Review Test Submission: MBC638 Quiz #11 - Confidence intervals (due Tuesday, Nov. 20, 10:30pm)

User	David Forteguerre
Course	MBC.638.M001.FALL18.Data Anls & Decisn Making
Test	MBC638 Quiz #11 - Confidence intervals (due Tuesday, Nov. 20, 10:30pm)
Started	11/19/18 5:55 PM
Submitted	11/19/18 6:41 PM
Status	Completed
Attempt Score	100 out of 100 points
Time Elapsed	45 minutes out of 1 hour
Results Displayed	All Answers, Submitted Answers, Correct Answers, Feedback, Incorrectly Answered Questions

Question 1

10 out of 10 points

Benji was driving to campus but realized that his car is out of gas. He started driving around Syracuse looking for a gas station and comparing their gas prices, searching for the cheapest one. He drove by 30 different gas stations. He computed their average gas price to be $2.73 per gallon.

Last year around the same time in November, the average gas price in the entire of Syracuse was $2.68 per gallon with the standard deviation of $0.11. Let's assume that the standard deviation hasn't changed since last year.

Benji called his mom:

Benji: "Mom, did you know that average gas prices have gone up since the same time last year?"

Mom: "Are you sure Benji?"

Benji: "Yes, mom, 99% sure."

Mom: "I don't think you're right."

Who is right?

Selected Answer: Mom is right

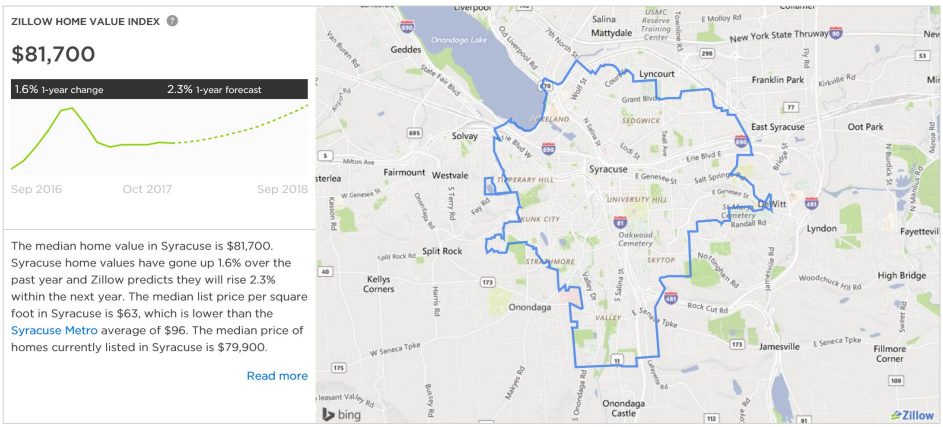
Answers: Mom is right

Benji is right

Response Feedback:

Question 2

15 out of 15 points



We want to estimate the average house price in Upstate New York. Given time and budget constraints, we collect a limited sample of homes in the Syracuse area. The 95% confidence interval is: [95,700 ; 112,700]. This interval uses historic data for the standard deviation of all Upstate New York homes, estimated to be 30,500.

MATCH THE FOLLOWING:

Question	Correct Match	Selected Match
The width of the interval =	D. 17,000	D. 17,000
The margin of error =	K. 8500	K. 8500
Approximately, how many homes were used for the analysis?	I. 50	I. 50
The current margin of error is rather high. If the desired margin of error is 5,000, then how many homes need to be surveyed?	L. 143	L. 143
The current margin of error is rather high. To bring the margin of error down to 1/3 of its current size, how many homes need to be surveyed?	C. 9 times more than the current sample size	C. 9 times more than the current sample size
Interpret the confidence interval [95,700 ; 112,700] .	B. We are 95% certain that the average home price in all of Upstate New York is between $95,700 and $112,700.	B. We are 95% certain that the average home price in all of Upstate New York is between $95,700 and $112,700.

All Answer Choices

A. 6 times more than the current sample size

B. We are 95% certain that the average home price in all of Upstate New York is between $95,700 and $112,700.

C. 9 times more than the current sample size

D. 17,000

E. 8479.9

F. 7094.8

G. We are 95% certain that home prices in all of Upstate New York range between $95,700 and $112,700.

H. 3 times more than the current sample size

I. 50

J. 35

K. 8500

L. 143

Response Feedback:

Question 3

15 out of 15 points

Wendy's is one of the many fast-food chains that offers drive-through service. The manager of a local Wendy's is interested in improving the service provided to customers who use the restaurant's drive-up window. As a first step in this process, the manager asks an assistant to record the time (in seconds) it takes to serve a large number of customers at the final window in the facility's drive-up system. The Excel file **Fast food.xlsx** contains a sample of 200 service times collected by the assistant during the busiest hour of the day.

Based on these data, the manager then concludes that, with **90% confidence**, the mean service time of all customers arriving during the busiest hour of the day at this Wendy's location is between **[a][b]** and **[c][d][f]** seconds. (Perform your calculations in Excel; round your answers to the closest whole number.)

The manager's goal is that the average service time does not exceed 100 seconds. Based on his analysis, should he be concerned? (YES=1, NO=2) **[g]**

Let's agree that, whenever sigma is known we use Z and whenever sigma is unknown we use t.

Selected Answer:

Wendy's is one of the many fast-food chains that offers drive-through service. The manager of a local Wendy's is interested in improving the service provided to customers who use the restaurant's drive-up window. As a first step in this process, the manager asks an assistant to record the time (in seconds) it takes to serve a large number of customers at the final window in the facility's drive-up system. The Excel file **Fast food.xlsx** contains a sample of 200 service times collected by the assistant during the busiest hour of the day.

