LAB: INTRO TO STAT ANALYSIS

Zeyi Qian

zeqian@clarku.edu
Office Hours: JC 201, Tuesday 4-5 PM & Thursday 3-4 PM

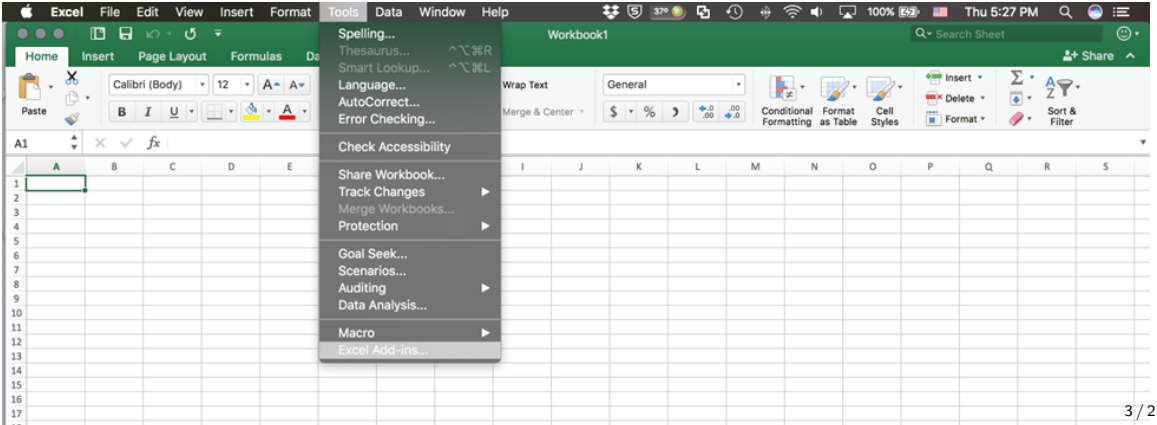
September 6, 2024

Lab schedule

- ECON160-901: Friday 09:00 AM - 09:50 AM JC103
- ECON160-902: Friday 01:25 PM - 02:15 PM JC103
- Materials will be posted on Canvas before sections
- Answers will be posted on Canvas after sections

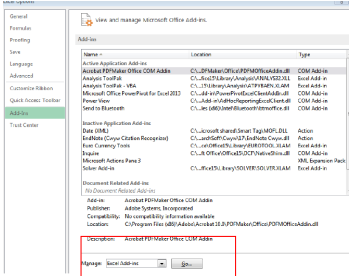
Install data analysis tool: MAC

- Click: tool-click “excel adds in”-select data analysis
- If you fail to do this on your Macbook, please use the computers in the lab (Those computers use Windows system)



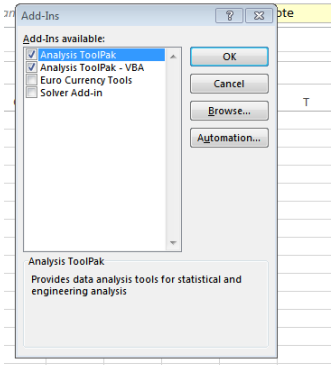
Install data analysis tool: Windows

- 1. Go to the Info screen by clicking on File from the main menu
- 2. Click on “options” at the base of the menu
- 3. Click on Add-Ins and highlight Analysis ToolPak (It will be in the list of “inactive application add-ins”)
- 4. Click “go” once you have ensured that manage “Excel Add-ins” appears in the menu and you have highlighted Analysis ToolPak



Install data analysis tool: Windows

- 5. In this menu, click on Analysis ToolPak and click on OK



- 6. You should see Data Analysis on your data menu

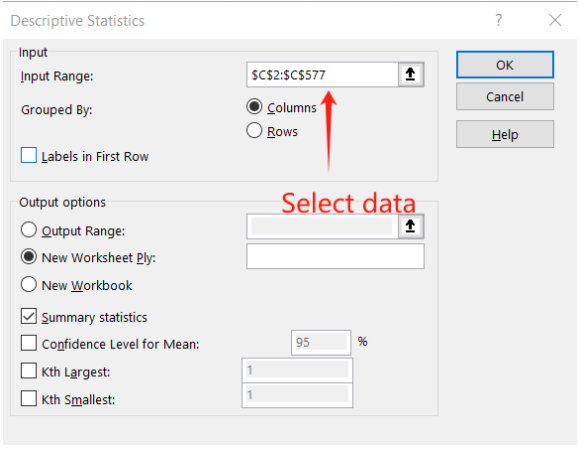
Question 1

Use ***Lab 1 data*** to answer this question. To help travelers know what to expect, researchers collected the prices of commodities in 576 cities around the world (prices are all in US$ as of August 2016).

- a. Use Excel to find the mean, median, range, IQR and stdev of cappuccino prices.
- b. Based on your findings in a), discuss what the distribution looks like. Is the mean or median a better measure of centre in this case? Is the range or IQR a better measure of variation?
- c. Now use Excel to draw the histogram of cappuccino prices around the world and then discuss the distribution of cappuccino prices.

