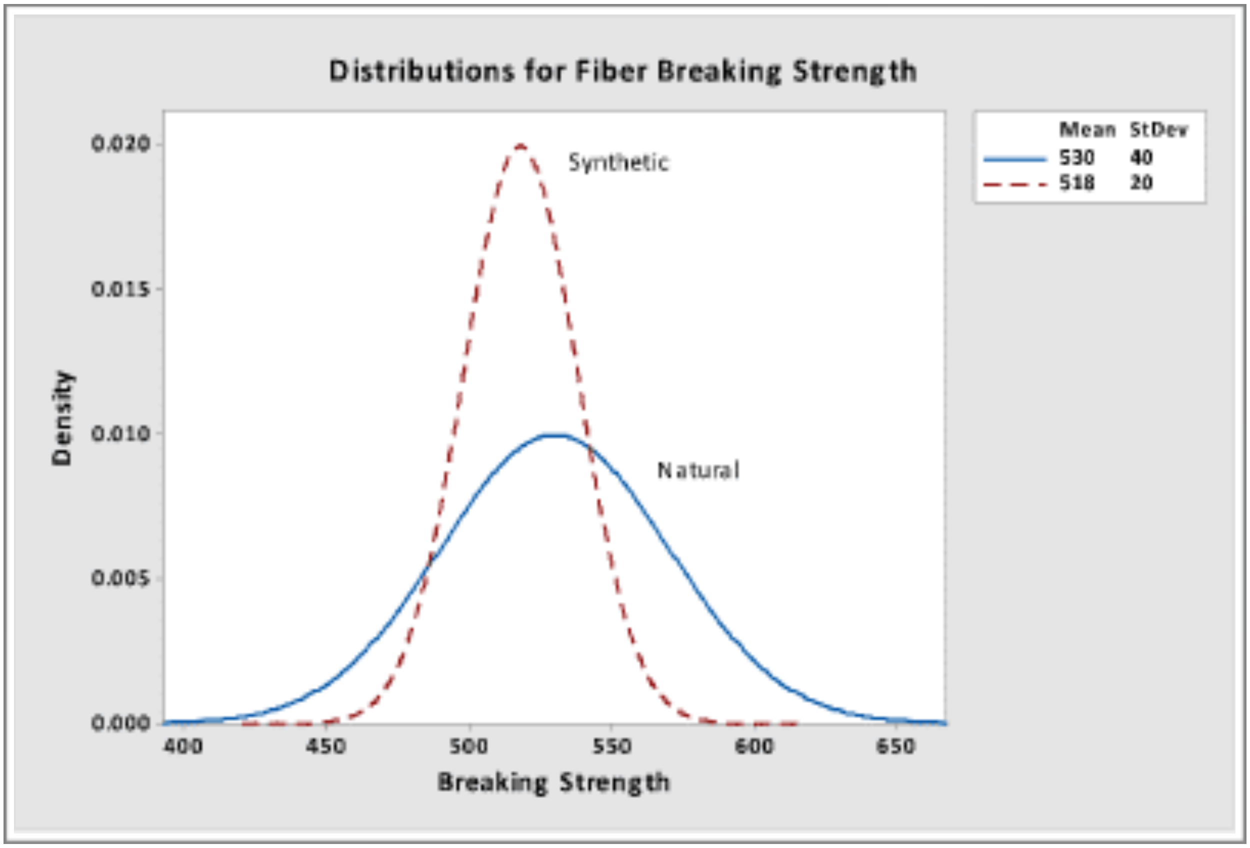
Project week 5

Two-sample Confidence Intervals & Hypothesis Testing



Maryam Heidari

winter 2019

Project week 5

Two-sample Confidence Intervals & Hypothesis Testing

Introduction

In this assignment, we are going to work with z-test and t-test for two samples and compare their mean or variances, what is a confidence interval for them and how we can calculate p-value in order to test our hypothesizes.