Based on these data, the manager then concludes that, with **90% confidence**, the mean service time of all customers arriving during the busiest hour of the day at this Wendy's location is between **8 9** and **1 0 5** seconds. (Perform your calculations in Excel; round your answers to the closest whole number.)

The manager's goal is that the average service time does not exceed 100 seconds. Based on his analysis, should he be concerned? (YES=1, NO=2) **1**

Let's agree that, whenever sigma is known we use Z and whenever sigma is unknown we use t.

Answers:

Wendy's is one of the many fast-food chains that offers drive-through service. The manager of a local Wendy's is interested in improving the service provided to customers who use the restaurant's drive-up window. As a first step in this process, the manager asks an assistant to record the time (in seconds) it takes to serve a large number of customers at the final window in the facility's drive-up system. The Excel file **Fast food.xlsx** contains a sample of 200 service times collected by the assistant during the busiest hour of the day.

Based on these data, the manager then concludes that, with **90% confidence**, the mean service time of all customers arriving during the busiest hour of the day at this Wendy's location is between **8 9** and **1 0 5** seconds. (Perform your calculations in Excel; round your answers to the closest whole number.)

The manager's goal is that the average service time does not exceed 100 seconds. Based on his analysis, should he be concerned? (YES=1, NO=2) **1**

Let's agree that, whenever sigma is known we use Z and whenever sigma is unknown we use t.

All Answer Choices

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

Response Feedback:

Question 4

15 out of 15 points

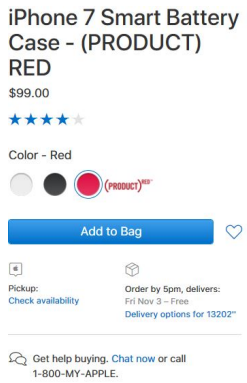
Your iPhone runs out of battery super fast and you are considering buying a battery backup case. You will either purchase it directly on Apple's website, or you will do so on Amazon.com that offers the same or a similar product from multiple vendors and frequently at lower prices. You go to **Amazon.com** and enter keywords: "iPhone 7 plus battery backup case ultra slim." **This file (click here)** contains the products, along with their listed prices, that show up on the first page of your search output, but there are over 10 pages of the results in total.

You want to **estimate the average price of all battery backup cases on Amazon.com**. You'd like to be **90%** certain that your estimation is correct. You construct an interval estimate of the average price. (Assume that, if you plot a histogram of all prices of battery backup cases, then the distribution is approximately Normal.) The margin of error is **[a]**. With 90% certainty, the price of all battery backup cases on Amazon averages between $**[b]** and $**[c]**.

The **Apple's website** offers only one model of battery backup cases, as shown on the picture on the right.

.

Should you buy a battery backup case from Amazon.com or from Apple? (Amazon / Apple) **[d]**



Selected Answer:

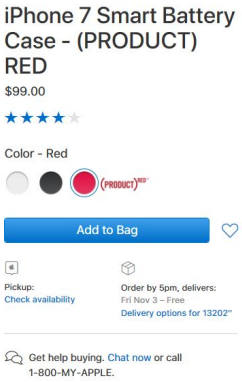
Your iPhone runs out of battery super fast and you are considering buying a battery backup case. You will either purchase it directly on Apple's website, or you will do so on Amazon.com that offers the same or a similar product from multiple vendors and frequently at lower prices. You go to **Amazon.com** and enter keywords: "iPhone 7 plus battery backup case ultra slim." This file (click here) contains the products, along with their listed prices, that show up on the first page of your search output, but there are over 10 pages of the results in total.

You want to **estimate the average price of all battery backup cases on Amazon.com**. You'd like to be **90%** certain that your estimation is correct. You construct an interval estimate of the average price. (Assume that, if you plot a histogram of all prices of battery backup cases, then the distribution is approximately Normal.) The margin of error is **4.2215**. With 90% certainty, the price of all battery backup cases on Amazon averages between $**23.30** and $**31.74**.

The **Apple's website** offers only one model of battery backup cases, as shown on the picture on the right.

.

Should you buy a battery backup case from Amazon.com or from Apple? (Amazon / Apple)
Amazon



Answers:

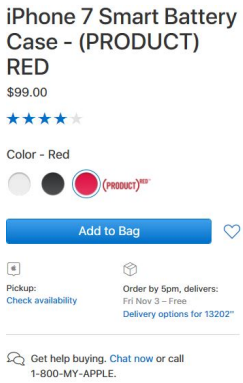
Your iPhone runs out of battery super fast and you are considering buying a battery backup case. You will either purchase it directly on Apple's website, or you will do so on Amazon.com that offers the same or a similar product from multiple vendors and frequently at lower prices. You go to **Amazon.com** and enter keywords: "iPhone 7 plus battery backup case ultra slim." This file (click here) contains the products, along with their listed prices, that show up on the first page of your search output, but there are over 10 pages of the results in total.

You want to **estimate the average price of all battery backup cases on Amazon.com**. You'd like to be **90%** certain that your estimation is correct. You construct an interval estimate of the average price. (Assume that, if you plot a histogram of all prices of battery backup cases, then the distribution is approximately Normal.) The margin of error is **4.2215**. With 90% certainty, the price of all battery backup cases on Amazon averages between $**23.30** and $**31.74**.

The **Apple's website** offers only one model of battery backup cases, as shown on the picture on the right.

.

