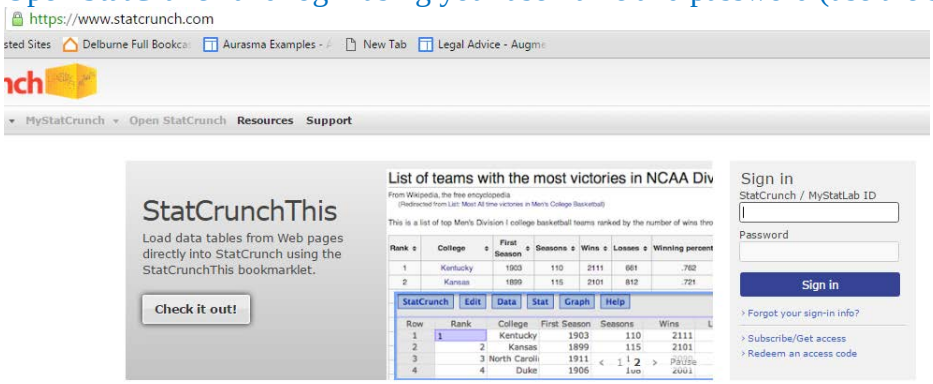
Step-by-step StatCrunch Guide

This also demonstrates using examples how to go through the steps. Some examples include links to data in StatCrunch.

Content

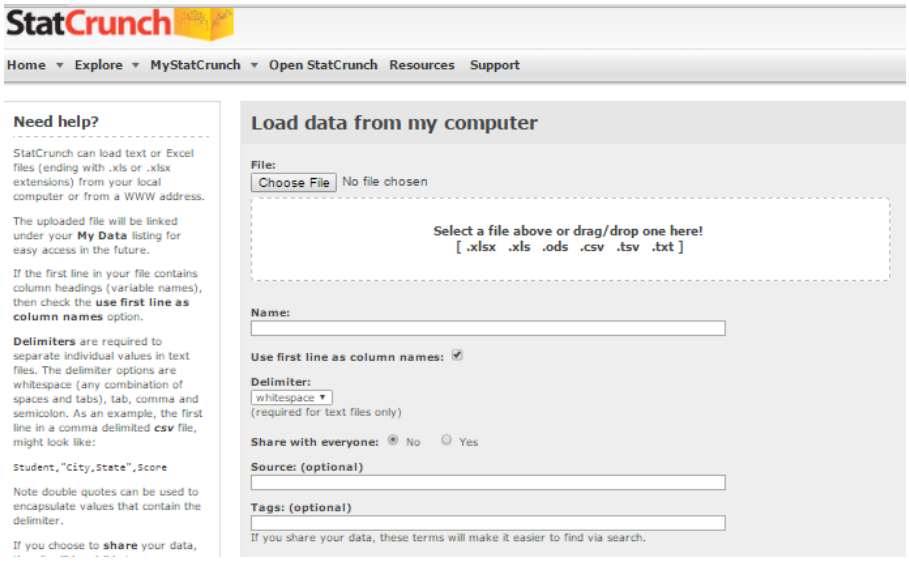
1. Upload an Excel file from your computer in StatCrunch

1. Open StatCrunch and login using your username and password (use the same login from MyStatLab).



2. Click on the tab “MyStatCrunch”
3. Click on the link “Select a file on my computer”

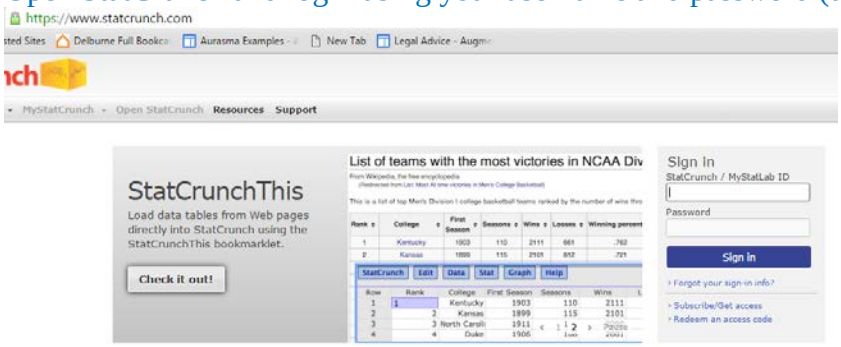
4. Click “Choose File”



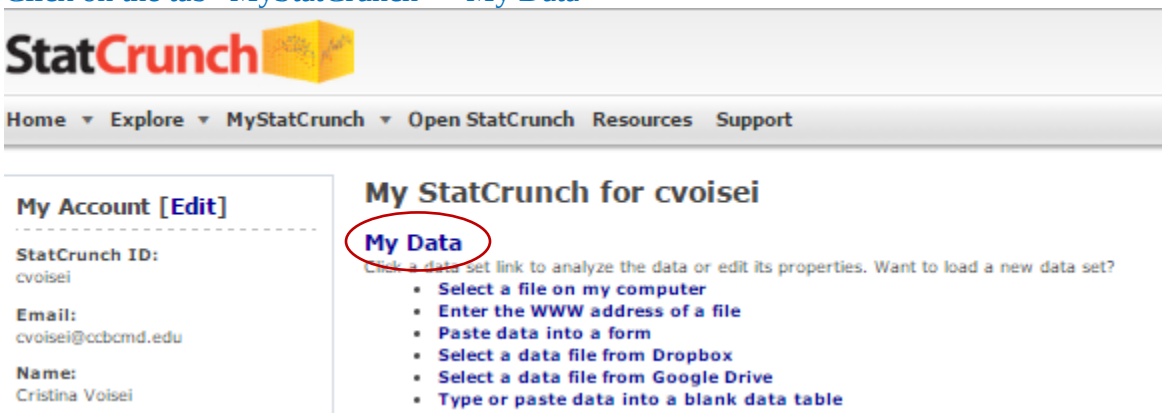
5. Select button “Share with everyone” if you want to share the file with others if not select “No”
6. Scroll to the bottom of the page and click on “Load File”

2. Select saved data in StatCrunch

1. Open StatCrunch and login using your username and password (use the same login from MyStatLab).



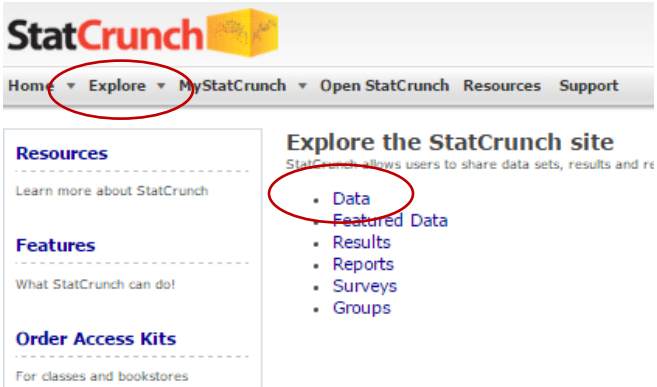
2. Click on the tab "MyStatCrunch" > My Data



