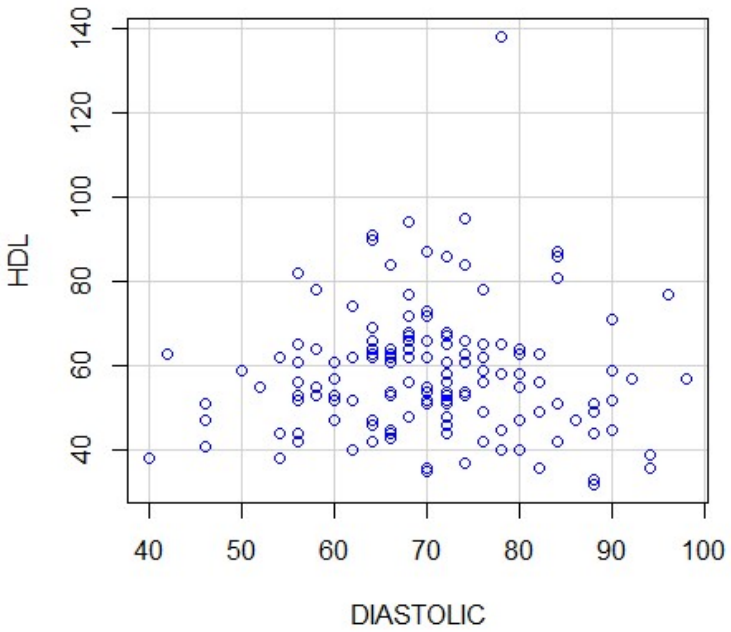
**We encourage students to use *RCommander.***

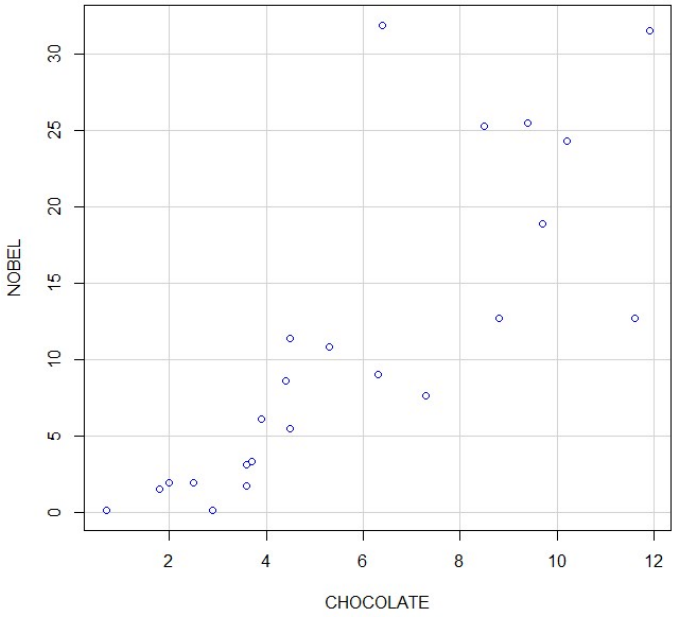
RCommander can be installed on Windows as well as on MAC OS, it is free, fairly easy to use, and contains many tools and methods for statistical analysis. It can also be used as an intermediate step towards using R.

# Using RCommander




R Commander
File Edit Data Statistics Graphs Models Distributions Tools Help
Scatterplot
Data Options
x-variable (pick one)
AGE
ARM.CIRC
BMI
DIASTOLIC
GENDER..1.M.
HDL
y-variable (pick one)
AGE
ARM.CIRC
BMI
DIASTOLIC
GENDER..1.M.
HDL
Plot by groups...
Subset expression
<all valid cases>


HDL
40
60
80
100
120
140
40
50
60
70
80
90
100
DIASTOLIC




Correlation Test

Variables (pick two)

