Updated 2017-08-06

MTH 113: Introduction to Probability and Statistics with Computer Applications

Required Materials:

- *Elementary Statistics*, 13th edition, by Mario F. Triola. ISBN 978-0-13-446245-5.
- *MyStatLab*, Access code for this course management system is packaged with the new textbook from the CSI bookstore or sold separately at http://www.coursecompass.com. This also provides access to the *StatCrunch* software used in the computer labs.
- *Computer Laboratory Assignments*: The projects can be downloaded at http://www.math.csi.cuny.edu/Courses/MTH113/.
- *TI 83 Plus* or *TI 84 Plus Calculator*. These two models have functions that do substantial amounts of calculation in exactly the manner needed for this class. The textbook contains purple boxes detailing how to use these models to do computations.

Computer Laboratory Classes:
Each semester has 14 laboratory hours scheduled. The dates will vary by section, as will the use of the Q&A periods, which your instructor may use for exam prep or homework help, or to give you extra time to complete the laboratory assignments.

Lesson	Date	Topic	Completed?
1		Q & A (period for asking questions, getting caught up on, exam prep, etc)	
2		Lab #1: Getting Started with StatCrunch	
3		Lab #2: Describing, Exploring and Comparing Data - Graphically	
4		Lab #3: Describing, Exploring and Comparing Data - Numerically	
5		Q & A	
6		Lab #4: Exploring Probabilities	
7		Lab #5: The Normal Distribution	
8		Q & A	
9		Lab #6: Confidence Intervals with known standard deviation	
10		Lab #7: Confidence Intervals with estimated standard deviation	
11		Q & A	
12		Lab #8: Hypothesis Testing	
13		Q & A	
14		Lab #9: Goodness-of-Fit, Contingency Tables, Linear Regression	

Lecture Schedule:
Each semester has 42 lecture classes scheduled. The precise timing of the exams varies from section to section to accommodate weekends, holidays, and the lab schedule, and may be impacted by catastrophic weather or other unpredicted emergencies. The reading and pen-and-paper homework is listed here; your instructor will assign additional problems through *MyStatLab*. For optimal learning, make an attempt at the reading the day before the lecture, attend the lecture, and then start the homework (online or paper) in the hours after the lecture.

Lesson	Date	Topic	Reading	Paper Homework	Completed?
1		1.1: Statistical and Critical Thinking	1–9	P11: 1–20 odd	
2		1.2: Basic Types of Data	13–22	P23: 5–12,29–32	
3		2.1: Frequency Distributions	42–47	P48: 1,4,5,6	
4		2.2: Histograms	51–54	P55: 1–8	
5		2.4: Scatterplots, Correlation, Regression	67–74	P74: 1–4,6,10	
6		3.1: Measures of Center	82–89	P91: 1–5,7,9,12,29,30	
7		3.2: Measures of Variation	97–107	P107: 1–4,7,9,12,37,38	
8		3.3: Relative standing, Boxplots	112–22	P124: 1–4,31	
9		Review for Exam 1			
10		**First Exam, Chapters 1, 2, 3**			
11		**First Exam, Chapters 1, 2, 3**			
12		4.1: Basic Concepts of Probability	131–42	P143: 1–27	
13		4.2: Addition Rule, Multiplication Rule	147–55	P155: 1–4,9,11,21	
14		4.3: Complements, Conditional Probability	159–65	P166: 1–3,7,17–20	
15		4.4: Counting	169–74	P174: 1–4,29	
16		5.1: Probability Distributions	186–95	P195: 1–14	
17		5.2: Binomial Distribution	199–207	P209: 1–4,21,23,25	
18		6.1: Standard Normal Distribution	226-39	P240: 1–20 odds,37	
19		6.2: Applications of Normal Distribution	242–9	P251: 1–20 odds	
20		6.3: Sampling Distributions	254–62	P262: 1–6,11	
21		6.4: CLT, First Hypothesis Test	265–71	P272: 1–4,5,18	
22		Review for Exam 2			
23		**Second Exam, Chapters 4, 5, 6**			
24		**Second Exam, Chapters 4, 5, 6**			
25		7.1: Estimating p	297–310	P311: 1–4,9,13,19	
26		7.2: Estimating μ	316–26	P327: 1–9,11	
27		8.1: Hypothesis Testing	356–70	P371: 1–4,7,8,11,12,15, 16,27,28	
28		8.2: Testing p	373–81	P382: 1–4,13,14,19,27	