3. Click on My Data: You would see your StatCrunch saved data files.

3. Find shared data on StatCrunch site

Click on the tab Explore > Data



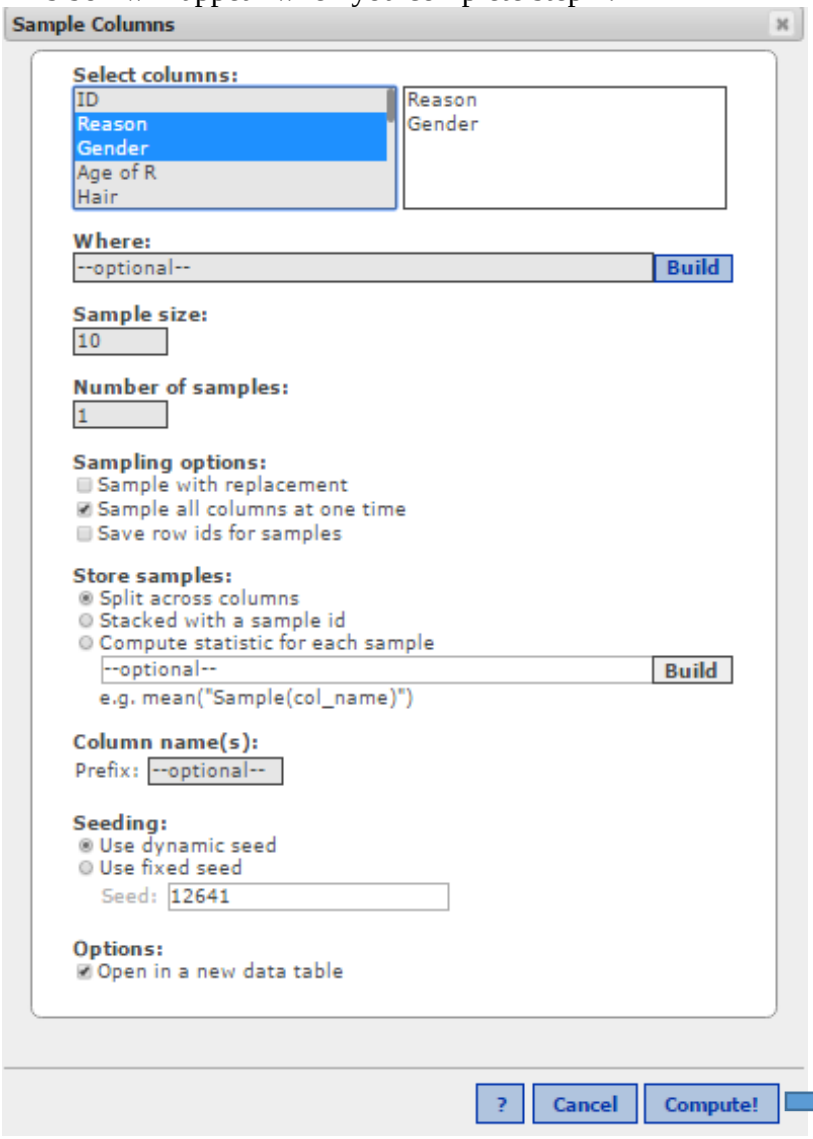
4. Take Random Samples

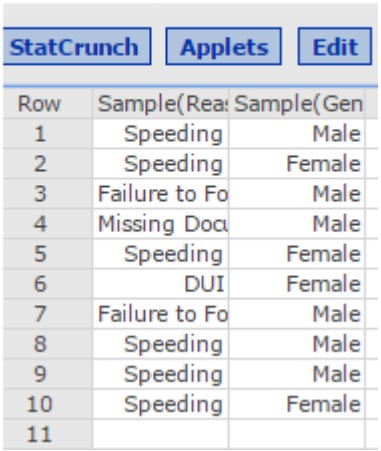
1. Choose: Data → Sample
2. Select the columns you want to sample
3. Type the desired sample size
4. Sampling options: Check 'Sample all columns at one time'
5. Store samples: Split across columns
6. Check Option "Open in a new data table" if you want the samples to open in a new page
7. Press: Compute!

Example: Traffic tickets

To access the data in StatCrunch click here. Must log in to be able to analyze data.

This box will appear when you complete step 1.





StatCrunch | Applets | Edit

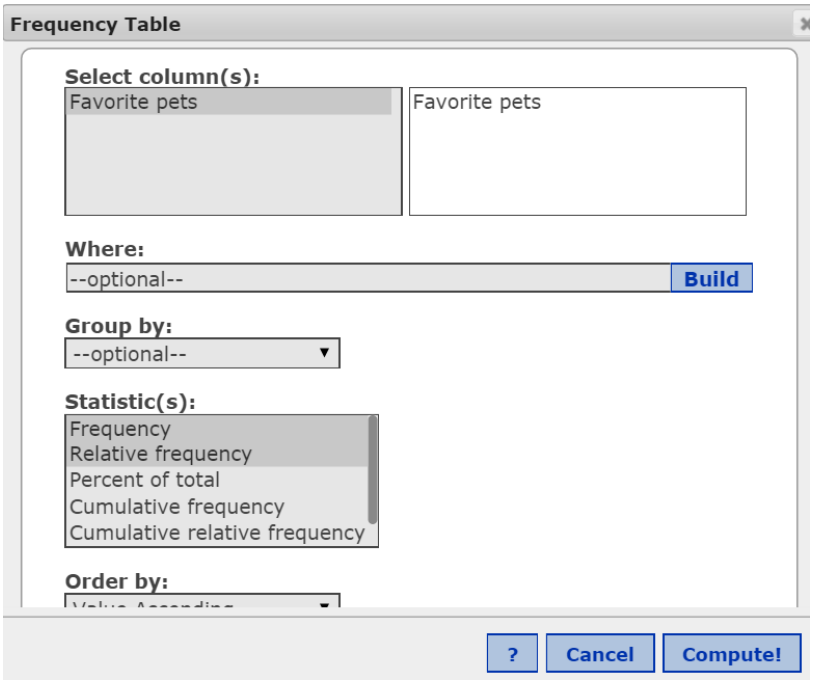
Row	Sample(Rea	Sample(Gen
1	Speeding	Male
2	Speeding	Female
3	Failure to Fo	Male
4	Missing Doc	Male
5	Speeding	Female
6	DUI	Female
7	Failure to Fo	Male
8	Speeding	Male
9	Speeding	Male
10	Speeding	Female
11		

Follow steps 2 through 7 and the samples will be selected.

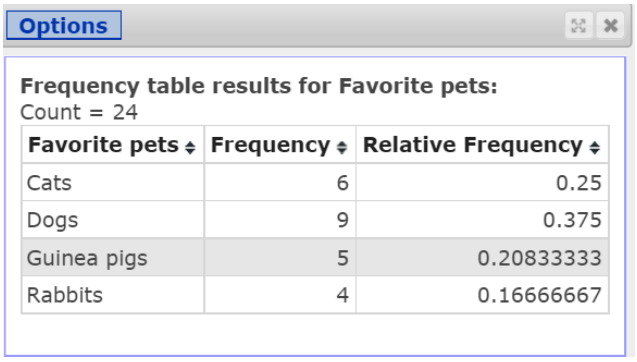
5. Create frequency and relative frequency tables

1. Select Stat > Tables > Frequency.
2. Select the column(s) you want to summarize.
3. Highlight the Statistic(s)
4. Click Compute!
5. You can then choose Options > Copy to copy the output for use elsewhere.

Example: Pet preferences
To access the data in StatCrunch click here. Must log in to be able to analyze data.
This box will appear when you complete step 1.



Follow steps 2 through 4 to get a new window with these numbers calculated.

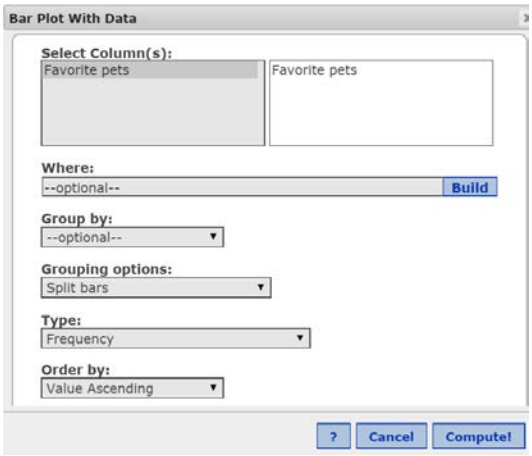


Favorite pets	Frequency	Relative Frequency
Cats	6	0.25
Dogs	9	0.375
Guinea pigs	5	0.20833333
Rabbits	4	0.16666667

6. Create frequency and relative frequency bar graphs

1. Select Graph > Bar Plot, then choose with data or with summary.
2. If you chose *with data*, select the column(s) you wish to use. If you chose *with summary*, set the columns containing the categories and counts.
3. Choose the type (*Frequency* or *Relative Frequency*).
4. Click Compute!

Example: Pet preferences
To access the data in StatCrunch click here. Must log in to be able to analyze data.
This box will appear when you complete step 1and 2.

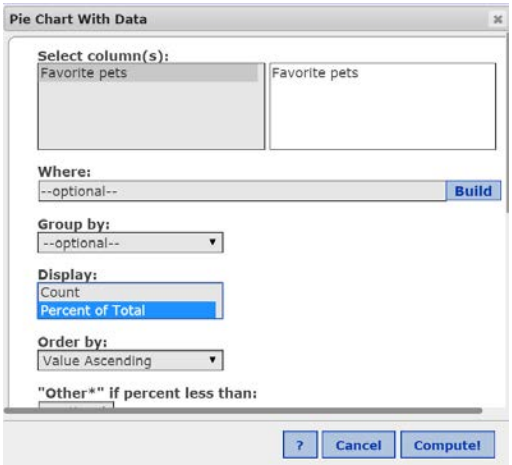


Complete the remaining steps and the graph will appear in a new window.