Discussion

In part one, by using sampling in data analytics package, I randomly draw a sample of size 36 from the 2015 vehicles per household data and a sample of size 34 from the 2016 vehicles per household data. Then, by using =average, = var, =sqrt, and =countif functions, I calculated the mean, variance, standard deviation and proportion of the population that are greater than 2 for both of the samples. You can see the result below:

Table A

	2015 Vehicles per Household (Population 1)	2016 Vehicles per Household (Population 2)
Population Mean:	1.69	1.72
Population Variance:	0.09	0.09
Population Standard deviation:	0.29	0.30
Proportion of the population that are greater than 2:	0.142	0.155

picture 1

In the past two, I copy the samples of part one and calculated the mean, variance and sample size. After that, by assuming that the two population's variances are known, I am going to compare two samples' means. for the different confidence interval. Because the sample size of both of them are large, I used the z-test.

You can see the formula and the results in below:

Confidence Interval for $\mu_1 - \mu_2$:

Large Independent Samples ($n_1 \geq 30$ and $n_2 \geq 30$)

σ_1 and σ_2 are known

Sample Mean1= $\overline{x}_1$, Sample Mean2 = $\overline{x}_2$

Standard Error:

$$\sigma_{\overline{x}_1-\overline{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Confidence Level = c

$$z_c : \quad = \mathbf{NORM.S.INV}\left(\frac{1+c}{2}\right)$$

Margin of Error $E = z_c \sigma_{\overline{x}_1-\overline{x}_2}$

CI Lower Limit = $(\overline{x}_1 - \overline{x}_2) - E$

CI Upper Limit = $(\overline{x}_1 - \overline{x}_2) + E$

picture 2

Table A

	Sample 1 (2015)	Sample 2 (2016)
Sample Mean	1.71	1.73
Sample Variance	0.07	0.08
Sample Size	36	34

	Population 1 (2015)	Population 2 (2016)
Population Variance	0.0852	0.0894

Sample Means Difference	-0.02	
Sampling Error	0.07	$\sigma_{\overline{x}_1-\overline{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Table B

Confidence Level CI	z_c	Margin of Error	CI Lower Limit for the Means Difference	CI Upper Limit for the Means Difference	CI Width
92%	1.751	0.114	-0.139	1.726	1.865
95%	1.960	0.128	-0.152	1.935	2.087
98%	2.326	0.151	-0.176	2.302	2.478

Table C

	Population 1 (2015)	Population 2 (2016)	Populations Mean Difference
Population Mean	1.69	1.72	-0.02

picture 3

As you can see, for all of the confidence interval the population mean difference is in the CI width.

In part 3, by using sampling in data analytics package, I randomly draw a sample of size 20 from the 2015 vehicles per household data and a sample of size 25 from the 2016 vehicles per household data. Then, I calculated the mean, variance and size by using =average, =var and =count function. After that, I compared the two samples' means. After that, by assuming that the two population's variances are

unknown, I am going to compare two samples' means. for the different confidence interval. Because the size of both samples is small, I used the t-test in here.
You can see the formula and the result is below:

T-Test Hypothesis Testing for $\mu_1 - \mu_2$:

Two Tailed problem

Samples are independent
σ_1 and σ_2 are unknown, but are assumed to be unequal
If any of the two samples is small, then the two populations are normally or approximately normally distributed

Null Hypothesis: H_0: $\mu_1 - \mu_2 = \mu_0$

Alternative Hypothesis: H_A: $\mu_1 - \mu_2 \neq \mu_0$

μ_0 = *Hypothesized Difference*
Sample Mean1= $\bar{x}_1$, Sample Mean2 = $\bar{x}_2$
Sample Variance1= s_1^2, Sample Variance2 = s_2^2

Degrees of Freedom $df = \dfrac{(s_1^2/n_1+s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$

Standard Error:

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

Test Statistic $t = \dfrac{(\bar{x}_1-\bar{x}_2)-\mu_0}{\sigma_{\bar{x}_1-\bar{x}_2}}$

picture 4

Table A

	Sample 1 (2015)	Sample 2 (2016)
Sample Mean	1.67	1.63
Sample Variance	0.04	0.13
Sample Size	20	25

Sample Means Difference	0.04
Sampling (standard) Error	0.08
Degrees of Freedom (DF)	38.80

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

$$df = \frac{(s_1^2/n_1+s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$$

Table B

Confidence Level CI	t_c	Margin of Error	CI Lower Limit for the Means Difference	CI Upper Limit for the Means Difference	CI Width
92%	1.799	0.152	-0.117	0.187	0.303
95%	2.024	0.171	-0.136	0.206	0.341
98%	2.429	0.205	-0.170	0.240	0.409

Table C

	Population 1 (2015)	Population 2 (2016)	Populations Mean Difference
Population Mean	1.69	1.72	-0.02

picture 5

Based on the result above, for all of the confidence interval, the population mean difference is in the CI width.