Lesson	Date	Topic	Reading	Paper Homework	Completed?
29		8.3: Testing μ	387–94	P396: 1–6,9,12,20,21	
30		9.1: Testing p_1, p_2	414–23	P423: 1–5,7–10,12	
31		9.2: Testing μ_1, μ_2, (independent samples)	428–35	P437: 1–4,7–10,20	
32		9.3: Testing μ_1, μ_2, (matched pairs)	442–7	P449: 1–6,11,12,18	
33		Review CIs			
34		Review Hypothesis Testing			
35		**Third Exam, Chapters 7, 8, 9**			
36		**Third Exam, Chapters 7, 8, 9**			
37		10.1: Correlation	468–81	P483: 1–4,9,17,18	
38		10.2: Regression	489–97	P499: 1–5,17,26	
39		11.1: Goodness-of-Fit	533–41	P542: 1–4,8,11,16	
40		11.2: Contingency Tables	546–55	P556: 1–5,10	
41		Review for Final Exam			
42		Review for Final Exam			

Formulas and Tables

Pages 5 and 6 of the "Formulas and Tables" pages are available in your textbook and also here:

https://media.pearsoncmg.com/aw/aw_triola_elemstats_13_2018/website/stat13t_barrelfold.pdf

They (and *only those two* pages) will be made available to you on the final exam. Their usage on the midterm exams depends on your instructor. The first four pages, the tables, are to be avoided in favor of using a calculator to do the computations.

Pearson's MyLab & Mastering Student Registration Instructions

Go to www.pearsonmylabandmastering.com.

1. Under Register, select Student.
2. Confirm you have the information needed, then select OK! Register now.
3. Enter your instructor's course ID and continue.
4. Enter your existing Pearson account username and password to Sign In. You have an account if you have used a Pearson product, for example: MyMathLab, MyITLab, MyPsychLab, MySpanishLab or Mastering, such as MasteringBiology. If you don't have an account, select Create and complete the required fields.
5. Select an access option. Use the access code that came with your textbook or that you purchased separately from the bookstore. Buy access using a credit card or PayPal account. If available, get 14 days temporary access. (The link is near the bottom of the screen.)
6. From the confirmation page, select Go To My Courses.
7. On the My Courses page, select the course tile Test to start your work.

To sign in later:

1. Go to www.pearsonmylabandmastering.com.
2. Select Sign In.
3. Enter your Pearson account username and password, and Sign In.
4. Select the course tile to start your work.

To upgrade temporary access to full access:

1. Go to www.pearsonmylabandmastering.com.
2. Select Sign In.
3. Enter your Pearson account username and password, and Sign In.
4. Select Upgrade access from the course tile.
5. Enter an access code or purchase access with a credit card or PayPal account.

For a registration overview, go to www.pearsonmylabandmastering.com/students/get-registered. Scroll down to Need a little help? and select a video.

Computer Lab Project No. 1

Getting Started with StatCrunch

StatCrunch is a statistical software package that can be accessed via the internet from any computer. We will use this package to analyze data during computer laboratory sessions in MTH 113. The purpose of this first lab assignment is to provide a brief introduction to StatCrunch and familiarize you with several useful features. [1]

1. You can access StatCrunch by clicking the link found on the left toolbar in MyStatLab. [2]

2. Note, you can either click "data sets from your textbook" to open a version of StatCrunch that enables you to work with data sets from the Triola textbook or "StatCrunch website" to access the more general version. We'll mainly work with the first option in this course.

3. Once you are within the StatCrunch page (this should open in a separate window), you can begin to analyze data as you wish. In this course, you will be provided with detailed instructions for each lab project.