7. Create pie charts

1. Select Graph > Pie Chart, then choose with data or with summary.
2. If you chose *with data*, select the column(s) you wish to use. If you chose *with summary*, set the columns containing the categories and counts.
3. Enter any modifications (labels, title, color scheme, etc)
4. Click Compute!

Example: Pet preferences
To access the data in StatCrunch click here. This box will appear when you complete step 1and 2.



Select the Display a “Percent of Total” and click Compute and the graph will appear in a new window.

8. Create a boxplot

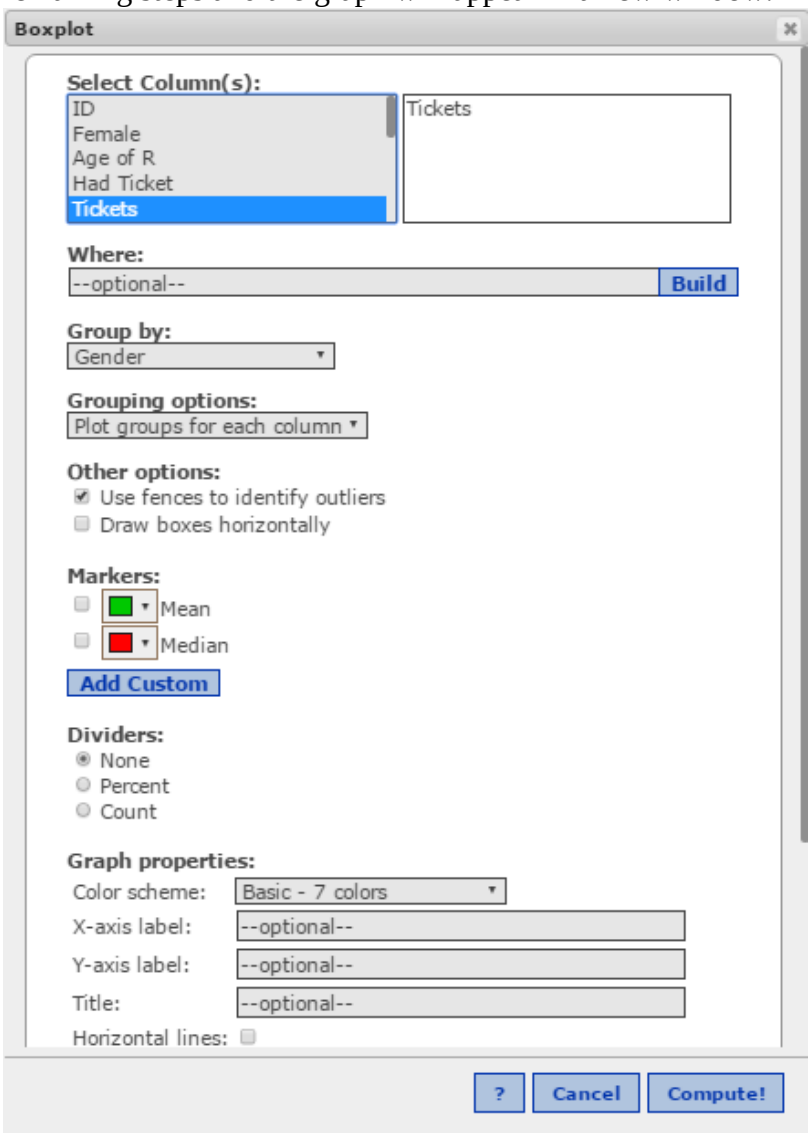
1. Select Graph > Boxplot
2. Select the column variable you'll be using.
3. Choose on “Other options” > use fences to identify outliers
4. Click Compute!

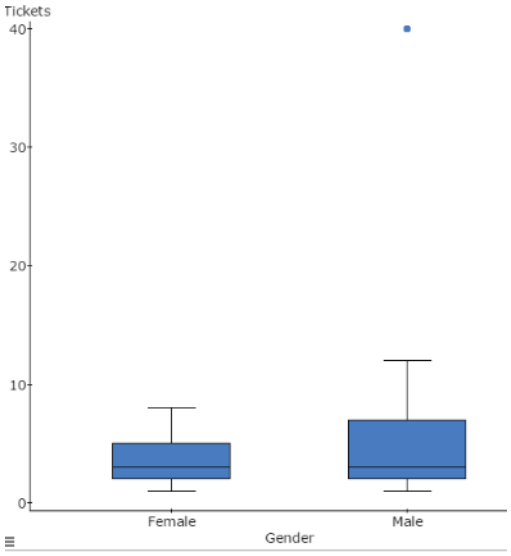
9. Create a side-by-side boxplot

1. Select Graph > Boxplot
2. Select the column variable you'll be using (in here we have Number of Tickets).
3. Under “Group by” select the variable you want to have the data grouped by (in our case is Gender).
4. Choose on “Other options” > use fences to identify outliers
5. Click Compute!

Example: Number of Traffic Tickets by Gender

To access the data in StatCrunch click here. This box will appear when you complete step 1. Complete the remaining steps and the graph will appear in a new window.



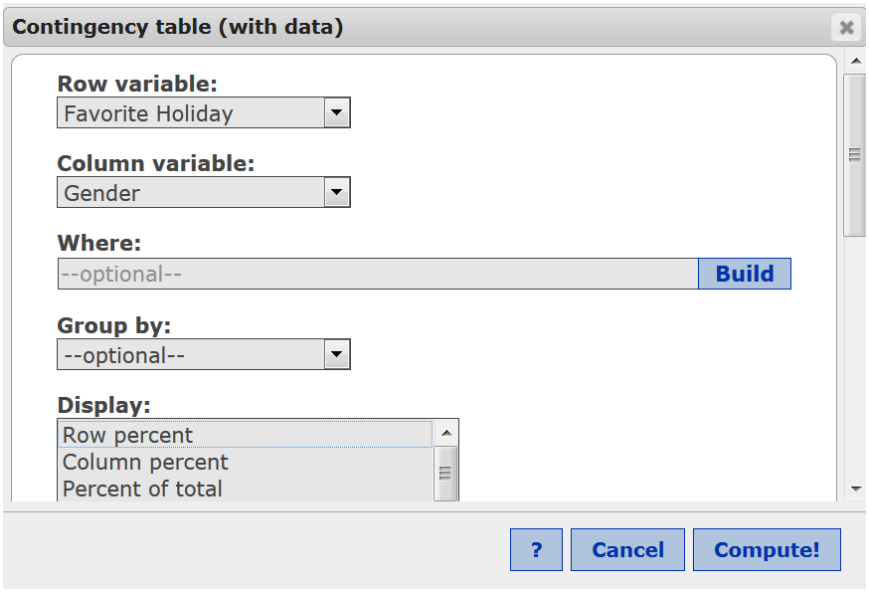


10. Create a contingency table

1. Select Stats > Tables > Contingency > With Data
2. Select the row variable you'll be using.
3. Select the column variable you'll be using.
4. Choose how you want the answer displayed: frequency or percent.
5. Click Compute!

Example: Favorite holidays by gender

To access the data in StatCrunch click here.
This box will appear when you complete step 1.