CHOCOLATE
INTERNET
NOBEL
POPULATION

Type of Correlation

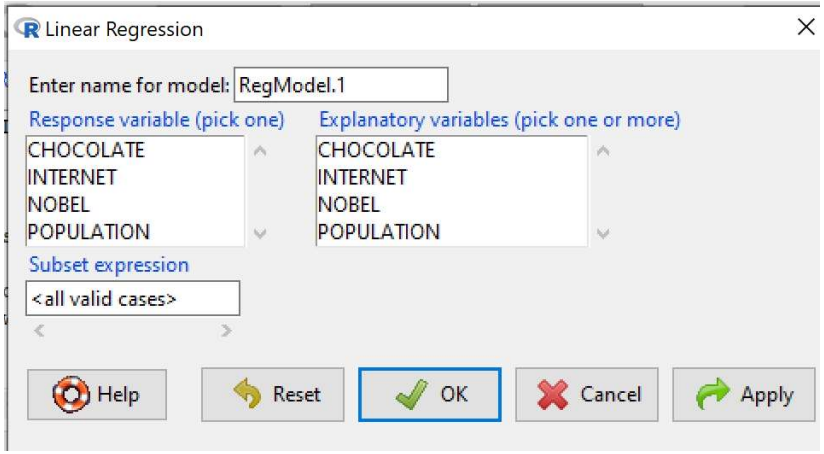
Pearson product-moment

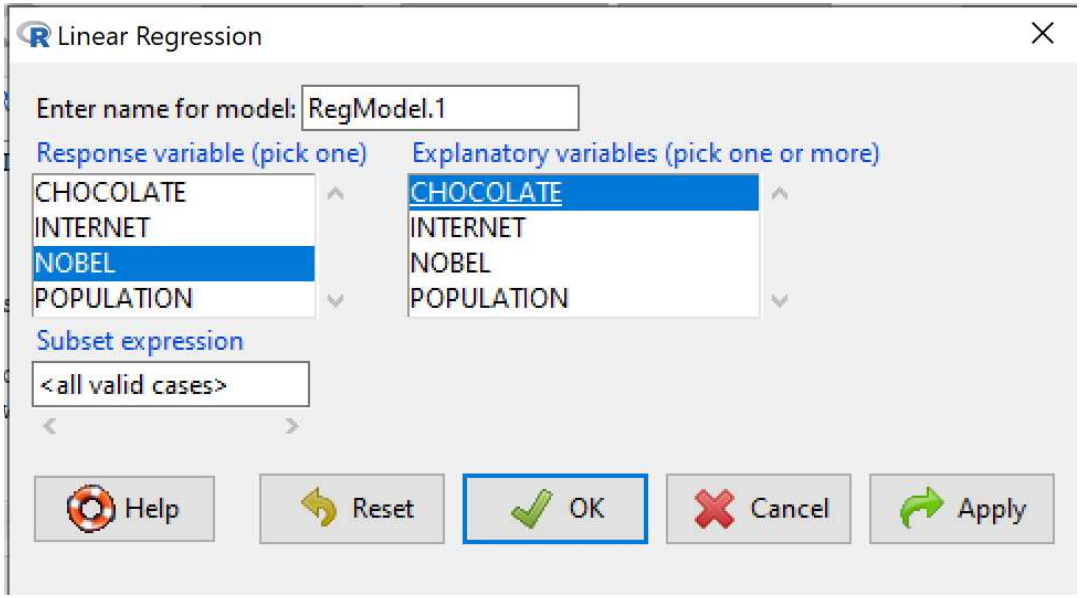
Alternative Hypothesis

Two-sided

```
        Pearson's product-moment correlation

data:  CHOCOLATE and NOBEL
t = 6.123, df = 21, p-value = 0.000004477
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5797205 0.9118788
sample estimates:
      cor
0.8006078
```

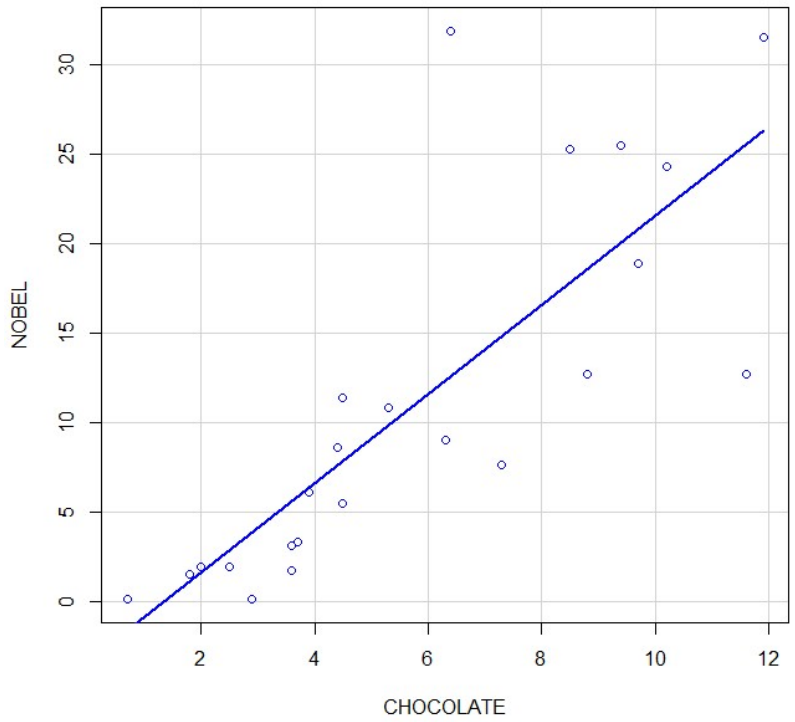
```
> cor(Dataset[,c("CHOCOLATE","NOBEL")], use="complete")
           CHOCOLATE     NOBEL
CHOCOLATE 1.0000000 0.8006078
NOBEL     0.8006078 1.0000000
```

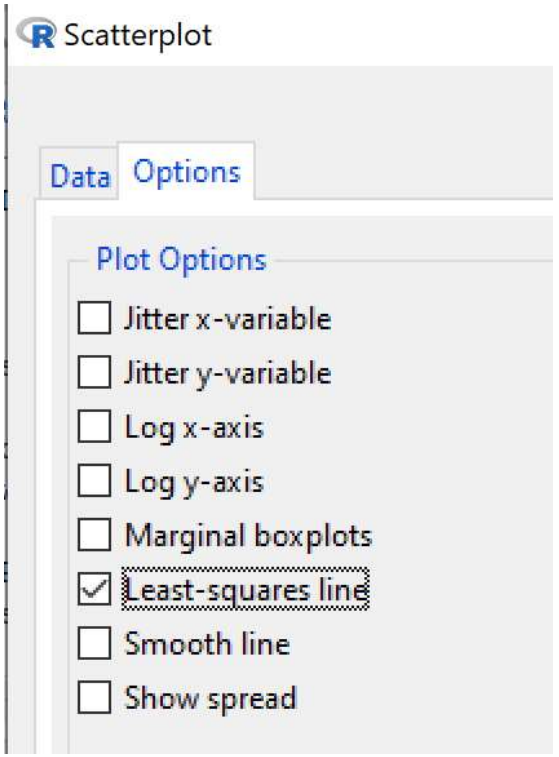

Linear Regression
Enter name for model: RegModel.1
Response variable (pick one)
Explanatory variables (pick one or more)
CHOCOLATE
INTERNET
NOBEL
POPULATION
CHOCOLATE
INTERNET
NOBEL
POPULATION
Subset expression
<all valid cases>
Help
Reset
OK
Cancel
Apply


Linear Regression
Enter name for model: RegModel.1
Response variable (pick one)
Explanatory variables (pick one or more)
CHOCOLATE
INTERNET
NOBEL
POPULATION
CHOCOLATE
INTERNET
NOBEL
POPULATION
Subset expression
<all valid cases>
Help
Reset
OK
Cancel
Apply