4. All commands in StatCrunch can be accessed from the menu bar at the top of the page. The StatCrunch menu bar consists of six tabs : *StatCrunch, Edit, Data, Stat, Graph,* and *Help.* When a data set is uploaded, you will also see the name of the file to the right side of the menu bar.

5. We will learn more about each command tab during later labs but in summary: "StatCrunch" allows you to organize the way your data and results are displayed. "Edit" enables you to manipulate the data in various ways. "Data" allows you to bring data into StatCrunch. "Stat" gives you access to all statistical procedures available in StatCrunch that will enable you to invoke the techniques and concepts learned in class. "Graph" provides several graphical techniques to visualize your data. Lastly, "Help" connects you to the StatCrunch help page to get an overview of all StatCrunch features.

6. The "Options" button in your StatCrunch results area can be used to edit, copy, print, or save your work. You will be able to save your work to either your computer or in your account on the StatCrunch website. We suggest saving your work to the website so you can access it from any computer. Access any saved data by clicking Data→Load Data→My Data. To load results, click StatCrunch→Open statcrunch.com. In the new window, click MyStatCrunch. The results are available under "My Results".

7. A few notes of **caution** about StatCrunch: This is a web-based application so be careful about how you handle your browser window during your session. For instance, you should not use the back button and you should always access other websites in another browser. Otherwise, you will lose all your data and results in your StatCrunch session without any warning!

Here is what you should do today to begin familiarizing yourself with StatCrunch:

1. Let's first practice loading a data set from the textbook. Click on the data named *Alcohol and Tobacco in Movies*. Notice, the rows and columns are now filled with data. How many rows and columns are there? Use the scroll bar to check this. What types of variables are included in this data set? Which are *quantitative*?

[1] More detailed information can also be found on the StatCrunch website at: www.statcrunch.com

[2] See syllabus for instructions on how to access the online course management system, MyStatLab. Access to StatCrunch can also be purchased separately on their website.

2. You can clear data by selecting "Edit" → "Select All" followed by "Edit" → "Cut". Specific rows or columns can be deleted by selecting "Edit" → "Column" (or Row)→ "Delete". Next click on the name of the column or row you wish to delete and click "Delete". You can choose several columns or rows to delete by pressing the control key while clicking on them. Try this. You may have to reload the data.

3. The first thing you'll often want to do during data analysis is visualize your data. The **GRAPH** tab will assist with this. Click on "Graph" → "Pie Chart" → "With Data" and then click *Tobacco (sec)*. Now, click "Compute!". A pie graph should appear in a separate window (this is the default option, see below for viewing results differently). You will learn about the appropriate graphs to use with which data types next lab.

4. Most frequently you will want to make a descriptive summary of data. The **STAT** command will allow you to do this. Click on "Stat" → "Summary Stats" → "Columns" to achieve this. Notice only the quantitative variables in the data set are available for selection here. Select each variable (by control-clicking in case you want to choose several), and click "Compute!" when you're done. A table of summary statistics should pop up. We will discuss how to interpret these values later in the course and revisit this application in Lab 3. Many of the other **STAT** features will also be explored in future labs.

5. Practice saving and reopening your results using the **OPTIONS** Tab in your results window. Click "Options" → "Save". In the popup window, you can choose a title under which to save your results, and you can choose whether the results should be public or only accessible to you. You can also add a tag and a note to your results. Your results are now saved at statcrunch.com. To access them later, follow the procedures described above: click StatCrunch→Open statcrunch.com. In the new window, click MyStatCrunch. The results are available under "My Results". Click on the file to open. You can now copy, print, or email your results.

6. Continue to explore StatCrunch. Feel free to try out a new data set! As with any statistical software, you will learn more and become more comfortable the more you utilize StatCrunch.

Computer Lab Project No. 2

Describing, Exploring, and Comparing Data - Graphically

In today's lab we will explore creating and interpreting graphs such as bar plots, pie graphs, and histograms, as well as frequency tables, in StatCrunch.