In the past 4, we are going to collect a random sample of size 120 from the 2015 vehicles per household data and a random sample of size 125 from the 2016 vehicles per household data. After counting the number of the household now has more than 2 vehicles, I calculated each sample proportion of success and failure. Then by using formulas in below, I calculated the z-test for each of them.

Sample sizes: n_1 and n_2
Samples are randomly selected and independent of each other.
Sample Proportions: $\hat{p}_1 = \frac{x_1}{n_1}$ and $\hat{p}_2 = \frac{x_2}{n_2}$
Requirements: $n_1\hat{p}_1 \geq 10$, $n_1\hat{q}_1 \geq 10$, $n_2\hat{p}_2 \geq 10$, and $n_2\hat{q}_2 \geq 10$
Pooled Estimate of proportions:

$$\overline{p} = \frac{n_1\,\hat{p}_1 + n_2\,\hat{p}_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\overline{q} = 1 - \overline{p}$$

Null Hypothesis: $H_0\!: p_1 - p_2 \geq p_0$

Alternative Hypothesis: $H_A\!: p_1 - p_2 < p_0$

Standard Error $\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\overline{p}\,\overline{q}\,\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

Test Statistic $z = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sigma_{\hat{p}_1 - \hat{p}_2}}$

picture 6

You can see the result below:

Table A

	Sample 1 (2015)	Sample 2 (2016)
Sample number of Success *	19	18
Sample Proportion of Success *	0.1583	0.1440
Sample Proportion of Failure	0.8417	0.8560
Sample Size	120	125

* Note: A proportion of Success is the proportion of households with more than two vehicles.

Table B

Difference of Sample Proportions	0.0143
Sampling (standard) Error	0.0458

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1\,\hat{q}_1}{n_1} + \frac{\hat{p}_2\,\hat{q}_2}{n_2}}$$

Table C

Confidence Level CI	z_c	Margin of Error	CI Lower Limit	CI Upper Limit	CI Width
92%	1.751	0.080	-0.066	0.094	0.160
96%	2.054	0.094	-0.080	0.140	0.220
99%	2.576	0.118	-0.104	0.118	0.222

Table D

	Population 1 (2015)	Population 2 (2016)
Population Proportion of Success	0.1420	0.1546

Table E

Difference of Population Proportions	-0.0126

picture 7

Based on picture 7, only in the confidence level of 69 %, the population proportion of success is in the CI width.

In part 5, We are going to use the null hypothesis in order to test our assumption. In here, by using alph=0.05, I tested the hypothesis that the standard deviations of the two populations are different. You can find the formulas that I used below:

Two Tailed problem

Use the *F distribution*

Sample sizes: n_1 and n_2

$df_1 = n_1 - 1,\ df_2 = n_2 - 1$

Sample variances s_1^2 (larger sample variance), and s_2^2 (smaller sample variance)

Null Hypothesis: $H_0\colon \sigma_1^2 = \sigma_2^2, \quad or \quad \sigma_1^2/\sigma_2^2 = 1$

Alternative Hypothesis: $H_A\colon \sigma_1^2 \neq \sigma_2^2, \quad or \quad \sigma_1^2/\sigma_2^2 \neq 1$

Test Statistic $f = \frac{s_1^2}{s_2^2}$

Level of significance $= \alpha$

Critical *F* value (F^*) = *F* with an area of $\alpha/2$ on its right.

