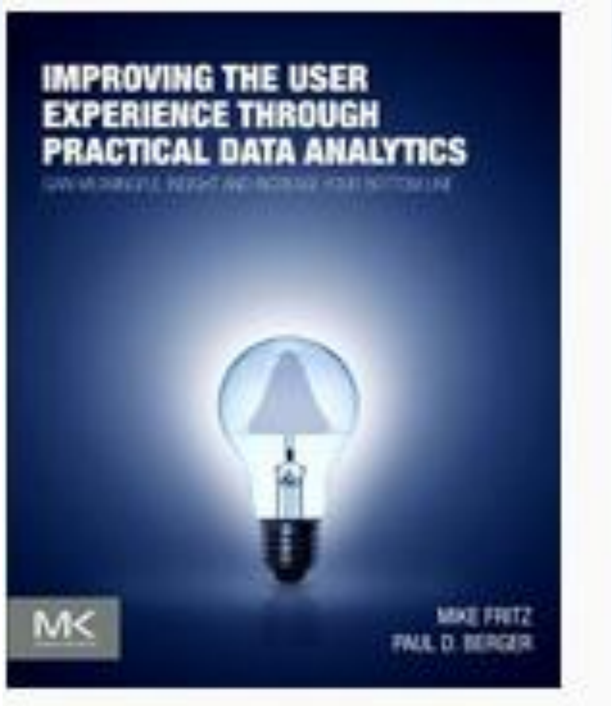
Lab 7

Data analysis

COMP 350

User Interface Design and Programming

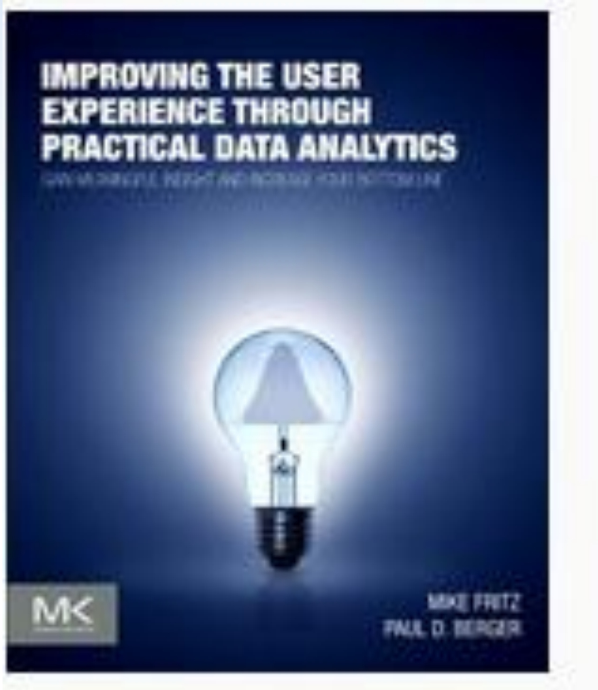
Instructor
Kyungjae Lee (just call me KJ)
KyungJae.Lee@ufv.ca


IMPROVING THE USER EXPERIENCE THROUGH PRACTICAL DATA ANALYTICS
MK
MIKE FRITZ
PAUL D. BERGER

Team project submission including survey/analysis

Chapter 1.
Introduction to a variety of useful statistical ideas and techniques

Improving the User Experience through Practical Data Analytics:
Gain Meaningful Insight and Increase Your Bottom Line
by Mike FritzPaul D. Berger


IMPROVING THE USER EXPERIENCE THROUGH PRACTICAL DATA ANALYTICS
MK
MIKE FRITZ
PAUL D. BERGER

Lab 7 (Individual)

Review statistical procedures covered in Page 14 - 29

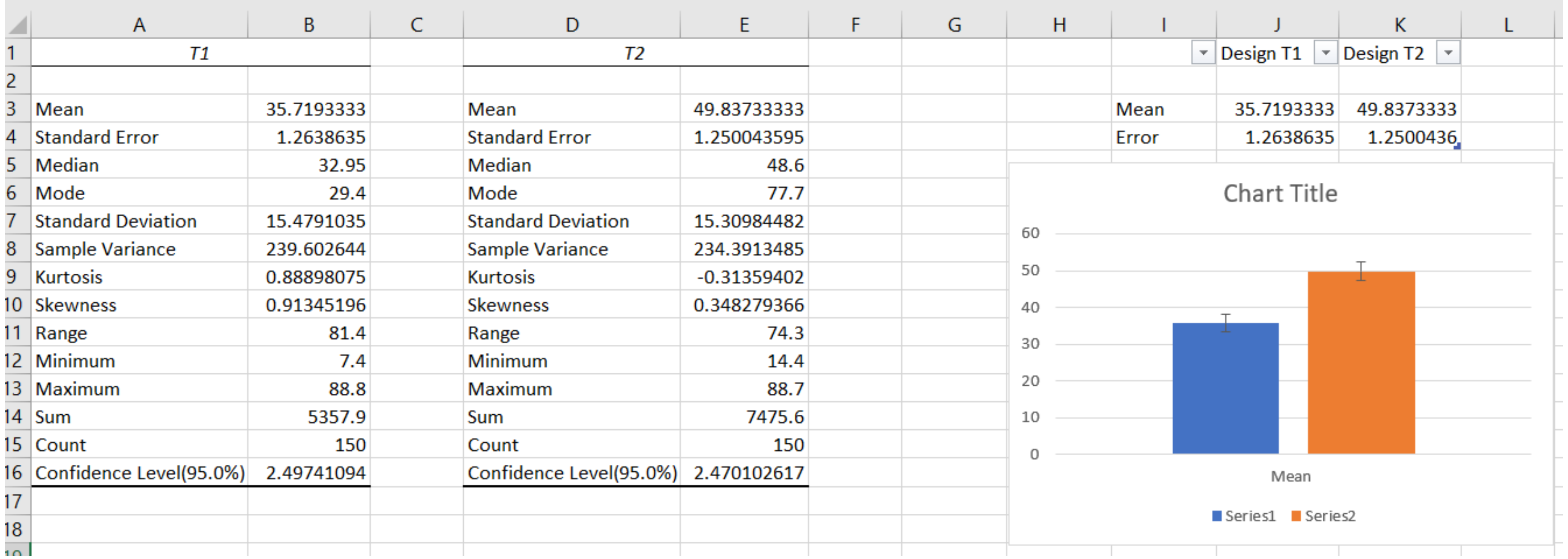
7.1 MS Excel

7.2 IBM SPSS

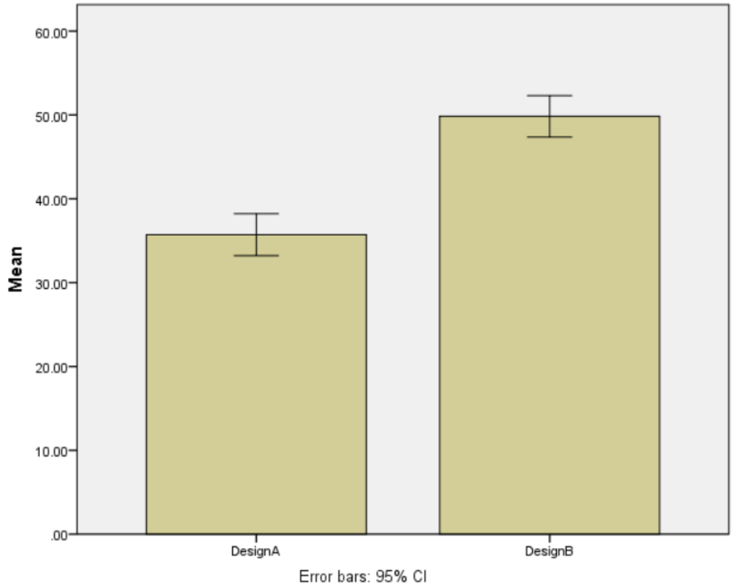
Improving the User Experience through Practical Data Analytics: Gain Meaningful Insight and Increase Your Bottom Line
by Mike FritzPaul D. Berger

MS Excel tips

- https://www.ablebits.com/office-addins-blog/2019/04/10/error-bars-excel-standard-custom/
- https://www.brightcarbon.com/blog/powerpoint-error-bars/
- https://youtu.be/DqeBe0rilAs



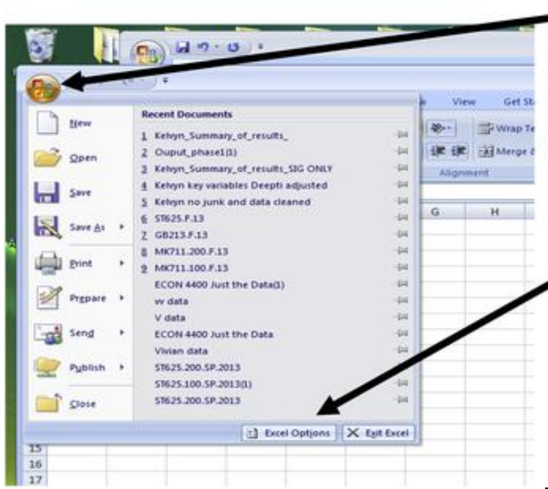
	A	B	C	D	E	F	G	H	I	J	K	L
1	*T1*			*T2*						Design T1	Design T2	
2												
3	Mean	35.7193333		Mean	49.83733333				Mean	35.7193333	49.8373333	
4	Standard Error	1.2638635		Standard Error	1.250043595				Error	1.2638635	1.2500436	
5	Median	32.95		Median	48.6							
6	Mode	29.4		Mode	77.7							
7	Standard Deviation	15.4791035		Standard Deviation	15.30984482							
8	Sample Variance	239.602644		Sample Variance	234.3913485							
9	Kurtosis	0.88898075		Kurtosis	-0.31359402							
10	Skewness	0.91345196		Skewness	0.348279366							
11	Range	81.4		Range	74.3							
12	Minimum	7.4		Minimum	14.4							
13	Maximum	88.8		Maximum	88.7							
14	Sum	5357.9		Sum	7475.6							
15	Count	150		Count	150							
16	Confidence Level(95.0%)	2.49741094		Confidence Level(95.0%)	2.470102617							
17												
18												

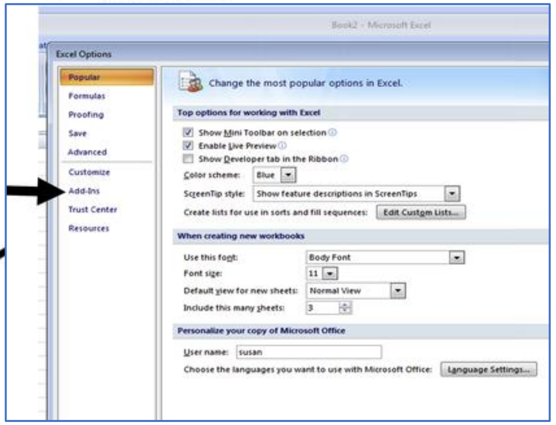


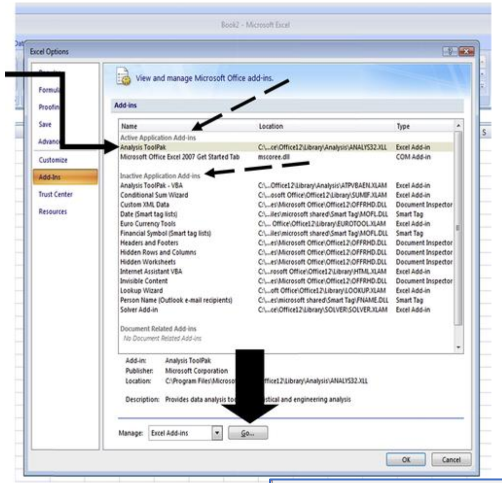
Homework 3 example

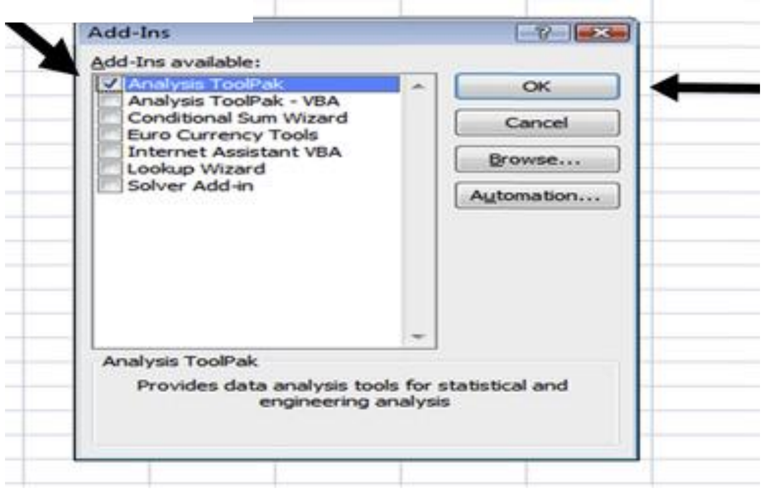
7.1 To find a confidence interval for the true (population) mean in Excel

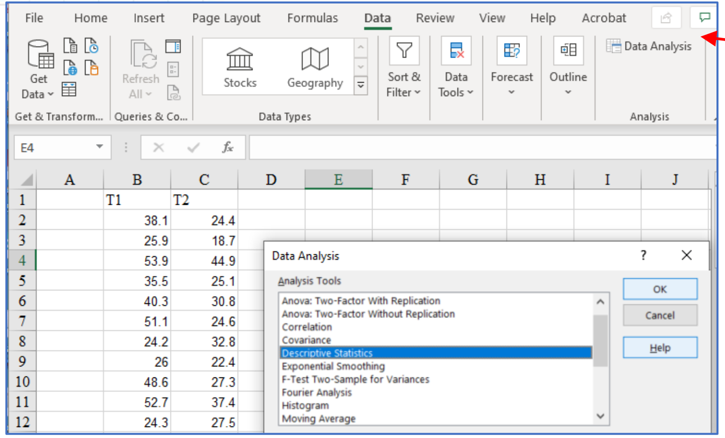
Office Button to




Book2 - Microsoft Excel
Excel Options
Popular
Formulas
Proofing
Save
Advanced
Customize
Add-Ins
Trust Center
Resources
Change the most popular options in Excel.
Top options for working with Excel
Show Mini Toolbar on selection
Enable Live Preview
Show Developer tab in the Ribbon
Color scheme: Blue
ScreenTip style: Show feature descriptions in ScreenTips
Create lists for use in sorts and fill sequences: Edit Custom Lists...
When creating new workbooks
Use this font: Body Font
Font size: 11
Default view for new sheets: Normal View
Include this many sheets: 3
Personalize your copy of Microsoft Office
User name: susan
Choose the languages you want to use with Microsoft Office: Language Settings...







	A	B	C
1		T1	T2
2		38.1	24.4
3		25.9	18.7
4		53.9	44.9
5		35.5	25.1
6		40.3	30.8
7		51.1	24.6
8		24.2	32.8
9		26	22.4
10		48.6	27.3
11		52.7	37.4
12		24.3	27.5

1. Open the file (ConfidenceInterval_yourLastnameFirstnameInitial.xlsx), and rename it with your last name and first name initial (e.g., **ConfidenceIntervalData_LeeKJ.xlsx**)

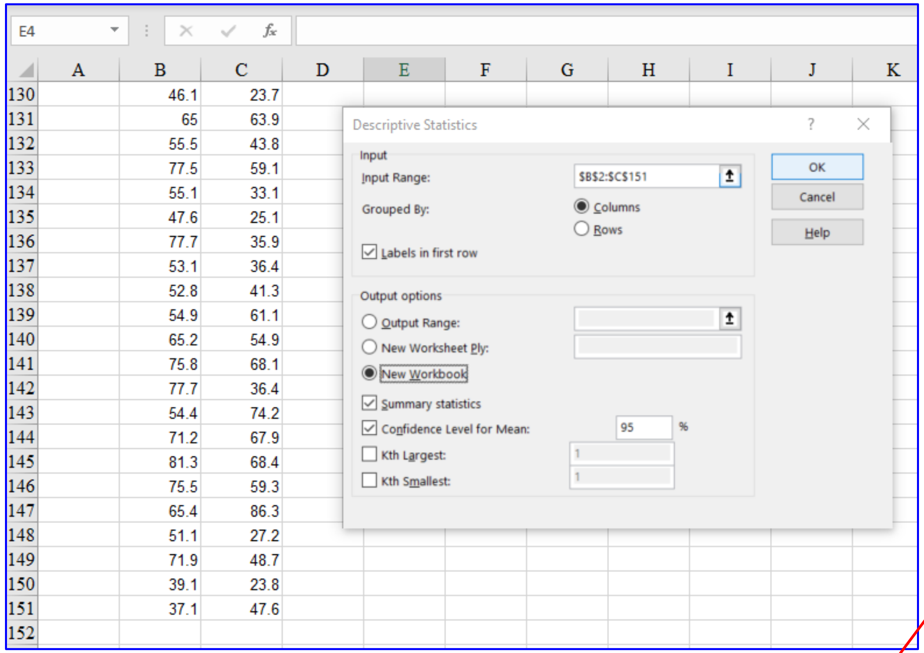
2. Go to File > Options.

3. Choose Add-Ins.

4. To activate Analysis tool, highlight "Analysis ToolPak" and then click on "Go" and make Analysis Toolpak option checked.

5. To find a confidence interval for the true (population) mean, select Data > click Data Analysis, and then highlight Descriptive Statistics and click "OK".

Sample from different data set



	A	B	C	D	E	F	G	H	I	J	K
130		46.1	23.7								
131		65	63.9								
132		55.5	43.8								
133		77.5	59.1								
134		55.1	33.1								
135		47.6	25.1								
136		77.7	35.9								
137		53.1	36.4								
138		52.8	41.3								
139		54.9	61.1								
140		65.2	54.9								
141		75.8	68.1								
142		77.7	36.4								
143		54.4	74.2								
144		71.2	67.9								
145		81.3	68.4								
146		75.5	59.3								
147		65.4	86.3								
148		51.1	27.2								
149		71.9	48.7								
150		39.1	23.8								
151		37.1	47.6								
152											

Descriptive Statistics

Input

Input Range: $B$2:$C$151

Grouped By: ● Columns ○ Rows

☑ Labels in first row

Output options

○ Output Range:

○ New Worksheet Ply:

● New Workbook

☑ Summary statistics

☑ Confidence Level for Mean: 95 %

☐ Kth Largest: 1

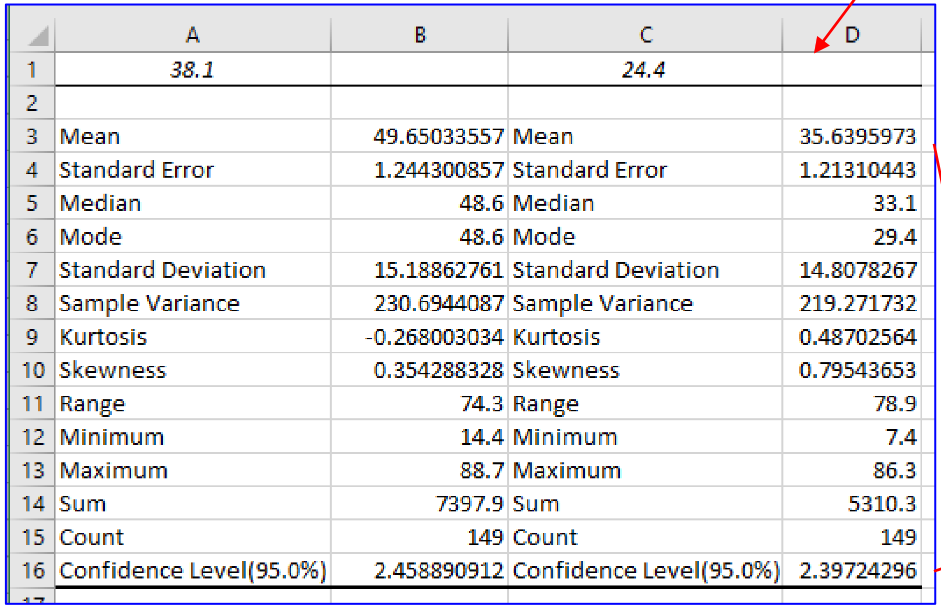
☐ Kth Smallest: 1

6. In the “Input Range” section, select an entire data range in T1 where the data values are B2 to B151, and we check off “Summary Statistics” and “Confidence Interval for Mean” (which has a default of 95%; traditionally, 95% is the choice used very often in industry).

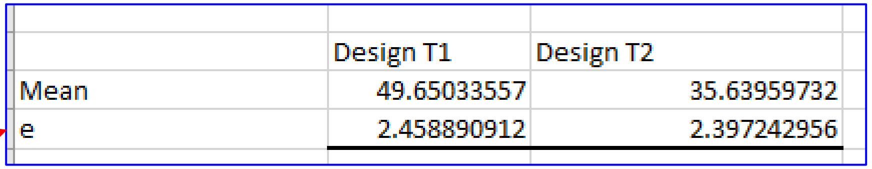
7. Also select options for Labels in first row and New Workbook. Click OK to generate the output.

8. Repeat previous steps for T2 column.

9. Save/rename a new worksheet using your name; **ConfidenceIntervalBarChart_yourLastnameFirstnameInitial.xlsx**. Design a table showing following labels across horizontal/vertical cells (your calculation may show different values from the sample image shown here). Copy both mean and error (Confidence level data cell) of both T1 and T2 design and create a new table with these values.



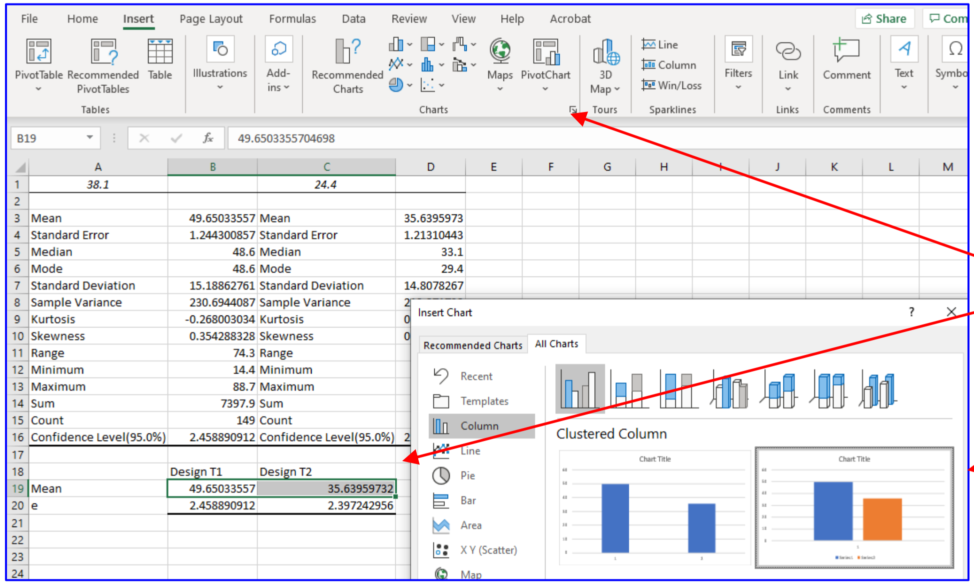
	A	B	C	D
1	*38.1*		*24.4*	
2				
3	Mean	49.65033557	Mean	35.6395973
4	Standard Error	1.244300857	Standard Error	1.21310443
5	Median	48.6	Median	33.1
6	Mode	48.6	Mode	29.4
7	Standard Deviation	15.18862761	Standard Deviation	14.8078267
8	Sample Variance	230.6944087	Sample Variance	219.271732
9	Kurtosis	-0.268003034	Kurtosis	0.48702564
10	Skewness	0.354288328	Skewness	0.79543653
11	Range	74.3	Range	78.9
12	Minimum	14.4	Minimum	7.4
13	Maximum	88.7	Maximum	86.3
14	Sum	7397.9	Sum	5310.3
15	Count	149	Count	149
16	Confidence Level(95.0%)	2.458890912	Confidence Level(95.0%)	2.39724296



	Design T1	Design T2
Mean	49.65033557	35.63959732
e	2.458890912	2.397242956

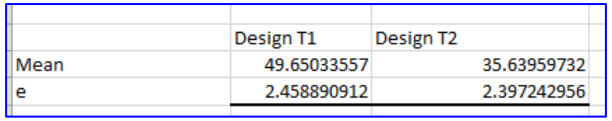
Sample output from different data set

How to Display Confidence Intervals On a Bar Chart in Excel



B19 49.6503355704698

	A	B	C	D
1	*38.1*		*24.4*	
2				
3	Mean	49.65033557	Mean	35.6395973
4	Standard Error	1.244300857	Standard Error	1.21310443
5	Median	48.6	Median	33.1
6	Mode	48.6	Mode	29.4
7	Standard Deviation	15.18862761	Standard Deviation	14.8078267
8	Sample Variance	230.6944087	Sample Variance	2
9	Kurtosis	-0.268003034	Kurtosis	0
10	Skewness	0.354288328	Skewness	0
11	Range	74.3	Range	
12	Minimum	14.4	Minimum	
13	Maximum	88.7	Maximum	
14	Sum	7397.9	Sum	
15	Count	149	Count	
16	Confidence Level(95.0%)	2.458890912	Confidence Level(95.0%)	2
17				
18		Design T1	Design T2	
19	Mean	49.65033557	35.63959732	
20	e	2.458890912	2.397242956	

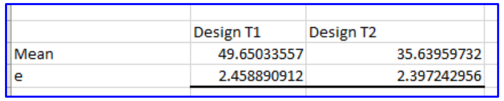


	Design T1	Design T2
Mean	49.65033557	35.63959732
e	2.458890912	2.397242956

11. After selecting Mean row, go to Insert > Chart > choose See All Charts arrow icon > Column > select 2nd option.

12. To add an error bar, click on each bar > click **+** symbol > check Error bar option > bring the mouse cursor on Error bar to show an arrow symbol (**>**), click **>** and choose More options.