Bar Plots:

1. Start StatCrunch.
2. Load relevant data into StatCrunch.
3. Click on "Graph" in the menu bar.
4. Hover over "Bar Plot" in the pulldown menu.
5. Click the "with data" option to use data consisting of individual outcomes in the data table.
6. Select the column(s) to be displayed. If you want to choose several columns, you have to control-click them (hold down the Ctrl key while clicking the columns you want to display - this method works whenever you want to choose several items from a list, not only in the context of bar plots).
7. There are some additional options. For example, "Group by" column allows you to make side by side bar plots, separating the data according to the chosen category. You can also choose the type (Frequency or Relative Frequency).
8. Click "Compute!" to construct the bar plot(s).

Frequency Tables:

1. Load Data into StatCrunch.
2. Click on "Stat" in the menu bar.
3. Hover over "Tables".
4. Click on "Frequency".
5. Select the column(s) for which summary statistics are to be computed.
6. Click "Compute!" to generate the frequency table(s).

Histograms:

1. Load Data into StatCrunch.
2. Click on "Graph" in the menu bar.
3. Click on "Histogram".
4. Select column(s) to be displayed.
5. The optional "Group by" column allows you to create side by side histograms, separating the data according to the chosen category.
6. You may set additional options such as "Type" (Frequency or Relative Frequency), number of bins, and bin width.
7. You can choose labels for the X/Y-axis, a title, a color scheme, etc., under "Graph properties".
8. If you create several histograms simultaneously, you may specify how many will go on one page (rows/columns)
9. Click "Graph!" to construct the histogram(s).

Pie Charts:

1. Start StatCrunch.
2. Load relevant data into StatCrunch.
3. Click on "Graph" in the menu bar.
4. Hover over "Pie Chart" in the pulldown menu.
5. The remaining steps are almost identical to the ones described under "Bar Plots".

Here is what you should do today:

Part 1:

1. Load the data set titled *Births* into MyStatlab.
2. Use both a pie chart and bar plot to display the days of the week at which the patients were released from hospital (the column *Discharged*).
3. Which of these two ways of representing the distribution of the discharge weekdays do you think is more useful?

Part 2:

1. Load the data set titled *Freshman 15*. It is a sample of weights and body mass index measurements from students taken at the beginning and end of the freshman year.
2. Use StatCrunch to create a frequency table and histogram of the variables *WT SEPT* and *WT APR*.
3. What relative frequency of people start their freshman year weighing less than 70 kilograms? How does this compare for men and women? What relative frequency of people end their freshman year weighing less than 70 kilograms?
4. Describe the distributions of these two variables. Are the data Normally distributed?
5. To analyze the weight gain/loss, let's create a new calculated column containing the difference between the weight in April and the weight in September. Follow these steps:
   (a) Click "Data" in the menu.
   (b) Click "Compute Expression" in the pulldown menu.
   (c) Type "WTAPR-WTSEP" in the text field.
   (d) Type a title for the new column. For example "WTGAIN".
   (e) Click "Compute!" To create the column.
6. Create a histogram of the calculated weight gain column.
7. Create histograms of that column, grouped by the students' sex, using the same x-axis and y-axis, with both histograms displayed side by side on one page.

Computer Lab Project No. 3

Describing, Exploring, and Comparing Data - Numerically

In today's lab we will explore obtaining and analyzing numerical summary statistics such as mean and standard deviation in StatCrunch. We will also study boxplots.

Obtaining Summary Statistics:

1. Start StatCrunch.
2. Load relevant data into StatCrunch.
3. Click on "Stat" in the menu bar.
4. Click on "Summary Stats".
5. Identify the column(s) which contain the data of interest.
6. Click "Calculate" to obtain the descriptive statistics.

Boxplots:

1. Start StatCrunch.
2. Load relevant data into StatCrunch.
3. Click on "Graphics" in the menu bar.
4. Click on "Boxplot".
5. The optional "Group by" column can be used to compare boxplots across groups on a single graph.
6. Click "Next" to indicate whether or not to use fences for the boxplots. Note: The five-number summary is used by default.
7. Click "Create Graph" to construct the boxplot(s).