Use Excel for *critical F^* value*:

$= \mathrm{F.INV}(1 - \alpha/2\, , df_1\, , df_2)$

P value = $2\ P(F \geq f)$

Use Excel for *P value*:

$= 2 * (1 - \mathrm{F.DIST}(f\ , df_1\, , df_2\, , 1))$

Decision:

Method 1: ***Reject Ho if*** $sample\ f \geq Critical\ F^*$

Method 2: ***Reject Ho if*** $Pvalue \leq \alpha$

picture 8

You can see the result below:

Table A			
	Sample from 2015 Data	Sample from 2016 Data	
Sample Variance	0.04	0.13	
Sample Standard deviation	0.20	0.36	
Sample Size	20	25	
Degrees of Freedom (DF)	19	24	
Hypothesis Testing:			
	Parameter:	Inequality Type	Hypothesized Quotient
Null Hypothesis H_o in terms of Variance:	σ_1^2/σ_2^2	"="	1
Alternative Hypothesis H_a in terms of Variance:	σ_1^2/σ_2^2	"!="	1
Test Statistic:	0.309		
P-value	1.989		
Significance Level:	0.050		
Ctitical Value(s):	0.966		
Decision: Reject H_o?	No	Explain why: Because p-value is bigger than alpha	
	Answer "YES" or "NO"		
Decision in the context of the problem:			
There is not enough evidence to conclude that the two standard deviation are different.			

picture 9

We can also do all of these steps by using F-test tow-sample for variances. The result is:

Table B: Hypothesis Testing by Using the Data Analysis ToolPak:		
F-Test Two-Sample for Variances		
	Variable 1	*Variable 2*
Mean	1.667	1.632
Variance	0.039517	0.12801667
Observations	20	25
df	19	24
F	0.308685	
P(F<=f) one-tail	0.005627	
F Critical one-tail	0.473005	
F Critical two-tail	0.94601	

picture 10

Based on picture 10, because the F critical two-tail is bigger than F, the null hypothesis is not rejected.

In part 6, We are going to use the null hypothesis in order to test our assumption. In here, by using alpha=0.1and the two population variances are known, I tested the hypothesis that the mean of a vehicle per household in 2015 was less than that in 2016. You can find the formulas that I used below:

Z-Test Hypothesis Testing for $\mu_1 - \mu_2$:

Right Tailed problem

Large Independent Samples ($n_1 \geq 30$ and $n_2 \geq 30$)

σ_1 and σ_2 are known

Null Hypothesis: H_0: $\mu_1 - \mu_2 \leq \mu_0$

Alternative Hypothesis: H_A: $\mu_1 - \mu_2 > \mu_0$

μ_0 = ***Hypothesized Difference***

Sample Mean₁= $\bar{x}_1$, Sample Mean₂ = $\bar{x}_2$

Standard Error:

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Test Statistic $z = \frac{(\bar{x}_1-\bar{x}_2)-\mu_0}{\sigma_{\bar{x}_1-\bar{x}_2}}$

Level of significance $= \alpha$

Critical *Z* value (z^*) = *z* with an area of α on its right.

Use Excel for ***critical z^* value***:

$= \text{NORM.S.INV}(1-\alpha)$

P value = $P(Z \geq z)$

Use Excel for ***P value***:

$= 1 - \text{NORM.S.DIST}(z, 1)$

Decision:

Method 1: ***Reject Ho if*** $sample\ z \geq Critical\ z^*$

Method 2: ***Reject Ho if*** $Pvalue \leq \alpha$

picture 11

You can see the result below:

Table A			
	Sample from 2015 Data	Sample from 2016 Data	
Sample Mean	1.71	1.73	
Sample Variance	0.07	0.08	
Sample Size	36.00	34.00	
	Population 1 (NY)	Population 2 (LA)	
Population Variance	0.0852	0.0894	
Sample Means Difference	-0.02	$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$	
Sampling (standard) Error	0.07		
Hypothesis Testing:			
	Parameter:	Inequality Type	Hypothesized Difference
Null Hypothesis H_o:	$\mu_1 - \mu_2$	>=	0
Alternative Hypothesis H_a:	$\mu_1 - \mu_2$	<	0
Test Statistic:	-0.3787		
P-value	0.1		
Significance Level:	0.1000		
Ctitical Value(s):	1.2816		
Decision: Reject H_o?	NO	Explain why: T < T*	
	Answer "YES" or "NO"		
Decision in the context of the problem:			
There is not enough evidence to conclude that the two standard deviation are different.			

picture 12

We can also do all of these steps by using the z-test two sample for mean in the data analytics package:

Table C: Hypothesis Testing by Using the Data Analysis ToolPak:		
z-Test: Two Sample for Means		
	Variable 1	*Variable 2*
Mean	1.708888889	1.733529412
Known Variance	0.07	0.08
Observations	36	34
Hypothesized Mean Difference	0	
z	-0.375878735	
P(Z<=z) one-tail	0.353503525	
z Critical one-tail	1.281551566	
P(Z<=z) two-tail	0.70700705	
z Critical two-tail	1.644853627	

picture 13

Based on picture 13, because the Z is smaller than the Z critical one-tail, the null hypothesis is not rejected.