13. Select Custom > click Specify Value , and enter a number(e.g., 2.41) shown on error(e) value for Design A.

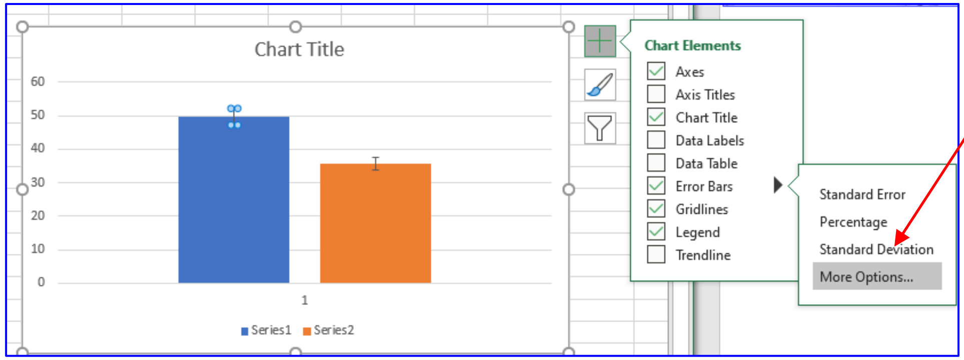


	Design T1	Design T2
Mean	49.65033557	35.63959732
e	2.458890912	2.397242956

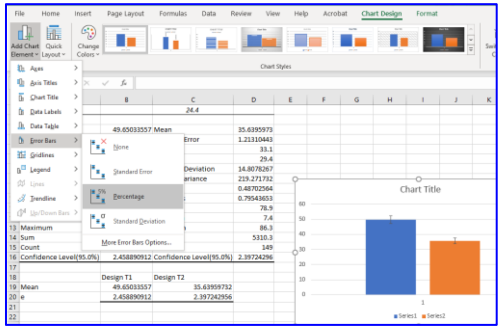
14. Repeat the previous process (step 12 & 13)to create an error bar for Design B.

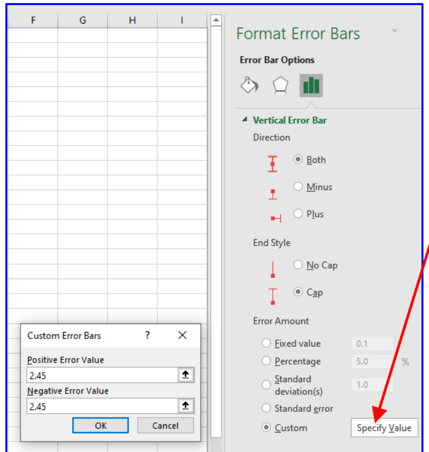
15. Another way to add an error bar is by clicking Add Chart Element under Chart Design.

16. **Save and submit both files for 7.1 (Must rename both files by adding your last/first name initial)**
    - **ConfidenceIntervalBarChart_yourLastnameFirstnameInitial.xlsx**
    - **ConfidenceIntervalData_yourLastnameFirstnameInitial.xlsx**



Sample output from different data set

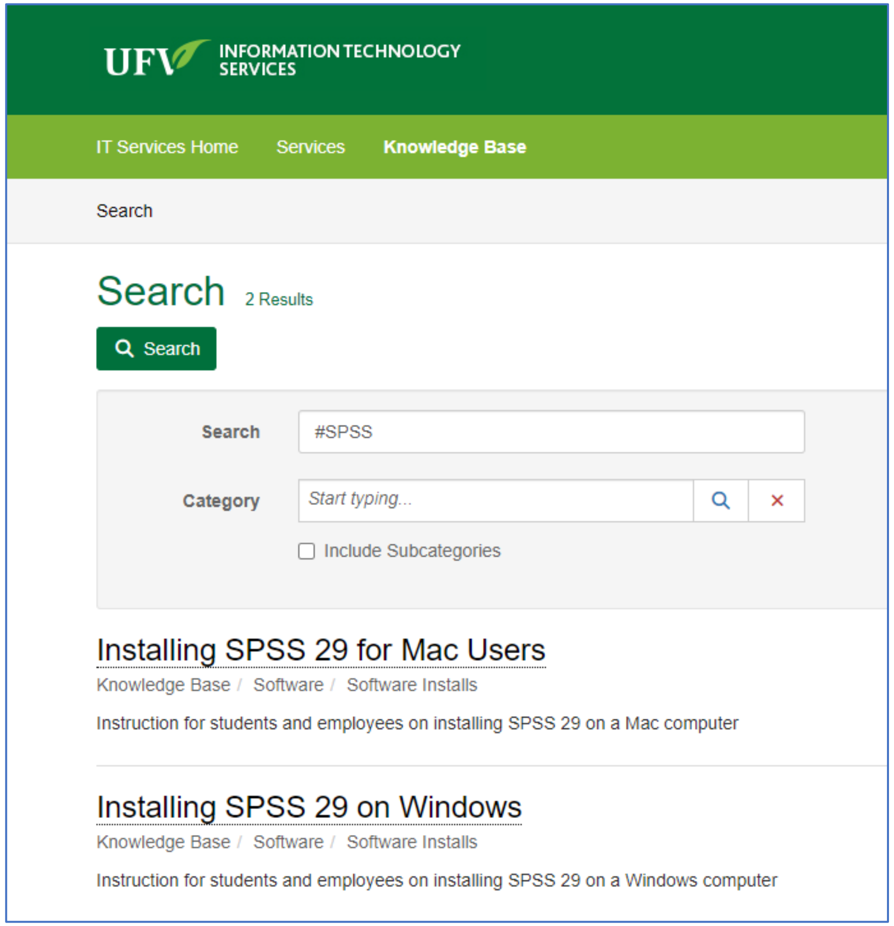




Positive Error Value: 2.45
Negative Error Value: 2.45

Installing SPSS 29 for Lab 7.2
(Or Run SPSS on school computer)

https://itservicedesk.ufv.ca/TDClient/52/ITServicesPortal/KB/Search?SearchText=%2523SPSS



Search 2 Results

Search

Search: #SPSS

Category: Start typing...

□ Include Subcategories

Installing SPSS 29 for Mac Users

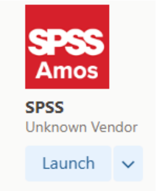
Knowledge Base / Software / Software Installs

Instruction for students and employees on installing SPSS 29 on a Mac computer

Installing SPSS 29 on Windows

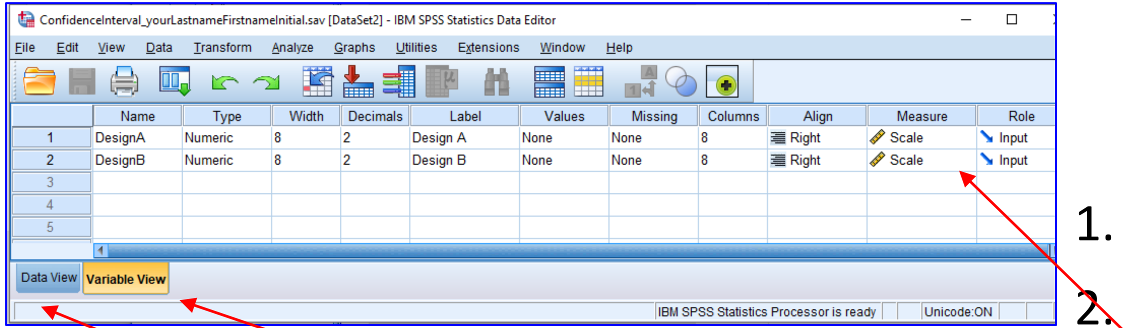
Knowledge Base / Software / Software Installs

Instruction for students and employees on installing SPSS 29 on a Windows computer

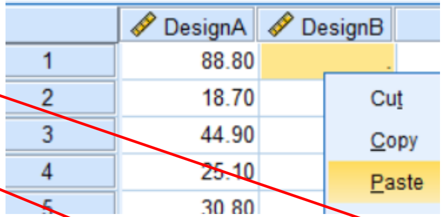


7.2 To find a confidence interval for the true (population) mean in SPSS

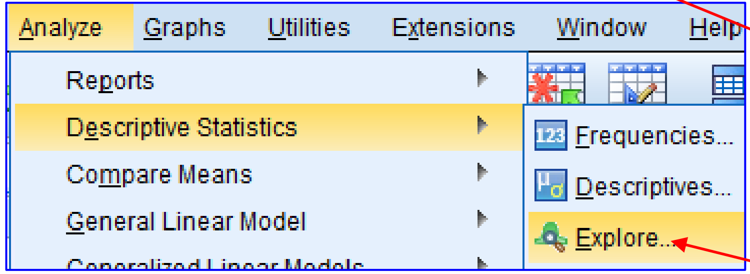
1. To access SPSS, either open it using Lab computer or go to
2. Create a new SPSS file, and name it with your last name and first name initial (**ConfidenceInterval_yourLastnameFirstnameInitial.sav**)
3. To copy data from MS Excel file, click Variable View tab, and create two variables (DesignA and DesignB). For Measure types, choose Scale option.
4. Then, copy data from two columns from Excel into Data View tab.
5. To find Mean and Error, go to Analyze > Descriptive Statistics > select Explore.
6. Using an arrow, put two Variables inside Dependent List.
7. Click Statistics > make Descriptives option checked with 95% Confidence Interval for the Mean.
8. Click Continue > OK to generate the result.

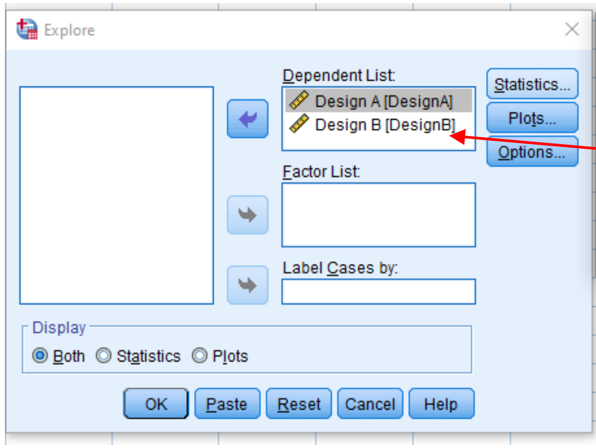

ConfidenceInterval_yourLastnameFirstnameInitial.sav [DataSet2] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

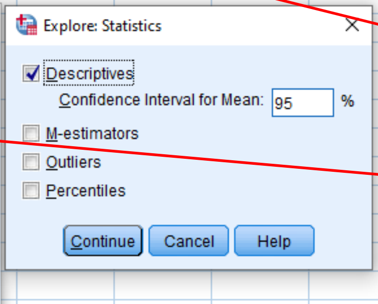
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	DesignA	Numeric	8	2	Design A	None	None	8	Right	Scale	Input
2	DesignB	Numeric	8	2	Design B	None	None	8	Right	Scale	Input
3											
4											
5											



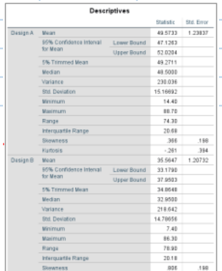
	DesignA	DesignB
1	88.80	
2	18.70	
3	44.90	
4	25.10	
5	30.80	





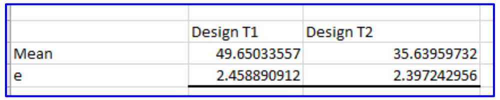


☑ Descriptives
Confidence Interval for Mean: 95 %
□ M-estimators
□ Outliers
□ Percentiles

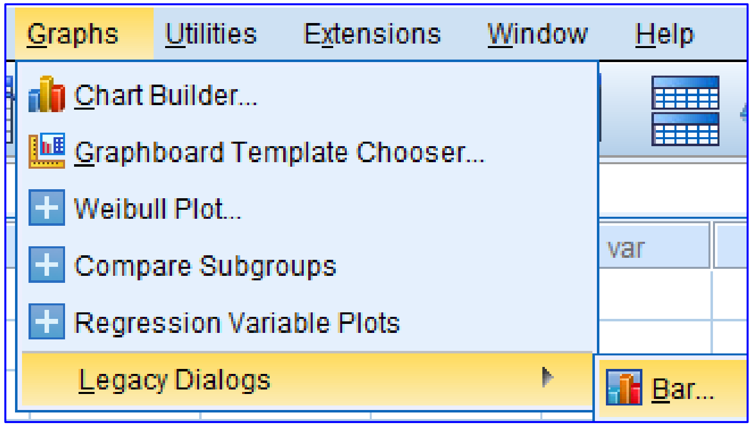


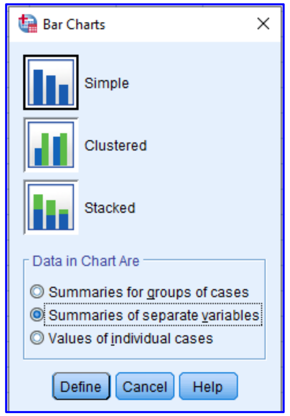
Descriptives

			Statistic	Std. Error
Design A	Mean		49.5733	1.23837
	95% Confidence Interval for Mean	Lower Bound	47.1263	
		Upper Bound	52.0204	
	5% Trimmed Mean		49.2711	
	Median		48.5000	
	Variance		230.036	
	Std. Deviation		15.16692	
	Minimum		14.40	
	Maximum		88.70	
	Range		74.30	
	Interquartile Range		20.68	
	Skewness		.366	.198
	Kurtosis		-.261	.394
Design B	Mean		35.5647	1.20732
	95% Confidence Interval for Mean	Lower Bound	33.1790	
		Upper Bound	37.9503	
	5% Trimmed Mean		34.8648	
	Median		32.9500	
	Variance		218.642	
	Std. Deviation		14.78656	
	Minimum		7.40	
	Maximum		86.30	
	Range		78.90	
	Interquartile Range		20.18	
	Skewness		.806	.198



	Design T1	Design T2
Mean	49.65033557	35.63959732
e	2.458890912	2.397242956

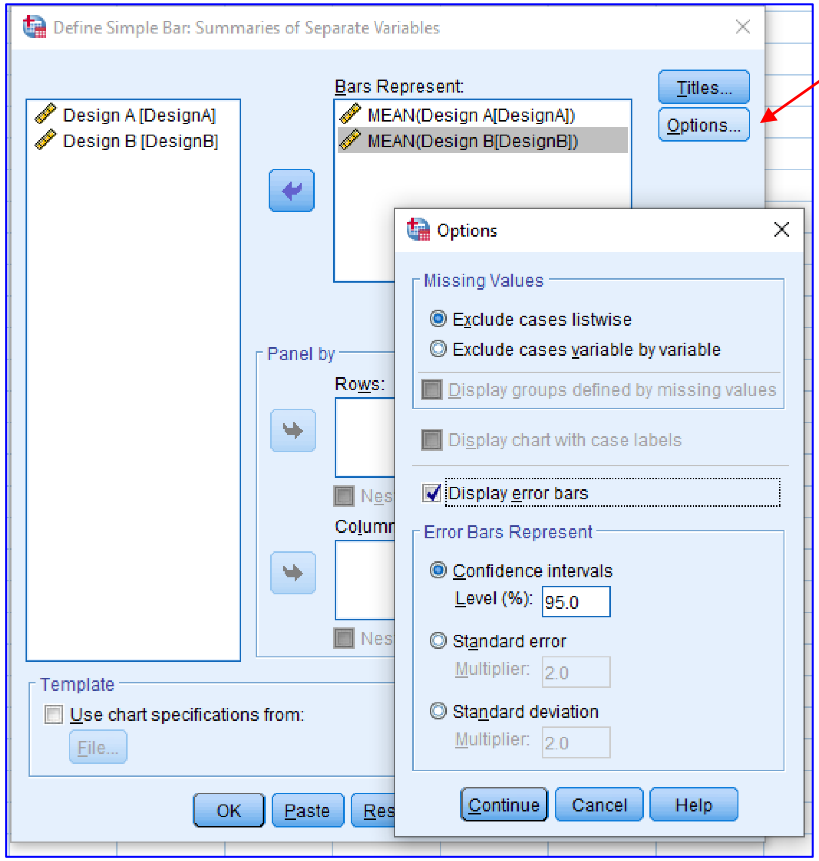



Bar Charts

- ● Simple
- Clustered
- Stacked

Data in Chart Are

- ○ Summaries for groups of cases
- ● Summaries of separate variables
- ○ Values of individual cases


Define Simple Bar: Summaries of Separate Variables

Variables: Design A [DesignA], Design B [DesignB]

Bars Represent: MEAN(Design A[DesignA]), MEAN(Design B[DesignB])

Panel by — Rows: — Columns:

□ Nest variables (no empty rows)

Template

□ Use chart specifications from:

Options

Missing Values

- ● Exclude cases listwise
- ○ Exclude cases variable by variable
- □ Display groups defined by missing values

□ Display chart with case labels

☑ Display error bars

Error Bars Represent

- ● Confidence intervals — Level (%): 95.0
- ○ Standard error — Multiplier: 2.0
- ○ Standard deviation — Multiplier: 2.0

8. To generate a bar graph, go to Graphs > Legacy Dialogs > Bar. Choose Simple and Summaries of separate variables options. Click Define.

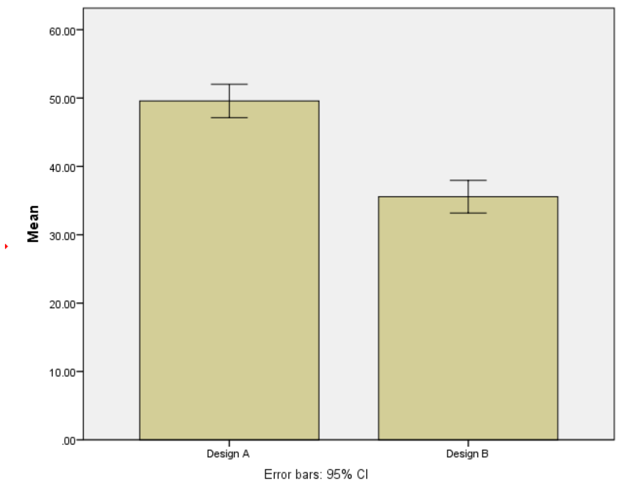
9. Select both Design and click the arrow to put it inside Bars Represent box.

10. Click Options and choose Display error bars option. Click Continue.

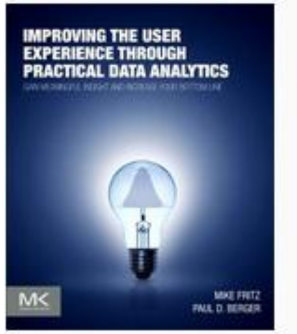
11. Press OK to generate the output. Save this output as **ConfidenceInterval_yourLastnameFirstnameInitial.spv** file format, which is different from spss data format sav.

12. In total, you have 4 files to submit (Two Excel files from 7.1 and two files for 7.2). Zip all four files to name **Lab07sp25_yourLastnameFirstnameInitial.zip**.

Graph

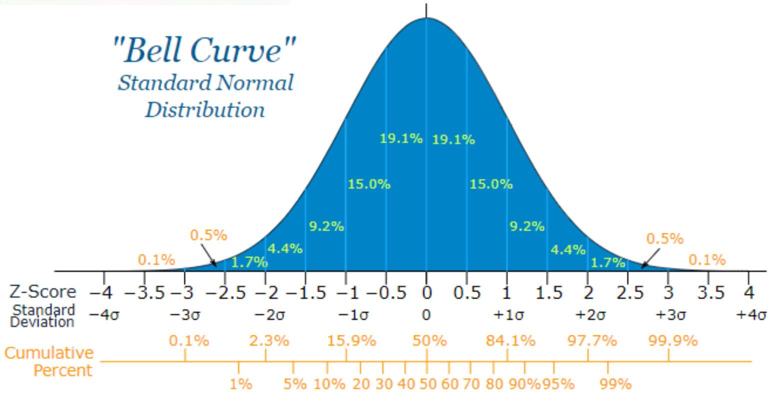


Sample output from different data set


IMPROVING THE USER EXPERIENCE THROUGH PRACTICAL DATA ANALYTICS
MK
MIKE FRITZ
PAUL D. BERGER

KEYWORDS

Bell-curve; Confidence intervals; Hypothesis testing; Normal distribution; Probability distribution of the mean



Review 'Normal Distribution' from

https://www.mathsisfun.com/data/standard-normal-distribution.html

Recognizing commonly used statistical symbols

Math Symbols List

Mathematical and scientific symbols

Basic statistical techniques.

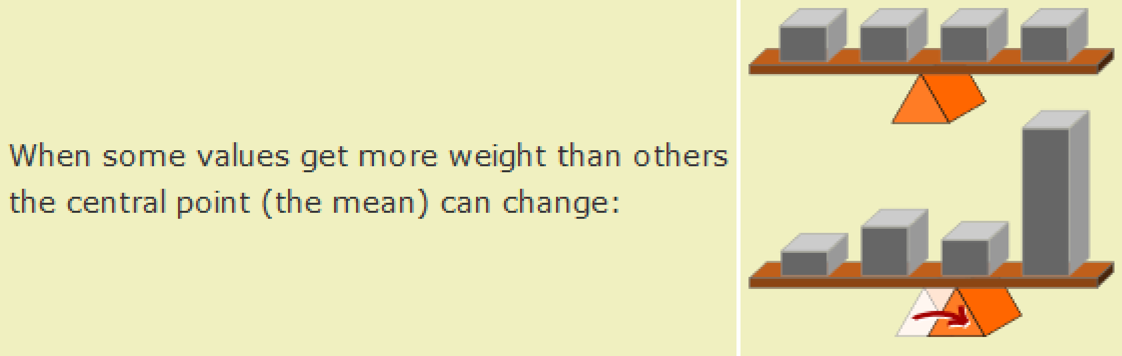
- the normal distribution, also called the bell curve, the bell-shaped curve, and the Gaussian curve.
- This probability distribution is the root of several other probability distributions, such as the t-distribution, chi-square distribution, and F-distribution.
- the probability behavior of a *mean* or *average* of many data points (e.g., satisfaction ratings).
- the basic techniques of confidence interval and hypothesis testing.

Descriptive statistics

Standard deviation is a measure of the dispersion of the dataset. Its square, s^2, is called the **sample variance**.

The **median** of a dataset is the middle value, with the same number of values above as below. If the number of values is even, then the median is the mean of the two middle scores.

The **mode** is the most frequently occurring value in the dataset. This is a more specialized statistic, applying only to datasets that have many repeating values.



When some values get more weight than others the central point (the mean) can change:

If **w** = $(w_1, w_2, \ldots, w_n)$ is a vector with the same number of components as the dataset, then we can use it to define the **weighted mean**:

$$\bar{x}_w = \frac{1}{n}\sum_{i=1}^{n} w_i x_i$$

In linear algebra, this expression is called the inner product of the two vectors, **w** and **x** = $(x_1, x_2, \ldots, x_n)$. Note that if we choose all the weights to be *1/n*, then the resulting weighted mean is just the sample mean.