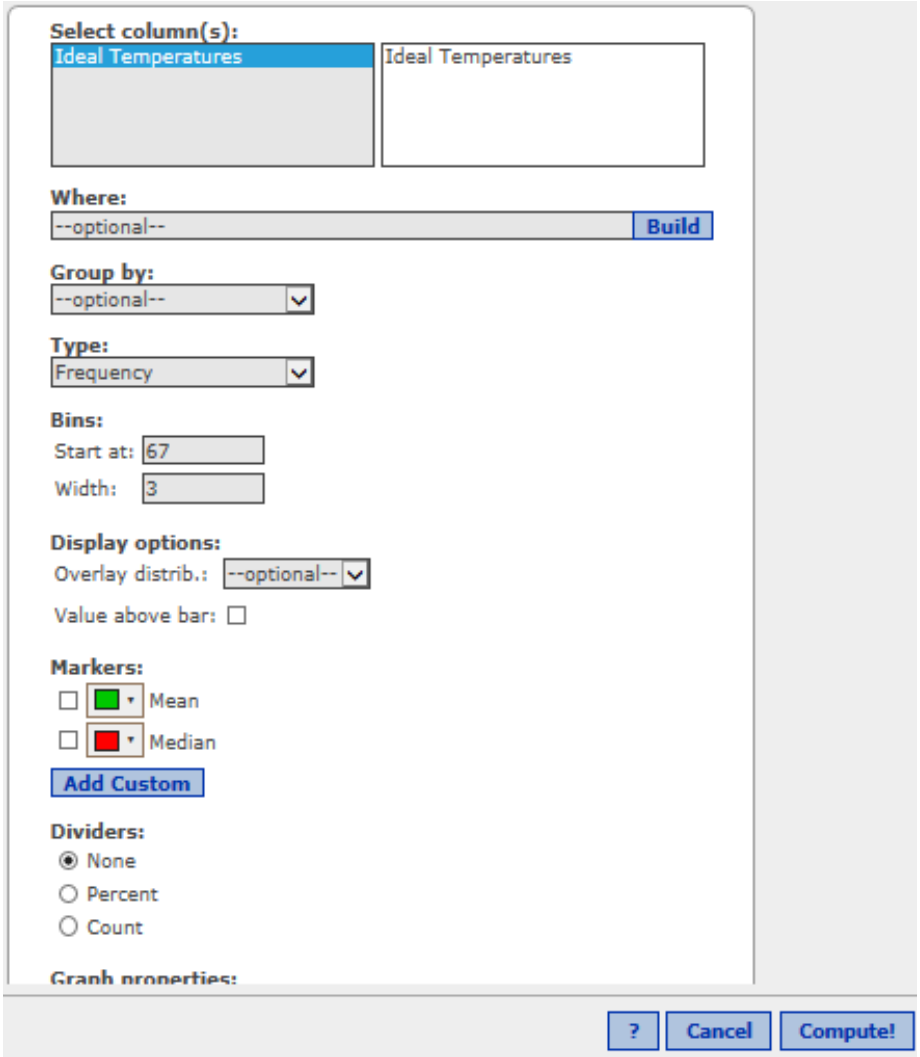


Complete the remaining steps and a new window will show the contingency table results.

11. Create histograms

1. Select Graph > Histogram
2. Select the column(s) you want to summarize
3. Set the *Type*. Set the Bins (*Start at:*) and (*width*)
4. Click Compute!

Example: Ideal Summer Temperatures
To access the data in StatCrunch click here. Must sign in StatCrunch to be able to analyze data.
This box will appear when you complete step 1.



Complete the remaining steps and the graph will appear in a new window.

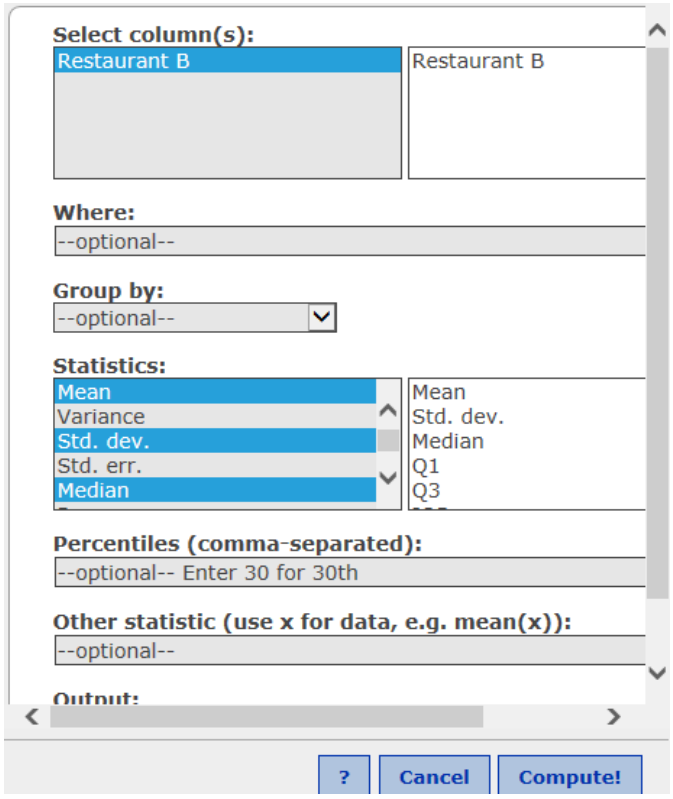
12. Find descriptive statistics measures (mean, median, mode, standard deviation, quartiles, etc.)

1. Select Stat > Summary Stat > Columns.
2. Select the variable you want to summarize.
3. Select any statistics that you want calculated.
4. Click "Compute!"

Example: Restaurant waiting times

To access the data in StatCrunch click here. Must sign in StatCrunch to be able to analyze data.

This box will appear when you complete step 1.



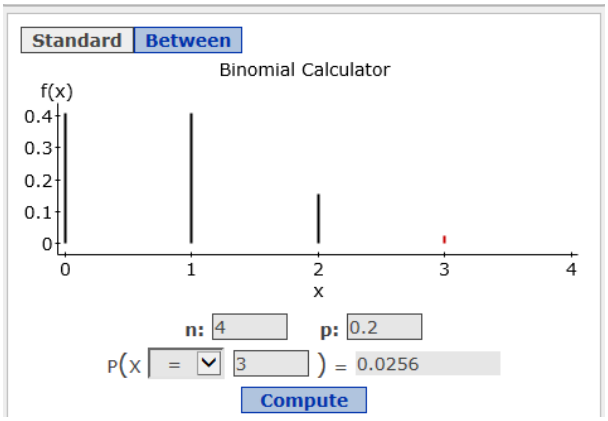
Complete the remaining steps and a new window with the desired descriptive statistics will pop up.

13. Find binomial probabilities

1. Select Stat > Calculators > Binomial
2. Type in the values for n, p, and x
3. Set the box after X to "equals." if you need a probability for one value.
   Set the box after X to “greater than and equal to” if you need to find a probability for “at least”
4. Click compute

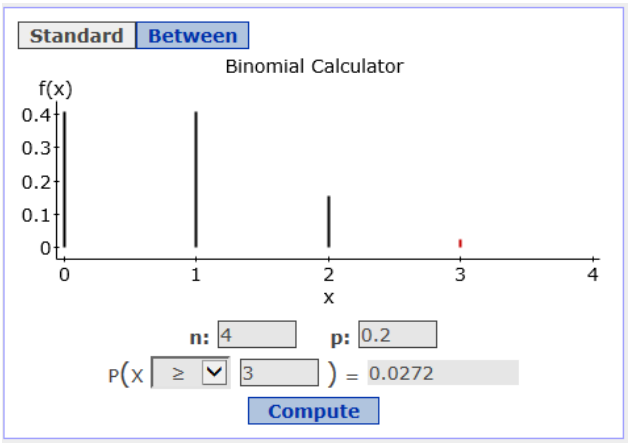
Example: A brand name has a 20% recognition rate. Assume the owner of the brand wants to verify that rate by beginning with a small sample of 4 randomly selected consumers.
a) Find the probability that exactly 3 of the selected consumers recognize the brand name.
b) Find the probability that at least 3 of the selected consumers recognize the brand name.

Solution: a) This box will appear when you complete step 1.



After typing in the values for n, p, and x from our example and press compute we get the answer for part (a) 0.0256.

b) Repeat the steps but make sure to set the box after X to “greater than and equal to” to find the probability for “at least 3”

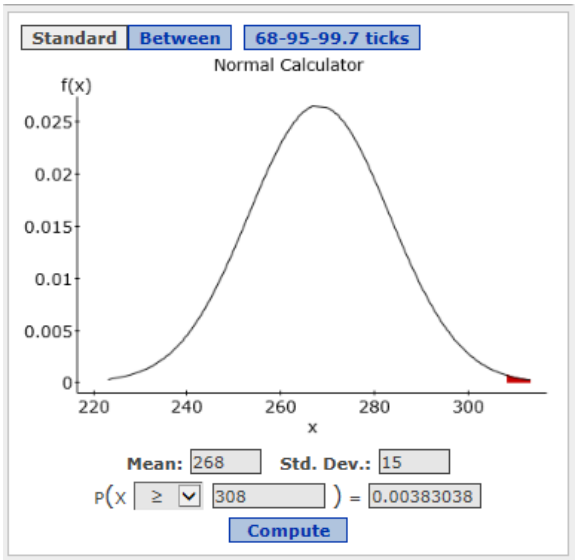


14. Find probabilities for Normal Distribution

1. Select Stat > Calculators > Normal
2. Type in the values for mean and std. dev.
3. Set the box after X to "≥" or "≤"
4. Click compute!