Should you buy a battery backup case from Amazon.com or from Apple? (Amazon / Apple)
Amazon



All Answer Choices

- 3.2283
- 3.9610
- 4.2215
- 1.6449
- 1.3406
- 1.7531
- 23.30
- 23.56
- 24.29
- 30.75
- 31.48
- 31.74
- Amazon
- Apple

Response Feedback:

Question 5 15 out of 15 points

Recall the gender discrimination suit data that we've used in lectures: **Bank Salaries.xlsx** . Perform all necessary calculations in Excel and answer the following questions. Also, in all cases when sigma is known use *Z*, and in all cases when sigma is unknown use *t*.

	
We want to estimate the average salary of all **female** bank employees, using 99.9% confidence level.	We want to estimate the average salary of all **male** bank employees, using 99.9% confidence level.
$t_{\alpha/2}$ = **[a] . [b][c][d][f][g]**	$t_{\alpha/2}$ = **[a1] . [b1][c1][d1][f1][g1]**
Margin of error = **[h][i][j][jj] . [k]**	Margin of error = **[h1][i1][j1][j2] . [k1]**
Confidence interval lower limit = **$ [l][m],[n][o][p]**	Confidence interval lower limit = **$ [l1][m1],[n1][o1][p1]**
Confidence interval upper limit = **$ [q][r],[s][t][u]**	Confidence interval upper limit = **$ [q1][r1],[s1][t1][u1]**
Based on these results, would you conclude that female employees at this bank are underpaid? *(NO=0, YES=1)* **[v]**	

Selected Answer: Recall the gender discrimination suit data that we've used in lectures: **Bank Salaries.xlsx** . Perform all

necessary calculations in Excel and answer the following questions. Also, in all cases when sigma is known use *Z*, and in all cases when sigma is unknown use *t*.

	
We want to estimate the average salary of all **female** bank employees, using 99.9% confidence level.	We want to estimate the average salary of all **male** bank employees, using 99.9% confidence level.
$t_{\alpha/2}$ = **3 . 3 6 1 9 0**	$t_{\alpha/2}$ = **3 . 4 4 1 7 5**
Margin of error = **1 9 0 6 . 8**	Margin of error = **6 6 1 2 . 5**
Confidence interval lower limit = **$ 3 5, 3 0 3**	Confidence interval lower limit = **$ 3 8, 8 9 3**
Confidence interval upper limit = **$ 3 9, 1 1 7**	Confidence interval upper limit = **$ 5 2, 1 1 8**
Based on these results, would you conclude that female employees at this bank are underpaid? *(NO=0, YES=1)* **0**	

Answers:

Recall the gender discrimination suit data that we've used in lectures: **Bank Salaries.xlsx** . Perform all necessary calculations in Excel and answer the following questions. Also, in all cases when sigma is known use *Z*, and in all cases when sigma is unknown use *t*.

	
We want to estimate the average salary of all **female** bank employees, using 99.9% confidence level.	We want to estimate the average salary of all **male** bank employees, using 99.9% confidence level.
$t_{\alpha/2}$ = **3 . 3 6 1 9 0**	$t_{\alpha/2}$ = **3 . 4 4 1 7 5**
Margin of error = **1 9 0 6 . 8**	Margin of error = **6 6 1 2 . 5**
Confidence interval lower limit = **$ 3 5, 3 0 3**	Confidence interval lower limit = **$ 3 8, 8 9 3**
Confidence interval upper limit = **$ 3 9, 1 1 7**	Confidence interval upper limit = **$ 5 2, 1 1 8**
Based on these results, would you conclude that female employees at this bank are underpaid? *(NO=0, YES=1)* **0**	

All Answer Choices

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

Response Feedback:

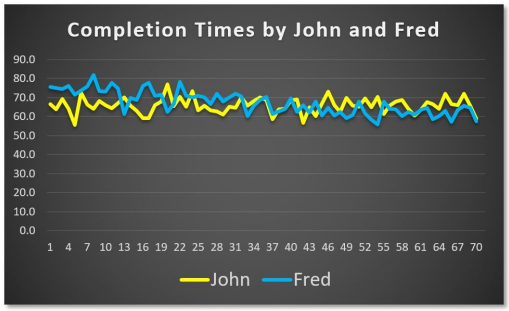
Question 6 5 out of 5 points

LIMITATIONS OF CONFIDENCE INTERVALS. On 11/14/2018, we had the following problem in lecture:

> John and Fred are two workers. The sample data shows repetitive task times for each of the two workers. John has been doing this task for months, whereas Fred has just started. Each time listed is the time (in seconds) to perform a routine task on an assembly line. The times shown are in chronological order (column A: 1=oldest, 70=most recent). Use 95% confidence intervals to determine the overall average time that it takes each worker to complete the task, and then decide which of the two workers you would rather keep.

Here is the data along with the solutions from 11/14/2018. The two samples (John and Fred) are an example of **#[a]** _____ (#1 = independent; #2 = paired) samples. If you recall, using the confidence interval for the difference in population means, we concluded that **#[b]** _____ (#1 = it's inconclusive which worker performs the task faster; #2 = John performs the task faster; #3 = Fred performs the task faster).

Let us now look at some limitations of the confidence interval that we constructed in class. Take a look at the following time series plot of John's and Fred's completion times over the sample period:



Which of the two workers would you rather keep? What is your conclusion after looking at this time series plot? **#[c]** (#1 = keep John and fire Fred; #2 = keep Fred and fire John) As you can see, a confidence interval for the difference in population means fails to capture any time variations in the time series data. Because of this, confidence intervals are more applicable for cross-sectional data (i.e., data that occurs in a single point of time).

Another limitation of the confidence interval that we constructed is that we don't know which factors may affect John and Fred's productivity and the completion times. For example, we don't know whether John and Fred may have different skills, experience, or education background. For this reason, any inferences made from confidence intervals analysis should be treated with caution because confidence intervals are **univariate** in nature and don't provide us with the complete picture of all variables that might be relevant for the estimation of population mean difference. (Other examples of univariate analyses are scatterplots, one-sample descriptive statistics, and simple regressions.) A better, **multivariate**, approach to study the differences in average completion times by John and Fred would be to build a multiple regression model and include all relevant factors.