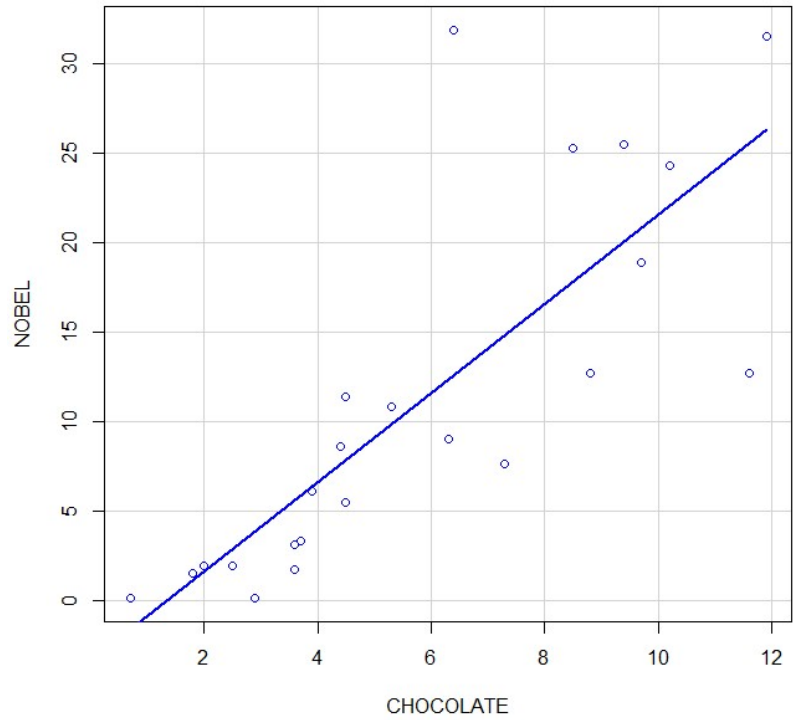
```
> summary(RegModel.1)

Call:
lm(formula = NOBEL ~ CHOCOLATE, data = Dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-12.8537  -2.9490  -0.2566   1.9791  19.3106

Coefficients:
            Estimate Std. Error t value   Pr(>|t|)
(Intercept)  -3.3667     2.7002  -1.247      0.226
CHOCOLATE     2.4931     0.4072   6.123 0.00000448 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.263 on 21 degrees of freedom
Multiple R-squared:  0.641, Adjusted R-squared:  0.6239
F-statistic: 37.49 on 1 and 21 DF,  p-value: 0.00004477
```

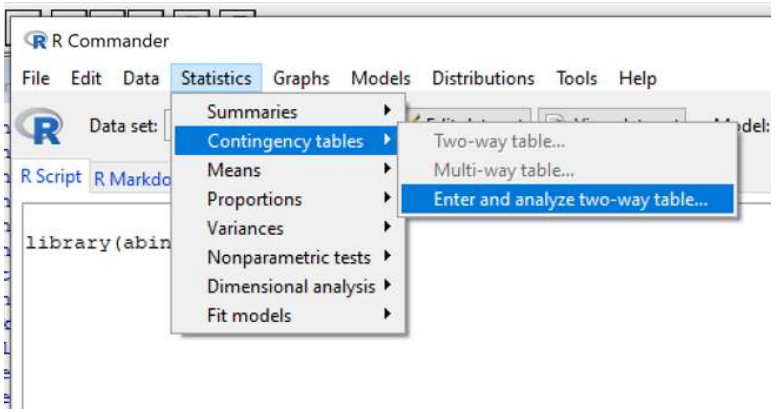
```
Coefficients:
             Estimate Std. Error t value   Pr(>|t|)
(Intercept)  -3.3667     2.7002  -1.247      0.226
CHOCOLATE     2.4931     0.4072   6.123 0.00000448 ***
```

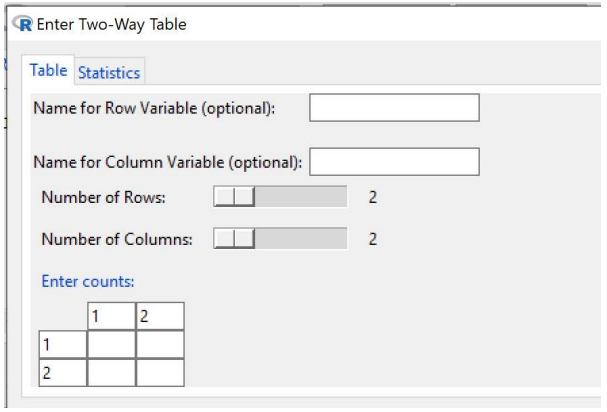


```
Coefficients:
             Estimate Std. Error t value    Pr(>|t|)
(Intercept)  -3.3667     2.7002  -1.247       0.226
CHOCOLATE     2.4931     0.4072   6.123 0.00000448 ***
```




R Commander
File Edit Data Statistics Graphs Models Distributions Tools Help
Data set:
R Script R Markdo
library(abin
Summaries
Contingency tables
Means
Proportions
Variances
Nonparametric tests
Dimensional analysis
Fit models
Two-way table...
Multi-way table...
Enter and analyze two-way table...


Enter Two-Way Table
Table Statistics
Name for Row Variable (optional):
Name for Column Variable (optional):
Number of Rows: 2
Number of Columns: 2
Enter counts:
1 2
1
2

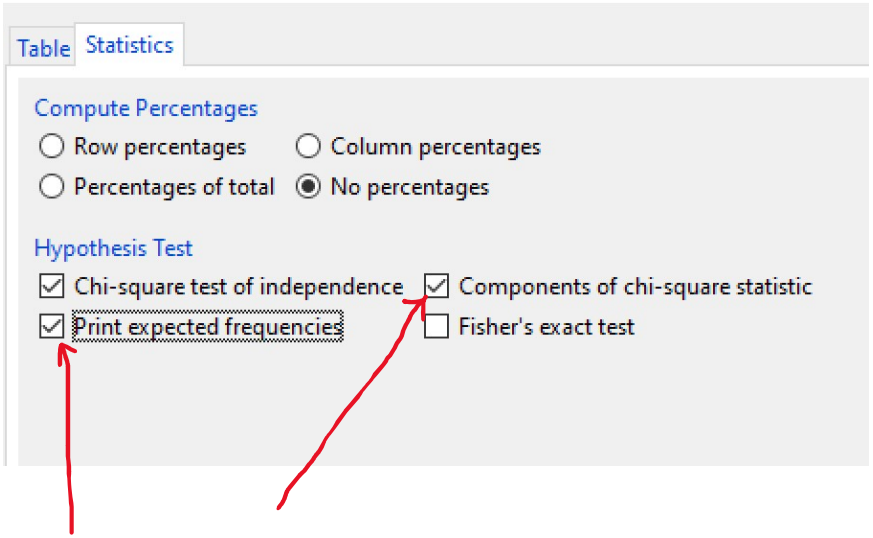


Table Statistics
Compute Percentages
Row percentages
Column percentages
Percentages of total
No percentages
Hypothesis Test
Chi-square test of independence
Components of chi-square statistic
Print expected frequencies
Fisher's exact test