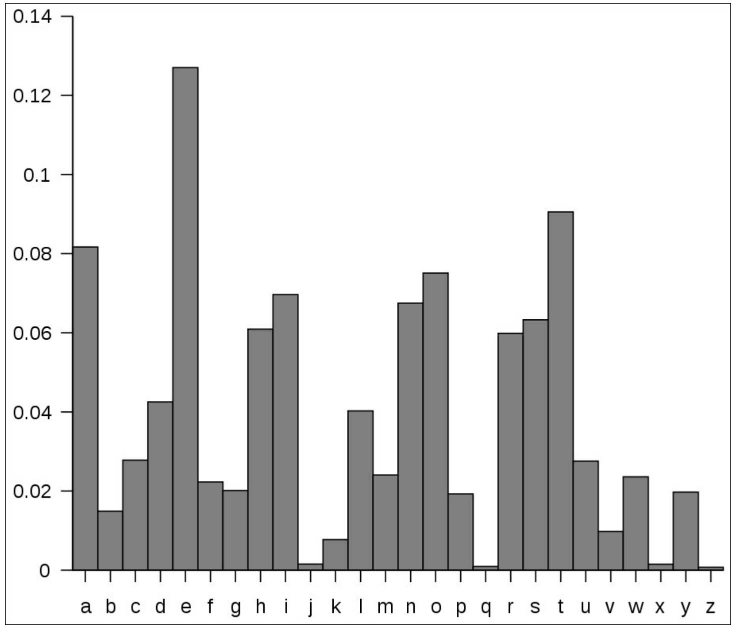


Figure 3-20. Frequency distribution of letters in English

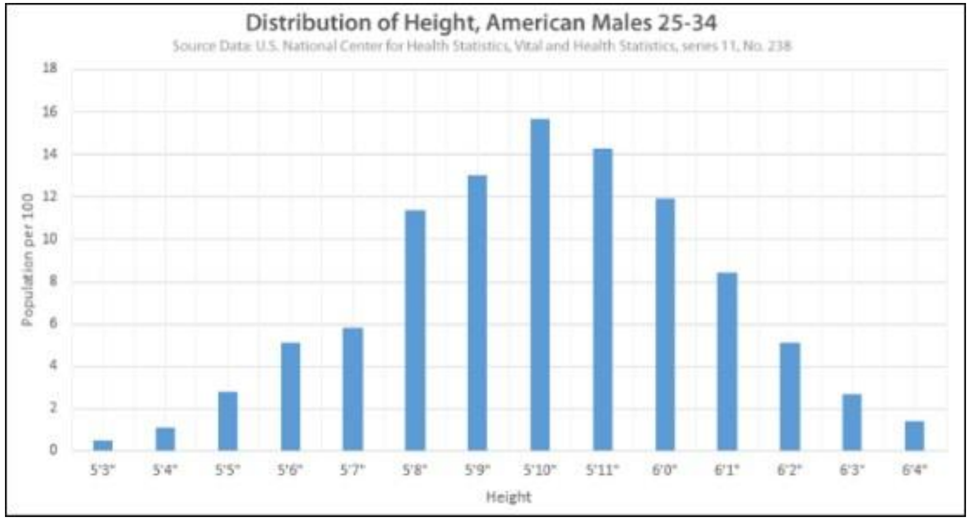
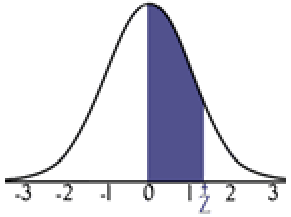


Figure 3-21. Heights of American males

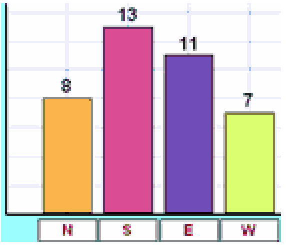
- The **sample mean** is the average x-value, and the **sample standard deviation** is a measure of how widely spread out the x-values are

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Data Index

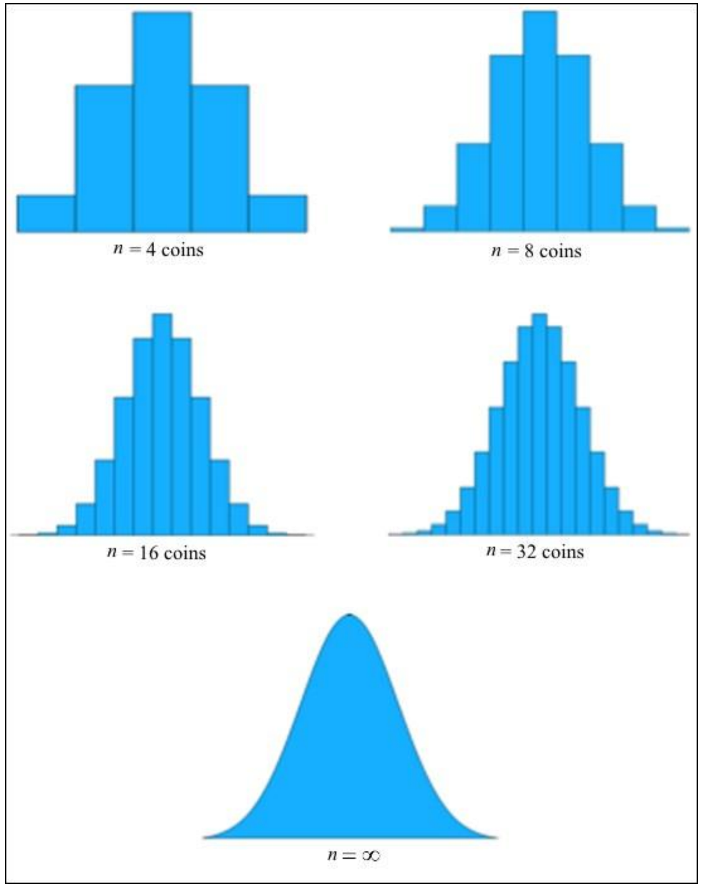


Probability and Statistics Index



Graphs Index

The normal distribution

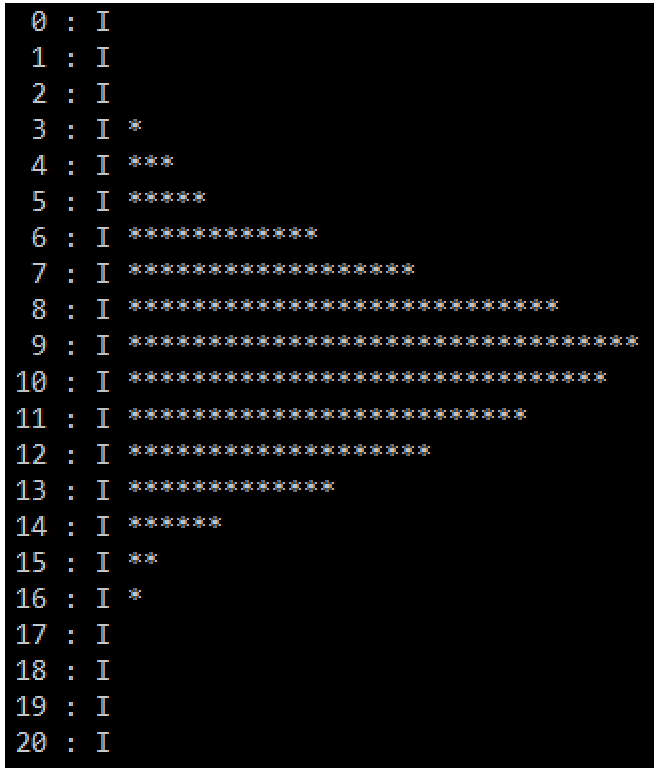


- bell shape curve or the Gaussian distribution, after its discoverer, Carl Friedrich Gauss (1777-1855)
- m is the mean and s is the standard deviation. The symbols e and are a mathematical constant: e = 2.7182818 and = 3.14159265. This function is called the probability density function for the (theoretical) distribution.

$$f(x)=\frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma}}$$

- four symbols $\bar{\mathrm{x}}$, s, m, and s. The first two are computed from the actual sample values; the second two are parameters used to define a theoretical distribution.

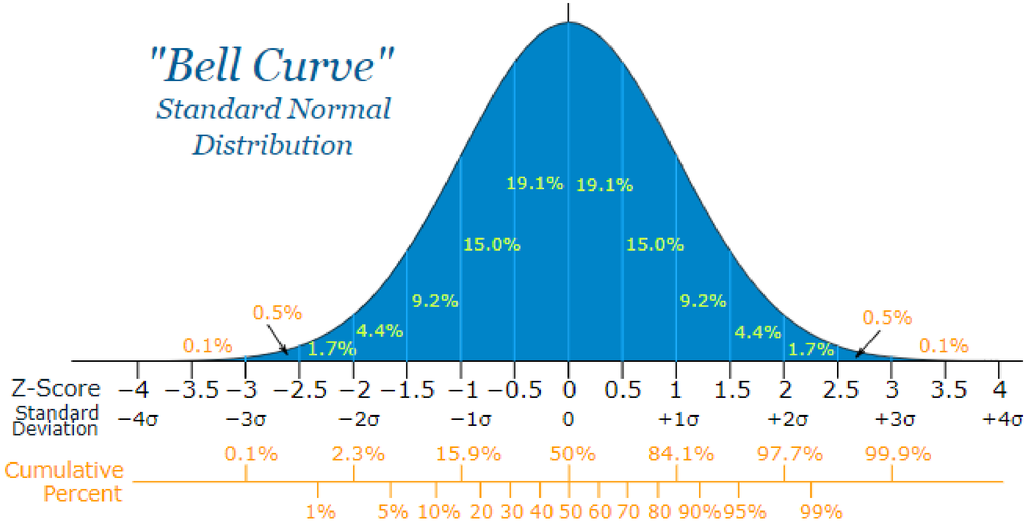
When phenomena can be attributed to a combination of many unbiased binary influences, such as 1,024 coins being flipped, their histograms tend to be normal.



```
 0 : I
 1 : I
 2 : I
 3 : I *
 4 : I ***
 5 : I *****
 6 : I ************
 7 : I ******************
 8 : I ***************************
 9 : I ********************************
10 : I ******************************
11 : I *************************
12 : I *******************
13 : I *************
14 : I ******
15 : I **
16 : I *
17 : I
18 : I
19 : I
20 : I
```

Review ‘Normal Distribution’ from

https://www.mathsisfun.com/data/standard-normal-distribution.html



$$f_N\left(x \,\middle|\, \begin{matrix}\mu \\ \sigma\end{matrix}\right) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Figure 1.3 The mathematical expression for the normal curve.

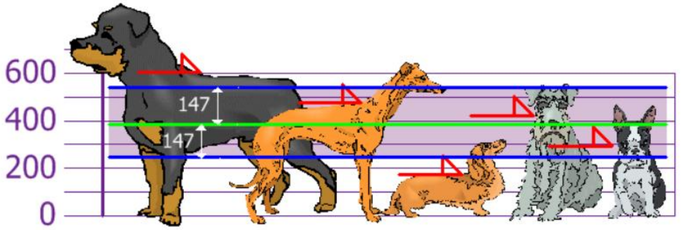
Experimental Design in Game Testing

Normal Distribution of Random Numbers

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

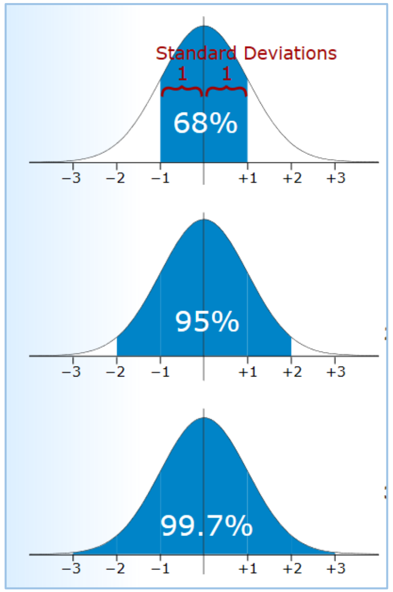
85, 82, 88, 86, 85, 93, 98, 40, 73, 83

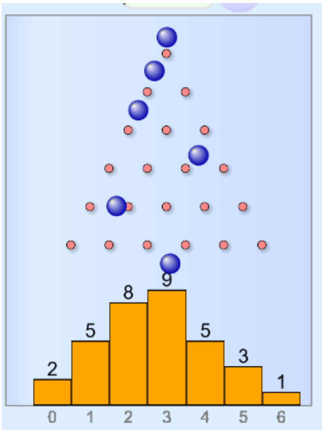
The mean is the average: 81.3



The standard deviation is calculated as the square root of the average of the squares of deviations around the mean. In other words, take the difference from the mean for each person and square it (variance). Calculate the average of all these values and take the square root as the standard deviation.

The standard deviation is the square root of the average variance:





Score	Difference from Mean	Variance
	Average Variance:	254.23
85	85-81.3 = 3.7	$(3.7)^2$ = 13.69
40	40-81.3 = -41.3	$(-41.3)^2$ = 1705.69
etc.		

Normal Curve

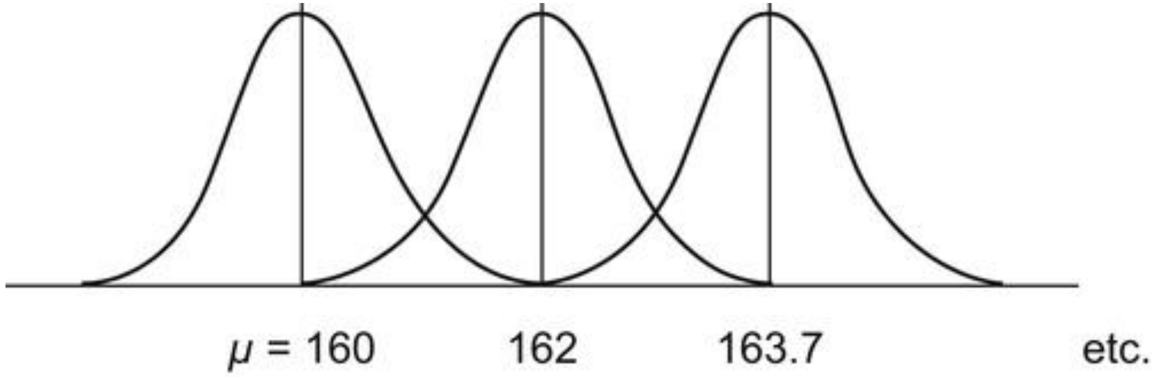


Figure 1.1 A family of normal curves.

Center point Greek letter, μ, pronounced "myu."

- The normal curve is symmetric around this center point (i.e., the left and right sides of the curve are mirror images of one another)

- *The area under the curve between any two points on the horizontal x axis equals the probability of getting a result in between these two points*. So, finding the area under a normal curve between two points is the same as finding the probability that you will get a result between two points.

- The area under the curve must always equal 1—although we don't specify the units of the "1." This means, essentially, that 100% of the data possibilities are represented under the curve.

- to keep a constant area of 1, any curve that is taller must be thinner and, if shorter, must be fatter.
  - An analogy is that a man weighs 160 pounds, but can still be either tall and thin, or short and fat, or any intermediate combination.

- if a curve is relatively taller and thinner, it represents a situation where there is relatively little variability in the result that comes out—nearly all of the area is near the center.

- if the curve is relatively shorter and fatter, it represents a situation where there is relatively more variability—more of the area is farther away from the center.
  - In responding to a Likert-scale question, larger variability would represent the fact that there is less agreement among the respondents (whether the pattern of responses was bell-shaped or not).

Flexibility of a normal curve

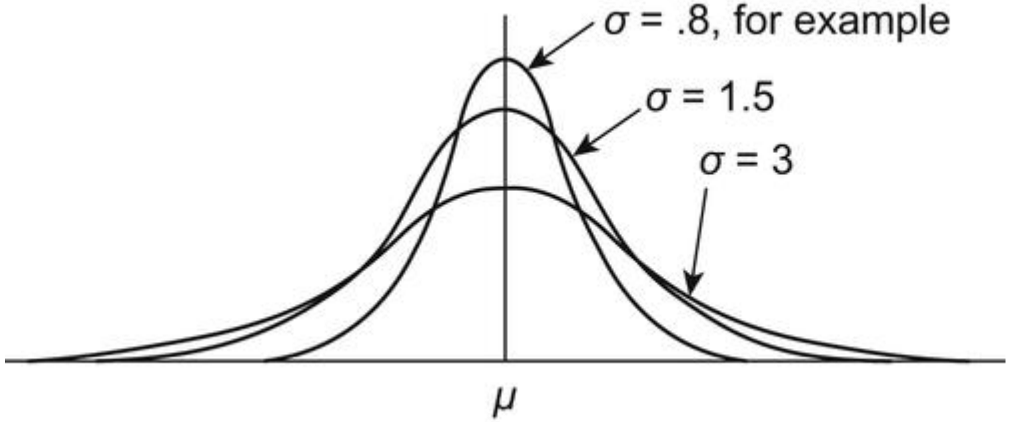


Figure 1.2 A family of normal curves with different values of σ.

Standard Deviation (Greek letter, σ, which is the "Greek s.")

- measure how tall and thin or short and fat a normal curve is by a quantity called the "standard deviation."

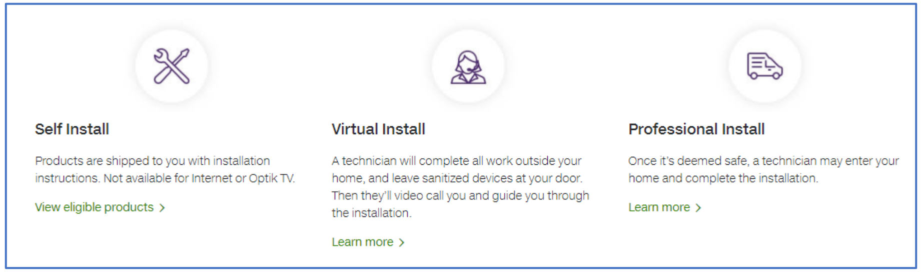
a normal curve is determined by its "μ" and its "σ," and

smaller σ ↔ taller, thinner curve
larger σ ↔ shorter, fatter curve

Center point Greek letter, μ, pronounced "myu."

1. Curve shape:
   a) if a curve is relatively taller and thinner, it represents a situation where there is relatively little variability in the result that comes out—nearly all of the area is near the center.
   b) if the curve is relatively shorter and fatter, it represents a situation where there is relatively more variability—more of the area is farther away from the center.
      - In responding to a Likert-scale question, larger variability would represent the fact that there is less agreement among the respondents (whether the pattern of responses was bell-shaped or not).
2. different phenomena have more variability than other phenomena, but the normal curve can accommodate these differences.
   - For example, the time it takes to perform certain usability test tasks will vary from person to person to a different extent from task to task. But the normal curve has the flexibility to accommodate all the different degrees of variability.

1.2.1. Finding Probabilities of Completion Times or Satisfaction Levels, or Anything Else, on a Normal Curve



Self Install

Products are shipped to you with installation instructions. Not available for Internet or Optik TV.

View eligible products ›

Virtual Install

A technician will complete all work outside your home, and leave sanitized devices at your door. Then they'll video call you and guide you through the installation.

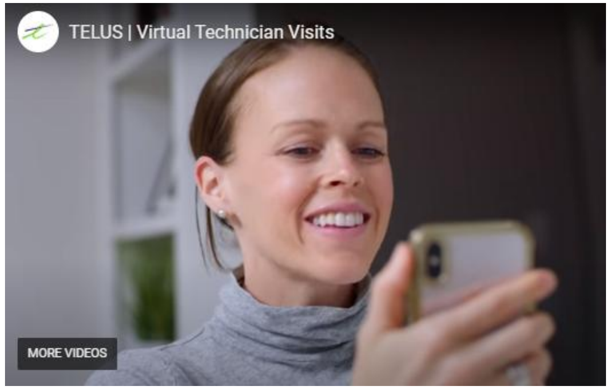
Learn more ›

Professional Install

Once it's deemed safe, a technician may enter your home and complete the installation.

Learn more ›

1.2.1.1. Vignette: how long does it take to hook up DSL Internet service?



You're the UX researcher at a gigantic telecommunications company that provides broadband service to hundreds of thousands of customers in the United States. You're part of the user experience team that is dedicated to improving the user experience for those broadband clients. One of the key projects you're working on is trying to decrease the time it takes for a typical new customer to install DSL Internet service. (In the past the company sent out technicians to complete the installation, but has recently gone to a "customer self-serve" model of sending out DSL installation kits. The cost savings have been dramatic.) Shorter installation times usually indicate higher satisfaction, and fewer calls to technical support, which means significant cost savings. During home visits to watch folks installing DSL, you take many measurements—time to download necessary software, time to activate the account, etc

But one of the most important metrics is simply the time to hook everything up. That is, plugging a phone line to the modem, plugging the modem to the computer, filtering the existing phone line, etc.

$$= \text{NORMDIST}\,(X\,,\,\mu\,,\,\sigma\,,\,1)\,,$$

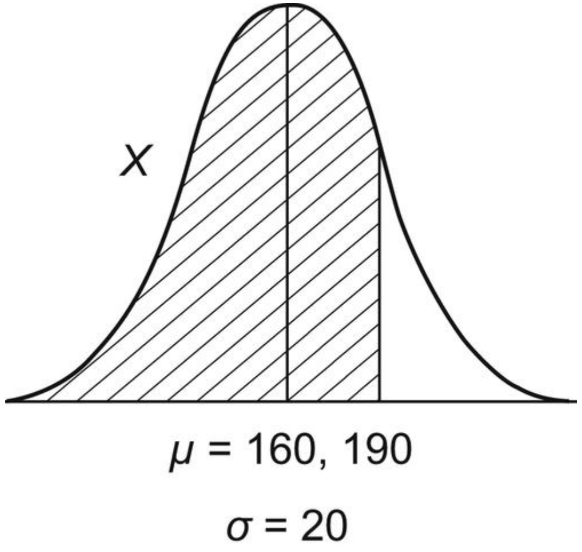


Figure 1.4 A depiction of the area under a normal curve.

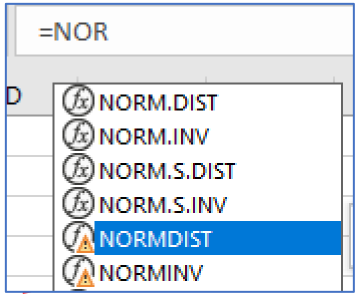
Your collected data indicates that the average time, μ, to hook everything up is 160 seconds (2 minutes, 40 seconds), and that the variability factor (which, remember, is a measure of how different the time required is from person to person), is represented by a standard deviation, σ, of 20 seconds. You know that the "hook up" time is viewed historically as around 180 seconds (3 minutes) and, to beef up the report—and anticipating the question during your presentation—you decide to calculate the percentage of the users who will require less than 190 seconds to hook everything up. OK, so we have the average task completion time value (μ) = 160 seconds with a standard deviation (σ) of 20 seconds. And we assume that we have a set of data that follows a normal curve (although not every process in the world follows a normal curve; more about the prevalence of the normal curve later), and a picture of what we want is the shaded-in area of the normal curve in Figure 1.4. The letter, X, here represents the time to hook everything up. In essence, X is a traditional statistical symbol; statisticians would refer to it as a "random variable." We would now find the shaded-in area (i.e., the probability that the time to hook everything up is less than 190 seconds)—or percentage—by using the following **Excel** command:

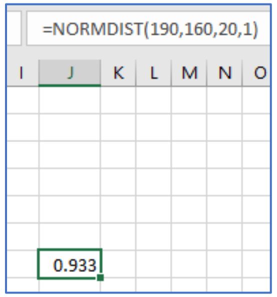
$$= \text{NORMDIST}\,(X\,,\,\mu\,,\,\sigma\,,\,1)\,, \qquad = P\,(\text{result} < [\text{specific value of}]\;X)\,.$$

$$= \text{NORMDIST}\,(190, 160, 20, 1)\,,$$

probability that the result is less than X, when the mean and standard deviation are the values they are. The "1" in the last position is simply always a "1" and is there for technical reasons that need not concern us—in a sense the "1" is simply indicating a "left-tail cumulative."