Here is what you should do today:

1. Load the data set titled *Word Counts*. It is a count of the number of words spoken in a day for both men and women, taken in six different samples.
2. First for a bit of preprocessing, let's combine all of the counts for males in one column and all of the counts for females into another column. This can be accomplished through the following:
   (a) Click on "Data" in the menu bar.
   (b) Select "Stack Columns" and select all the columns with (M) for male. Store the labels in var13 and the data in var14.
   (c) Click "Stack Columns" and verify that your data is now combined on your spreadsheet. You should change the name of the var14 column label to Male.
   (d) Repeat this by putting the female data in columns var15 and var16.
3. Using what you learned in Lab 2, create and compare a histogram of the Male and Female word data. Do they appear to have the same general shape? Would you consider these samples to have come from a population that was Normally distributed?
4. Calculate the summary statistics and use them to comment on the central tendency and variation of the data. Does it appear that women are more talkative than men?

5. Create side-by-side boxplots for both variables. Do these boxplots reinforce the description you gave above? How many outliers are present? Did you notice those outliers earlier?

6. Now repeat the steps above separately for sample 3 (which was taken on students in Mexico) and sample 5. How would you describe and compare these two samples using relevant displays and summary statistics?

Computer Lab Project No. 4

Exploring Probabilities

Today's lab will allow us to further explore the concept of probability through the use of simulation.

1. Simulation allows you to replicate or produce data similar to the way a true physical process would behave. We can simulate the result of a birth by generating a list of 0s and 1s, 0 representing a boy and 1 representing a girl. Here are the steps to simulate 152 births:

   (a) Click Data → Simulate → Discrete Uniform.

   (b) In the pop-up window, put "152" in the text field labeled "Rows".

   (c) Put "1" in for "Columns".

   (d) Put "0" in for minimum and "1" in for maximum.

   (e) Under "Seeding", choose "Dynamic".

   (f) You may choose a name for the column (such as "boy/girl").

   (g) Click "Compute!".

   Use your simulation to estimate the probability of having a boy. Is is reasonable to expect at least 127 boys in 152 births? Why or why not? You may make a frequency distribution in order to analyze the results of the simulation.

2. Shaquille O'Neal is a professional basketball star who had a reputation for being a poor free throw shooter with a success rate of 0.528. We can simulate 200 free throws as follows:

   (a) Click Data → Simulate → Bernoulli.

   (b) In the pop-up window, put "200" in the text field labeled "Rows".

   (c) Put "1" in for "Columns".

   (d) Put "0.528" in for **p**.

   (e) Click "Compute!".

   Repeat the simulation of free throws five times and record the number of times that the free throw was made. Is the percentage of successful free throws from the simulation reasonably close to 0.528 in each case? You may use an appropriate frequency table for this. Study the sequences of hits and misses, how long is the longest run of misses? How long is the longest run of hits? Compare this with your classmates.

3. The probability of randomly selecting an adult who recognizes the brand name of McDonald's is 0.95. Conduct a simulation of size 10 and record the number of consumers who recognize the brand name of McDonald's. Is the proportion of those who recognize McDonald's reasonably close to the 0.95? Try another simulation this time with sample size 75. How do the results compare?

4. Lastly, let's consider a well know probability example called *The Birthday Problem*. Similarly to number 1 above, simulate 30 random birthdays (Hint: birthdays can be represented by whole numbers between 1 and 365). Examine your 30 simulated birthdays, how many people have the same birthday? Now perform 10 replications of simulating 30 birthdays, and estimate the probability of getting at least two people sharing the same birthday when you have 30 people. Compare with your classmates.