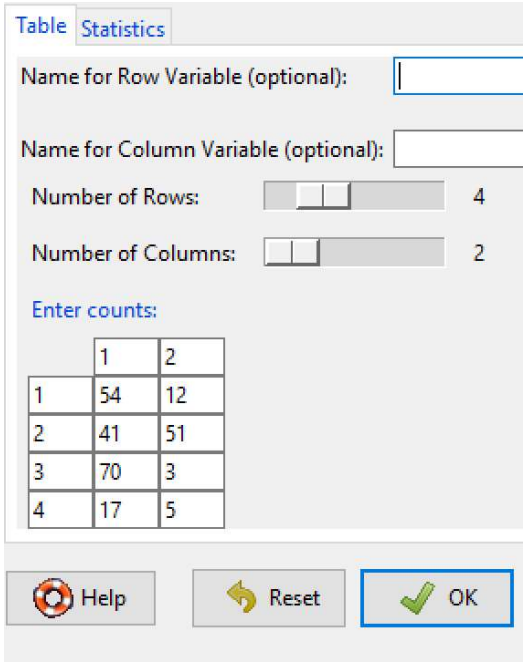


Table Statistics
Name for Row Variable (optional):
Name for Column Variable (optional):
Number of Rows: 4
Number of Columns: 2
Enter counts:
1 2
1 54 12
2 41 51
3 70 3
4 17 5
Help
Reset
OK

# Alternatives:

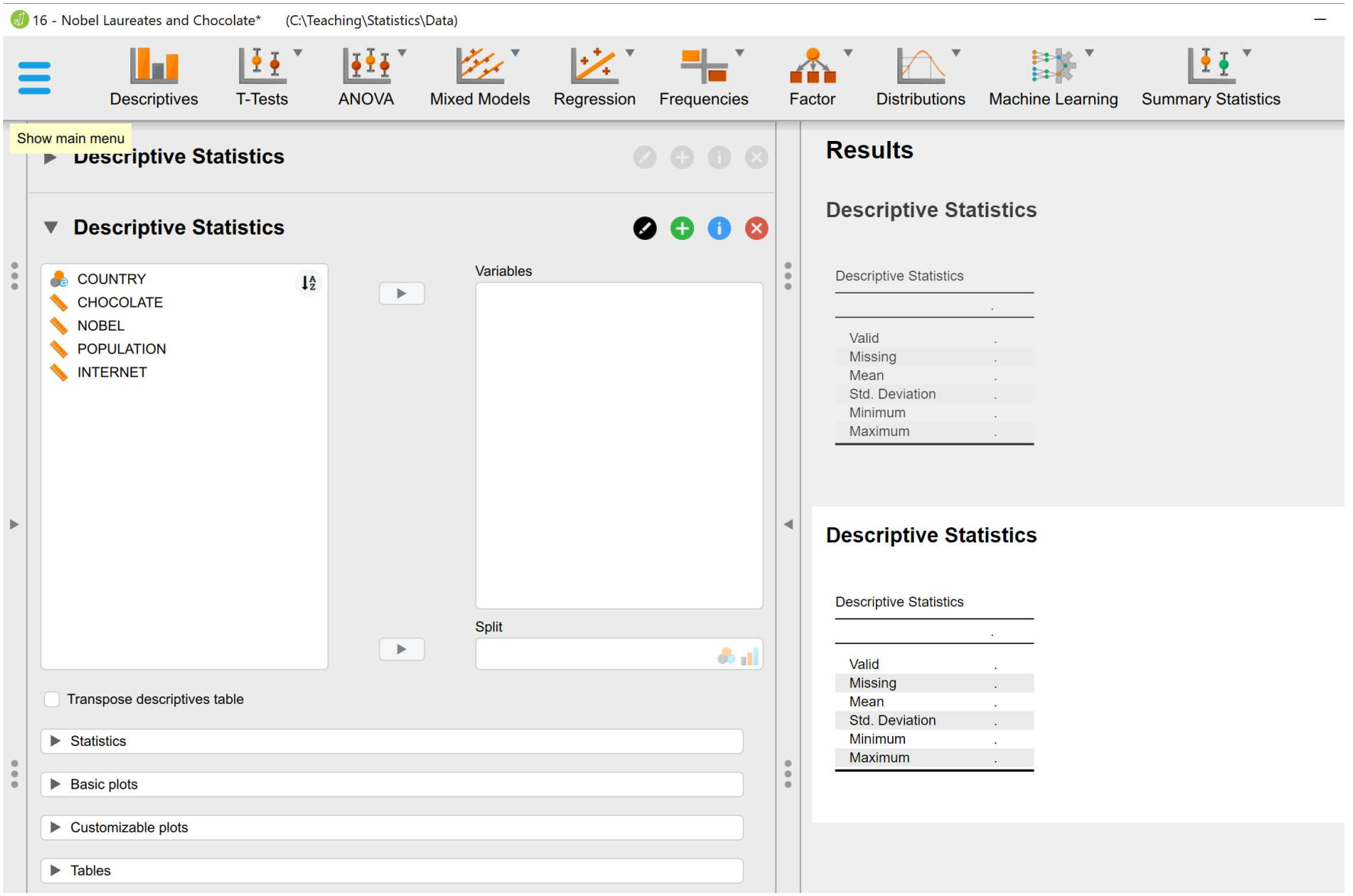
## *JASP*

JASP is open-source, free, ***very easy to install*** and has ***very nice GUI***.

# JASP

https://jasp-stats.org/


Note: you'll need to export data as a .csv – file.