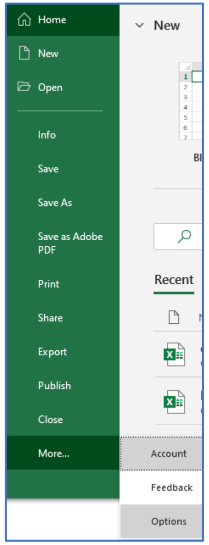
=NORMDIST(190,160,20,1)   0.933192799


=NOR


=NORMDIST(190,160,20,1)

0.933

Data analysis using Excel



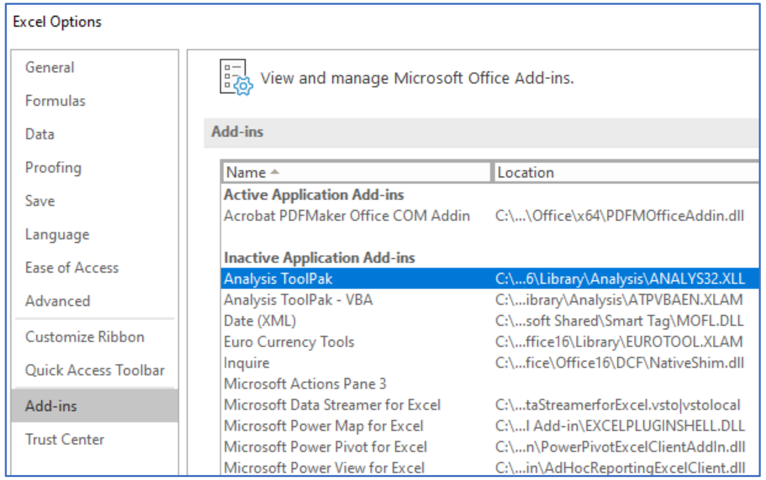
To find a sample mean of data set (say, 100 data points in the first 100 cells of column A of an Excel spreadsheet), we use the Excel command:

$$= \mathrm{AVERAGE}\,(\mathrm{A1} : \mathrm{A100})$$

finding a standard deviation

$$= \mathrm{STDEV}\,(\mathrm{A1} : \mathrm{A100})$$

- Load the Analysis ToolPak in Excel
- Create a simple formula in Excel
- Standard Deviation
- How to Calculate Standard Deviation in Excel (Step-by-Step)
- Calculating the mean and standard deviation with excel
- STDEV function
- Use the Analysis ToolPak to perform complex data analysis



View and manage Microsoft Office Add-ins.

Add-ins

Name	Location
Active Application Add-ins	
Acrobat PDFMaker Office COM Addin	C:\...\Office\x64\PDFMOfficeAddin.dll
Inactive Application Add-ins	
Analysis ToolPak	C:\...6\Library\Analysis\ANALYS32.XLL
Analysis ToolPak - VBA	C:\...ibrary\Analysis\ATPVBAEN.XLAM
Date (XML)	C:\...soft Shared\Smart Tag\MOFL.DLL
Euro Currency Tools	C:\...ffice16\Library\EUROTOOL.XLAM
Inquire	C:\...fice\Office16\DCF\NativeShim.dll
Microsoft Actions Pane 3	
Microsoft Data Streamer for Excel	C:\...taStreamerforExcel.vsto|vstolocal
Microsoft Power Map for Excel	C:\...l Add-in\EXCELPLUGINSHELL.DLL
Microsoft Power Pivot for Excel	C:\...n\PowerPivotExcelClientAddIn.dll
Microsoft Power View for Excel	C:\...in\AdHocReportingExcelClient.dll

$$= \mathrm{NORMDIST}\,(X\,,\ \mu\,,\ \sigma\,,\ 1)\,, \qquad = P\,(\mathrm{result} < \text{[specific value of]}\ X)\,.$$

=NORMDIST(190,160,20,1) ⇨ 0.933192799

- The probability is 0.9332, or, to answer the question specifically posed, 93.32% of the users will hook everything up within 190 or less seconds.
- If we wanted to find the percentage of users who would require more than 190 seconds to hook everything up, we would simply subtract 93.32% from 100%, to get 6.68%.

1.2.2. Finding Completion Times or Satisfaction Levels, or Anything Else, on a Normal Curve

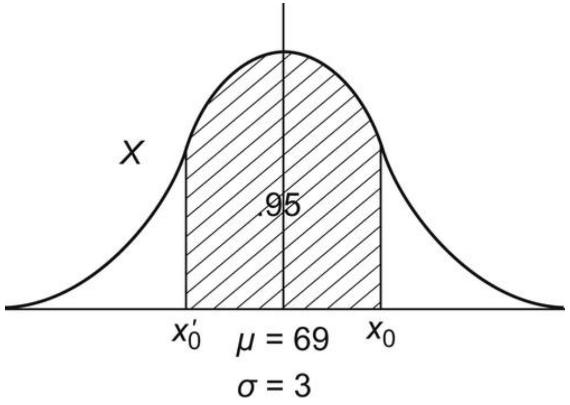
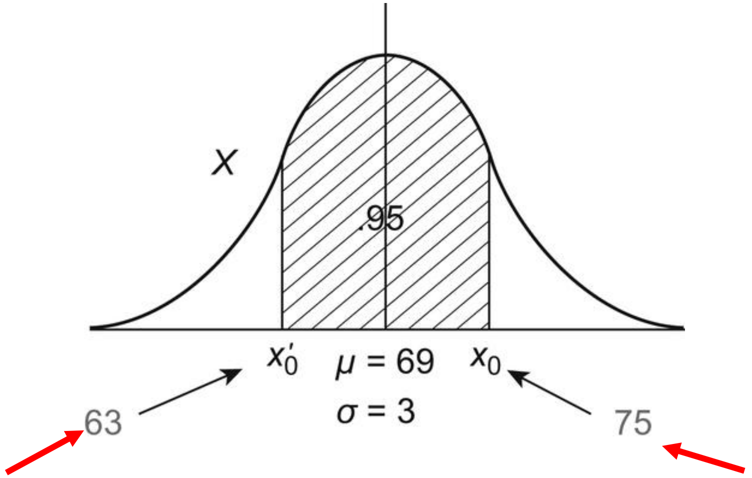


Figure 1.7 Depiction of values sought on a normal curve, given the value of the area.



$= \text{NORMINV}\,(2.5\%\,,\,69, 3)$ $= \text{NORMINV}\,(97.5\%\,,\,69, 3)$

$$= \text{NORMINV}\;(p\;,\;\mu\;,\;\sigma)$$

Figure 1.8 The range of values determined to be from 63 to 75 seconds.

- Another task: the time it takes to open up the installation kit and verify that all the necessary installation components have been included in the kit. Again, we assume that we have a normal distribution (curve) of times required to complete the task.
- The question about how long it takes most people to open the box and find all the parts, you decide to determine the range of values that 95% of the users will require (symmetric around the mean of 69 seconds) to open up the box and find everything. In essence, you want to find two values of X, x′0 and x0, along the horizontal axis (such that the area under the curve between the two values equals 0.95), as depicted in Figure 1.7.
- This command finds for us the value of X (again, here, the time to complete a task) so that "p" proportion of the time (or, with probability = p) the result will come out under X.This time, you're going to have to use the ole noggin a bit more. Due to the symmetry of the normal curve, each white (i.e., not shaded-in) corner of Figure 1.7 is 2.5% (or, 0.025). With 95% in the middle, the white must add to 5% (remember, the total area under the curve must be 1, or, 100%), and hence, each corner is 2.5%. So, the total area below x′0 is 2.5% and the total area below x0 is 97.5%. Therefore, the answer to our question can be determined by the commands:
- 95% of the user task completion times will be between about 63 seconds (=NORMINV(2.5%, 69, 3)) and 75 seconds (=NORMINV(97.5%, 69, 3)).

1.2.3. The Probability Curve for the Mean of Many Results

mean of n users' completion times

The adjustment is simply to replace the value of the standard deviation, σ, which is known to be 20, by what we can call the "adjusted standard deviation" (officially called "the standard deviation of the mean") which is the original standard deviation (i.e., for one person/time), divided by the positive square root of the sample size comprising the mean—here, the sample size equals 4. So, the adjusted standard deviation is 20 divided by the positive square root of 4, which equals, of course, 2.The resulting adjusted standard deviation is, hence, 20/2 = 10. Now we use the same Excel command we used earlier, but with a "10" instead of a "20." We repeat the earlier command for reference, and then the new, revised command:

$$= \text{NORMDIST}\,(190, 160, 20, 1) = 0.9332 \text{ or } 93.32\%$$

$$= \text{NORMDIST}\,(190, 160, \mathbf{10}, 1) = 0.9987 \text{ or } 99.87\%$$

sample mean. $\overline{X}$

The Not-So-Magic Number 30

many statistics instructors have used 30 as the “magic number”—the number you need to obtain any meaningful result about anything.In truth, for most of the probability distributions/curves that you will encounter in a UX research setting (say, using data from Likert scales), the curve of the sample mean will converge fairly quickly to what is very close to a normal-shaped curve, even for sample sizes less than 15. So, even with a sample size of 10, or even 5, you should apply the central limit theorem and assume that the probability curve for the mean, X-bar, is bell-shaped, or close enough to it for any practical purpose.

1.2.4. The Central Limit Theorem

- Central Limit Theorem: as we average a larger and larger number of sample values, the resulting sample mean, denoted “X-bar,” converges to follow a normal curve, no matter what the shape of the individual X curve before we take means.
- We do need to clarify that if the sample size is only 10, the resulting histogram for the means may deviate somewhat from a normal curve, since n = 10 is not that large a sample, even though it is what you may be dealing with as a UX researcher; i.e., usability studies are usually conducted with a sample size of 5–8. However, as a practical matter, you should assume that the mean, even for n = 10, will follow a normal distribution/curve, since the difference from a normal curve is unlikely to be material.

1.3. Confidence Intervals

In the field of data analysis, we never use the five-letter “dirty word,” guess. We never guess!! We estimate!!

A point estimate is just a fancy name for a single value as an estimate; for estimating the true mean, μ, the best point estimate (based on various criteria to define the word, “best”) is the sample mean, X-bar.

an interval estimate, which is a confidence interval. This usually amounts to providing a range of values (an interval) that has a probability of 95% of containing the true population mean, μ.

- This measure of uncertainty is best (and most often) conveyed by a confidence interval. Put simply, a confidence interval is an interval, which contains a population value, such as the population mean, with some specified probability, usually, 0.95 or 95%.
- In turn, interval estimation is one of two topics under a major heading of what we call “statistical inference.” This refers to the basic idea of being able to infer what we can about the true mean (or other true quantity, such as the true standard deviation), based on the data.
- When we “turn it around,” and assume the true (population) values are the unknowns (which is the heart of data analysis and predictive analytics), and use data to make inference about the true values, we enter the world of statistics.

reporting the sample mean, X-bar, but we also report an interval: where “e” can be (for the moment) called “error”

$$\overline{X} - e \text{ to } \overline{X} + e$$

$$P\left(\overline{X} - e < \mu < \overline{X} + e\right) = 1 - \alpha.$$

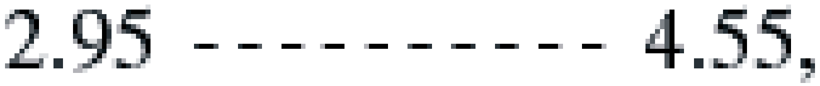
This says that the probability (“P” stands for “probability”) that the interval (image − e) to (image + e) contains the true value, μ, equals (1 − α). By tradition, industry generally uses 0.95 (i.e., 95%) for the value of (1 − α). “Alpha” is the amount of probability outside the confidence interval. So, if the confidence level (1 − α), is 0.95, then α = 0.05. We will see this “0.05” used more directly in the next section.

“If we constructed a large number of 95% confidence intervals from a large number of replicates of a given experiment, 95% of these intervals would contain the true mean.”

1.3.1. The Logic and Meaning of a Confidence Interval

- Here’s an example. Suppose that after running a usability study with 4 participants, we have these 4 data points on a 5-point (1–5) Likert-scale question for the satisfaction of a new design: 2, 3, 4, 3. The sample mean, X-bar, is, of course, (2 + 3 + 4 + 3)/4 = 3.0. As can be determined (we’ll see how, very soon), we get a 95% confidence interval for the true mean of 1.70–4.30. The interval has 3.0 as its center; it is pretty wide (in practice, we always prefer a narrower confidence interval, since it means we have homed in more closely to the true mean), but that is primarily because we have sampled only four satisfaction values, and there is a fair amount of variability in the four results.
- We have 95% confidence that the true mean satisfaction of all people who have experienced the design is between 1.70 and 4.30.

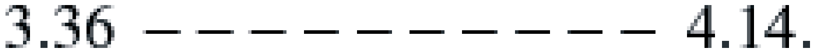
Now, if the data had been 3, 4, 4, 4 (still, only four data points, but less variable from one to another!!), the 95% confidence interval would be much narrower (which we like better!), and would be,



2.95 ----------- 4.55,

centered around the X-bar value of 3.75.

If the data consisted of eight data points: 3, 4, 4, 4, 3, 4, 4, 4 (purposely having the last four data points duplicating the first four data points, for comparison's sake), the X-bar still is 3.75, but the 95% confidence interval is now



3.36 ———————————— 4.14.

Now, this is an interval that is less than half the width as when there were four data points, meaning more than twice as accurate! Getting better all the time! (Thanks Lennon and McCartney.) You can see that, ceteris paribus, the confidence interval gets narrower (more accurate) as the sample size increases.The formula for finding confidence intervals depends on three different things: (1) the sample size, (2) the variability among the data values, and (3) your chosen value of (1-α), or level of confidence, usually 0.95 by convention, and its corresponding value on a normal curve.

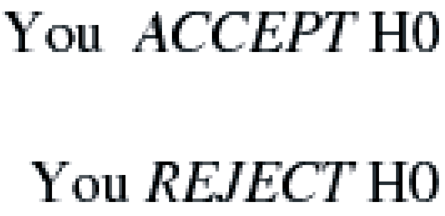
1.4. Hypothesis Testing:
Deciding whether something is true or not, based on analyzing the data. Prove it!

- An example may help to illustrate the concept of hypothesis testing. Suppose that you're working as a UX researcher at a company where the CEO wants to completely change the design of the company's Web home page. Let's suspend reality a moment and agree that we know, for the current Web home page, that the mean satisfaction rating is 4.10 on a scale of 1–5, where 1 = Not At All Satisfied to 5 = Extremely Satisfied.

- Recently, the design team has come up with a new home page design. So far, everyone seems to like it, but you want some empirical evidence that the new design is indeed an improvement over the current design in terms of mean satisfaction. You decide to run a 25 person survey, and probe on satisfaction of the new design. You **calculate the new satisfaction rating mean (the X-bar) for the new design from your sample of users and get** a 4.15.
- So, you need to decide—is an X-bar of 4.15 really enough above 4.10 to constitute convincing evidence that the new design has a true higher mean satisfaction? If 4.15 is not enough above 4.10 to be convincing, is 4.20 or 4.30?

- The answer depends on the sample size, the variability in the data, and how much chance you are willing to take when stating that the new design has a higher mean satisfaction rating when in truth, it doesn't! This "risk" has a formal name of "significance level," always denoted by "alpha," and is traditionally chosen to be 0.05–1 chance out of 20.

- Our goal is to choose between two hypotheses. Their formal names are the null hypothesis and the alternate hypothesis. Simply put, the null hypothesis typically states that there is no relationship between one or more things, or that the status quo has not changed.
- the null hypothesis provides a benchmark against which we can propose another hypothesis that indeed a change has taken place, or indeed, something is now different: an "alternate" hypothesis. In our case, our null hypothesis is that the true mean satisfaction of the new design is no higher than the 4.1 true mean of the current home page design. Conversely, the alternate hypothesis is that the new design has a higher mean rating than the 4.1 mean of the current design (and, hence, we should go with the new design!) The "μ" in the hypotheses below stands for the true mean of the new design
  - H0: $\mu \leq 4.1$ (mean satisfaction rating of new design is no higher than 4.1, the mean of the current design; stay with the current design!)
  - H1: $\mu > 4.1$ (mean satisfaction rating of new design is higher than 4.1, the mean of the current design; go with a smile with the new design!!) And, you phrase your conclusion by deciding whether



You *ACCEPT* H0

You *REJECT* H0

From a practical point of view, the two hypotheses should be such that one of them must be true, but both of them cannot be true.

1.4.1. p-Value

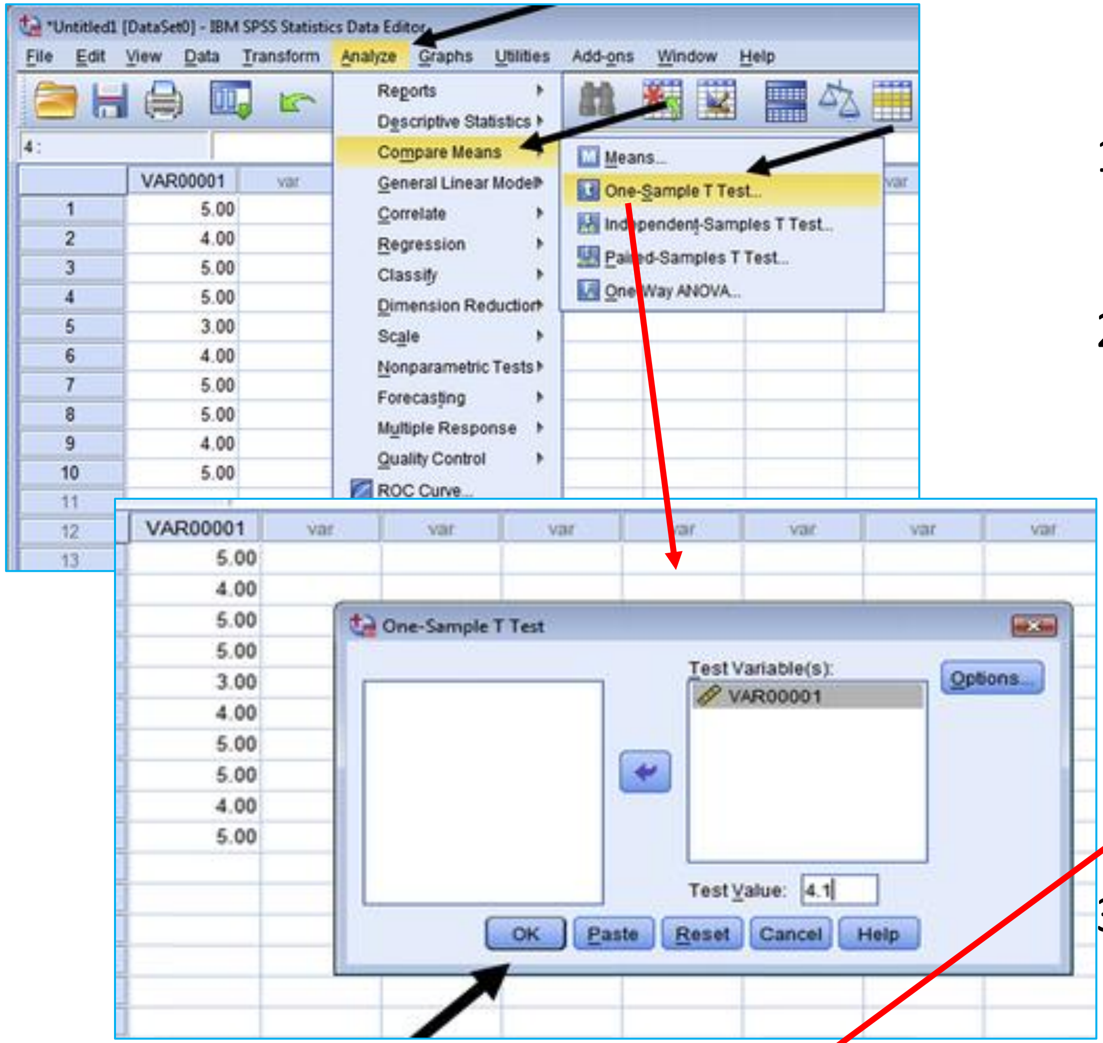
The p-value, in simple terms, is the probability that you would obtain a specific data result as far away or farther away from what H0 says, assuming that H0 is true. The logic is kind of like a "proof by contradiction"

If the data results you obtain have a very small chance of occurring under the assumption that H0 is true, then, indeed, we conclude that H0 is NOT true!!

Keep in mind that the term "accept H0" is really the same as "Insufficient reason to reject H0."

- Traditional cutoff point for "very small chance of occurring" is 0.05.
- if any event occurred only 5 out of 100 times, that's a 5% probability. Pretty small indeed. So, the p-value is a measure of the strength of your evidence against H0.

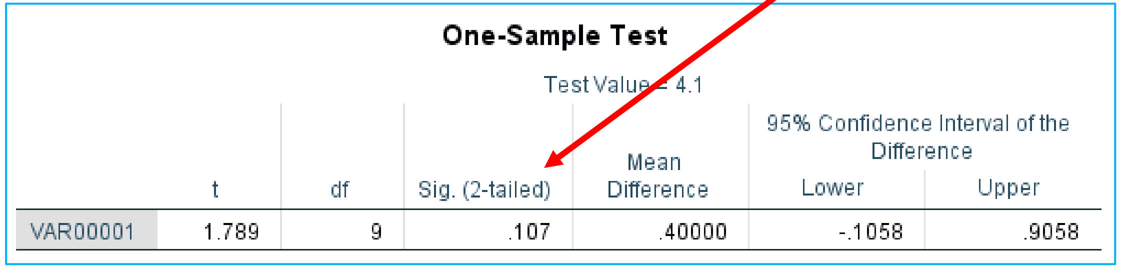
Getting p-Value using SPSS



	VAR00001
1	5.00
2	4.00
3	5.00
4	5.00
5	3.00
6	4.00
7	5.00
8	5.00
9	4.00
10	5.00

Test Variable(s): VAR00001

Test Value: 4.1



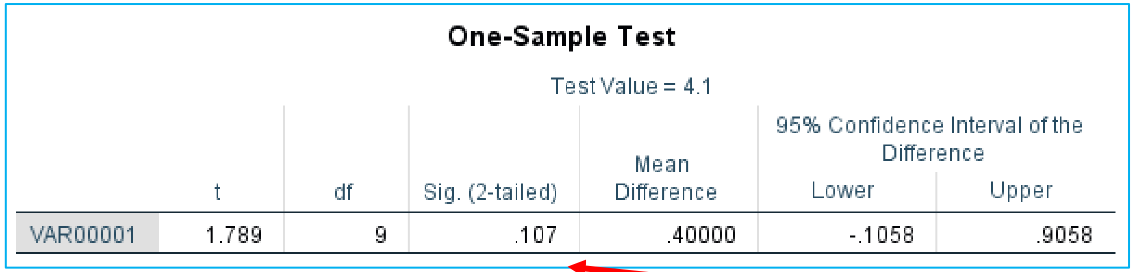
One-Sample Test

Test Value = 4.1

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
VAR00001	1.789	9	.107	.40000	-.1058	.9058

1. Open spss file (**ch01HypotheisTesting_p37.sav**) from the lecture note.
2. Now we pull down the “Analyze” menu item (see arrow in Figure 1.20) and go “compare means” and then to One-Sample T-test. On the resulting dialog box, drag over VAR00001 to become the Test Variable (see vertical arrow in Figure 1.22) and change “Test Value” from 0 to 4.10, the value stated in H0.
3. The p-value is thus given as 0.107. But, we do have a slight complication here. You will note that the “Sig.” (i.e., p-value) is labeled “two-tailed.” But our hypothesis test is one-tailed!

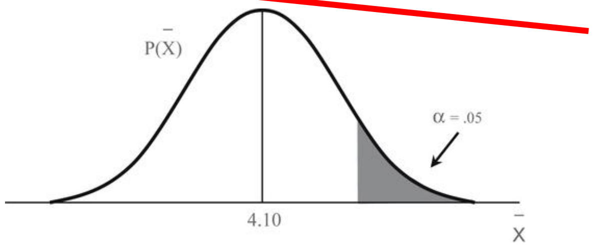
How Many Tails?