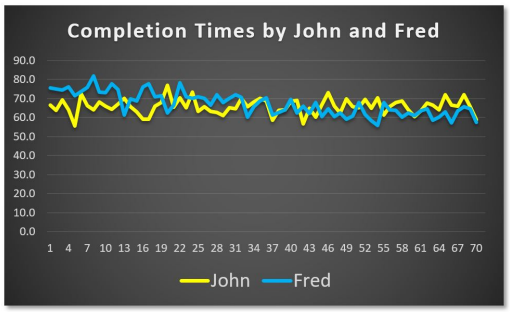
Selected Answer:

LIMITATIONS OF CONFIDENCE INTERVALS. On 11/14/2018, we had the following problem in lecture:

> John and Fred are two workers. The sample data shows repetitive task times for each of the two workers. John has been doing this task for months, whereas Fred has just started. Each time listed is the time (in seconds) to perform a routine task on an assembly line. The times shown are in chronological order (column A: 1=oldest, 70=most recent). Use 95% confidence intervals to determine the overall average time that it takes each worker to complete the task, and then decide which of the two workers you would rather keep.

Here is the data along with the solutions from 11/14/2018. The two samples (John and Fred) are an example of **#2** _____ (#1 = independent; #2 = paired) samples. If you recall, using the confidence interval for the difference in population means, we concluded that **#1** _____ (#1 = it's inconclusive which worker performs the task faster; #2 = John performs the task faster; #3 = Fred performs the task faster).

Let us now look at some limitations of the confidence interval that we constructed in class. Take a look at the following time series plot of John's and Fred's completion times over the sample period:



Which of the two workers would you rather keep? What is your conclusion after looking at this time series plot? # **2** (#1 = keep John and fire Fred; #2 = keep Fred and fire John) As you can see, a confidence interval for the difference in population means fails to capture any time variations in the time series data. Because of this, confidence intervals are more applicable for cross-sectional data (i.e., data that occurs in a single point of time).

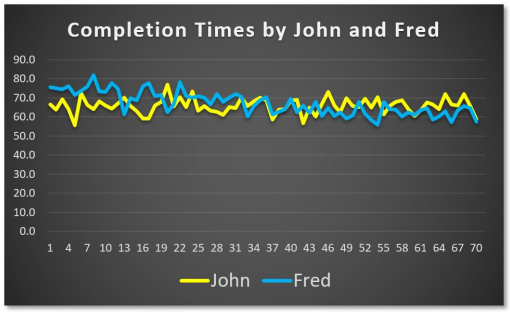
Another limitation of the confidence interval that we constructed is that we don't know which factors may affect John and Fred's productivity and the completion times. For example, we don't know whether John and Fred may have different skills, experience, or education background. For this reason, any inferences made from confidence intervals analysis should be treated with caution because confidence intervals are **univariate** in nature and don't provide us with the complete picture of all variables that might be relevant for the estimation of population mean difference. (Other examples of univariate analyses are scatterplots, one-sample descriptive statistics, and simple regressions.) A better, **multivariate**, approach to study the differences in average completion times by John and Fred would be to build a multiple regression model and include all relevant factors.

Answers: **LIMITATIONS OF CONFIDENCE INTERVALS.** On 11/14/2018, we had the following problem in lecture:

> John and Fred are two workers. The sample data shows repetitive task times for each of the two workers. John has been doing this task for months, whereas Fred has just started. Each time listed is the time (in seconds) to perform a routine task on an assembly line. The times shown are in chronological order (column A: 1=oldest, 70=most recent). Use 95% confidence intervals to determine the overall average time that it takes each worker to complete the task, and then decide which of the two workers you would rather keep.

Here is the data along with the solutions from 11/14/2018. The two samples (John and Fred) are an example of # **2** _____ (#1 = independent; #2 = paired) samples. If you recall, using the confidence interval for the difference in population means, we concluded that # **1** _____ (#1 = it's inconclusive which worker performs the task faster; #2 = John performs the task faster; #3 = Fred performs the task faster).

Let us now look at some limitations of the confidence interval that we constructed in class. Take a look at the following time series plot of John's and Fred's completion times over the sample period:



Which of the two workers would you rather keep? What is your conclusion after looking at this time series plot? # **2** (#1 = keep John and fire Fred; #2 = keep Fred and fire John) As you can see, a confidence interval for the difference in population means fails to capture any time variations in the time series data. Because of this, confidence intervals are more applicable for cross-sectional data (i.e., data that occurs in a single point of time).

Another limitation of the confidence interval that we constructed is that we don't know which factors may affect John and Fred's productivity and the completion times. For example, we don't know whether John and Fred may have different skills, experience, or education background. For this reason, any inferences made from confidence intervals analysis should be treated with caution because confidence intervals are **univariate** in nature and don't provide us with the complete picture of all variables that might be relevant for the estimation of population mean difference. (Other examples of univariate analyses are scatterplots, one-sample descriptive statistics, and simple regressions.) A better, **multivariate**, approach to study the differences in average completion times by John and Fred would be to build a multiple regression model and include all relevant factors.

All Answer Choices

- 1
- 2
- 3

Response Feedback:

Question 7

25 out of 25 points

On Sunday, November 5, 2017, a lone shooter killed over 25 people in the First Baptist Church in Sutherland Springs, Texas, making it the deadliest mass shooting in Texas history to date. On Monday, November 6, the President Trump issued a statement saying that this incident "isn't a guns situation".