Example: The lengths of pregnancies are normally distributed with a mean of 268 days and a standard deviation of 15 days. One classic use of the normal distribution is inspired by a letter to "Dear Abby" in which a wife claimed to have given birth 308 days after a brief visit from her husband, who was serving in the Navy. Given this information, find the percentage of pregnancies lasting 308 days or longer.

This box will appear when you complete step 1. Type in the values from our example (mean 268, std. dev. 15), select the symbol ≥ (because the problem says 308 or longer) and press compute to get the answer (see below).

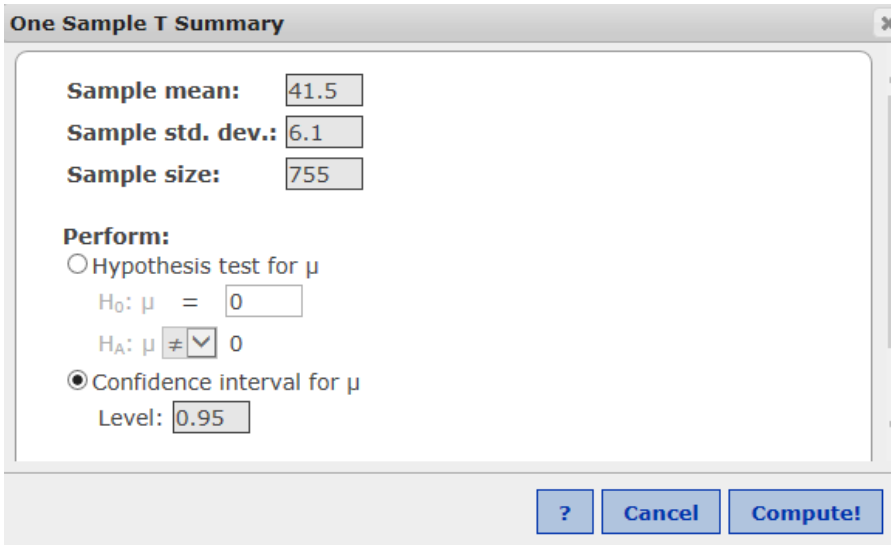


15. Find Confidence Intervals for Means

1. Stat > T Stats > One Sample > With Summary
2. Type in the values for sample mean, sample stand. dev., and sample size
3. Click on confidence interval
4. Enter the confidence level
5. Click Compute!

Example: A random sample of 755 US cell phone users age 18 and older in May 2000 found that the average number of text messages sent or received per day is 41.5 messages with a standard deviation of 6.1. Construct a 95% confidence interval for the population mean number of text messages.

Complete step 1 and the "One sample T Summary" box will pop up.



Type in the values for sample mean, sample stand. dev., and sample size from our example, select the confidence level and press compute to get the lower and upper limit for the confidence interval.

95% confidence interval results:
μ : Mean of population

Mean	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
μ	41.5	0.22200173	754	41.064185	41.935815

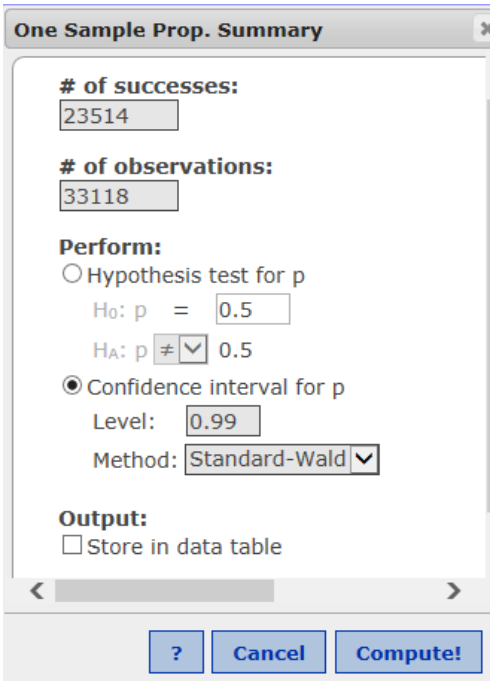
16. Find Confidence Intervals for Proportions

1. Stat > Proportion Stats > One Sample > With Summary
2. Type in the values for number of successes, number of observations (this is the sample size)
3. Check on confidence interval
4. Enter the confidence level
5. Click Compute!

Example: An online site presented this question "Would the recent norovirus outbreak deter you from taking a cruise"? Among the 33118 people who responded, 71% said "yes." Construct a 99% confidence interval estimate for the proportion of the population of all people who would respond "yes" to that question.

Complete step 1 and the "One sample Prop. Summary" box will pop up.Type in the values for values for number of successes (the number of people who said yes), number of observations (this is the sample size) from our example, select the confidence level.

observations = 33118 # successes = 0.71*33118 = 23514



Press compute to get the lower and upper limit for the confidence interval.

99% confidence interval results:
p : Proportion of successes
Method: Standard-Wald

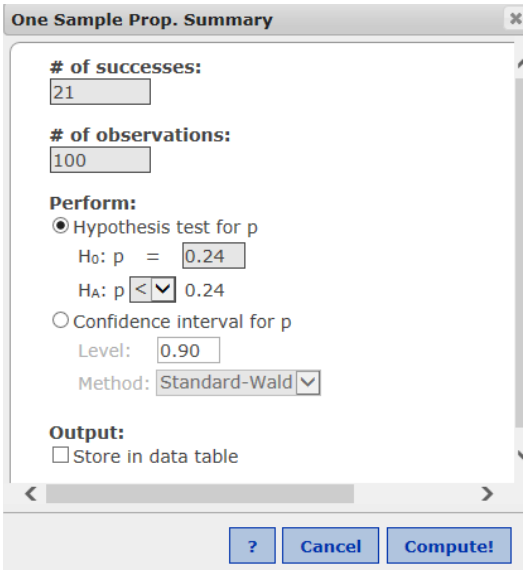
Proportion	Count	Total	Sample Prop.	Std. Err.	L. Limit	U. Limit
p	23514	33118	0.71000664	0.002493407	0.70358405	0.71642923

17. Hypothesis Testing (1 proportion)

1. Stat > Proportion Stats > One Sample > With Summary
2. Type in the values for number of successes, number of observations (this is the sample size)
3. Check on Perform hypothesis test for p
4. Enter the value for the null hypothesis and select the correct symbol in the alternative hypothesis
5. Enter the confidence level if different than 95%
6. Click Compute!

Example: According to the General Household Survey of 2005, 24% of individuals aged over 16 years smoked cigarettes in the United Kingdom. Among a random sample taken in 2008 of 100 individuals aged 16 and over, 21 smoked cigarettes. If appropriate, test using the p-value method at level of significance $\alpha = 0.10$ whether the population proportion of smokers in the United Kingdom has decreased since 2005.

Complete step 1 and the "One sample Prop. Summary" box will pop up.



- Type in the values for values for number of successes (the number of people who smoke), number of observations (this is the sample size) from our example
- Enter the value for the null hypothesis 0.24 and select the symbol "<" in the alternative hypothesis, enter the confidence level 0.90 because alpha is 0.10.
- Make sure the "Perform Hypothesis test for p is checked" and click "Compute".