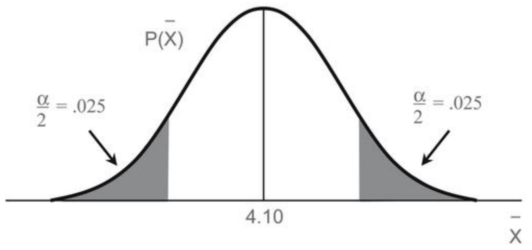


One-Sample Test

	Test Value = 4.1					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
VAR00001	1.789	9	.107	.40000	-.1058	.9058

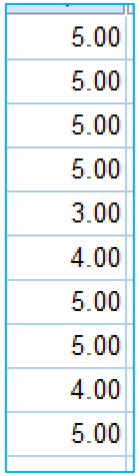
A one-tail test has a p-value that is half of the p-value of the corresponding two-tail test.



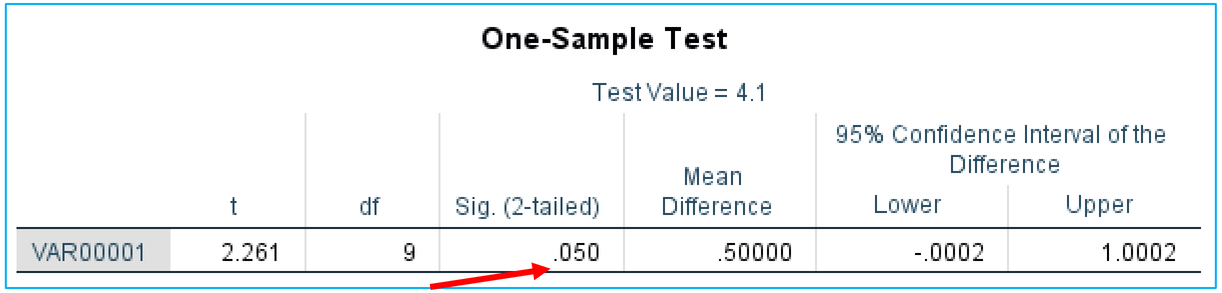


So, the relevant p-value for you to consider is half of 0.107, which equals 0.0535, which is a small amount above 0.05, telling us that, if H0 is true, the data we have (with its mean = 4.50, with n = 10, and the variability in the data) is not (by a tad!!) as rare an event as required to reject H0: μ ≤ 4.1, when using 0.05 as a cutoff point. Thus we can say that we do not have sufficient proof that the true mean satisfaction rating of the new design exceeds 4.10.

- When the issue is a “not over” versus “over,” which is the case here (H0: true average of new design not over 4.10, versus H1: true average of new design is over 4.10), or, less frequent in practice, a “not under” versus “under” form of comparison, we refer to the hypothesis test as “one-tailed,” because the “rejection values,” based on common sense, are in only ONE tail of the curve—here, the tail some amount above 4.10.
- Conversely, when we are testing whether a true mean equals a specific value, versus is not equal to that specific value (e.g., H0: μ = 4.1 versus H1: μ ≠ 4.1), that is called a two-tailed test, since we would reject H0 if X-bar is too much in the direction of either of the TWO tails.

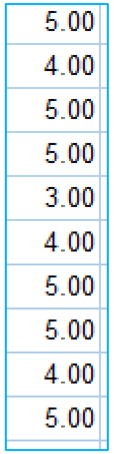


5.00
5.00
5.00
5.00
3.00
4.00
5.00
5.00
4.00
5.00

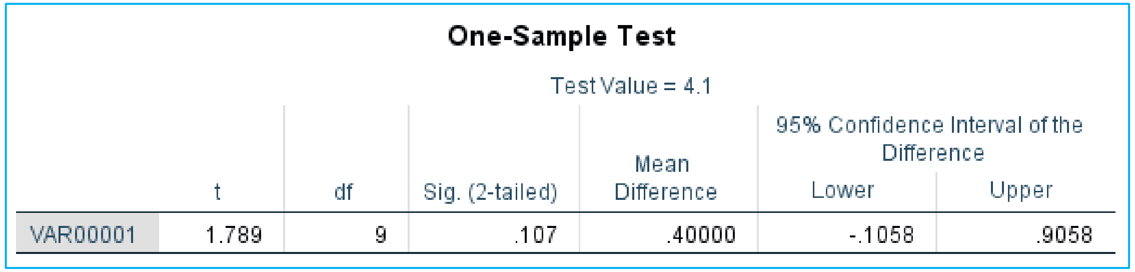


One-Sample Test

	Test Value = 4.1					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
VAR00001	2.261	9	.050	.50000	-.0002	1.0002



5.00
4.00
5.00
5.00
3.00
4.00
5.00
5.00
4.00
5.00



One-Sample Test

	Test Value = 4.1					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
VAR00001	1.789	9	.107	.40000	-.1058	.9058

- Consider a slightly different scenario that yields different results. If we go back to the original data and change one of the 4’s to a 5 (say, the second data value) getting data of the p-value for you to consider is 0.025 (half of the 0.050), and now, with your X-bar = 4.60, we would conclude that we do have sufficient evidence to reject H0, and conclude that the new design does, indeed, have a true mean satisfaction rating above 4.10, the known mean satisfaction rating of the current design.

2.1 INTRODUCTION

- As the researcher on a UX team, you are probably frequently asked to determine which version of a design or feature is more useful or usable (in essence, better).

- In addition, you maybe asked to determine which design is preferable on a variety of attributes, such as aesthetic appeal, trust, emotional connection, and of course, commercial potential.

Why compare different versions?

- Lots of reasons. If you're at an agency with outside clients, the client will often ask to see a "couple of different passes."

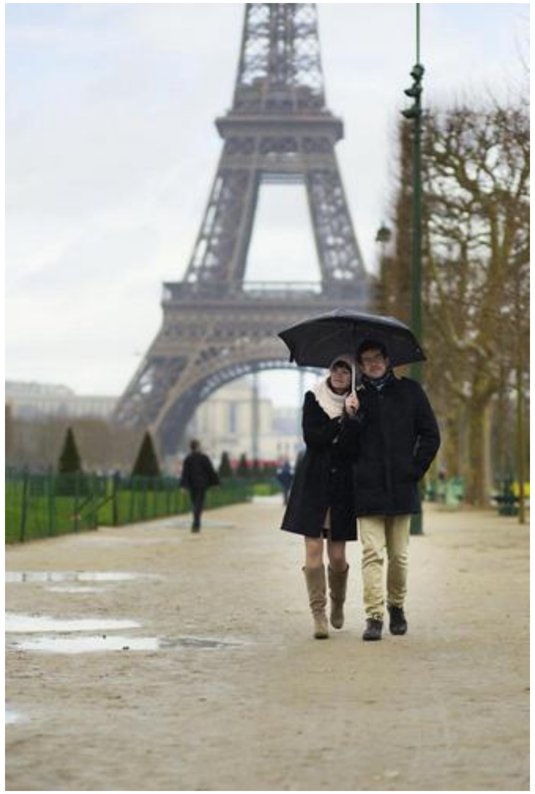
- If you're at a company with its own UX team, you may still have "clients" whom you're still beholden to—they're just internal, like the Chief Executive Officer or VP of Marketing. And they want to see a "couple of takes" as well.

- And for good reason. Different versions mean that the team has explored different approaches to improve a user's perception of usefulness and usability, and that they're not satisfied with the first approach that someone comes up with. The team wants to satisfy both the clients and themselves that they've left no stone unturned (i.e., considered all alternatives) in their pursuit of the most intuitive design.

2.2 CASE STUDY: COMPARING DESIGNS AT MADEMOISELLE LA LA

- You've been hired as the UX researcher at Mademoiselle La La, a high-end online apparel merchant aimed at urbane women from 18–55 years of age with well-above-average disposable income.
- Within a couple of days of your start date, you're able to determine the source of the nervous energy that permeates the Creative Group: the rumor mill has it that CEO Massimo Ranieri thinks the home page needs a complete revamp to increase the "sophistication" factor of the brand—and he wants it fast.

Creative Director Kristen McCarthey calls a team meeting to give out the marching orders. Thin, short-coiffed, over-caffeinated, with a penchant for body piercings, McCarthey has been with "La La" for over 5 years, and has been terrorizing her information architects, designers, researchers, and front-end coders for just as long.



Creative Director Kristen McCarthey says, "our brand is about sophistication, but we're appealing to bargain basement shoppers in our design!" She glares at the group, as her shrill voice bounces off the exposed brick: "I'm giving you lightweights one more chance with this new website design. Make it sophisticated, make it hip—and make it pop. “Grabbing her iPad and *latte*, she heads out the door, but screams one last time: "And I want it by the end of the week!"

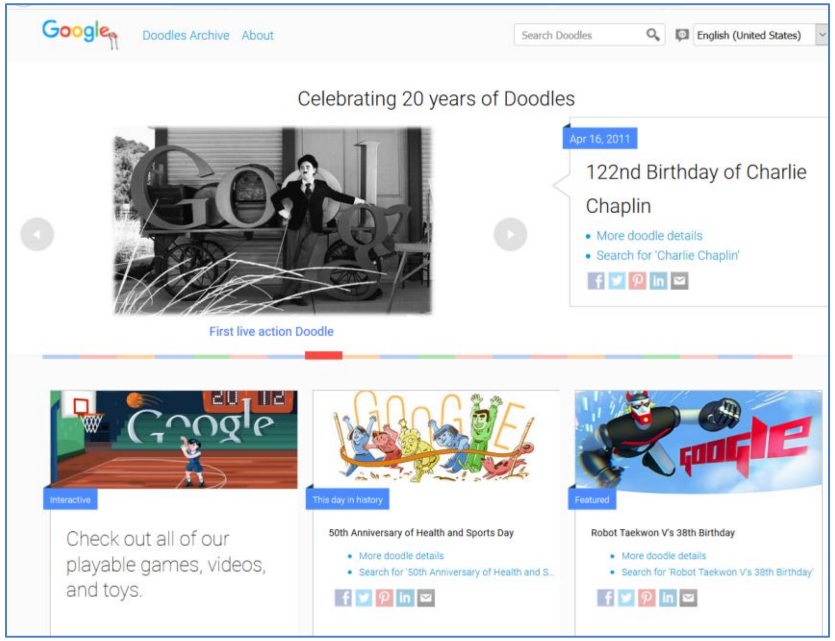
The team has its marching orders, and within a couple of days, the designers come back with two designs:

Branding?

[These are the world's most valuable brands]

[The 10 Most Valuable Brands in 2018]

[How to Calculate Your Brand's Equity]



Celebrating 20 years of Doodles

Apr 16, 2011

122nd Birthday of Charlie Chaplin

- More doodle details
- Search for 'Charlie Chaplin'

First live action Doodle

Interactive

Check out all of our playable games, videos, and toys.

This day in history

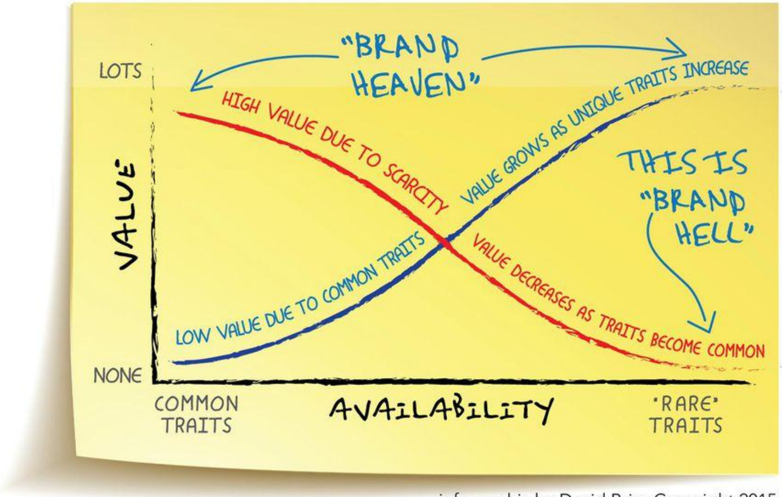
50th Anniversary of Health and Sports Day

- More doodle details
- Search for '50th Anniversary of Health and S...

Featured

Robot Taekwon V's 38th Birthday

- More doodle details
- Search for 'Robot Taekwon V's 38th Birthday'

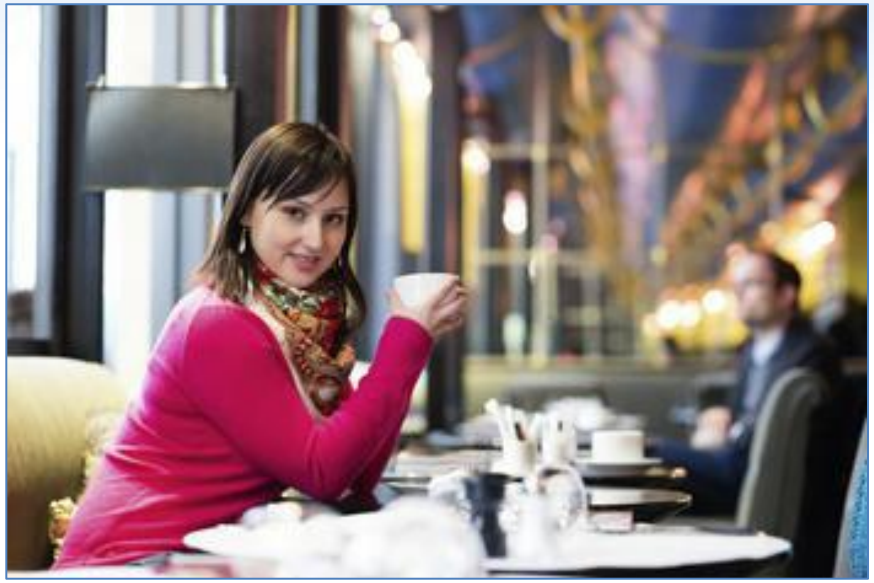


"BRAND HEAVEN"
LOTS
NONE
VALUE
HIGH VALUE DUE TO SCARCITY
VALUE GROWS AS UNIQUE TRAITS INCREASE
THIS IS "BRAND HELL"
LOW VALUE DUE TO COMMON TRAITS
VALUE DECREASES AS TRAITS BECOME COMMON
COMMON TRAITS
AVAILABILITY
'RARE' TRAITS

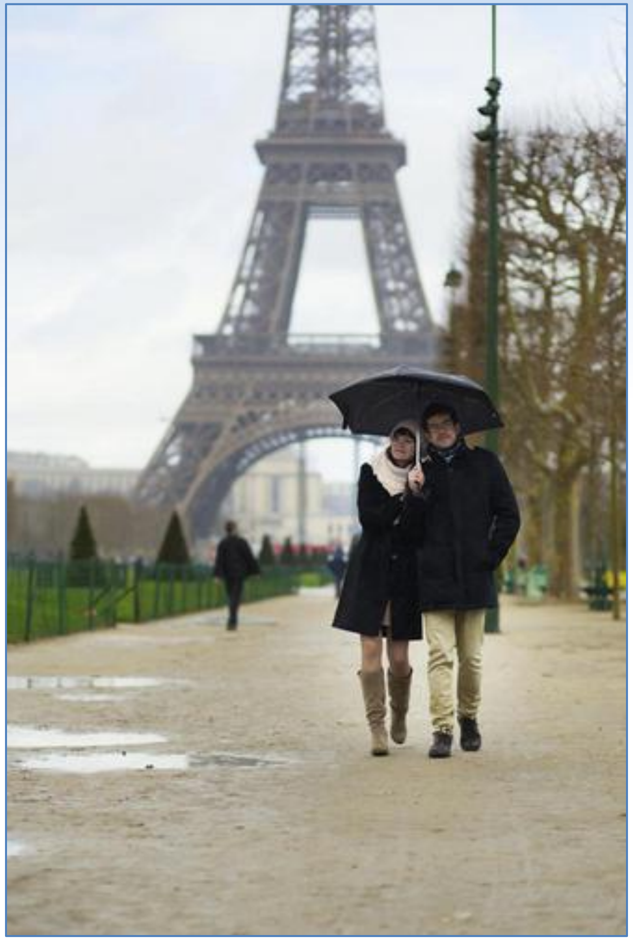
infographic by David Brier Copyright 2015

Brand Value

Which design do you prefer?
Which design do your customers prefer?
Can you differentiate?

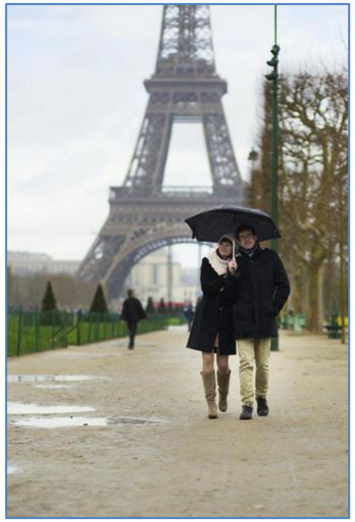


Design 1 in Figure 2.1 features a scene of a young *demoiselle* sipping a coffee at an outdoor French café, ignoring the adoring eyes of a nearby young man.



Design 2 in Figure 2.2 features a young couple snuggling together under one umbrella during a shower, with the Eiffel Tower bisecting a slate-gray sky. One version is aloof, mysterious, unresolved, but there's a promise of romance. The other version is resolved, with true love trumping the odds.

Invariably, some folks like Design 1 and other folks like Design 2, and they're passionate about their preferences. During multiple reviews, the advocates for each design become more strident and dig in. Consensus is elusive, if not impossible.



During what is supposed to be the "final review," tempers rise, and the two camps are getting increasingly agitated, extolling the sophistication of their design over the other.

There are other subplots, because designer Josh Cheysak (Design 1) is hoping his design is chosen to increase his chances of a raise, and designer Autumn Taylor (Designer 2) is hoping her design is chosen because she's bucking for the Creative Director position. Nobody will budge. Think Boehner and Obama during the Great Government Shutdown of 2013.

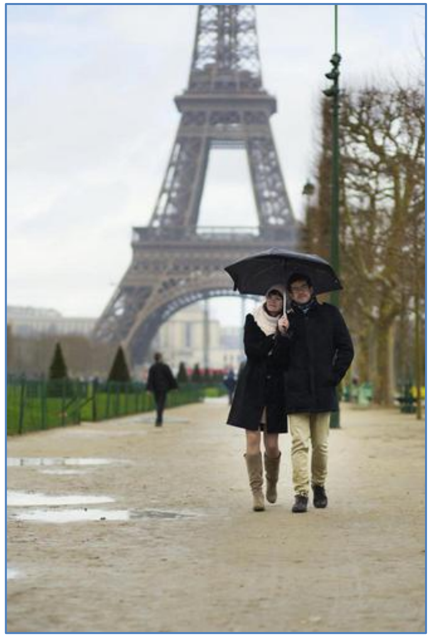
Just as Cheysak is about to pour his Red Bull all over Taylor's Moleskine sketchbook, you cautiously offer to perform a "head-to-head comparison survey with independent samples."

You calmly explain that by running a survey featuring two different designs with two different groups of people, you may be able to determine differences between the two designs in terms of perceived sophistication and preference ratings. In other words, you can determine which one is best, based on user feedback. Trying not to sound too professorial, you add that "proper statistical analysis and good survey design can guard against obtaining misleading results."

- The attendees simultaneously nod in agreement but McCarthey won't agree to anything without "pushback.“
- "Where are you getting these participants?" McCarthey asks with skepticism in her voice.
- "We can use our client list, and offer them a $25 gift certificate for completing a survey. After all, they are the target audience, but they haven't seen any one of these two designs. Of course, we'll collect demographic data about each participant."
- "Will everyone see both designs?" McCarthey asks. "Doesn't sound right."
- "In this case, it's probably better that one group sees one design and another group sees the other design. This eliminates any bias from seeing a second design after seeing the first. I can randomize which participant sees which design."
- "Hmmm," McCarthey says, staring down at her iPhone, half-listening. "I guess it's worth a try.“ Ah, a peaceful resolution is in sight. Before you can say "Camp David Peace Accords," the tension in the conference room evaporates. No more bruised egos, no more subterfuge. Just an easy way to let data do the talkin’ and to determine the correct design based on objective evidence, not hunches and egos.

2.3 COMPARING TWO MEANS

We considered whether to believe that the mean satisfaction with a modified design was higher than the mean satisfaction with the old design, which was 4.1 on a scale of 1 (not at all satisfied) to 5 (extremely satisfied). Formally, we were testing two hypotheses, the null hypothesis (H0) and the alternate hypothesis (H1):



•H0: The true mean satisfaction with the modified design is no higher than 4.1.

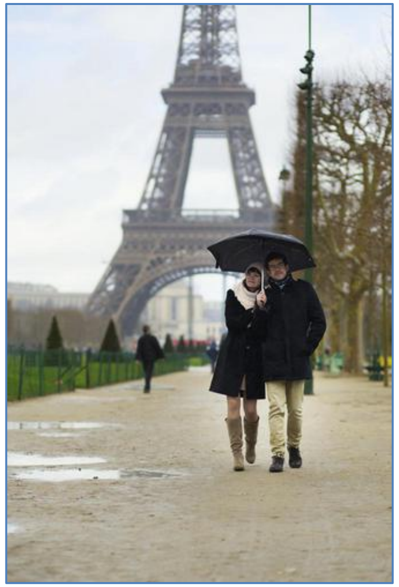
•H1: The true mean satisfaction with the modified design is greater than 4.1, that of the original design.

the true mean satisfaction with the modified design as "mu"
(the Greek letter, μ)

•H0: $\mu \leq 4.1$ (the modified design mean satisfaction is not above that of the original design—weep!!)
•H1: $\mu > 4.1$ (indeed, the modified design mean satisfaction is above 4.1; there has been an increase in mean satisfaction—yeah!!)

2.4 INDEPENDENT SAMPLES:

Two-sample t-test (two unknown means that we wish to compare, based on two sets of data)



In the case of Mademoiselle La La, you have to compare the means of two independent samples, and determine the one, if either, that has the higher perception of sophistication. To do so, we will create two hypotheses in an effort to decide between the two:

H0: μ1 = μ2

(The two designs *do not differ with respect to mean sophistication)*

H1: μ1 ≠ μ2

(The two designs do, indeed, differ with respect to mean sophistication)

If the second hypothesis is determined to be true, then we can conclude that the means are indeed different. As a consequence, we can conclude that the mean perceptions of satisfactions are different. Thus, the design with the higher ratings of satisfaction is the winner—and is the one to use for the site's home page.

Two-sample t-test for independent samples

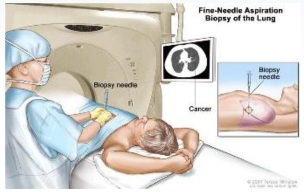
- *there are two different groups of people involved, each evaluating one of the two designs.*

- Having two different groups of people is the reason this approach is called "independent samples;" no one person is in both groups, and the mean score for one group of people is totally *independent* from (i.e., unrelated to) the mean score for the other group.

WHY INDEPENDENT GROUPS?

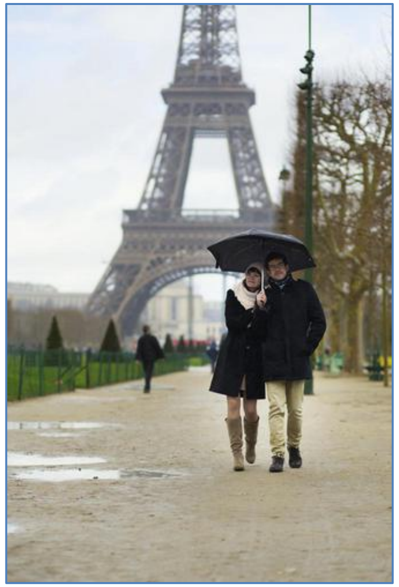
WHY INDEPENDENT GROUPS?

- Usually two reasons:
  - (1) One person cannot be a member of both groups—for example, the perception of sophistication garnered from women of 18–25 years of age versus the perception of sophistication of women ranged 26–32 years of age. In that case, there would likely be *one design* that each group would experience (profiling/target user factor).
  - *(2) Compare two designs with similar groups of people, and it is not appropriate for the same person to experience both designs*.
    - Sometimes you'll want to eliminate the "learning curve" from having experienced the first design affecting the person, so he/she cannot give an objective evaluation of the second design.
    - In a medical experiment, you have two different medicines, both of which might be a safe and effective cure for a medical problem. Although there are rare exceptions (crossover designs—two or more treatments are applied to the same subject; there is the advantage of needing fewer subjects, but the disadvantage is that there may be a carryover effect from the first treatment administered to the second treatment administered), you would not give the same person each of the two medications to determine which is more effective, since, after a person has taken one medication, his/her medical condition has likely changed.
    - if the two designs are like the two medications—in this case, it is the "perspective" or "experience" that has changed and would not allow an independent evaluation to be made of a second design experienced by the same person.