16 - Nobel Laureates and Chocolate* (C:\Teaching\Statistics\Data)
Descriptives
T-Tests
ANOVA
Mixed Models
Regression
Frequencies
Factor
Distributions
Machine Learning
Summary Statistics
Show main menu
Descriptive Statistics
Descriptive Statistics
COUNTRY
CHOCOLATE
NOBEL
POPULATION
INTERNET
Variables
Split
Transpose descriptives table
Statistics
Basic plots
Customizable plots
Tables
Results
Descriptive Statistics
Descriptive Statistics
Valid .
Missing .
Mean .
Std. Deviation .
Minimum .
Maximum .
Descriptive Statistics
Descriptive Statistics
Valid .
Missing .
Mean .
Std. Deviation .
Minimum .
Maximum .

16 - Nobel Laureates and Chocolate* (C:\Teaching\Statistics\Data)

Descriptives | T-Tests | ANOVA | Mixed Models | Regression | Frequencies | Factor | Distributions | Machine Learning | Summary Statistics

## Linear Regression

COUNTRY
POPULATION
INTERNET

Dependent Variable: NOBEL

Method: Enter

Covariates: CHOCOLATE

Factors:

WLS Weights (optional):

▶ Model

▶ Statistics

▶ Method Specification

▼ Plots

Residuals Plots
- ☑ Residuals vs. dependent
- ☐ Residuals vs. covariates
- ☐ Residuals vs. predicted

Other Plots
- ☐ Marginal effects plots
- ☐ Confidence intervals 95.0 %
- ☐ Prediction intervals 95.0 %

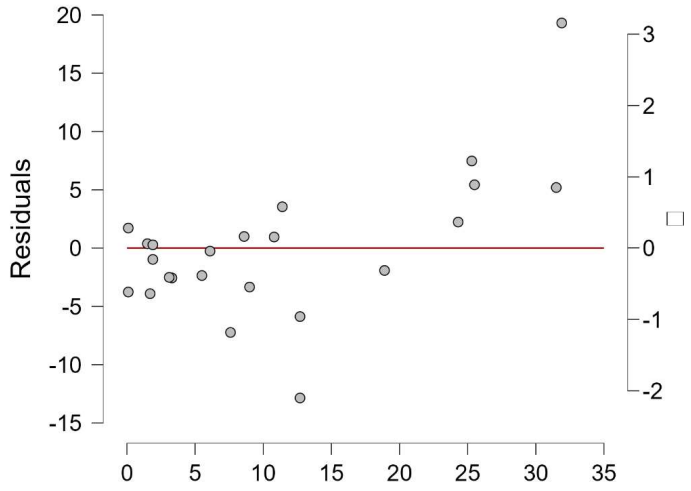
| | | | | |
|---|---|---|---|---|
| $H_1$ | Regression | 1476.449 | 1 | 1476.449 |
| | Residual | 823.640 | 21 | 39.221 |
| | Total | 2294.090 | 22 | |

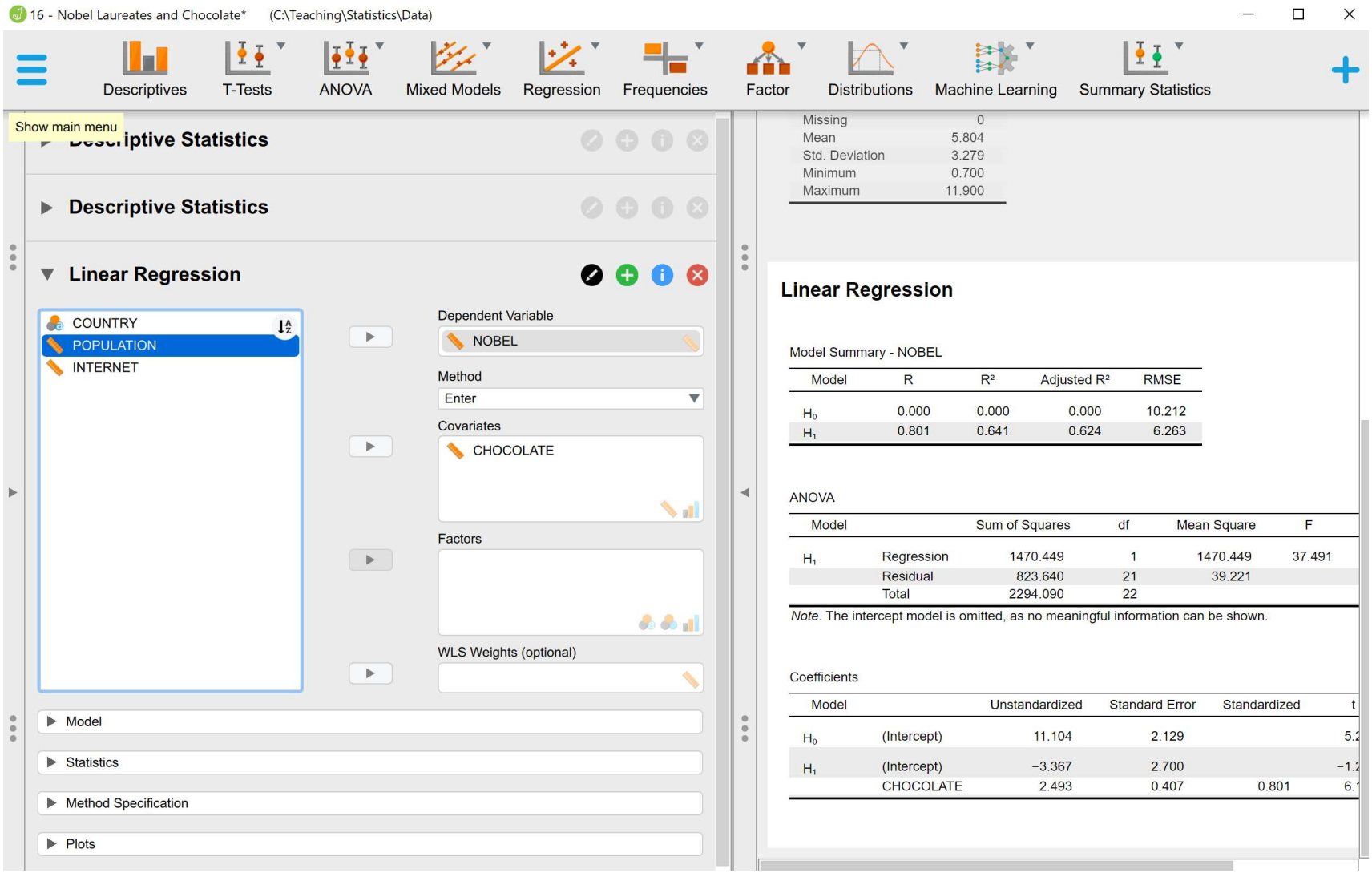
*Note.* The intercept model is omitted, as no meaningful information can be shown.