TRUMP: DEADLY TEXAS SHOOTING ISN'T A GUNS ISSUE

Trump Describes Worst Mass Killing In Texas History As A Mental Health Problem

Source: Huffington Post, 11/6/2017 .

Let's look at some real data and try to shed some light on this matter. Has gun violence and resulting casualties increased in recent years? The CNN article from 11/6/2017 "Mass shootings in America are a serious problem -- and these 9 charts show just why" (click to see the article) illustrates the trends in gun violence in the U.S. in recent years and refers to http://www.gunviolencearchive.org/ for the source of the data. I have compiled the data that this CNN article is based on into a single Excel file: **mass_shootings_Jan2014-Dec2016.xlsx** . Use this data to perform the following analysis and answer the following questions.

STEP 1: Save the Excel data file on your computer, then open the file. Insert an empty column after Column A. Create a new variable "Year" in the empty Column B that shows the year of the incident. Use the `year( )` command in Excel. You might need to change the format of the Year variable to 'number' after you create it. The resulting data would look like this (the first 10 rows):

	A	B	C	D	E	F	G
1	Incident Date	Year	State	City Or Co	Address	# Killed	# Injured
2	1-Jan-14	2014	Virginia	Norfolk	Rockingha	2	2
3	3-Jan-14	2014	New York	Queens	Farmers B	1	3
4	11-Jan-14	2014	Louisiana	Tallulah	3600 block	0	6
5	11-Jan-14	2014	Oregon	Portland	Southeast	0	5
6	12-Jan-14	2014	Louisiana	Tallulah	N/A	0	6
7	12-Jan-14	2014	Alabama	Huntsville	University	0	5
8	12-Jan-14	2014	Illinois	Elgin	300 block	0	5
9	13-Jan-14	2014	Michigan	Detroit	800 block	0	4
10	14-Jan-14	2014	California	Los Angel	West 86th	0	4

STEP 2: Copy and paste the "State" variable (Column C) into Column J and remove duplicates. You will have a total of 46 states represented in the database. Create headers in Columns K, L, and M: 2014, 2015, and 2016. Then, for each of these three years, and for each state, compute the total number of mass shooting incidents. Then, repeat the same in Columns O, P, Q, R and then T, U, V, W, but instead of the number of mass shooting incidents by year and state, compute the total number of injuries and total number of fatalities (killed), respectively, by year and state. If you do all computations correctly, your end result will look like this (first 10 rows):

I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
#events:	State	2014	2015	2016	#injured:	State	2014	2015	2016	#killed:	State	2014	2015	2016
	Virginia	8	5	11		Virginia	28	21	41		Virginia	12	2	14
	New York	14	21	15		New York	50	90	69		New York	9	19	5
	Louisiana	12	14	13		Louisiana	61	73	54		Louisiana	5	11	15
	Oregon	3	2	2		Oregon	13	13	5		Oregon	1	11	3
	Alabama	2	4	15		Alabama	9	13	53		Alabama	1	4	16
	Illinois	25	25	4		Illinois	108	100	150		Illinois	16	17	32
	Michigan					Michigan					Michigan	10	3	15
	California					California					California	45	38	39
	Utah					Utah					Utah	5	0	0

In **Texas**, in **2014**, the total number of mass shooting incidents was **[a][b]** , the total number of injuries was **[c][d]**, and the total number killed was **[f][g]**.The same numbers in **2016** were **[aa][bb]**, **[cc][dd][ee]**, and **[ff][gg]**.

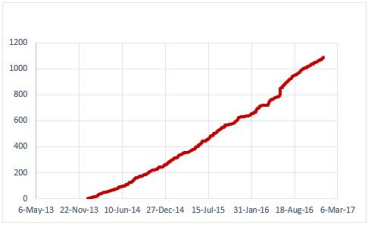
STEP 3: In each state, did the number of mass shooting incidents, total number of injuries, and total number killed change between **2014** and **2016**? Let's use confidence intervals to answer these questions. Perform all analysis in StatTools. Using **98%** confidence level, construct appropriate confidence intervals that would help shed light on the following three questions: overall, have the average (by state) number of mass shooting incidents, the average number of injuries, and the average number killed changed in 2016 relative to 2014? Using 2016 as your Population 1 and 2014 as your Population 2, construct 3 appropriate intervals for the difference in means:

- The confidence interval for the difference in the overall average (by state) **number of mass shooting incidents** is: **[h] . [i][j][k][l]** to **[m] . [n][o][p][q]** (round to 4 decimal places).
- The confidence interval for the difference in the overall average (by state) **number of injuries** is: **[hh] . [ii][jj][kk][ll]** to **[mm][rr] . [nn][oo][pp][qq]** (round to 4 decimal places).
- The confidence interval for the difference in the overall average (by state) **number killed** is: **[hhh] . [iii][jjj][kkk][lll]** to **[mmm] . [nnn][ooo][ppp][qqq]** (round to 4 decimal places).