2.5 MADEMOISELLE LA LA REDUX(revive)

You're ready to jump into action as the UX researcher to determine the winner through a t-test with independent samples. Now, there are lots of ways to collect this data, but probably the most economical and most efficient way to collect the data is using online surveys.



An online survey is one of the easiest ways to collect attitudinal data from your target audience. Typically, surveys contain some combination of open-ended comments, binary yes/no responses, Likert-type rating scales, Net Promoter scales, and more.

In the case of Mademoiselle La La, you'll probably want to construct two different surveys. Each is identical except for the design shown. One will collect basic demographic data (gender, age, income, etc.) and then go on to reveal a new design.

Net Promoter Score definition

The survey will then probe on several different variables: organization of the page, aesthetics, whether the page evokes attributes (in this case, sophistication), and then rate agreement with some kind of bottom line question, like "This home page is sophisticated."

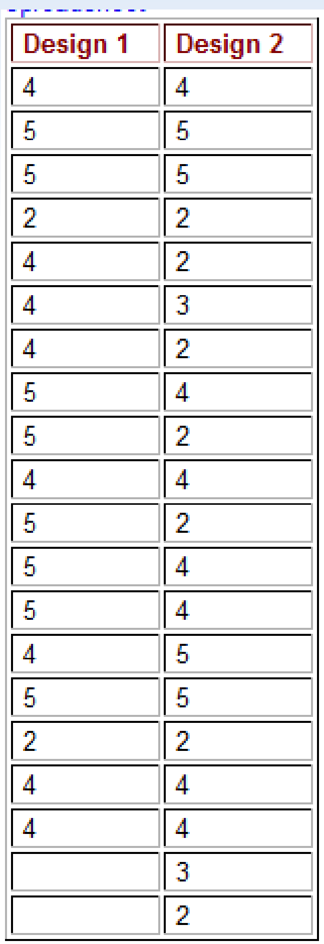
YOU DON'T NEED THE SAME NUMBER OF PARTICIPANTS IN EACH GROUP

- it's important to note that we do *NOT* need to have the same number of people evaluating each design, although if one were allocating people to each design, it would be most efficient to split the folks in half, or as close as you can get to a 50/50 split.

LOW SAMPLE SIZE MAY MEAN LOW POWER http://www.gpower.hhu.de/en.html

We can use the same hypothesis test framework, and fix the significance level at whatever you wish (usually, 0.05), regardless of the sample size; this controls the probability of rejecting H0 when it is true (called a "type 1 error")—one way an incorrect conclusion can be reached. However, there is another way of reaching an incorrect conclusion, and that is to accept H0, when, indeed, it is false. This is called a "type 2 error." The probability of this happening is one we usually do not control, and cannot even compute unless we pick some very specific H1 scenario that would typically be arbitrary (after all, if we don't even know for sure if the means are equal or unequal, it is very unlikely we would ever know exactly how unequal they are if they are unequal!!). The probability of this type of accepting H0 when it is false decreases as the sample size increases (ceteris paribus). The complement of this probability of incorrect conclusion, which would be the *probability of rejecting H0 when it is false* (a good thing!), is called the *power* of the (hypothesis) test. With a small sample size, the power of the test is often smaller than you might like (and correspondingly, the probability of accepting H0 when it is false, the type 2 error probability, is higher than one might like it to be). Of course, it is difficult to quantify this notion of power when we cannot sensibly determine the actual probabilities. Nevertheless, we repeat the key point of the sidebar. If the sample size is small, it is possible that, although you can still control (say, at 0.05) the probability of rejecting H0 when it is true, you may be conducting a hypothesis test with a low amount of power.

Survey design & Data Analysis



Design 1	Design 2
4	4
5	5
5	5
2	2
4	2
4	3
4	2
5	4
5	2
4	4
5	2
5	4
5	4
4	5
5	5
2	2
4	4
4	4
	3
	2

You check your latest survey results and see you have a sample size of 18. That is, 18 people have rated Design 1 on a 1–5 Likert scale, where you ask respondents to rate agreement or disagreement with statements like: "This is a sophisticated design," and "1" represents "strongly disagree," and "5" represents "strongly agree." (The 5-point scale is: Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree). You also have collected a sample size of 20 data points on Design 2; i.e., these 20 people saw Design 2 and responded to the same statement and using the same scale.

The (sample) means for the ratings of the two designs (columns of Table 2.1) are 4.22 and 3.40, respectively.

38 different people in this study. We are testing the hypotheses:

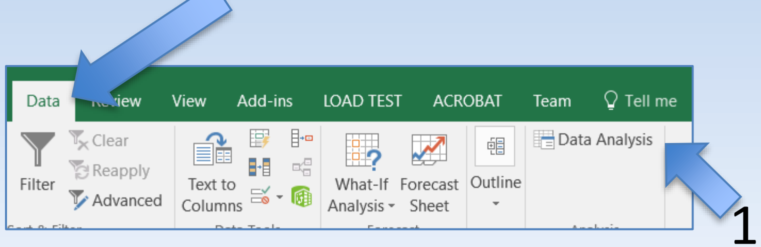
$$H0: \mu 1 = \mu 2$$
$$H1: \mu 1 \neq \mu 2$$

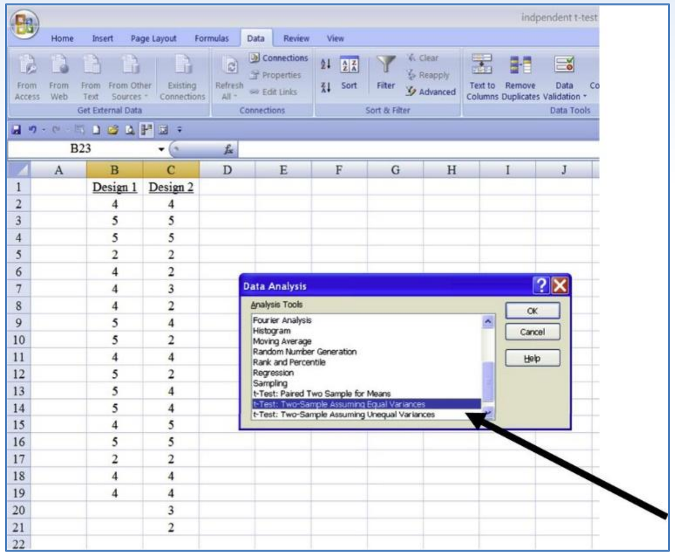
where μ1 = the true mean of the sophistication level with Design 1, and μ2 = the true mean of the sophistication level with Design 2.

Before we do the analysis, we are not sure whether the difference in the sample means (4.22 − 3.40 = 0.82) is indicative of a real difference in the true means if we could magically collect data from *all* clients of Mademoiselle La La (current and future!)

We're left with the question of whether the difference of 0.82, given the variability among the values within each group/column and the sample sizes of 18 and 20, is a large enough difference to believe in H1—that there is a real difference!! This is exactly what an independent samples t-test will determine.

ch02_t-test_base.xlsx



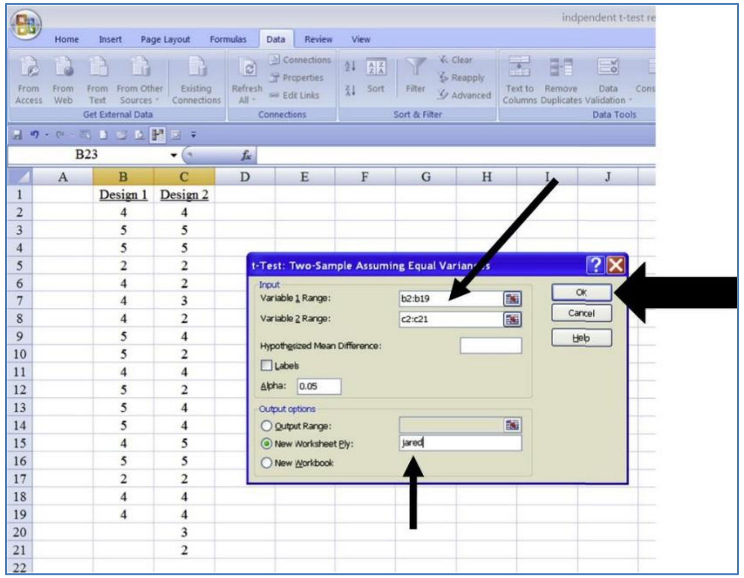


	A	B	C	D	E	F	G	H	I	J
1		Design 1	Design 2							
2		4	4							
3		5	5							
4		5	5							
5		2	2							
6		4	2							
7		4	3							
8		4	2							
9		5	4							
10		5	2							
11		4	4							
12		5	2							
13		5	4							
14		5	4							
15		4	5							
16		5	5							
17		2	2							
18		4	4							
19		4	4							
20			3							
21			2							
22										

Data Analysis

Analysis Tools

- Fourier Analysis
- Histogram
- Moving Average
- Random Number Generation
- Rank and Percentile
- Regression
- Sampling
- t-Test: Paired Two Sample for Means
- t-Test: Two-Sample Assuming Equal Variances
- t-Test: Two-Sample Assuming Unequal Variances



	A	B	C	D	E	F	G	H	I	J
1		Design 1	Design 2							
2		4	4							
3		5	5							
4		5	5							
5		2	2							
6		4	2							
7		4	3							
8		4	2							
9		5	4							
10		5	2							
11		4	4							
12		5	2							
13		5	4							
14		5	4							
15		4	5							
16		5	5							
17		2	2							
18		4	4							
19		4	4							
20			3							
21			2							
22										

t-Test: Two-Sample Assuming Equal Varian…s

Input

Variable 1 Range: b2:b19

Variable 2 Range: c2:c21

Hypothesized Mean Difference:

□ Labels

Alpha: 0.05

Output options

○ Output Range:

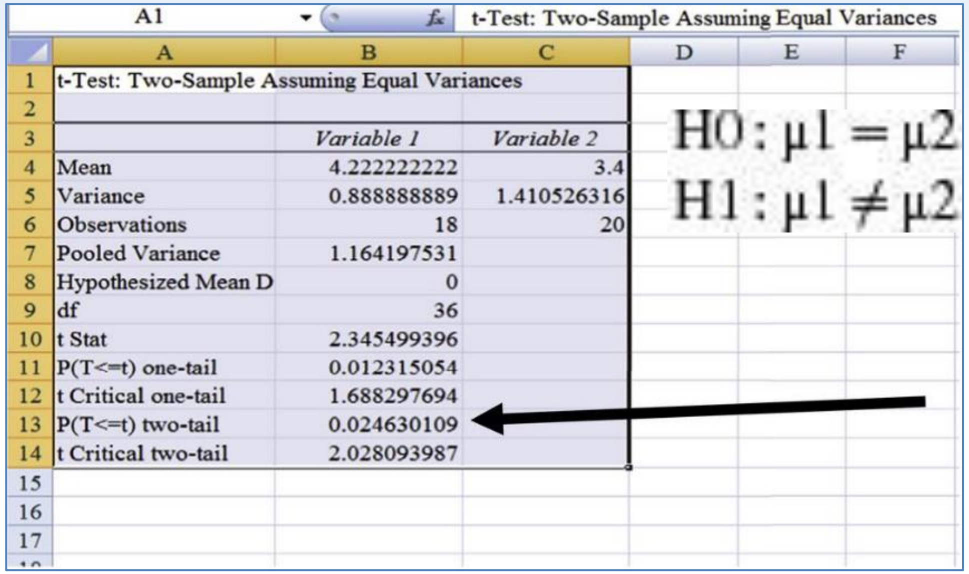
● New Worksheet Ply: jared

○ New Workbook

1. First, you go to the "data" tab on the Excel ribbon and click on "Data Analysis.
2. Then highlight "t-Test: Two-Sample Assuming Equal Variances."
3. Then, you click on the highlighted command and consider the dialog box that comes up. Fill in the location of each "variable"—i.e., tell Excel where each column of data you wish to use is located.
4. You can see that the first design's data ("variable 1") is b2–b19, while the data from the second design ("variable 2") is located in c2–c21.
5. The output be put on a new worksheet (i.e., page!!). Finally, you click on "OK" in the upper right corner

Result:

P-value (link 1, link 2,)



A1 t-Test: Two-Sample Assuming Equal Variances

	A	B	C
1	t-Test: Two-Sample Assuming Equal Variances		
2			
3		*Variable 1*	*Variable 2*
4	Mean	4.222222222	3.4
5	Variance	0.888888889	1.410526316
6	Observations	18	20
7	Pooled Variance	1.164197531	
8	Hypothesized Mean D	0	
9	df	36	
10	t Stat	2.345499396	
11	P(T<=t) one-tail	0.012315054	
12	t Critical one-tail	1.688297694	
13	P(T<=t) two-tail	0.024630109	
14	t Critical two-tail	2.028093987	

$$H0: \mu 1 = \mu 2$$

$$H1: \mu 1 \neq \mu 2$$

Excel calls the p-value by "P(T < = t) two-tail." See arrow in Figure 2.5. And, we see that the value is 0.0246. This is below 0.05, the traditional standard for significance, so we reject H0 and go with H1, and thus conclude that there is sufficient evidence to be convinced that the true mean sophistication ratings of the two designs are different.

Choosing which statistical test to use

Hypothesis tests, p-value

2.6 BUT WHAT IF WE CONCLUDE THAT THE MEANS AREN'T DIFFERENT?

What if we did not find a significant result (i.e., a significant difference between the means for the two designs) and thus, the data did not indicate a clear winner between the two designs? And what if increasing the sample sizes with another survey still produced no significant difference?

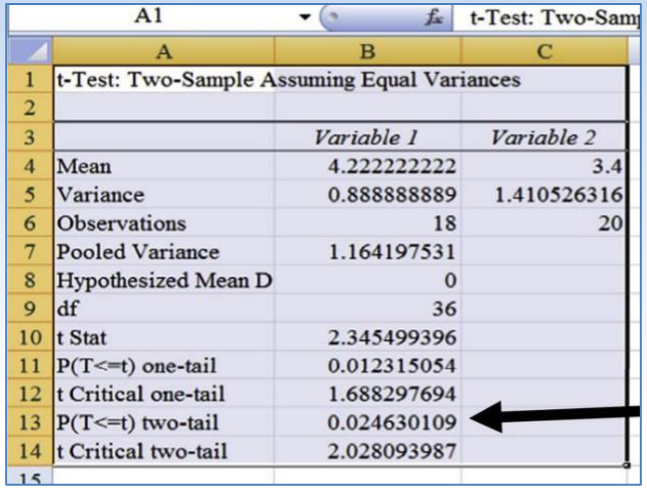
If there is no clear winner, there's no need to run for the hills with your tail between your legs. Invariably, other factors will help your team make a decision.

- For example: one design may be **cheaper** to implement because it contains original artwork, whereas the other design relies on expensive licensed photography.
- Or one design is more expensive because it will mean 2 weeks of Flash programming work, while the other design requires none.
- Perhaps you use Design 2 (the couple under the umbrella) because you're launching the new home page in the fall, and you switch to Design 1 (the café scene) come spring.

SIDEBAR: CAN I USE A HIGHER ALPHA?

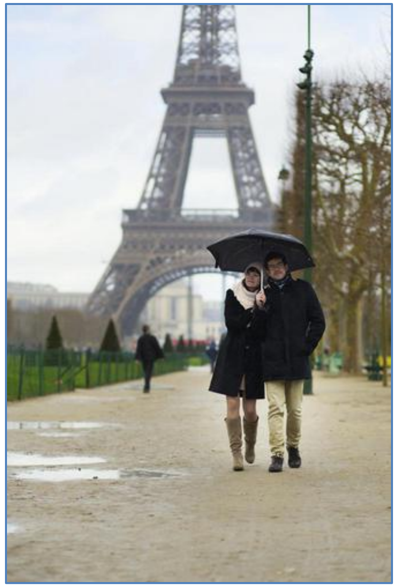
- If the p-value is, for example, 0.04, this says that assuming H0 is true, the probability of finding the result we found from the data, or a result further away from what H0 says (at the equal to point), is 0.04. We typically set a benchmark of 0.05, and if the probability expressed by the p-value is under 0.05, we call it "significant," and reject H0.
- However, what if the *p*-value is 0.20? This would not be called a significant result. Still, it does indicate that if H0 is true, the aforementioned probability is 0.20; some might view this as indicating that there is an 80% chance that H0 is false, and further, view that as a reason to "bet on" H1. This reasoning is faulty for some subtle reasons way beyond the scope of the text, that have to do with Bayesian statistics and prior probabilities.

2.7 FINAL OUTCOME AT MADEMOISELLE LA LA



	A	B	C
1	t-Test: Two-Sample Assuming Equal Variances		
2			
3		*Variable 1*	*Variable 2*
4	Mean	4.222222222	3.4
5	Variance	0.888888889	1.410526316
6	Observations	18	20
7	Pooled Variance	1.164197531	
8	Hypothesized Mean D	0	
9	df	36	
10	t Stat	2.345499396	
11	P(T<=t) one-tail	0.012315054	
12	t Critical one-tail	1.688297694	
13	P(T<=t) two-tail	0.024630109	
14	t Critical two-tail	2.028093987	
15			

After all, the p-value of 0.023 is saying that a difference of 0.82 (or more) between the means of the two designs has only a 2.3% probability of occurring if the true means are actually the same. That's pretty low, so you're able to deliver the news: the higher sophistication mean of 4.22 for Design 1 over Design 2's 3.40 clearly makes Design 1 the better choice for new home page.



Your analysis has shown that perceptions for the image of the young girl being adored by the handsome young man are more "sophisticated" than that of the couple in the rain, clutching the single umbrella. The reason for that perception difference is another matter, and perhaps one that you'll be asked to explore. But for now, just knowing the perception difference exists is tangible progress.

At the next creative staff meeting, you're ready for the inevitable question from Creative Director Kristen McCarthey:

"So Sherlock, what did the survey tell us about the designs? You got a winner?"

"Yes," you calmly reply. "I have some reliable results."

"What the survey data shows is that Design 1 is perceived as more sophisticated than Design 2 by representative members of our target audience of women ages 18–55 with well-above-average disposable income. Design 1 got a 4.22 compared to 3.40 for Design 2. Furthermore, the low p-value of 0.023 means that we have a statistically significant difference between the two designs. We should launch with Design 1."

A hush settles over the conference room. Your colleagues seem impressed, but McCarthey isn't ready to concede anything yet.

"What's your sample size?" she asks.

"18 for Design 1 and 20 for Design 2," you reply.

"How can you make any conclusions with such a small sample size?" she retorts with a huff.

"The p-value of .023 is saying that a difference of .82 (or more) in the sample means of the two designs has only a 2.3% probability of occurring if the true means are, indeed, the same."

She pauses, but has one more salvo: What do you mean by "true means?"

"The true means are the ones we could obtain if we were magically able to poll every customer we have. At last count, our database has roughly 33,000 customers, so it's not likely we'll able to get to them all. Thus, we have to use a representative sample and statistical techniques to make predictions about the true means."

"Hmm…," McCarthey says, staring down at her ubiquitous iPhone. You sense retreat. She punches in a quick text to an unknown recipient, then gets up and bolts for the door, saying: "OK, let's go with Design 1."

As the team streams out of the meeting, Autumn Taylor seems to be taking the news in stride despite the fact that her design lost. She approaches you and whispers: "Hey, I'm really glad we did that survey, for a number of reasons. First, at least I know we're choosing the design based on data and statistics, and not political maneuvering."

"No problem," you reply. "What's the other reason?"

"The fact that you were able to get McCarthey to back down. That's a first!"

Since a sample mean does not come out equal to the true mean, a sample mean is always more valuable to a decision maker if some measure of how far it might be off from the actual population mean is provided. Indeed, the degree of uncertainty is usually provided by a confidence interval.

A confidence interval for the mean is a statement about the value of the true mean. We take the sample average/mean and add and subtract an amount, "e," to form the confidence interval, where the software uses the sample mean, the variability in the data, and the sample size, along with the confidence level (which we talked about in Chapter 1 and noted that traditionally, the confidence level is 95%). So, if the sample mean is 4 (representing, say, the mean satisfaction with a design for some sample size), and the 95% confidence interval is 2.8–5.2 (note: the center of the interval is 4), this says that we are 95% confident that the interval 2.8–5.2 contains the true mean (i.e., the mean for the entire population of people who might use the design and fill out the satisfaction questionnaire). In looser terms, we are 95% confident that the true mean is between 2.8 and 5.2.

Design 1 Satisfaction Values	Design 2 Satisfaction Values
1	4
3	6
4	5
2	7

Let us refer to the sample means as (X-bar1) and (X-bar2) for Designs 1 and 2, respectively, and suppose a 10-point scale, 1–10. We have X-bar1 = 2.5 and X-bar2 = 5.5.

The 95% confidence intervals for the true means, μ1 and μ2 (we introduced the "μ notation" earlier), are:

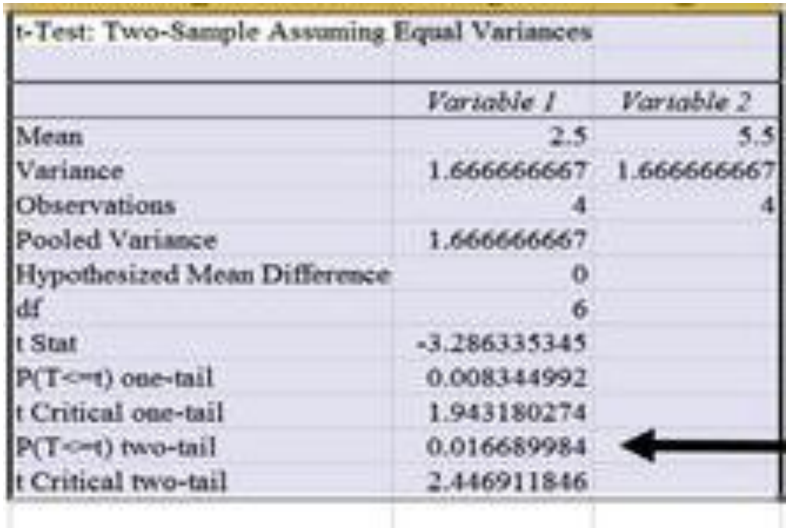
$$\text{Design1}: 0.45 - 4.55$$
$$\text{Design2}: 3.45 - 7.55$$

And we can see that, indeed, the two confidence intervals overlap and not just by a tad—there is an overlap of 3.45–4.55.

So, you may be prone to accept H0 and conclude that the *true means cannot be said to be different*.

However, when we perform the t-test for two independent samples, testing:

$$H0: \mu 1 = \mu 2,$$
$$H1: \mu 1 \neq \mu 2.$$



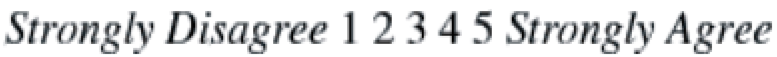
t-Test: Two-Sample Assuming Equal Variances		
	Variable 1	*Variable 2*
Mean	2.5	5.5
Variance	1.666666667	1.666666667
Observations	4	4
Pooled Variance	1.666666667	
Hypothesized Mean Difference	0	
df	6	
t Stat	-3.286335345	
P(T<=t) one-tail	0.008344992	
t Critical one-tail	1.943180274	
P(T<=t) two-tail	0.016689984	
t Critical two-tail	2.446911846	

We can see that the (two-sided) *p*-value is 0.0167 (see arrow in Figure 2.12), well below the traditional 0.05 cutoff point (and recall—a 0.05 cutoff point corresponds with a 95% confidence interval), and thus, based on this output, we would reject H0 (and, it's not even close!!), and conclude that *there is a difference in the true means for two designs*.

6.1 INTRODUCTION

- This chapter explains how to compare means when you have more than two, one-factor analysis of variance (ANOVA).
  - For example, you may be comparing the average ratings of ease-of-use for the different tasks that you just asked participants to complete in a usability test.
  - Or you could be comparing attractiveness of a specific design across different age brackets.
  - Or you may be comparing more than two task-completion times from a usability test.

- Often, the most common scenario of comparing more than two means that a UX researcher confronts is assessing the scores from Likert scales. A typical Likert scale is a statement to which the respondents rate their level of agreement. Usually a five-point scale is used, and the two ends are labeled.