Question 1: a & b (use data analysis tool)

- Use tool: "Descriptive statistics"

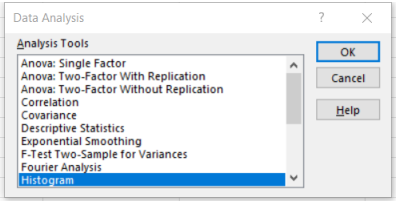


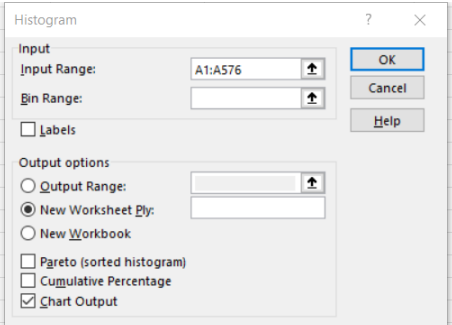
Question 1: a & b (use formulas)

		Excel
mean	2.720087	=AVERAGE(A1:A576)
median	2.85	=MEDIAN(A1:A576)
max	5.22	=MAX(A1:A576)
min	0.6	=MIN(A1:A576)
range	4.62	=MAX(A1:A576)-MIN(A1:A576)
Q1	1.6725	=QUARTILE.EXC(A1:A576,1)
Q3	3.56	=QUARTILE.EXC(A1:A576,3)
IQR	1.8875	=QUARTILE.EXC(A1:A576,3)-QUARTILE.EXC(A1:A576,1)
stdev	1.075143	=STDEV(A1:A576)

- Mean<median: left-skewed; Median & IQR are more appropriate measures of central tendency because median & IQR are not affected by extreme values

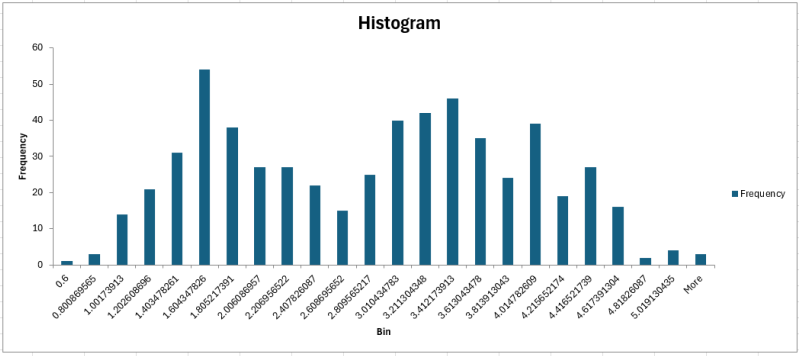
Question 1: c (use data analysis tool)





Question 1: c (use data analysis tool)

Bin	Frequency
0.6	1
0.80087	3
1.00174	14
1.20261	21
1.40348	31
1.60435	54
1.80522	38
2.00609	27
2.20696	27
2.40783	22
2.6087	15
2.80957	25
3.01043	40
3.2113	42
3.41217	46
3.61304	35
3.81391	24
4.01478	39
4.21565	19
4.41652	27
4.61739	16
4.81826	2
5.01913	4
More	3



- Bin: upper limits for several values (Bin=0.6 means Minimum to 0.6)

Question 1: c (use data analysis tool)

- Prices of cappuccino are concentrated in the range of 1.6 to 3.6 worldwide
- The majority of prices are relatively moderate, with very high or very low prices being less common

Question 2

Prior to graduation, a high school class was surveyed about its plans. The following table displays the results for white and minority students (the "Minority" group included African-American, Asian, Hispanic, and Native American students).

	White	Minority
4-Year College	198	44
2-Year College	36	6
Military	4	1
Employment	14	3
Other	16	3

- a. What percent of the seniors are planning to attend a 2-year college?
- b. What percent of the seniors are white and planning to attend a 2-year college?
- c. What percent of the white seniors are planning to attend a 2-year college?
- d. What percent of the seniors planning to attend a 2-year college are white?
- e. Create a conditional distribution by race (draw bar chart).

Question 2

	White	Minority
4-Year College	198	44
2-Year College	36	6
Military	4	1
Employment	14	3
Other	16	3
Total	268	57

- 268 "=SUM(B3:B7)"
- 57 "=SUM(C3:C7)"

Question 2: a, b, c, & d

- a. What percent of the seniors are planning to attend a 2-year college?
  $12.92\% = \frac{36+6}{268+57}$ "=(B3+C3)/(B7+C7)"
- b. What percent of the seniors are white and planning to attend a 2-year college?
  $11.08\% = \frac{36}{268+57}$ "=B3/(B7+C7)"
- c. What percent of the white seniors are planning to attend a 2-year college?
  $13.43\% = \frac{36}{268}$ "=B3/B7"
- d. What percent of the seniors planning to attend a 2-year college are white?
  $85.71\% = \frac{36}{36+6}$ "=B3/(B3+C3)"