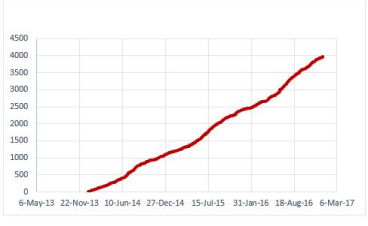
STEP 4: What is your overall conclusion: can you conclude that, on average, for all states, the mass shooting problem has gotten worse between 2014 and 2016? (0=No, 1=Yes) **[r]**

STEP 5: Time series plot. Copy and paste the Incident Date variable in an empty column and remove duplicates (the same day may have more than one incident). Sort the dates from oldest to most recent. For each distinct date, count the total number of incidents, injuries, and fatalities. If you use a line graph option in Excel to plot the dates on the x-axis and the 3 variables on the y-axes, then you will face two big problems. First, the graph will be too "jagged" and it will be difficult to see any trends. Second, on some dates there are no events (e.g., January 4 through 10 of 2014 have no incidents), but Excel will plot all distinct dates with equal intervals in between on the x-axis. It's an existing issue in Excel that you should be aware of: the x-values won't be appropriately distanced from each other on the x-axis according to their magnitudes (for example, if you have three consecutive dates: 1/1/2014, 1/3/2014, and 1/11/2014, they will be placed with equal intervals between them on the x-axis). So, to fix the second issue, you should use the scatter plot (with line connecting the points) option in Excel instead of a line graph option. To fix the first issue, one solution is to create a *cumulative* plot for the number of incidents, injuries, and fatalities, with each consecutive value being the sum of all previous values in the previous and current dates. For example, on 1/1/2014 the number of injuries is 2, on 1/3/2014 the cumulative number of injuries is 5, and so on. Create three time series plots that show the cumulative number of incidents, injuries, and fatalities. Of the following 3 charts, which one corresponds to the cumulative number of fatalities (killed)? **#[v]**

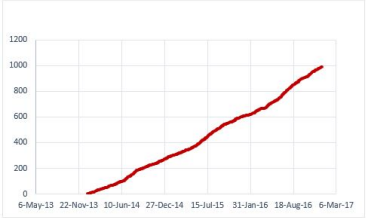
#1:



#2:



#3:



Selected Answer:

On Sunday, November 5, 2017, a lone shooter killed over 25 people in the First Baptist Church in Sutherland Springs, Texas, making it the deadliest mass shooting in Texas history to date. On Monday, November 6, the President Trump issued a statement saying that this incident "isn't a guns situation".

TRUMP: DEADLY TEXAS SHOOTING ISN'T A GUNS ISSUE

Trump Describes Worst Mass Killing In Texas History As A Mental Health Problem

Source: Huffington Post, 11/6/2017 .

.

Let's look at some real data and try to shed some light on this matter. Has gun violence and resulting casualties increased in recent years? The CNN article from 11/6/2017 "Mass shootings in America are a serious problem -- and these 9 charts show just why" (click to see the article) illustrates the trends in gun violence in the U.S. in recent years and refers to http://www.gunviolencearchive.org/ for the source of the data. I have compiled the data that this CNN article is based on into a single Excel file: **mass_shootings_Jan2014-Dec2016.xlsx** . Use this data to perform the following analysis and answer the following questions.

.

STEP 1: Save the Excel data file on your computer, then open the file. Insert an empty column after Column A. Create a new variable "Year" in the empty Column B that shows the year of the incident. Use the `year( )` command in Excel. You might need to change the format of the Year variable to 'number' after you create it. The resulting data would look like this (the first 10 rows):

	A	B	C	D	E	F	G
1	Incident Date	Year	State	City Or Co	Address	# Killed	# Injured
2	1-Jan-14	2014	Virginia	Norfolk	Rockingha	2	2
3	3-Jan-14	2014	New York	Queens	Farmers B	1	3
4	11-Jan-14	2014	Louisiana	Tallulah	3600 block	0	6
5	11-Jan-14	2014	Oregon	Portland	Southeast	0	5
6	12-Jan-14	2014	Louisiana	Tallulah	N/A	0	6
7	12-Jan-14	2014	Alabama	Huntsville	University	0	5
8	12-Jan-14	2014	Illinois	Elgin	300 block	0	5
9	13-Jan-14	2014	Michigan	Detroit	800 block	0	4
10	14-Jan-14	2014	California	Los Angel	West 86th	0	4

.

STEP 2: Copy and paste the "State" variable (Column C) into Column J and remove duplicates. You will have a total of 46 states represented in the database. Create headers in Columns K, L, and M: 2014, 2015, and 2016. Then, for each of these three years, and for each state, compute the total number of mass shooting incidents. Then, repeat the same in Columns O, P, Q, R and then T, U, V, W, but instead of the number of mass shooting incidents by year and state, compute the total number of injuries and total number of fatalities (killed), respectively, by year and state. If you do all computations correctly, your end result will look like this (first 10 rows):

#events:	State	2014	2015	2016	#injured:	State	2014	2015	2016	#killed:	State	2014	2015	2016
	Virginia	8	5	11		Virginia	28	21	41		Virginia	12	2	14
	New York	14	21	15		New York	50	90	69		New York	9	19	5
	Louisiana	12	14	13		Louisiana	61	73	54		Louisiana	5	11	15
	Oregon	3	2	2		Oregon	13	13	5		Oregon	1	11	3
	Alabama	2	4	15		Alabama	9	13	53		Alabama	1	4	16
	Illinois	25	25	42		Illinois	108	108	150		Illinois	16	17	32
	Michigan	13	11	11		Michigan	49	55	41		Michigan	10	3	15
	California	33	28	37		California	175	120	194		California	45	38	39
	Utah	3	3	3		Utah	6	9	6		Utah	3	0	

In **Texas**, in **2014**, the total number of mass shooting incidents was ✓ **1**✓ **4** , the total number of injuries was ✓ **6**✓ **5**, and the total number killed was ✓ **2**✓ **5**.The same numbers in **2016** were ✓ **3**✓ **1**, ✓ **1** ✓ **1**✓ **5**, and ✓ **4**✓ **2**.