Strongly Disagree 1 2 3 4 5 *Strongly Agree*

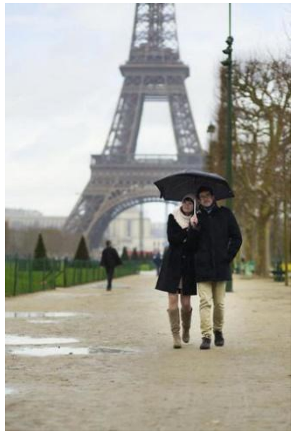
6.2. Case Study: Sophisticated for Whom?

Cinny Bittle, the brash new director of marketing, pokes her head in your office, waving a copy of the latest Forrester report: "Did you see this? **Older boomers** spend the most online. We're hosed!"

"Yeah…**older boomers**, **56–66**, spent an average of $367 in the last 3 months, double that of **Gen Z**, **18–25**, at $138. **Younger boomers, 46–55,** were next at $318, followed by $315 for **Gen X, 36–45**, closely followed by **Gen Y, 26–35**, at $311."“I guess it makes sense," you offer as you copy down the stats. "The older folks are in the prime of their money-making years."

FORRESTER
Technographics® Insight
Older Boomers are
Big Online Buyers
"Did you see this? Older boomers spend the most online! We're hosed!"

- UX researcher at Mademoiselle La La, a high-end online store aimed at urbane women from ages 18–55 years with well-above-average disposable income.
- Design 1—a scene of a young woman sipping a coffee at an outdoor French café, ignoring the adoring eyes of a nearby young man—was considered more "sophisticated" than Design 2, a scene of a young couple snuggling together under one umbrella during a shower, with the Eiffel Tower bisecting a slate-gray sky. Furthermore, you were able to report that the low p-value of 0.023 meant there was a statistically significant difference between the two designs.



6.2. Case Study: Sophisticated for Whom?

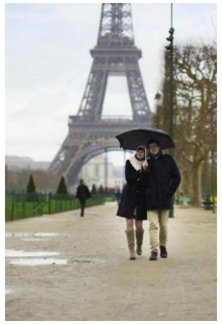
See if there are differences in sophistication by age for the chosen design

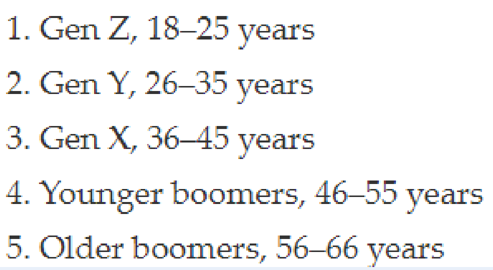
Cinny Bittle, the brash new director of marketing, pokes her head in your office, waving a copy of the latest Forrester report: “Did you see this? Older boomers spend the most online. We’re hosed!”

- But keep in mind that our biggest customers are in the 26–45 range, followed closely by the younger boomers, who are ages 46–55. Our styles are a little conservative for the 18–25 crowd, and not conservative enough for the 56–66 crowd. Since sales are low for those ‘outlier’ brackets, we don’t worry about them as much as the others.”
- she persists: “Can we just see which group thinks the design is sophisticated?”

“Yeah…**older boomers**, **56–66**, spent an average of $367 in the last 3 months, double that of **Gen Z**, **18–25**, at $138. **Younger boomers, 46–55,** were next at $318, followed by $315 for **Gen X, 36–45**, closely followed by **Gen Y, 26–35**, at $311.”“I guess it makes sense,” you offer as you copy down the stats. “The older folks are in the prime of their money-making years.”

“Did you see this? Older boomers spend the most online! We’re hosed!”

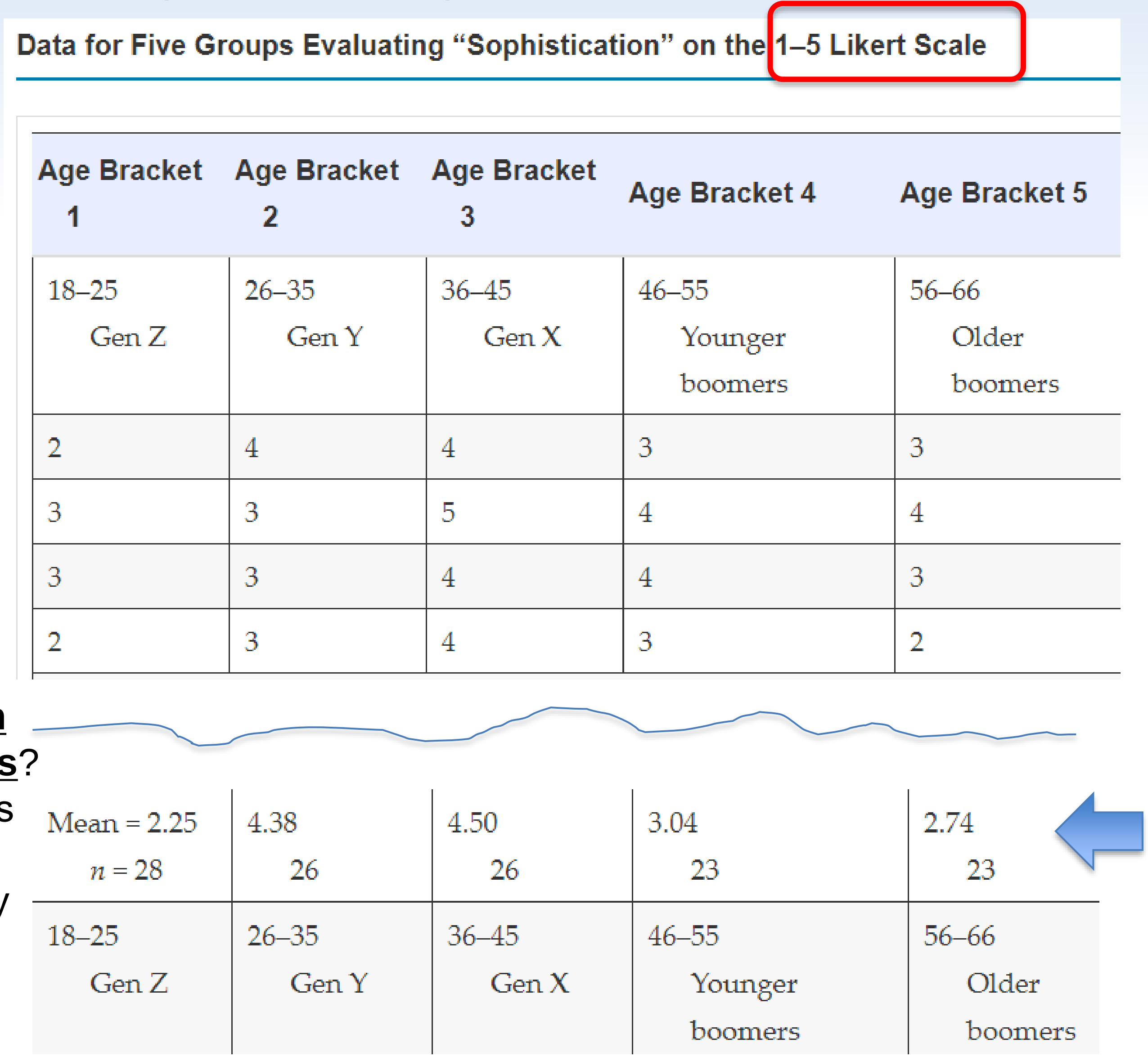




1. Gen Z, 18–25 years
2. Gen Y, 26–35 years
3. Gen X, 36–45 years
4. Younger boomers, 46–55 years
5. Older boomers, 56–66 years

Sort the data by youngest to oldest, using these age brackets:

Agreement with statement: “This page makes Mademoiselle La La seem sophisticated.”1 = Strongly disagree, 5 = strongly agree.



Data for Five Groups Evaluating “Sophistication” on the 1–5 Likert Scale

Age Bracket 1	Age Bracket 2	Age Bracket 3	Age Bracket 4	Age Bracket 5
18–25 Gen Z	26–35 Gen Y	36–45 Gen X	46–55 Younger boomers	56–66 Older boomers
2	4	4	3	3
3	3	5	4	4
3	3	4	4	3
2	3	4	3	2

Mean = 2.25 n = 28	4.38 26	4.50 26	3.04 23	2.74 23
18–25 Gen Z	26–35 Gen Y	36–45 Gen X	46–55 Younger boomers	56–66 Older boomers

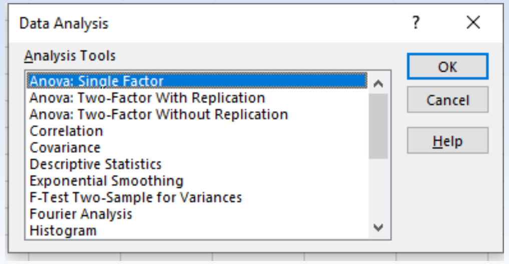
higher ratings for part of Mademoiselle La La’s target audience of 36–45 years. But you’re not sure what the other numbers mean. Is there really a **statistical difference in sophistication ratings between the older and younger boomers**? And what about those “punks,” as Bittle calls Gen Z? Is their perception of sophistication really different from those of their grandmothers? Analysis of variance (ANOVA) to the rescue!

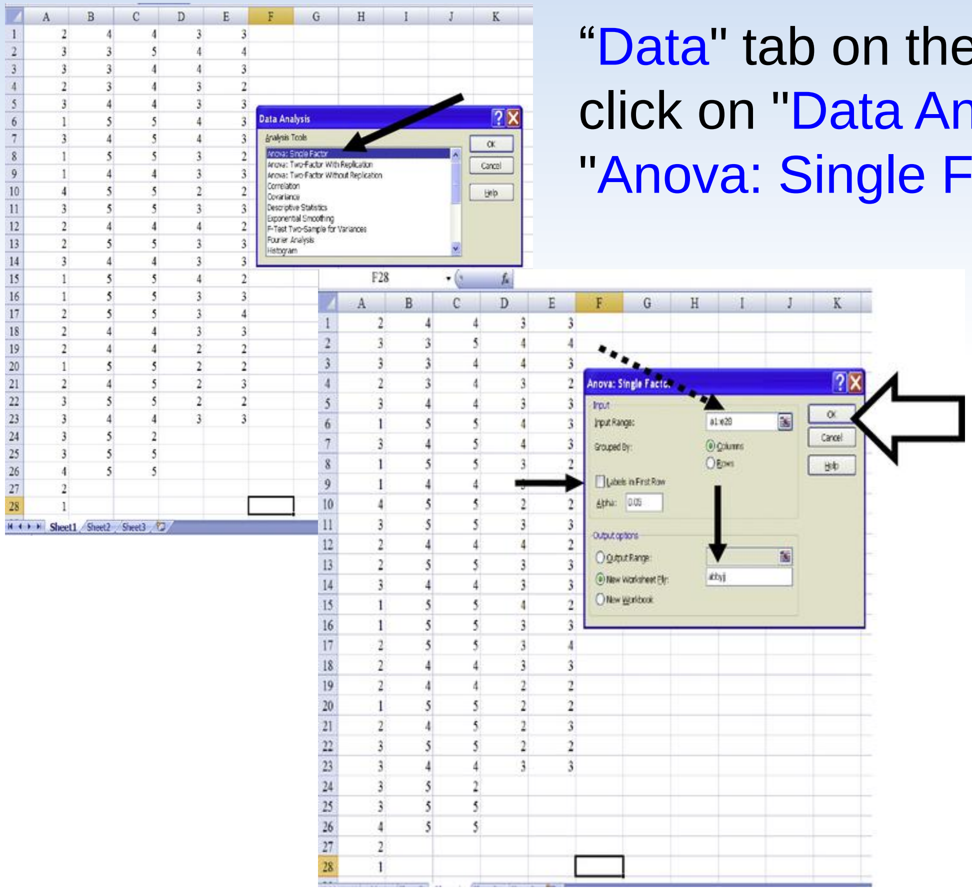
6.3. Independent Samples: One-Factor ANOVA

- To test the hypothesis that differences in perception of sophistication are different for different age-groups. In this kind of scenario, we'll use an independent sample ANOVA to test this hypothesis; the test statistic for ANOVA is the F-statistic.
- One task or design and we wish to compare groups of people who differ in some attribute(s). This is the case here: We have five age-groups (brackets), and various people in each age-group providing an evaluation of the same page.
- It is not important that we have the same number of people evaluating each of the designs or evaluating the ease-of-use of the different tasks being studied. And the sample sizes do not affect the analysis methodology and workings of the software at all; it would not matter if each age-group consisted of around 3 people or around 3000 people. Of course, the results are more accurate if we have larger sample sizes.
- "F-test." The F-distribution is a probability curve that does not look like a normal curve—indeed, it's not even symmetric—but it is based on the individual data points following a normal curve, and can be defined in terms of a function of a normal curve.

6.4.1. Excel (Chap6_Figure6_2_ANOVA.xlsx)

“Data" tab on the Excel ribbon > click on "Data Analysis. > highlight "Anova: Single Factor."





	A	B	C	D	E
1	2	4	4	3	3
2	3	3	5	4	4
3	3	3	4	4	3
4	2	3	4	3	2
5	3	4	4	3	3
6	1	5	5	4	3
7	3	4	5	4	3
8	1	5	5	3	2
9	1	4	4	3	3
10	4	5	5	2	2
11	3	5	5	3	3
12	2	4	4	4	2
13	2	5	5	3	3
14	3	4	4	3	3
15	1	5	5	4	2
16	1	5	5	3	3
17	2	5	5	3	4
18	2	4	4	3	3
19	2	4	4	2	2
20	1	5	5	2	2
21	2	4	5	2	3
22	3	5	5	2	2
23	3	4	4	3	3
24	3	5	2		
25	3	5	5		
26	4	5	5		
27	2				
28	1				

Anova: Single Factor

Input

Input Range: a1:e28

Grouped By: ● Columns ○ Rows

□ Labels in First Row

Alpha: 0.05

Output options

○ Output Range:

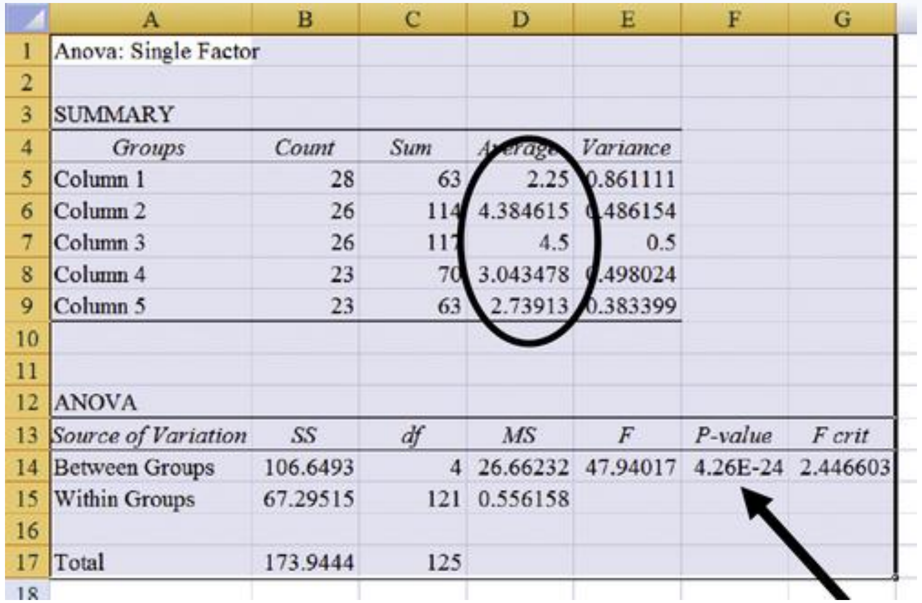
● New Worksheet Ply: abbyj

○ New Workbook

OK

Cancel

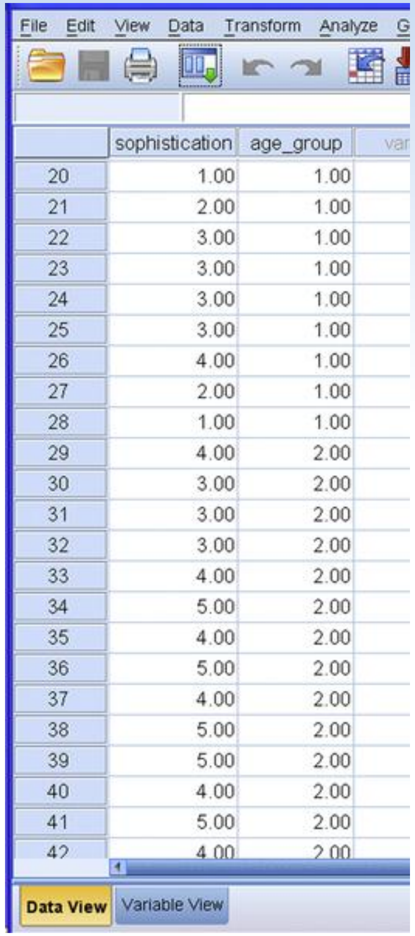
Help



	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3	SUMMARY						
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5	Column 1	28	63	2.25	0.861111		
6	Column 2	26	114	4.384615	0.486154		
7	Column 3	26	117	4.5	0.5		
8	Column 4	23	70	3.043478	0.498024		
9	Column 5	23	63	2.73913	0.383399		
10							
11							
12	ANOVA						
13	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
14	Between Groups	106.6493	4	26.66232	47.94017	4.26E-24	2.446603
15	Within Groups	67.29515	121	0.556158			
16							
17	Total	173.9444	125				
18							

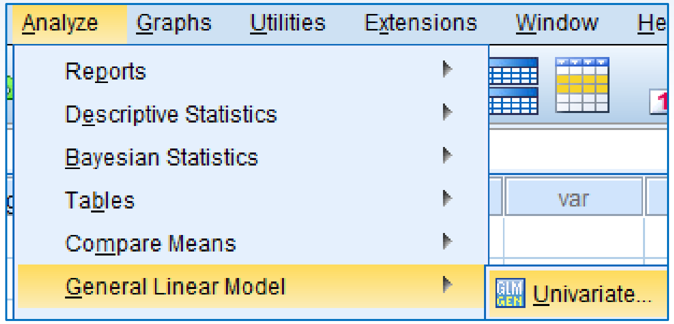
In this output, the p-value = 4.26E-24. As we have noted before, “E” means “exponential” and really means “power of 10.” In other words, the p-value = 4.26×10^{-24}, or (with 23 zeros) 0.00000000000000000000000424. Obviously, this is (way) below 0.05; in fact, it’s about as close to zero as you’ll ever see! So, we reject H0 and go with H1, and thus conclude that there is **sufficient evidence** to be convinced that the true averages of sophistication evaluation for the five age-groups are **not equal.**

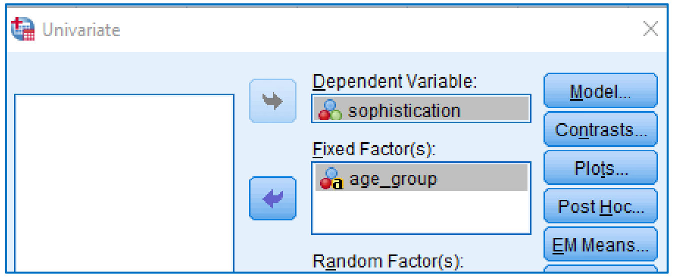
6.4.2. SPSS (Ch6_4_similarExampleDataFile.sav)



	sophistication	age_group	var
20	1.00	1.00	
21	2.00	1.00	
22	3.00	1.00	
23	3.00	1.00	
24	3.00	1.00	
25	3.00	1.00	
26	4.00	1.00	
27	2.00	1.00	
28	1.00	1.00	
29	4.00	2.00	
30	3.00	2.00	
31	3.00	2.00	
32	3.00	2.00	
33	4.00	2.00	
34	5.00	2.00	
35	4.00	2.00	
36	5.00	2.00	
37	4.00	2.00	
38	5.00	2.00	
39	5.00	2.00	
40	4.00	2.00	
41	5.00	2.00	
42	4.00	2.00	

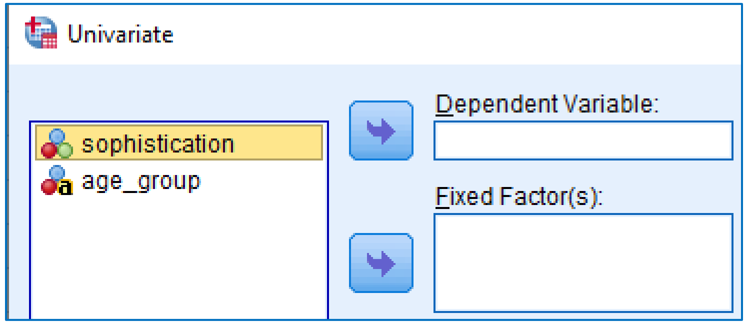
VAR00001 (column 1) is labeled "sophistication." VAR00002 (column 2) is labeled "age_group."





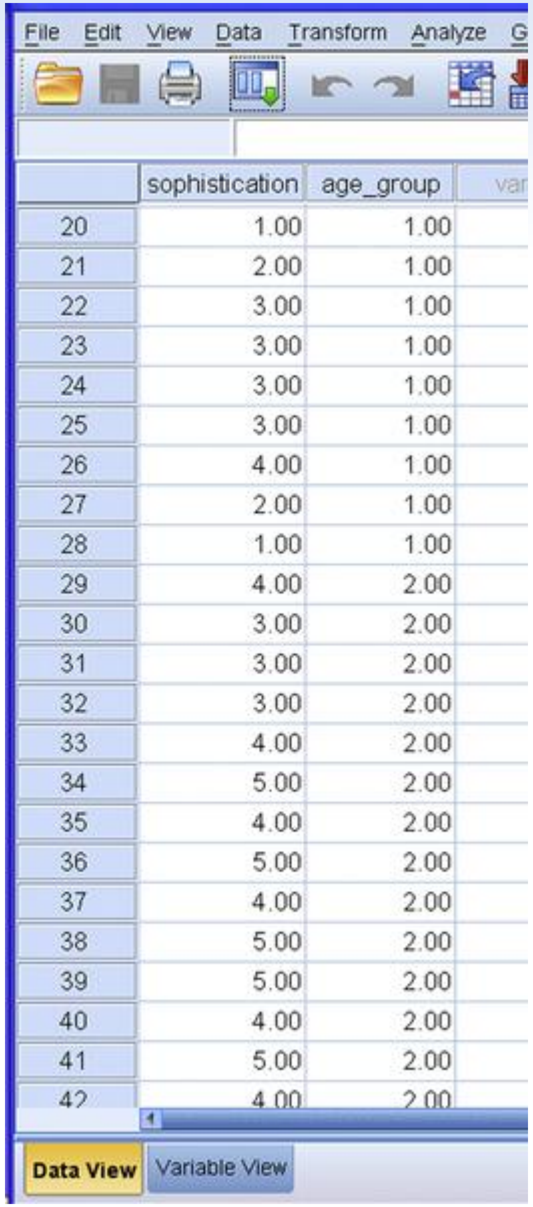
Dependent Variable: sophistication

Fixed Factor(s): age_group



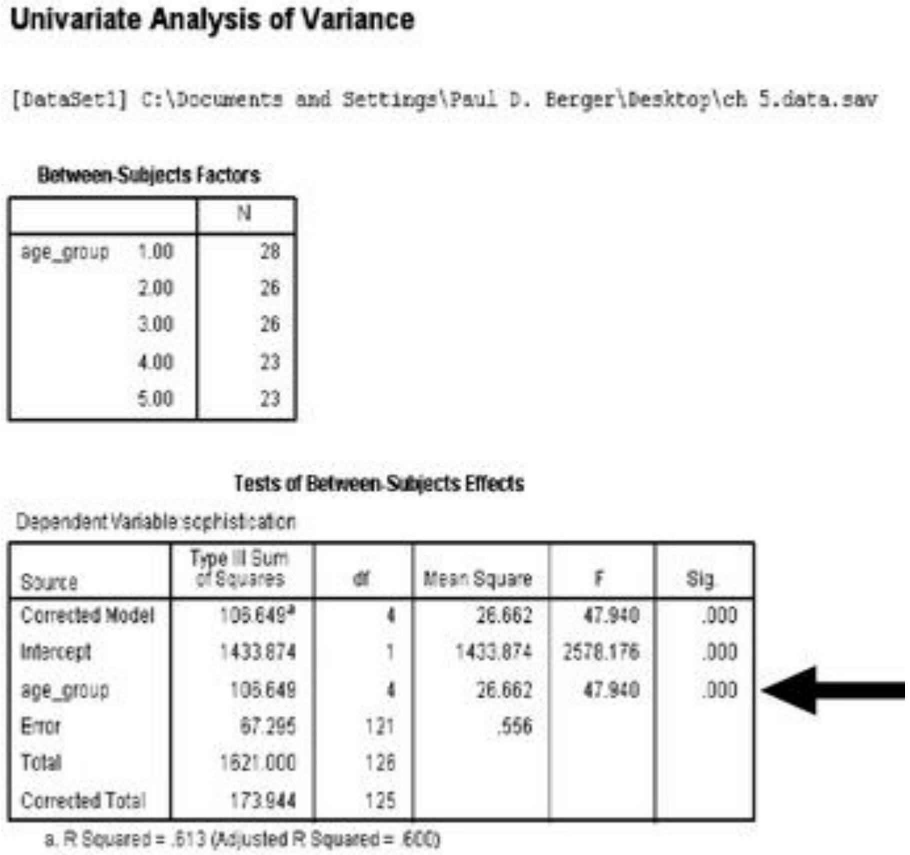
- pull down the "Analyze" dropdown, choose "General Linear Model" (horizontal arrow), and then on the submenu, "Univariate" (vertical arrow) even though there was no choice here
- drag "sophistication" to the "dependent variable" box and the "age_group" to the "Fixed Factor" box.