Hypothesis test results:
p : Proportion of successes
H_0 : p = 0.24
H_A : p < 0.24

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	21	100	0.21	0.042708313	-0.70243936	0.2412

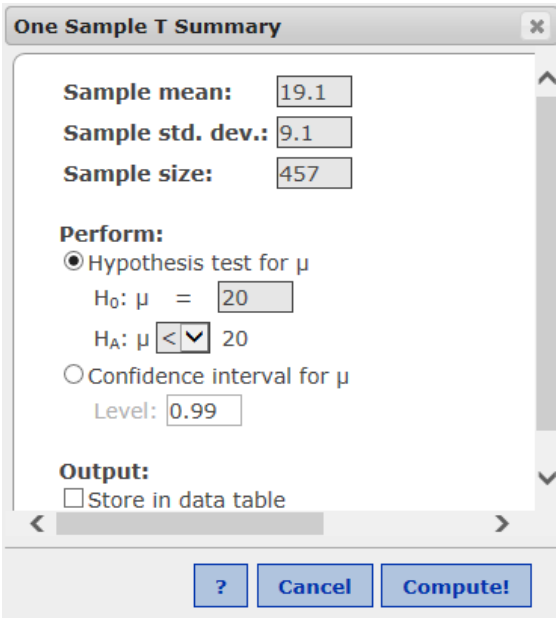
Use the obtained p-value to make a decision about the test.
p-value = 0.2412 is > alpha = 0.10 so we retain the null.

18. Hypothesis Testing (1 mean)

1. Stat > T Stats > One Sample > With Summary
2. Type in the values for sample mean, sample std. dev., sample size
3. Check on Perform hypothesis test for mean
4. Enter the value for the null hypothesis and select the correct symbol in the alternative hypothesis
5. Enter the confidence level if different than 95%
6. Click Compute!

Example: A nutritionist claims that the mean daily consumption of fiber for 20-39 –year-old males is less than 20 grams per day. In a survey of 457 males who were 20-39 years old, conducted by the U. S. Department of Agriculture, it was found that the mean daily intake of fiber was 19.1 grams, with standard deviation 9.1 grams. Test whether the mean daily consumption of fiber for 20-39 –year-old males is less than 20 grams per day using a significance level of 0.01

Complete step 1 and the One Sample T Summary box will pop up.



- Enter the values for sample mean 19.1, sample std. dev. 9.1, and sample size 457 from our example.
- Enter the value for the null hypothesis 20 and select the symbol "<" in the alternative hypothesis, enter the confidence level 0.99 because alpha is 0.01.
- Make sure the "Perform Hypothesis test for mean is checked" and click "Compute".

Hypothesis test results:
μ : Mean of population
H_0 : μ = 20
H_A : μ < 20

Mean	Sample Mean	Std. Err.	DF	T-Stat	P-value
μ	19.1	0.42568005	456	-2.114264	0.0175

Use the obtained p-value to make a decision about the test. p-value = 0.0175 is > alpha = 0.01 so we retain the null.

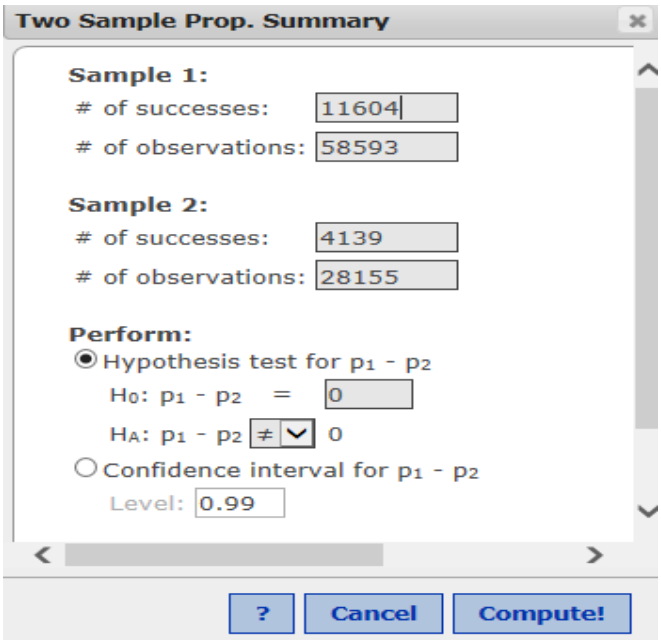
19. Hypothesis Testing (2 Proportions)

1. Stat > Proportion Stats > Two Samples > With Summary
2. Type in the values for sample 1 (the group that is on the left side of the hypothesis testing) the number of successes and the number of observations. Enter for sample 2 (the group that is on the right side of the hypothesis testing) the number of observations and the number of successes
3. Check on Perform hypothesis test
4. Select the correct symbol in the alternative hypothesis
5. Enter the confidence level if different than 95%
6. Click Compute!

Example: A study investigated survival rates for in-hospital patients who suffered cardiac arrest. Among 58,593 patients who had cardiac arrest during the day, 11,604 survived and were discharged. Among 28,155 patients who suffered cardiac arrest at night, 4139 survived and were discharged. Using the level of significance 0.01 test the claim that the survival rates are the same for day and night.

H_0: $p_d = p_n$
H_1: $p_d \neq p_n$

Complete step 1 and the Two Sample Prop. Summary box will pop up.



Enter the values for sample 1 (the day group)
Successes =11604
Observations = 58593

Enter the values for sample 2 (the night group)
Successes =4139
Observations = 28155

Select the symbol "≠" in the alternative hypothesis, enter the confidence level 0.99
Make sure the "Perform Hypothesis test" is checked and click "Compute".

Hypothesis test results:
p_1 : proportion of successes for population 1
p_2 : proportion of successes for population 2
$p_1 - p_2$: Difference in proportions
H_0 : $p_1 - p_2 = 0$
H_A : $p_1 - p_2 \neq 0$

Difference	Count1	Total1	Count2	Total2	Sample Diff.	Std. Err.	Z-Stat	P-value
$p_1 - p_2$	11604	58593	4139	28155	0.051036499	0.0027948458	18.260936	<0.0001

Use the obtained p-value to make a decision about the test. p-value is < 0.0001 which is < alpha = 0.01 so we reject the null.

20. Hypothesis Testing (2 means)

1. Stat > T Stats > Two Sample > With Summary
2. Type in for each sample the sample mean, sample std. dev., sample size
3. Uncheck the tab "Pool variances"
4. Check on Perform hypothesis test for mean. Select the correct symbol in the alternative hypothesis
5. Enter the confidence level if different than 95%
6. Click Compute!

Example: A study was done on body temperatures of men and women. The results are shown in the table below. Assume that the two samples are independent simple random samples selected from normally distributed populations. Use a 0.01 significance level, and test the claim that men have a higher mean body temperature than women.

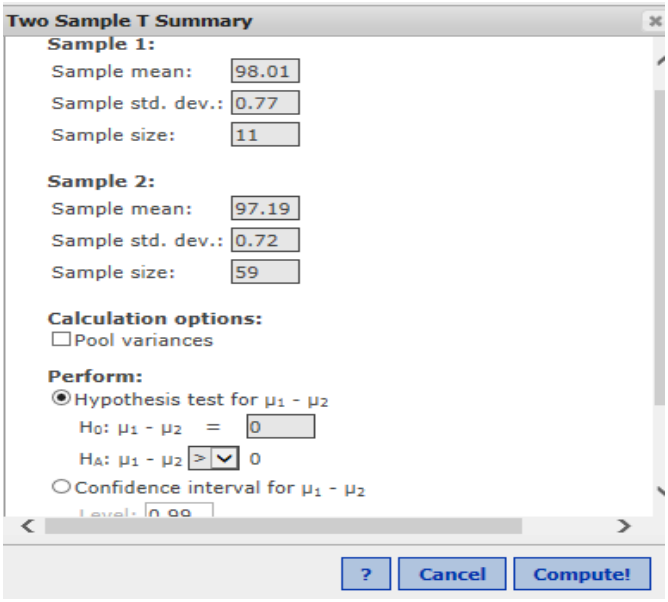
	Men	Women
μ	μ_1	μ_2
n	11	59
$\bar{x}$	98.01°F	97.19°F
s	0.77°F	0.72°F

The hypothesis are:

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$

Complete step 1 and the Two Sample T Summary box will pop up.



Enter the values for sample 1 (the men group)
Enter the values for sample 2 (the women group)

Select the symbol ">" in the alternative hypothesis, enter the confidence level 0.99 Make sure the "Perform Hypothesis test" is checked and click "Compute".