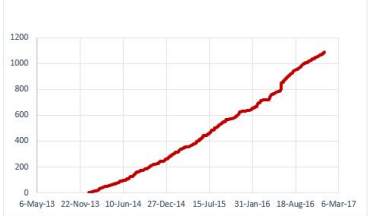
STEP 3: In each state, did the number of mass shooting incidents, total number of injuries, and total number killed change between **2014** and **2016**? Let's use confidence intervals to answer these questions. Perform all analysis in StatTools. Using **98%** confidence level, construct appropriate confidence intervals that would help shed light on the following three questions: overall, have the average (by state) number of mass shooting incidents, the average number of injuries, and the average number killed changed in 2016 relative to 2014? Using 2016 as your Population 1 and 2014 as your Population 2, construct 3 appropriate intervals for the difference in means:

- The confidence interval for the difference in the overall average (by state) **number of mass shooting incidents** is: ✓ **0 .** ✓ **6**✓ **7**✓ **5**✓ **8** to ✓ **4 .** ✓ **0**✓ **6**✓ **3**✓ **4** (round to 4 decimal places).
- The confidence interval for the difference in the overall average (by state) **number of injuries** is: ✓ **1 .** ✓ **3**✓ **1**✓ **0**✓ **0** to ✓ **1**✓ **7 .** ✓ **8**✓ **2**✓ **0**✓ **4** (round to 4 decimal places).
- The confidence interval for the difference in the overall average (by state) **number killed** is: ✓ **0 .** ✓ **6**✓ **8**✓ **7**✓ **8** to ✓ **7 .** ✓ **6**✓ **1**✓ **6**✓ **5** (round to 4 decimal places).

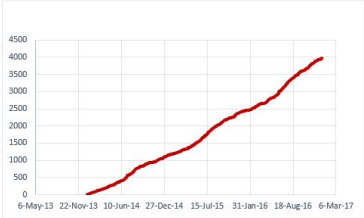
STEP 4: What is your overall conclusion: can you conclude that, on average, for all states, the mass shooting problem has gotten worse between 2014 and 2016? (0=No, 1=Yes) ✓ **1**

STEP 5: Time series plot. Copy and paste the Incident Date variable in an empty column and remove duplicates (the same day may have more than one incident). Sort the dates from oldest to most recent. For each distinct date, count the total number of incidents, injuries, and fatalities. If you use a line graph option in Excel to plot the dates on the x-axis and the 3 variables on the y-axes, then you will face two big problems. First, the graph will be too "jagged" and it will be difficult to see any trends. Second, on some dates there are no events (e.g., January 4 through 10 of 2014 have no incidents), but Excel will plot all distinct dates with equal intervals in between on the x-axis. It's an existing issue in Excel that you should be aware of: the x-values won't be appropriately distanced from each other on the x-axis according to their magnitudes (for example, if you have three consecutive dates: 1/1/2014, 1/3/2014, and 1/11/2014, they will be placed with equal intervals between them on the x-axis). So, to fix the second issue, you should use the scatter plot (with line connecting the points) option in Excel instead of a line graph option. To fix the first issue, one solution is to create a *cumulative* plot for the number of incidents, injuries, and fatalities, with each consecutive value being the sum of all previous values in the previous and current dates. For example, on 1/1/2014 the number of injuries is 2, on 1/3/2014 the cumulative number of injuries is 5, and so on. Create three time series plots that show the cumulative number of incidents, injuries, and fatalities. Of the following 3 charts, which one corresponds to the cumulative number of fatalities (killed)? #✓ **1**

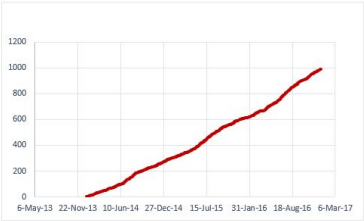
#1:



#2:



#3:



Answers:

On Sunday, November 5, 2017, a lone shooter killed over 25 people in the First Baptist Church in Sutherland Springs, Texas, making it the deadliest mass shooting in Texas history to date. On Monday, November 6, the President Trump issued a statement saying that this incident "isn't a guns situation".

TRUMP: DEADLY TEXAS SHOOTING ISN'T A GUNS ISSUE

Trump Describes Worst Mass Killing In Texas History As A Mental Health Problem

Source: Huffington Post, 11/6/2017 .

Let's look at some real data and try to shed some light on this matter. Has gun violence and resulting casualties increased in recent years? The CNN article from 11/6/2017 "Mass shootings in America are a serious problem -- and these 9 charts show just why" (click to see the article) illustrates the trends in gun violence in the U.S. in recent years and refers to http://www.gunviolencearchive.org/ for the source of the data. I have compiled the data that this CNN article is based on into a single Excel file: **mass_shootings_Jan2014-Dec2016.xlsx** . Use this data to perform the following analysis and answer the following questions.

STEP 1: Save the Excel data file on your computer, then open the file. Insert an empty column after Column A. Create a new variable "Year" in the empty Column B that shows the year of the incident. Use the `year( )` command in Excel. You might need to change the format of the Year variable to 'number' after you create it. The resulting data would look like this (the first 10 rows):

	A	B	C	D	E	F	G
1	Incident Date	Year	State	City Or Co	Address	# Killed	# Injured
2	1-Jan-14	2014	Virginia	Norfolk	Rockingha	2	2
3	3-Jan-14	2014	New York	Queens	Farmers B	1	3
4	11-Jan-14	2014	Louisiana	Tallulah	3600 block	0	6
5	11-Jan-14	2014	Oregon	Portland	Southeast	0	5
6	12-Jan-14	2014	Louisiana	Tallulah	N/A	0	6
7	12-Jan-14	2014	Alabama	Huntsville	University	0	5
8	12-Jan-14	2014	Illinois	Elgin	300 block	0	5
9	13-Jan-14	2014	Michigan	Detroit	800 block	0	4
10	14-Jan-14	2014	California	Los Angel	West 86th	0	4