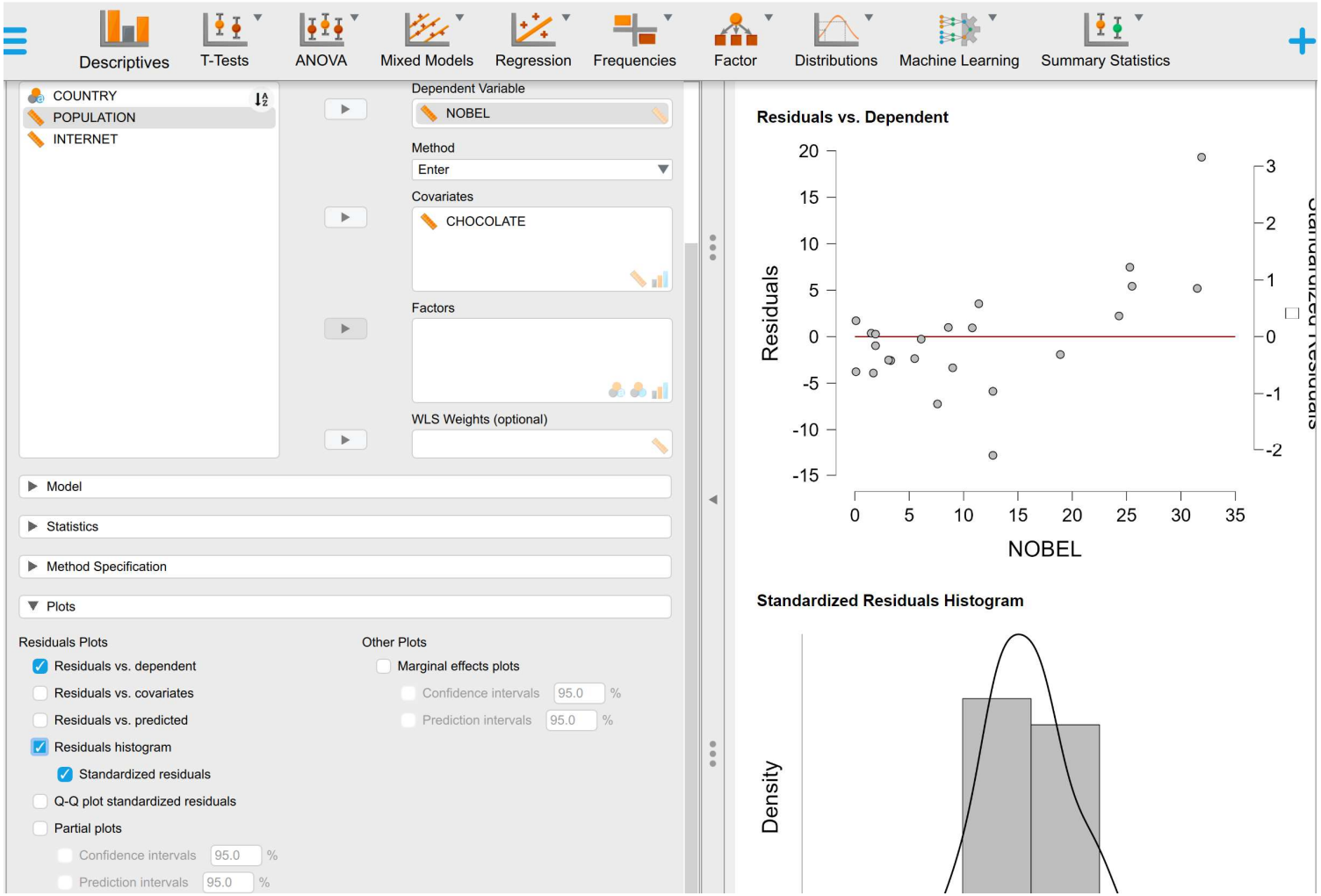
Coefficients

| Model | | Unstandardized | Standard Error | Standardized |
|---|---|---|---|---|
| $H_0$ | (Intercept) | 11.104 | 2.129 | |
| $H_1$ | (Intercept) | −3.367 | 2.700 | |
| | CHOCOLATE | 2.493 | 0.407 | 0.801 |

**Residuals vs. Dependent**




16 - Nobel Laureates and Chocolate* (C:\Teaching\Statistics\Data)
Descriptives
T-Tests
ANOVA
Mixed Models
Regression
Frequencies
Factor
Distributions
Machine Learning
Summary Statistics
Show main menu
Descriptive Statistics
Descriptive Statistics
Linear Regression
COUNTRY
POPULATION
INTERNET
Dependent Variable
NOBEL
Method
Enter
Covariates
CHOCOLATE
Factors
WLS Weights (optional)
Model
Statistics
Method Specification
Plots
Missing 0
Mean 5.804
Std. Deviation 3.279
Minimum 0.700
Maximum 11.900
Linear Regression
Model Summary - NOBEL
Model R R² Adjusted R² RMSE
H0 0.000 0.000 0.000 10.212
H1 0.801 0.641 0.624 6.263
ANOVA
Model Sum of Squares df Mean Square F
H1 Regression 1470.449 1 1470.449 37.491
Residual 823.640 21 39.221
Total 2294.090 22
Note. The intercept model is omitted, as no meaningful information can be shown.
Coefficients
Model Unstandardized Standard Error Standardized t
H0 (Intercept) 11.104 2.129 5.2
H1 (Intercept) −3.367 2.700 −1.2
CHOCOLATE 2.493 0.407 0.801 6.1