Computer Lab Project No. 5

The Normal Distribution

In this project, you are going to explore how to use software in order to calculate probabilities for normal distributions. You won't need to consult the probability table for the standard normal distribution, and you won't need to calculate z-scores to find probabilities for nonstandard normal distributions. Here is the procedure for finding probabilities or values (percentiles) for normal distributions using StatCrunch:[1]

1. Start StatCrunch.
2. Click on "Stat" in the menu bar.
3. Click or hover over the menu item "Calculators".
4. Click "Normal".
5. In the "Normal Calculator" window that opens, enter the parameters μ and σ of the desired normal distribution in the text fields "Mean" and "Std. Dev."
6. If you want to **find the probability** that X (a normally distributed random variable with the specified parameters) is less than (or equal) to a value, enter that value in the text box on the left of the second line.
7. If you know the probability that X is less than (or equal) to an unknown value and you want to **find the value**, then enter the probability in the textbox on the right of the second line.
8. When you're done entering the information, click the "Compute" button on the bottom right. The field you left open will be filled with the answer you were looking for.

If you are interested in probabilities of the form $P(X \geq x)$, then you can change the relation symbol on the bottom left by clicking it and choosing the one you want.

Here is what you should do today:

Part 1:

1. Load the data set titled *Body Temperatures*. It is a sample of body temperatures of healthy adults (male/female, smoker/nonsmoker), in degrees Fahrenheit, and taken at different times of day.
2. Use the "Graphics" menu of StatCrunch to make a histogram of the sample. Leave the window containing the histogram open.
3. Use the "Stat" menu of StatCrunch to calculate a summary of the column statistics of column *Day 1 - 8 am*.
4. Open the normal calculator, as described above. Enter the sample mean as μ and the sample standard deviation as σ. These values can be read off the summary statistics. Three decimal places suffice.
5. Use the normal calculator to calculate the (theoretical) probability $P(X \leq 97.85)$, if X is normally distributed with the parameters you entered.
6. In addition to the wanted probability, the normal calculator also displays a graph of the normal distribution with the parameters you entered. Compare it to the histogram. Do you think this normal distribution is a good approximation to the distribution of body temperatures of healthy adults?

[1] Instructions can also be found in the textbook on page 325.

7. What percentage of the sample data is less than 97.85? Compare that percentage to the probability you calculated. In order to tell what percentage of the data is less than 97.85, you can order the data (Data $\rightarrow$ Sort).

8. Find the 80th percentile of the normal distribution with the parameters you entered.

Part 2:

1. Open the data set titled *Cigarette Contents*. It contains data concerning tar, nicotine and carbon monoxide content of various types of cigarettes.

2. Use the "Stat" menu of StatCrunch to calculate a summary of the column statistics. We'll be interested in the column titled "MENTH NICOTINE". It represents the nicotine contents of menthol cigarettes.

3. Open the normal calculator, as described above. Enter the sample mean (the average nicotine content of non-menthol filtered cigarettes) as μ and the sample standard deviation as σ.

4. Use the normal calculator to calculate the (theoretical) probability $P(0.6 \leq X \leq 0.8)$, if X is normally distributed with the parameters you entered. Assuming the nicotine contents are roughly normally distributed with these parameters, that probability should be close to the relative frequency of scores lying between 0.6 and 0.8 observed in a sample of large size.

5. Look at the graph of the normal distribution displayed in the normal calculator. The maximal value of the curve is larger than 1.5. Discuss how this is possible, considering that probabilities can never be larger than 1.

Computer Lab Project No. 6

Confidence Intervals with known standard deviation

Today we are going to experiment with confidence intervals for the population mean. We are going to use the methods that were developed in class, for estimating the population mean when the *population standard deviation is known.*

1. Simulate rolling a fair die 100 times, using StatCrunch. You'll need 3 samples which should be put in columns 1 to 3. Here is how this can be done:
   (a) Click Data→Simulate Data→Discrete Uniform.
   (b) In the popup window, put "100" in the text field labeled "Rows".
   (c) Put "3" in the text field labeled "Columns".
   (d) Put "1" as minimum and "6" as maximum.
   (e) You can leave the other settings as they are. Click the "Simulate" button at the bottom right of the window.

2. Consider the experiment of rolling a fair die. Calculate the mean and the standard deviation. The random numbers you generated can be viewed as being samples taken from a population with that mean and standard deviation.