STEP 2: Copy and paste the "State" variable (Column C) into Column J and remove duplicates. You will have a total of 46 states represented in the database. Create headers in Columns K, L, and M: 2014, 2015, and 2016. Then, for each of these three years, and for each state, compute the total number of mass shooting incidents. Then, repeat the same in Columns O, P, Q, R and then T, U, V, W, but instead of

the number of mass shooting incidents by year and state, compute the total number of injuries and total number of fatalities (killed), respectively, by year and state. If you do all computations correctly, your end result will look like this (first 10 rows):

I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
#events:	State	2014	2015	2016	#injured:	State	2014	2015	2016	#killed:	State	2014	2015	2016
	Virginia	8	5	11		Virginia	28	21	41		Virginia	12	2	14
	New York	14	21	15		New York	50	90	69		New York	9	19	5
	Louisiana	12	14	13		Louisiana	61	73	54		Louisiana	5	11	15
	Oregon	3	2	2		Oregon	13	13	5		Oregon	1	11	3
	Alabama	2	4	15		Alabama	9	13	53		Alabama	1	4	16
	Illinois	25	25	32		Illinois	108	108	150		Illinois	16	17	32
	Michigan	13	11	11		Michigan	49	45	41		Michigan	10	3	15
	California	33	38	33		California	175	128	198		California	45	38	39
	Utah	3	3	3		Utah	6	9	6		Utah	9	0	

In **Texas**, in **2014**, the total number of mass shooting incidents was ✓ **1**✓ **4** , the total number of injuries was ✓ **6**✓ **5**, and the total number killed was ✓ **2**✓ **5**.The same numbers in **2016** were ✓ **3**✓ **1**, ✓ **1** ✓ **1**✓ **5**, and ✓ **4**✓ **2**.

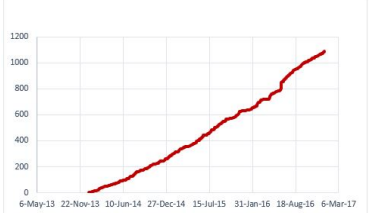
STEP 3: In each state, did the number of mass shooting incidents, total number of injuries, and total number killed change between **2014** and **2016**? Let's use confidence intervals to answer these questions. Perform all analysis in StatTools. Using **98%** confidence level, construct appropriate confidence intervals that would help shed light on the following three questions: overall, have the average (by state) number of mass shooting incidents, the average number of injuries, and the average number killed changed in 2016 relative to 2014? Using 2016 as your Population 1 and 2014 as your Population 2, construct 3 appropriate intervals for the difference in means:

- The confidence interval for the difference in the overall average (by state) **number of mass shooting incidents** is: ✓ **0 .** ✓ **6**✓ **7**✓ **5**✓ **8** to ✓ **4 .** ✓ **0**✓ **6**✓ **3**✓ **4** (round to 4 decimal places).
- The confidence interval for the difference in the overall average (by state) **number of injuries** is: ✓ **1 .** ✓ **3**✓ **1**✓ **0**✓ **0** to ✓ **1**✓ **7 .** ✓ **8**✓ **2**✓ **0**✓ **4** (round to 4 decimal places).
- The confidence interval for the difference in the overall average (by state) **number killed** is: ✓ **0 .** ✓ **6**✓ **8**✓ **7**✓ **8** to ✓ **7 .** ✓ **6**✓ **1**✓ **6**✓ **5** (round to 4 decimal places).

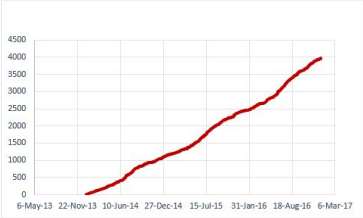
STEP 4: What is your overall conclusion: can you conclude that, on average, for all states, the mass shooting problem has gotten worse between 2014 and 2016? (0=No, 1=Yes) ✓ **1**

STEP 5: Time series plot. Copy and paste the Incident Date variable in an empty column and remove duplicates (the same day may have more than one incident). Sort the dates from oldest to most recent. For each distinct date, count the total number of incidents, injuries, and fatalities. If you use a line graph option in Excel to plot the dates on the x-axis and the 3 variables on the y-axes, then you will face two big problems. First, the graph will be too "jagged" and it will be difficult to see any trends. Second, on some dates there are no events (e.g., January 4 through 10 of 2014 have no incidents), but Excel will plot all distinct dates with equal intervals in between on the x-axis. It's an existing issue in Excel that you should be aware of: the x-values won't be appropriately distanced from each other on the x-axis according to their magnitudes (for example, if you have three consecutive dates: 1/1/2014, 1/3/2014, and 1/11/2014, they will be placed with equal intervals between them on the x-axis). So, to fix the second issue, you should use the scatter plot (with line connecting the points) option in Excel instead of a line graph option. To fix the first issue, one solution is to create a *cumulative* plot for the number of incidents, injuries, and fatalities, with each consecutive value being the sum of all previous values in the previous and current dates. For example, on 1/1/2014 the number of injuries is 2, on 1/3/2014 the cumulative number of injuries is 5, and so on. Create three time series plots that show the cumulative number of incidents, injuries, and fatalities. Of the following 3 charts, which one corresponds to the cumulative number of fatalities (killed)? #✓ **1**

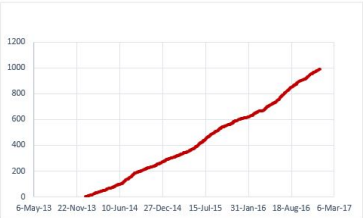
#1:



#2:



#3:



All Answer Choices

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

Response Feedback: **YOU GOT IT !!**

Wednesday, November 21, 2018 1:17:42 PM EST

← OK