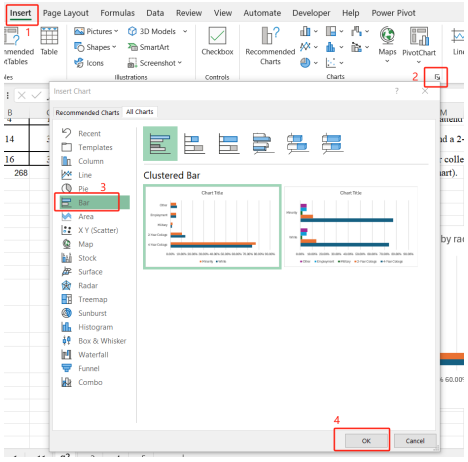
Question 2: e

- e. Create a conditional distribution by race (draw bar chart).
- Create a new table: we let the data for each race be divided by the total number of races (column %)

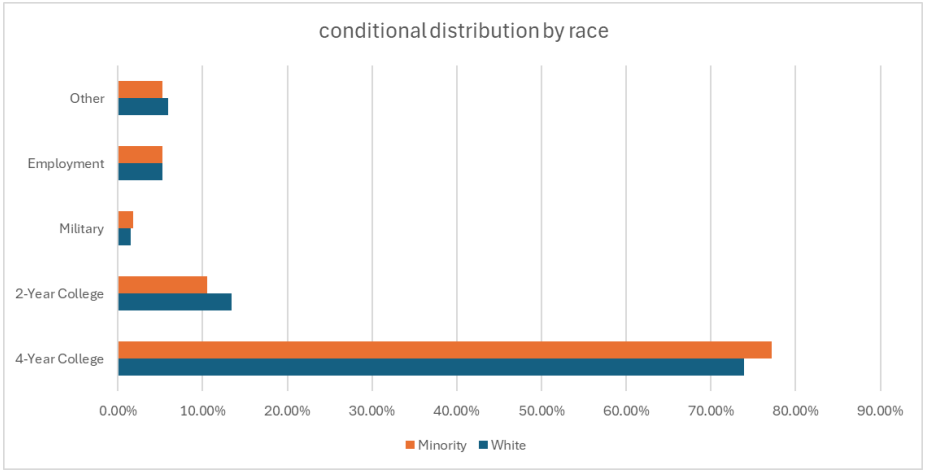
	White	Minority
4-Year College	73.88%	77.19%
2-Year College	13.43%	10.53%
Military	1.49%	1.75%
Employment	5.22%	5.26%
Other	5.97%	5.26%

Question 2: e

- Select the whole new table and click “Insert”
- Find “Bar” from “All Charts”



Question 2: e


conditional distribution by race
Other
Employment
Military
2-Year College
4-Year College
0.00%
10.00%
20.00%
30.00%
40.00%
50.00%
60.00%
70.00%
80.00%
90.00%
Minority
White

Question 3

For a period of five years, physicians at McGill University Health Center followed more than 5000 adults over the age of 50. The researchers were investigating whether people taking a certain class of antidepressants (SSRIs) might be at greater risk of bone fractures. Their observations are summarized in the table:

- Do these results suggest there's an association between taking SSRI antidepressants and experiencing bone fractures? Explain.

	Taking SSRI	No SSRI	Total
Experienced Fractures	14	244	**258**
No Fractures	123	4627	**4750**
Total	**137**	**4871**	**5008**

Question 3

	SSRI	No SSRI
Exp	$\frac{14}{137}$	$\frac{224}{4871}$
No exp	$\frac{123}{137}$	$\frac{4627}{4871}$

$\downarrow$ (column %)

	SSRI	No SSRI
Exp	10.22%	5.01%
No exp	89.78%	94.99%

- People taking SSRIs have a higher risk of fractures than those not taking SSRIs

Extra Practice Question: 4

A clerk entering salary data into a company spreadsheet accidentally put an extra "0" in the boss's salary, listing it as $2,000,000 (2 million!) instead of $200,000. Explain how this error will affect these summary statistics for the company payroll:

- a. measures of center:
  - Median: Little to no effect
  - Mean: Significantly increased (The maximum value has increased)
- b. measures of spread:
  - Range: Significantly increased (The maximum value has increased)
  - IQR: Little to no effect
  - Standard Deviation: Significantly increased (Outlier)

Extra Practice Question: 5

Students were asked to describe their politics as "Liberal", "Moderate", or "Conservative". Here are the results:

- a. What percent of the class is male? $115/192 = 59.89\%$
- b. What percent of the class considers themselves to be "Conservative"? $(6 + 21)/192 = 14.06\%$
- c. What percent of males in the class consider themselves to be "Conservative"? $21/115 = 18.26\%$
- d. What percent of all students in the class are males who consider themselves to be "Conservative"? $21/192 = 10.93\%$

	Liberal	Moderate	Conservative	Total
Female	35	36	6	77
Male	50	44	21	115
Total				192