3. Use what you learned in class to calculate confidence intervals for the population mean with confidence level 0.95 based on these samples. Note that even though these samples are realizations of the same distribution (are taken from the same population), the confidence intervals are not the same. The confidence intervals are calculated from, and thus dependent on, the specific sample used. Recall the formula for the confidence interval for the mean at confidence level α:
$$(\bar{x} - E, \bar{x} + E), \text{ where } E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$
Recall that $z_{\alpha/2}$ is the score such that the area under the normal curve and to the right of the score is $\alpha/2$. In other words, $z_{\alpha/2}$ is the $1 - \frac{\alpha}{2}$-th percentile of the normal distribution. Also, α is 1−the confidence level, so in our case, $\alpha = 0.05$. You can use StatCrunch to calculate $z_{\alpha/2}$, using the normal calculator, as discussed last time.

4. In how many of your samples was the population mean contained in the confidence interval you calculated?

5. Collect all the results obtained in the class. What percentage of confidence intervals calculated by you and your peers contained the population mean?

6. Repeat these steps from the beginning, but this time, calculate confidence intervals at confidence level 0.99.

Computer Lab Project No. 7

Confidence Intervals, continued

In today's project, you'll learn how to use software in order to calculate confidence intervals for the mean, using the normal distribution and the t-distribution. Here is how it works:

1. Start StatCrunch.
2. If you want to calculate the confidence interval of a sample, then load it into StatCrunch.
3. Click on "Stat" in the menu bar.
4. If you want to use the t-distribution for the calculation of the confidence interval, then choose "T statistics" in the submenu, and if you want to use the normal distribution, then choose "Z statistics".
5. In the next submenu, choose "One sample".
6. In the following submenu, choose either "with data" or "with summary". Note that the confidence interval for the mean of a sample at a fixed confidence level depends only on the sample mean, the sample size and the standard deviation (either of the sample or of the population). If you choose "with summary", you'll have to enter these values, instead of providing the entire sample.
7. If you chose "with data", then you'll have to select the column that contains the data in the next popup window. You can ignore the other text fields and click "Next>".

   If you chose "with summary", then you'll have to enter the measurements of the sample mentioned above in the next popup window.
8. If you wish, you can check the box "Store output in data table".
9. Choose "Confidence Interval", and enter the desired confidence level.
10. Click "Compute!".

Here is what you should do today:

1. Load the data set titled *Body Data* into StatCrunch.
2. We'll only be interested in the pulse, but we will want to be able to analyze the data by gender. So you can delete all columns except for *Gender (1=Male)* and *Pulse*. This data set assumes gender is binary, which no doubt is outdated and inaccurate.
3. Calculate 90% and 95% confidence intervals for the population mean of the male population, using the appropriate method. To focus only on the male population, after choosing the variable "Pulse", use the text field labeled "Where:". Click "Build", and put together the condition "GENDER = 1". Don't add the results to the table.
4. In the window displaying the results, click "Options" and save them to your computer.
5. Proceed as above, but this time, the "where" condition is that the gender is 0, corresponding to "female."
6. Calculate and save a 90% and 95% confidence interval for the population mean again. Compare the confidence intervals you calculated for the pulses of males and females.

Computer Lab Project No. 8

Hypothesis Testing

Today's topic is hypothesis testing. Here is the general procedure, very similar to the calculation of confidence intervals. The built-in hypothesis tests of StatCrunch use the P-value method instead of the critical value method.

1. Start StatCrunch.
2. If you want to test a hypothesis concerning a given sample, then load it into StatCrunch.
3. Click on "Stat" in the menu bar.
4. If you want to use the t-distribution for the hypothesis test, then choose "T statistics" in the submenu, and if you want to use the normal distribution, then choose "Z statistics". These choices apply if you want to test hypotheses concerning the mean of a population, with σ unknown or known. If you want to test a hypotheses about a proportion of a population, choose "Proportions".
5. In the next submenu, choose "One sample", "Two sample" or "Paired" (only in the T-statistics submenu).
6. In the following submenu, choose either "with data" or "with summary". Note that the test statistic depends only on the sample mean, the sample size and the standard deviation (either of the sample or of the population). If you choose "with summary", you'll have to enter these values, instead of providing the entire sample.
7. If you chose "with data", then you'll have to select the column(s) that contain(s) the data in the next popup window. You can ignore the other text fields and click "Next>".
8. Choose "Hypothesis Test" (this is the default), enter the desired significance level and the null hypothesis, and choose the relation symbol in the alternative hypothesis.
9. If you wish, you can check the box "Store output in data table".
10. Click "Compute!".