Hypothesis test results:

μ_1 : Mean of Population 1
μ_2 : Mean of Population 2
$\mu_1 - \mu_2$: Difference between two means
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 > 0$
(without pooled variances)

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	0.82	0.2503726	13.464321	3.2751187	0.0029

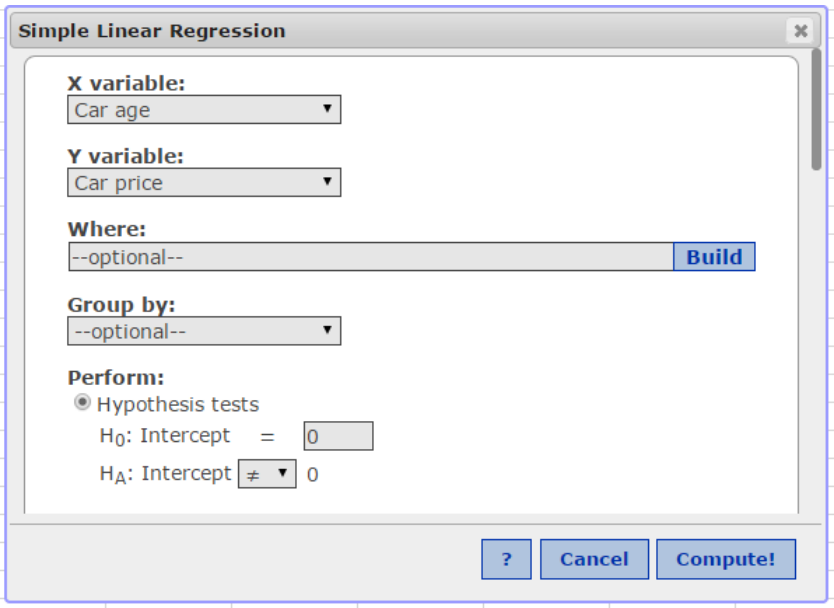
Since the p-value =0.0029 is < alpha = 0.01, we reject the null hypothesis.

21. The linear correlation coefficient (r) and the equation of the regression line

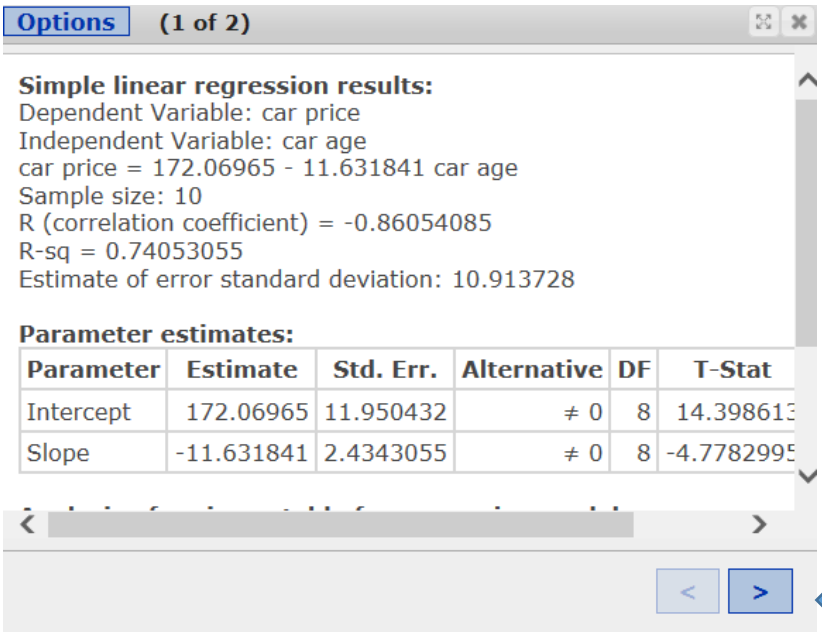
1. Stat > Regression > Simple Linear
2. Select the predictor variable for X and the response variable for Y
3. Click Compute!

Example: Car prices (in hundreds of dollars) and car age (in years)

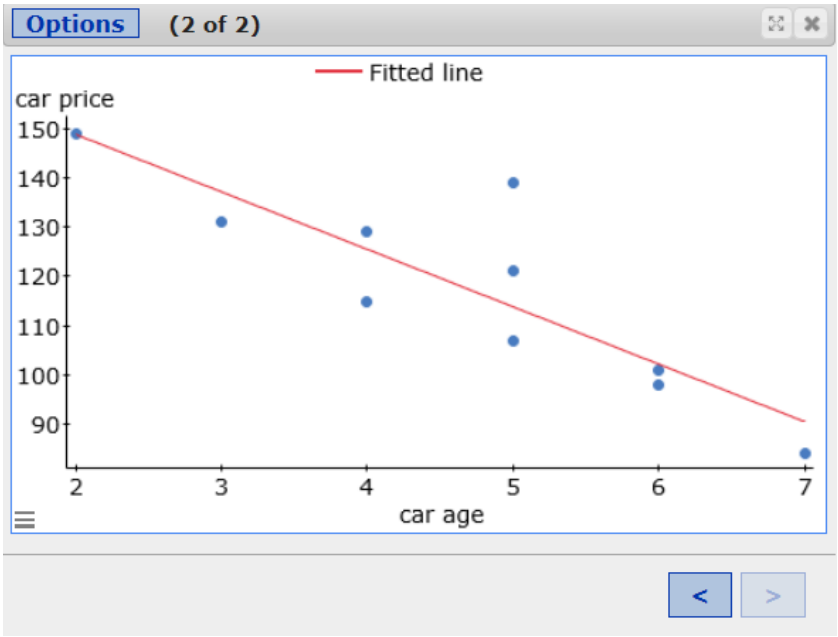
To access the data in StatCrunch click here. Must sign in StatCrunch to be able to analyze data.
This box will appear when you complete step 1.



The answers appears in a new window like the one below.
R = -0.86
Regression equation
Car Price = 172.07 – 11.63 Car Age



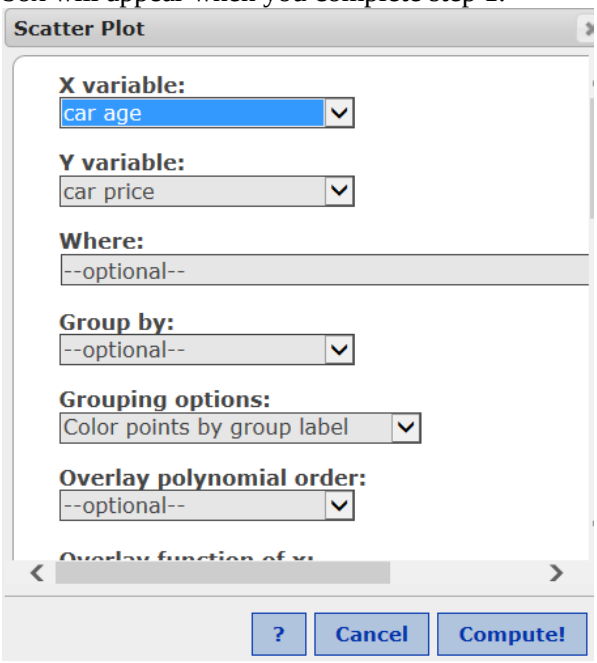
If we click on the arrow we see the scatter plot with the regression line.



22. Create a scatter plot

1. Select Graph > Scatter plot
2. Select the x –variable (row variable) you'll be using.
3. Select the y – variable (column variable) you'll be using.
4. Click Compute!

Example: To access the data in StatCrunch click here. Must sign in StatCrunch to be able to analyze data. This box will appear when you complete step 1.



23. Chi Square

1. Enter the contingency table in StatCrunch: Have your columns labeled and then enter the corresponding cell value.
2. Select Stat> Tables >Contingency > With Summary
3. Select your table Columns
4. Select your Row labels
5. Select as Display: Expected count
6. Click Compute!

Example: We have a random sample of 500 U.S. adults who are questioned regarding their political affiliation and opinion on a tax reform bill. Their answers are summarized in the table below. Test if the political affiliation and their opinion on a tax reform bill are dependent at a 1% level of significance.