Descriptives
T-Tests
ANOVA
Mixed Models
Regression
Frequencies
Factor
Distributions
Machine Learning
Summary Statistics
COUNTRY
POPULATION
INTERNET
Dependent Variable
NOBEL
Method
Enter
Covariates
CHOCOLATE
Factors
WLS Weights (optional)
Model
Statistics
Method Specification
Plots
Residuals Plots
Residuals vs. dependent
Residuals vs. covariates
Residuals vs. predicted
Residuals histogram
Standardized residuals
Q-Q plot standardized residuals
Partial plots
Confidence intervals 95.0 %
Prediction intervals 95.0 %
Other Plots
Marginal effects plots
Confidence intervals 95.0 %
Prediction intervals 95.0 %
Residuals vs. Dependent
Residuals
NOBEL
Standardized Residuals Histogram
Density

16 - Nobel Laureates and Chocolate* (C:\Teaching\Statistics\Data)

Descriptives | T-Tests | ANOVA | Mixed Models | Regression | Frequencies | Factor | Distributions | Machine Learning | Summary Statistics

Show main menu

TION
INTERNET

NOBEL
CHOCOLATE

Split

Transpose descriptives table

▶ Statistics

▼ Basic plots

Distribution plots ☑ Correlation plots

Display density
Display rug marks
Bin width type Sturges
Number of bins 30

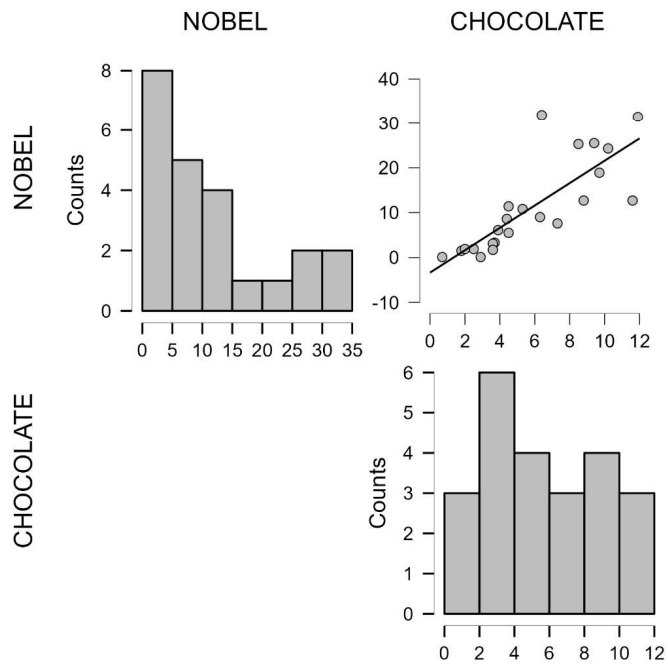
Interval plots
Q-Q plots
Pie charts
Dot plots

Categorical plots
Pareto plots
Pareto rule 95.0 %
Likert plots
Assume all variables share the same levels
Adjustable font size for vertical axis Normal

Descriptive Statistics

| | NOBEL | CHOCOLATE |
|---|---|---|
| Valid | 23 | 23 |
| Missing | 0 | 0 |
| Mean | 11.104 | 5.804 |
| Std. Deviation | 10.212 | 3.279 |
| Minimum | 0.100 | 0.700 |
| Maximum | 31.900 | 11.900 |

**Correlation plot**



16 - Nobel Laureates and Chocolate* (C:\Teaching\Statistics\Data)

Descriptives | T-Tests | ANOVA | Mixed Models | Regression | Frequencies | Factor | Distributions | Machine Learning | Summary Statistics

## Linear Regression

COUNTRY
POPULATION
INTERNET

Dependent Variable: NOBEL

Method: Enter

Covariates: CHOCOLATE

Factors:

WLS Weights (optional):

▶ Model

▶ Statistics

▶ Method Specification

▼ Plots

Residuals Plots
- ☑ Residuals vs. dependent
- ☐ Residuals vs. covariates
- ☐ Residuals vs. predicted

Other Plots
- ☐ Marginal effects plots
- ☐ Confidence intervals 95.0 %
- ☐ Prediction intervals 95.0 %

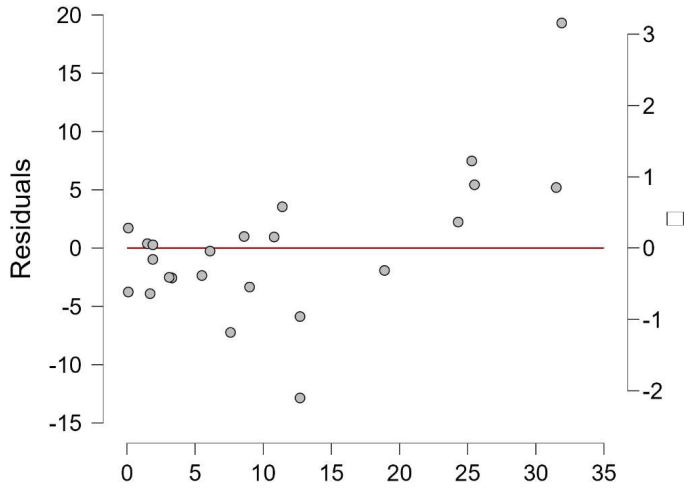
| | | | | |
|---|---|---|---|---|
| $H_1$ | Regression | 1476.449 | 1 | 1476.449 |
| | Residual | 823.640 | 21 | 39.221 |
| | Total | 2294.090 | 22 | |

*Note.* The intercept model is omitted, as no meaningful information can be shown.