Here is what you can work on today. When testing, always use significance level 0.05.

1. Test the claim that a proportion is 0.6, given a data sample of 40 successes among 100 trials, with the alternative hypothesis that the proportion is not 0.6.
2. Load the data set *Body Data* into the StatCrunch table. Test the claim that the mean body temperature is less than 98.6 degrees.
3. Test the hypothesis that the mean female pulse rate is less than the mean male pulse rate, using two-sample t-statistics. Compare to the confidence intervals for these rates that you calculated last time. Remember, you can focus on the male/female population by using the "where" condition GENDER = 0/1.

Computer Lab Project No. 9

Goodness-of-Fit, Contingency Tables, Linear Regression

Today's topics are Goodness-of-Fit, Contingency Tables and Linear Regression. First, here are the general procedures for a **Goodness-of-Fit** analysis:

1. Start StatCrunch.
2. Put the observed frequencies in one column and the hypothesized frequencies in another, so that they line up.
3. Click on Stat→Goodness-of-Fit→Chi-Square test.
4. In the popup-window, put the observed frequencies column in the text field labeled "Observed", and put the hypothesized/expected frequencies column in the field labeled "Expected".
5. Click on the "Calculate" button on the bottom right.

Contingency Tables:

1. Enter the row labels in a column, then enter the frequencies in separate columns. So the row labels go in the spreadsheet body, unlike column headings.
2. Click on "Stat" in the menu bar.
3. Click on "Tables", choose "Contingency", and click the option "with summary".
4. In the popup-window, under "Select column(s)", select all columns containing observed frequencies (by ctrl-clicking them in turn), and then in the next box titled "Row labels:", select the column with the row labels.
5. Make sure under "Hypothesis tests:", the Chi-Square test for independence is chosen.
6. Click on the "Compute!" button on the bottom right.

Here are the projects concerning goodness of fit and contingency tables you can work on today:

1. Two dice are rolled repeatedly. The frequency distribution of the results are as follows.

   Die 1:

Number	1	2	3	4	5	6
Frequency	20	12	11	8	9	0

   Die 2:

Number	1	2	3	4	5	6
Frequency	13	12	8	12	8	7

   For each die, test the hypothesis that it is fair.

2. Refer to your textbook for a discussion of the following data describing the fate of passengers and crew on the Titanic:

	Men	Women	Boys	Girls
Survived	332	318	29	27
Died	1360	104	35	18

   Perform a test for independence between the categories of surviving/dying and passenger type.

Now let's turn to Linear Regression. The general procedure, using StatCrunch, is as follows:

1. Load the (two-dimensional) data you want to analyze into the StatCrunch table.

2. Click on Stat→Regression, and select the option "Simple Linear".

3. In the popup window, select the columns you want to analyze as X- and Y-variable.

4. Scroll down until you reach the "Graphs" field. If you want to see the data as a scatter plot, fitted with the regression line, then choose "Fitted line plot".

5. Click the "Compute!" button on the bottom right. If you chose the graphics option, then you can go back and forth between the numerical linear regression results and the plot using the "Next>" and "<Back" buttons.

Here is what you can work on:

1. Load the Data Set titled *Bear Measurements* into StatCrunch. This data set contains measurements of wild bears such as age, sex, headlength, weight, etc.

2. Use linear regression to analyze the relationship between the age of a bear and at least two of the measurements (other than month or sex). Make plots.

3. Compare with the results of your peers. Which measurement seems to be the most reliable/accurate indicator of the age of a wild bear?

4. If you have time, you can try to analyze the data for male and female wild bears separately. Does this improve the linear approximation? Are there outliers? Can you guess why? If you take away potential outliers, what is the best correlation coefficient you get?