- VAR00001 (column 1) is labeled “sophistication.”
  VAR00002 (column 2) is labeled “age_group.”



	sophistication	age_group	var
20	1.00	1.00	
21	2.00	1.00	
22	3.00	1.00	
23	3.00	1.00	
24	3.00	1.00	
25	3.00	1.00	
26	4.00	1.00	
27	2.00	1.00	
28	1.00	1.00	
29	4.00	2.00	
30	3.00	2.00	
31	3.00	2.00	
32	3.00	2.00	
33	4.00	2.00	
34	5.00	2.00	
35	4.00	2.00	
36	5.00	2.00	
37	4.00	2.00	
38	5.00	2.00	
39	5.00	2.00	
40	4.00	2.00	
41	5.00	2.00	
42	4.00	2.00	

The key quantity is the p-value, which is 0.000 for age_group in Figure 6.8 (see arrow). Recall that SPSS rounds all p-values to three digits. And we remind the reader (for the last time) that a p-value in SPSS is called “Sig.” So, we conclude that the five age-groups do not have equal true mean sophistication levels.



Univariate Analysis of Variance

```
[DataSet1] C:\Documents and Settings\Paul D. Berger\Desktop\ch 5.data.sav
```

Between-Subjects Factors

		N
age_group	1.00	28
	2.00	26
	3.00	26
	4.00	23
	5.00	23

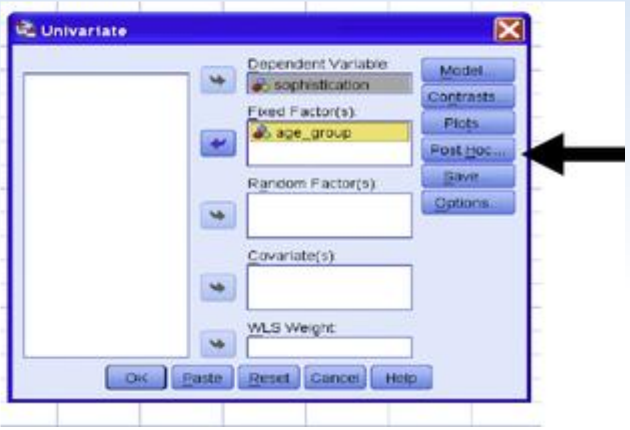
Tests of Between-Subjects Effects

Dependent Variable:sophistication

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	106.649[a]	4	26.662	47.940	.000
Intercept	1433.874	1	1433.874	2578.176	.000
age_group	106.649	4	26.662	47.940	.000
Error	67.295	121	.556		
Total	1621.000	126			
Corrected Total	173.944	125			

a. R Squared = .613 (Adjusted R Squared = .600)

6.5. Multiple Comparison Testing



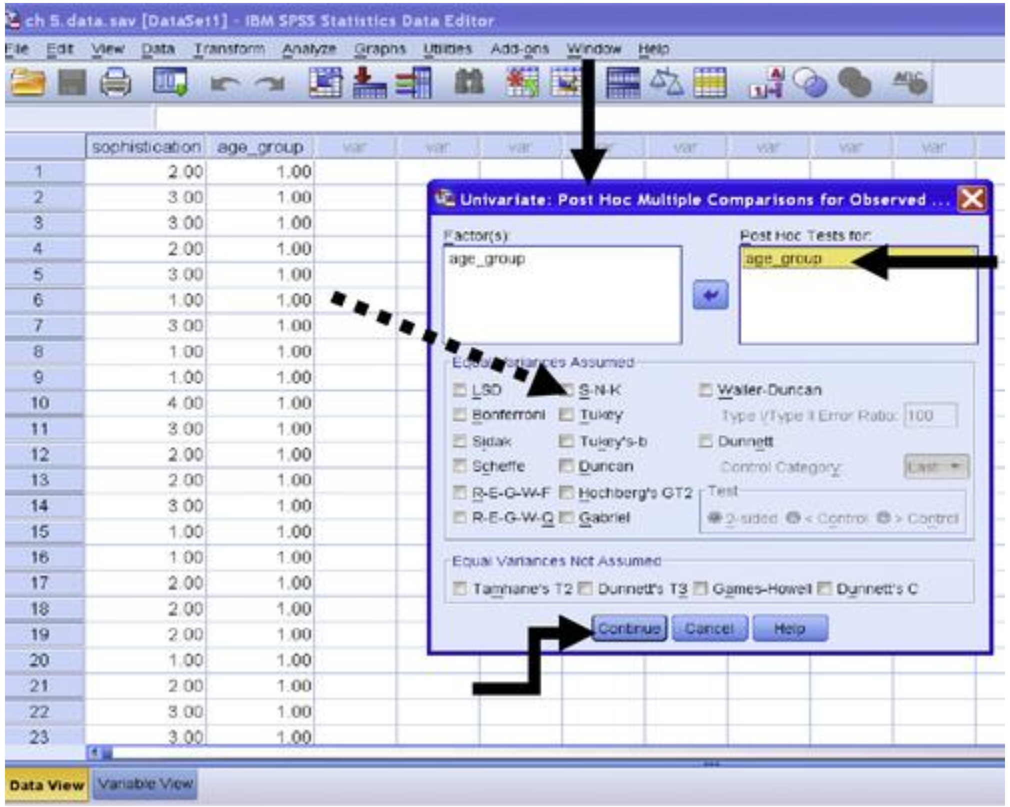
Dependent Variable: sophistication

Fixed Factor(s): age_group

Random Factor(s):

Covariate(s):

WLS Weight:



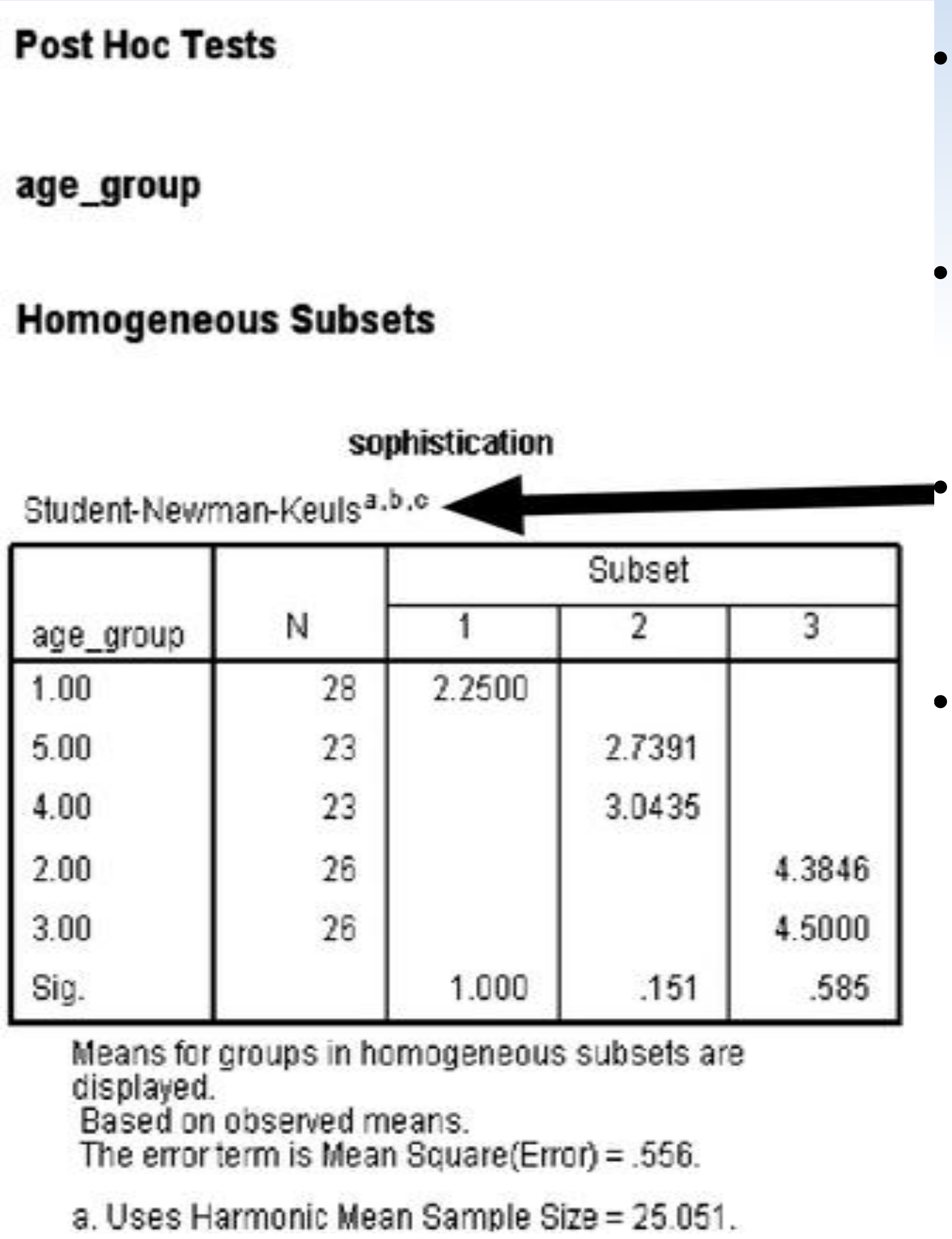
Factor(s): age_group

Post Hoc Tests for: age_group

Type I/Type II Error Ratio: 100

	sophistication	age_group	var
1	2.00	1.00	
2	3.00	1.00	
3	3.00	1.00	
4	2.00	1.00	
5	3.00	1.00	
6	1.00	1.00	
7	3.00	1.00	
8	1.00	1.00	
9	1.00	1.00	
10	4.00	1.00	
11	3.00	1.00	
12	2.00	1.00	
13	2.00	1.00	
14	3.00	1.00	
15	1.00	1.00	
16	1.00	1.00	
17	2.00	1.00	
18	2.00	1.00	
19	2.00	1.00	
20	1.00	1.00	
21	2.00	1.00	
22	3.00	1.00	
23	3.00	1.00	

- Click on “Post Hoc” The term, “post hoc” is Latin and literally means “after this,” but is generally translated in context as “after the fact.” In a manner of speaking, multiple comparison testing can conceptually be viewed as something we examine “after the fact” of having a significant F-test, to increase our knowledge of what the data’s message is.
- After obtaining the Post Hoc dialog box, we need to bring the factor (age_group) over to the right
- checked off S-N-K test
- click “Continue” > click on “OK.”



Post Hoc Tests

age_group

Homogeneous Subsets

sophistication

Student-Newman-Keuls[a,b,c]

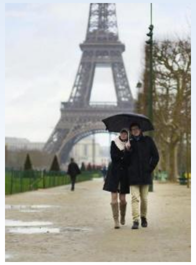
age_group	N	Subset		
		1	2	3
1.00	28	2.2500		
5.00	23		2.7391	
4.00	23		3.0435	
2.00	26			4.3846
3.00	26			4.5000
Sig.		1.000	.151	.585

Means for groups in homogeneous subsets are displayed.
Based on observed means.
The error term is Mean Square(Error) = .556.

a. Uses Harmonic Mean Sample Size = 25.051.

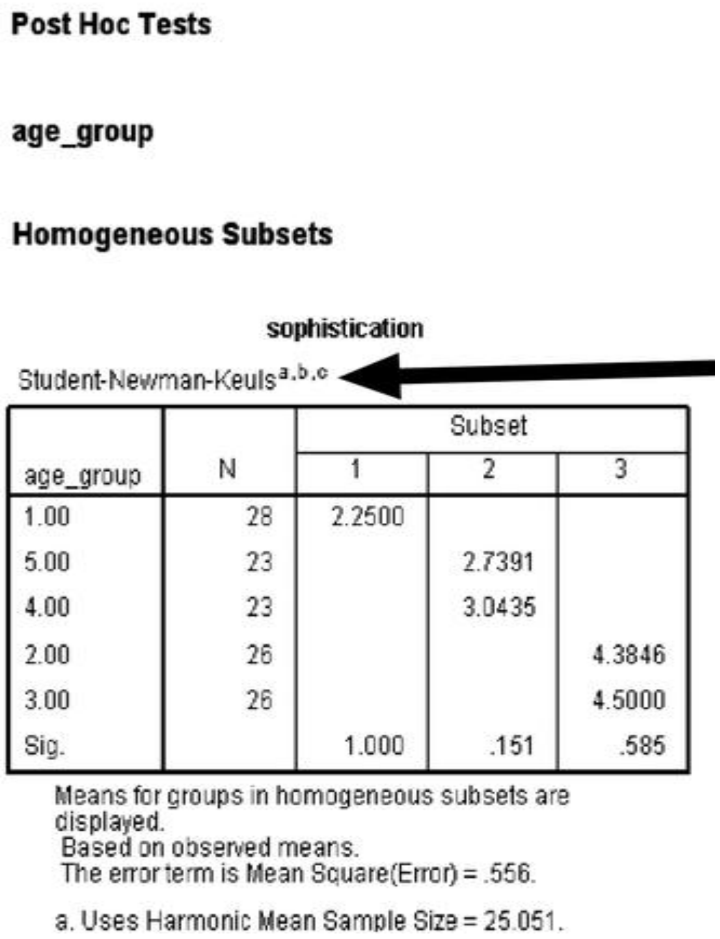
- "Homogeneous Subsets." Basically, this table is telling us that there are **three sets** of means that are significantly different from one another.
- First, the true (sophistication) mean of age-group 1 is judged as different from (lower than) the means of the other four columns.
- The means of age-groups 5 and 4 are judged as the same but higher than that of group 1, but lower than that of groups 2 and 3.
- Finally, the means of age-groups 2 and 3 are judged to be the same, but higher than the other three age-groups. If the only important issue is finding out if one of the age-group means is higher than all the others, then the last result mentioned above is the relevant result—the true mean of age-groups 2 and 3 cannot be said to be different, but the true mean of those two age-groups can be said to be **higher** than the true mean of the other three age-groups.

FORRESTER
Technographics® Insight
Older Boomers are
Big Online Buyers
"Did you see this? Older boomers spend the most online! We're hosed!"



Agreement with statement: “This page makes Mademoiselle La La seem sophisticated.”1 = Strongly disagree, 5 = strongly agree.

- Cinny Bittle, your new texting-mad marketing manager, got her hand on a Forrester report that claimed older boomers (ages 56–66 years) spend the most online of all generations. Since Bittle was worried that your new home page design was not considered sophisticated by this age bracket, you offered to slice and dice the survey data by age. You were trying to determine if there were different perceptions of sophistication by age, and perhaps appease Bittle.



Post Hoc Tests

age_group

Homogeneous Subsets

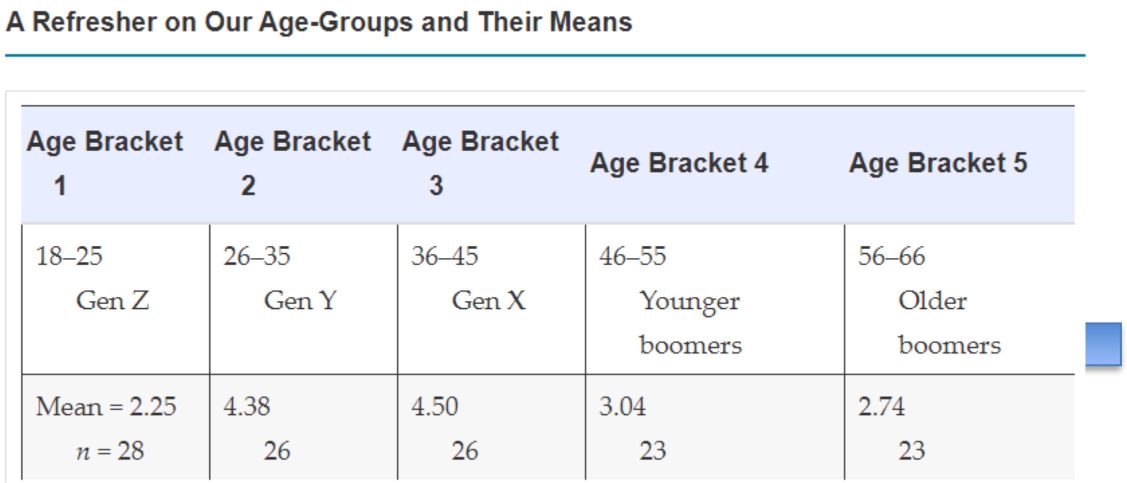
sophistication

Student-Newman-Keuls[a,b,c]

age_group	N	Subset		
		1	2	3
1.00	28	2.2500		
5.00	23		2.7391	
4.00	23		3.0435	
2.00	26			4.3846
3.00	26			4.5000
Sig.		1.000	.151	.585

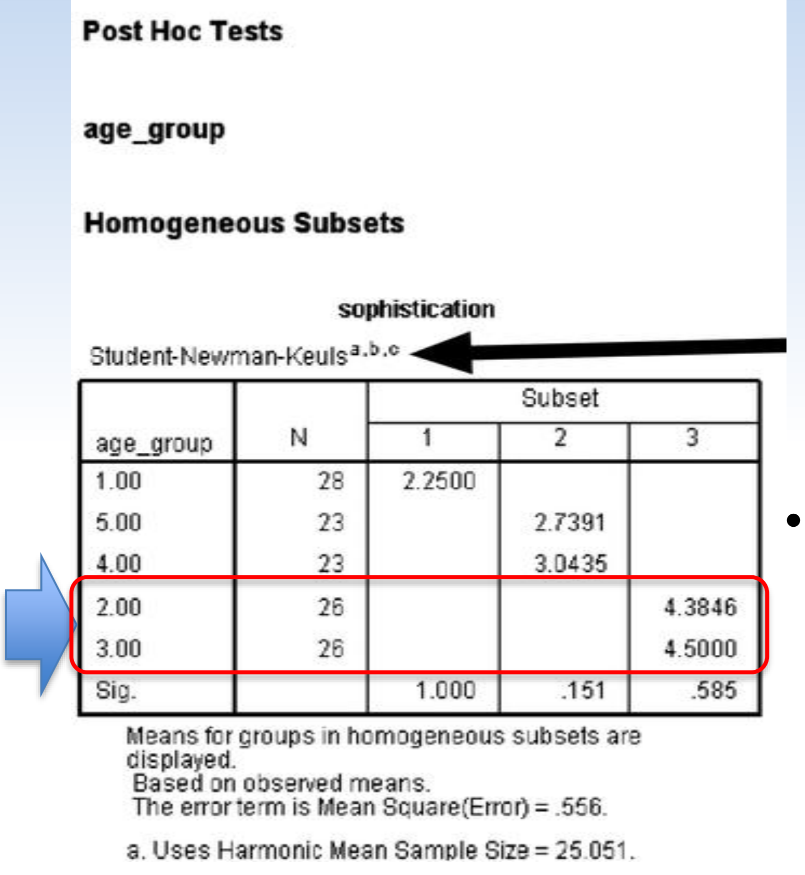
Means for groups in homogeneous subsets are displayed.
Based on observed means.
The error term is Mean Square(Error) = .556.

a. Uses Harmonic Mean Sample Size = 25.051.



A Refresher on Our Age-Groups and Their Means

Age Bracket 1	Age Bracket 2	Age Bracket 3	Age Bracket 4	Age Bracket 5
18–25 Gen Z	26–35 Gen Y	36–45 Gen X	46–55 Younger boomers	56–66 Older boomers
Mean = 2.25 *n* = 28	4.38 26	4.50 26	3.04 23	2.74 23



Post Hoc Tests

age_group

Homogeneous Subsets

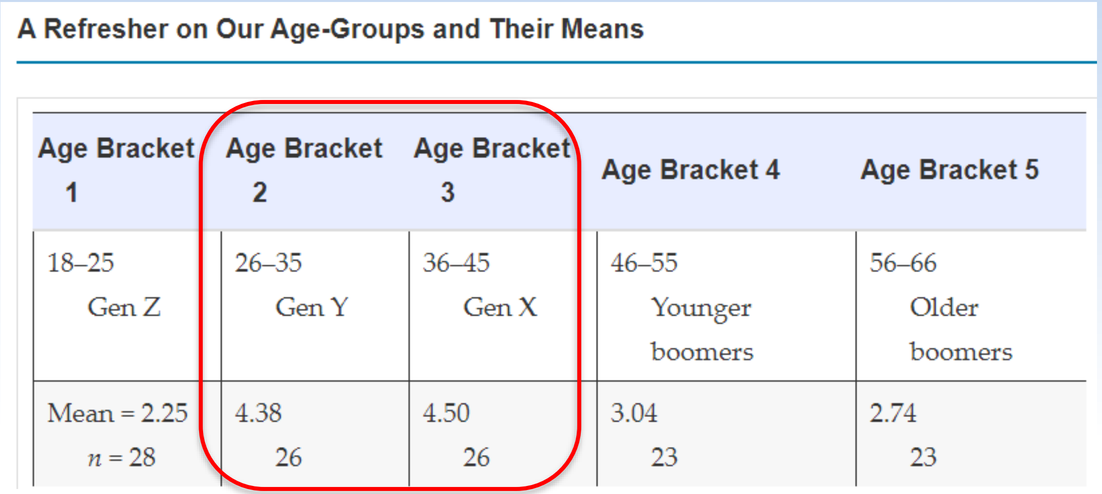
sophistication

Student-Newman-Keuls^{a,b,c}

age_group	N	Subset		
		1	2	3
1.00	28	2.2500		
5.00	23		2.7391	
4.00	23		3.0435	
2.00	26			4.3846
3.00	26			4.5000
Sig.		1.000	.151	.585

Means for groups in homogeneous subsets are displayed.
Based on observed means.
The error term is Mean Square(Error) = .556.

a. Uses Harmonic Mean Sample Size = 25.051.



A Refresher on Our Age-Groups and Their Means

Age Bracket 1	Age Bracket 2	Age Bracket 3	Age Bracket 4	Age Bracket 5
18–25 Gen Z	26–35 Gen Y	36–45 Gen X	46–55 Younger boomers	56–66 Older boomers
Mean = 2.25 n = 28	4.38 26	4.50 26	3.04 23	2.74 23

- **Means of age-groups 2 and 3 (Gen Y [26–35 years] and Gen X [36–45 years])** cannot be said to be different, but the means of those two age-groups can be said to be higher than the means of the other three age-groups. This is great news for Mademoiselle La La, since their greatest sales come from the 26–45 years age bracket. In other ones, the customers who are buying the most are the ones who think the new home page is most sophisticated.

“Yeah…older boomers, 56–66, spent an average of $367 in the last 3 months, double that of Gen Z, 18–25, at $138. Younger boomers, 46–55, were next at $318, followed by $315 for Gen X, 36–45, closely followed by Gen Y, 26–35, at $311.”“I guess it makes sense,” you offer as you copy down the stats. “The older folks are in the prime of their money-making years.”

FORRESTER
Technographics® Insight: Older Boomers are Big Online Buyers
"Did you see this? Older boomers spend the most online! We're hosed!"