The null and alternative hypotheses:

H_0 : Political affiliation and opinion on tax reform bill are independent

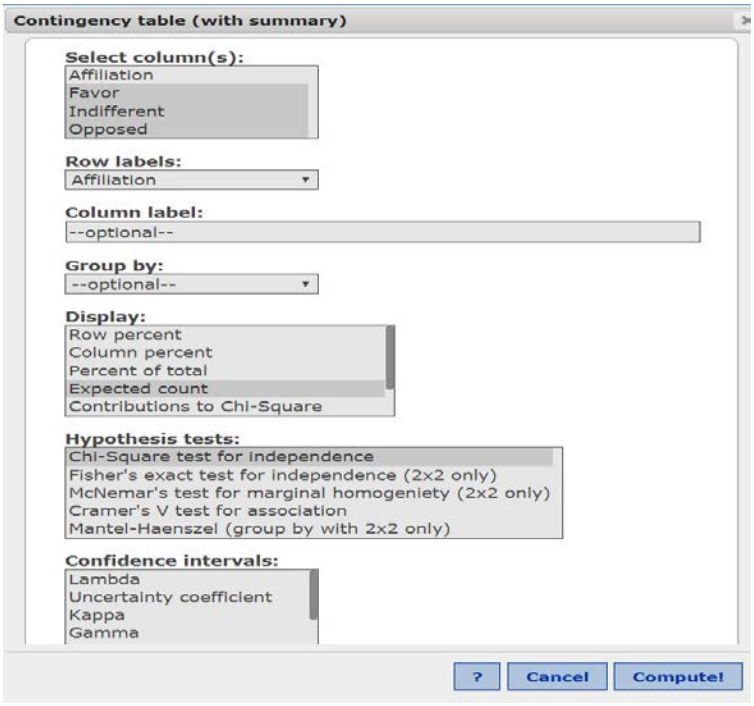
H_A : Political affiliation and opinion on tax reform bill are dependent

Affiliation	favor	indifferent	opposed	total
democrat	138	83	64	285
republican	64	67	84	215
total	202	150	148	500

Enter the contingency table in StatCrunch: Have your columns labeled as Affiliation, Favor, Indifferent, Opposed and then enter the corresponding cell value (see picture below).

Row	Affiliation	Favor	Indifferent	Opposed
1	Democrat	138	83	64
2	Republican	64	67	84
3				

Complete Step 2 and the Contingency table box will pop up. Complete step 3 through 5 and click Compute.



StatCrunch output:

Options

Contingency table results:
Rows: Affiliation
Columns: None

Cell format
Count (Expected count)

	Favor	Indifferent	Opposed	Total
Democrat	138 (115.14)	83 (85.5)	64 (84.36)	285
Republican	64 (86.86)	67 (64.5)	84 (63.64)	215
Total	202	150	148	500

Chi-Square test:

Statistic	DF	Value	P-value
Chi-square	2	22.152469	<0.0001

The test statistic $\chi 2 = 22.152$ and p-value <0.0001

Since p-value is less than level of significance 0.01, we reject the null hypothesis. We conclude that political affiliation and opinion on the tax reform are dependent.

24. One Way ANOVA

1. Enter the table in StatCrunch. Have your columns labeled accordingly.
2. Select Stat> ANOVA > One Way
3. Select your columns
4. Click Compute!

Example: The grade point averages of students participating in sports at a local college are to be compared. The data are listed in the table on the right.

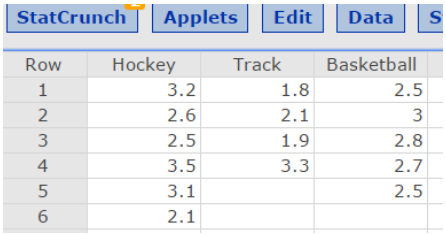
Hockey	Track	Basketball
3.2	1.8	2.5
2.6	2.1	3.0
2.5	1.9	2.8
3.5	3.3	2.7
3.1		2.5
2.1		

Test, at the level of significance 0.05, the hypothesis that there is a difference in the mean grade point averages of the three groups. Assume that the requirement for one-way ANOVA is satisfied.

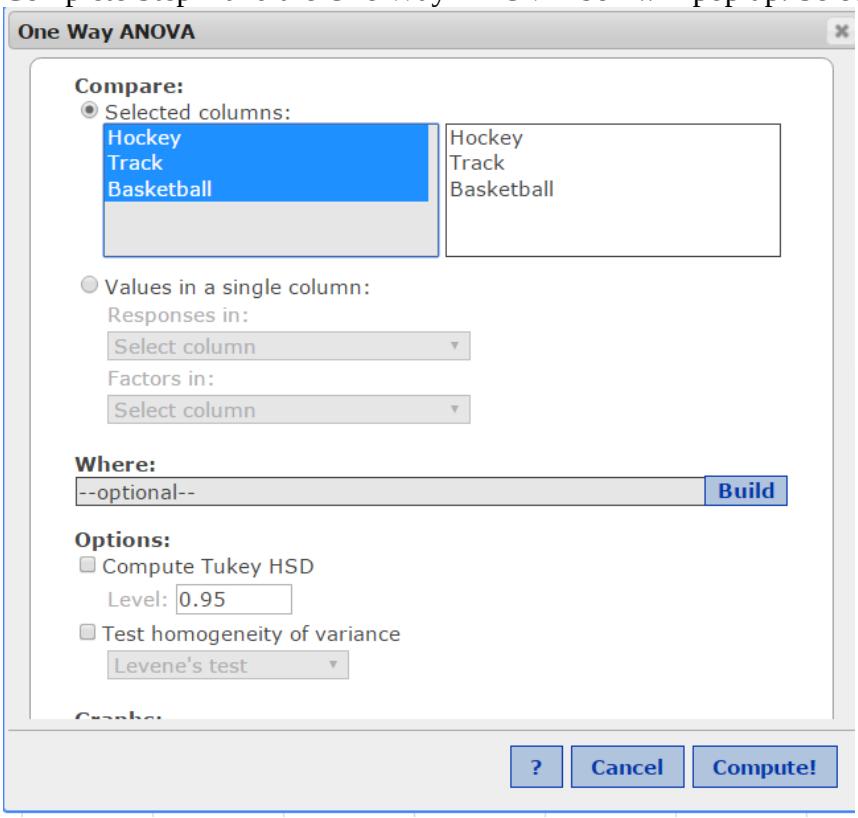
The null and alternative hypotheses:

H_o: $\mu_H = \mu_T = \mu_B$
H_A: At least one mean is different.

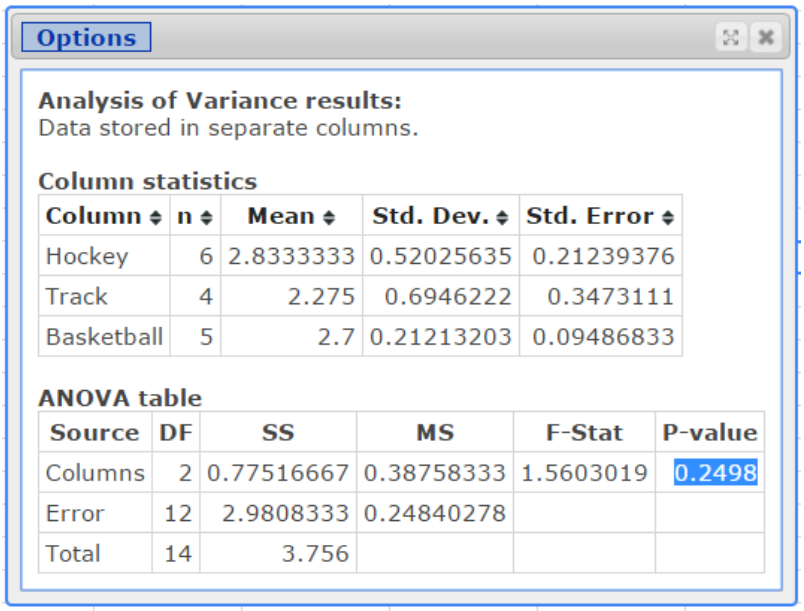
Enter the table in StatCrunch like in the picture below.



Complete Step 2 and the One Way ANOVA box will pop up. Select your columns and click Compute.



StatCrunch output:



Options

Analysis of Variance results:
Data stored in separate columns.

Column statistics

Column	n	Mean	Std. Dev.	Std. Error
Hockey	6	2.8333333	0.52025635	0.21239376
Track	4	2.275	0.6946222	0.3473111
Basketball	5	2.7	0.21213203	0.09486833

ANOVA table

Source	DF	SS	MS	F-Stat	P-value
Columns	2	0.77516667	0.38758333	1.5603019	0.2498
Error	12	2.9808333	0.24840278		
Total	14	3.756			

The test statistic F-stat = 1.56 and p-value = 0.2498

Do not reject H_o because p-value is not less than the significant level of 0.05. Thus, the sample data does not suggest that there is a difference in the mean grade point averages of the three groups.