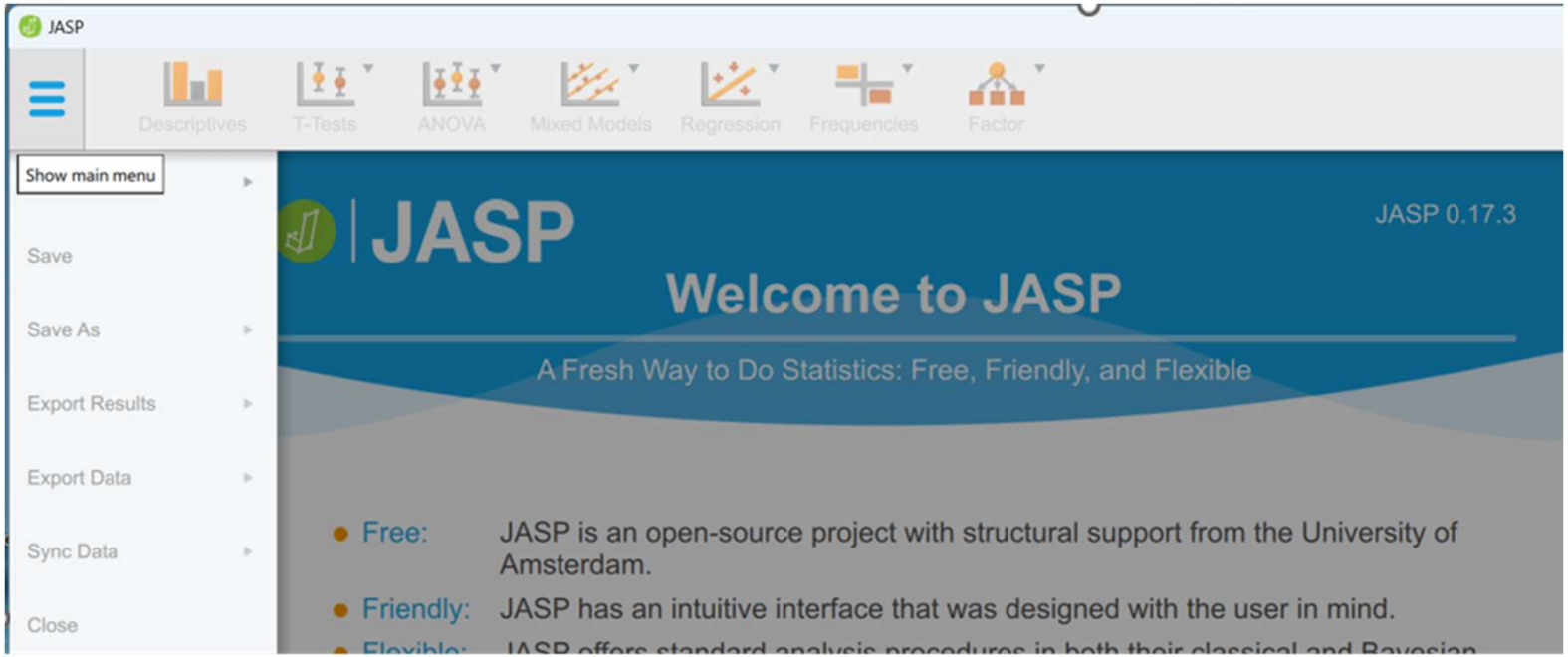
Coefficients

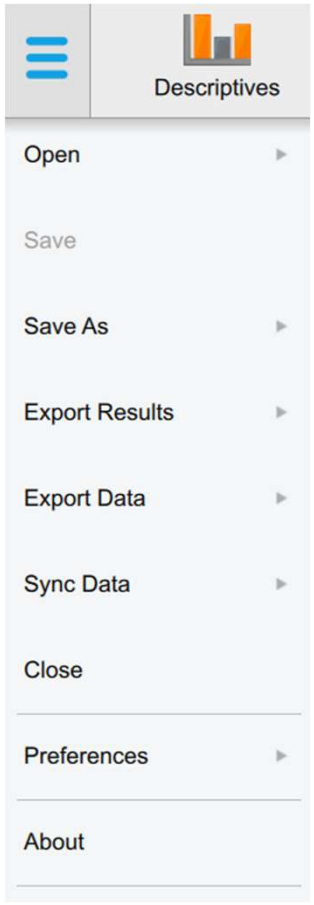
| Model | | Unstandardized | Standard Error | Standardized |
|---|---|---|---|---|
| $H_0$ | (Intercept) | 11.104 | 2.129 | |
| $H_1$ | (Intercept) | −3.367 | 2.700 | |
| | CHOCOLATE | 2.493 | 0.407 | 0.801 |

**Residuals vs. Dependent**



# Problems and “problems”


JASP
Descriptives
T-Tests
ANOVA
Mixed Models
Regression
Frequencies
Factor
Show main menu
Save
Save As
Export Results
Export Data
Sync Data
Close
JASP
JASP 0.17.3
Welcome to JASP
A Fresh Way to Do Statistics: Free, Friendly, and Flexible
Free: JASP is an open-source project with structural support from the University of Amsterdam.
Friendly: JASP has an intuitive interface that was designed with the user in mind.


Descriptives
Open
Save
Save As
Export Results
Export Data
Sync Data
Close
Preferences
About

## One Sample T-Test

One Sample T-Test

| | t | df | p | Mean Difference | 95% CI for Mean Difference | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| M | 8.970 | 183 | < .001 | 5726.897 | 4467.214 | 6986.579 |

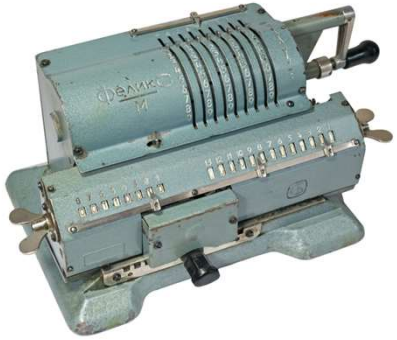
*Note*. For the Student t-test, location difference estimate is given by the sample mean difference *d*.

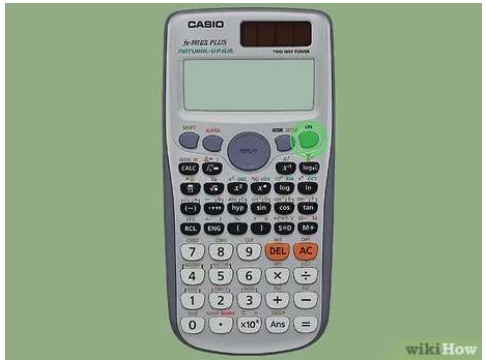
*Note*. For the Student t-test, the alternative hypothesis specifies that the mean is different from 10000.

*Note*. Student's t-test.

```
        One Sample t-test

data:  PULSE
t = -5.9145, df = 152, p-value = 0.0000000106
alternative hypothesis: true mean is less than 75
95 percent confidence interval:
     -Inf 71.09779
sample estimates:
mean of x
  69.5817
```

# Conclusions/Discussion



# ChatGPT???

*Thank you for attention!*