In the past 7, I am going to test the hypothesis that the mean number of vehicles per household in 2015 was less than that in 2016. And our assumptions are the two population variances are unequal and alpha= 0.1
You can see the formula is below:

T-Test Hypothesis Testing for $\mu_1 - \mu_2$:

Two Tailed problem

Samples are independent
σ_1 and σ_2 are unknown, but are assumed to be unequal
If any of the two samples is small, then the two populations are normally or approximately normally distributed

Null Hypothesis: $H_0\colon\ \mu_1 - \mu_2 = \mu_0$

Alternative Hypothesis: $H_A\colon\ \mu_1 - \mu_2 \neq \mu_0$

μ_0 = *Hypothesized Difference*
Sample Mean1= $\bar{x}_1$, Sample Mean2 = $\bar{x}_2$
Sample Variance1= s_1^2, Sample Variance2 = s_2^2

Degrees of Freedom $df = \dfrac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\dfrac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \dfrac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$

Standard Error:

$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

Test Statistic $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sigma_{\bar{x}_1 - \bar{x}_2}}$

Level of significance $= \alpha$
Critical T value 1 $(t_1^*) = t$ **with an area of $\alpha/2$ on its right.**
Critical T value 2 $(t_2^*) = t$ **with an area of $\alpha/2$ on its left.**

Use Excel for critical t_1^* and t_2^* values:
t_1^*: $= \mathrm{T.INV}\left(1 - \frac{\alpha}{2}, df\right)$
t_2^*: $= \mathrm{T.INV}\left(\frac{\alpha}{2}, df\right)$

P value = $2\,P(T \geq t)$ **if the test statistic t is positive**
P value = $2\,P(T \leq t)$ **if the test statistic t is negative**
Use Excel for *P values*:
If $t > 0$, $= 2 * \left(1 - \mathrm{T.DIST}(t, df, 1)\right)$
If $t < 0$, $= 2 * \mathrm{T.DIST}(t, df, 1)$

Decision:
Method 1: Reject Ho if:
test statistic $t \geq t_1^*$ for $t > 0$
test statistic $t \leq t_2^*$ for $t < 0$
Method 2: ***Reject Ho if*** $Pvalue \leq \alpha$

picture 14

The results are:

Table A			
	Sample from 2015 Data	Sample from 2016 Data	
Sample Mean	1.67	1.63	
Sample Variance	0.04	0.13	
Sample Size	20.00	25.00	
Sample Means Difference	0.04	$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$	
Sampling (standard) Error	0.08		
Degrees of Freedom (DF)	38.80	$df = \frac{(s_1^2/n_1+s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$	
Hypothesis Testing (Population variances are known to be unequal):			
	Parameter:	Inequality Type	Hypothesized Difference
Null Hypothesis H_o:	$\mu_1 - \mu_2$	>=	0
Alternative Hypothesis H_a:	$\mu_1 - \mu_2$	<	0
Test Statistic:	0.4155		
P-value	0.1000		
Significance Level:	0.1000		
Ctitical Value(s):	-1.3042		
Decision: Reject H_o?	No	Explain why: because t > t*	
	Answer "YES" or "NO"		
Decision in the context of the problem:			
There is not enough evidence to conclude that the two standard deviation are different.			

picture 15

we can also do these steps by using t-Test: Two-Sample Assuming Unequal Variances in data analytics package:

Table B: Hypothesis Testing by Using the Data Analysis ToolPak:		
t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	*Variable 2*
Mean	1.667	1.632
Variance	0.039516842	0.128016667
Observations	20	25
Hypothesized Mean Difference	0	
df	39	
t Stat	0.415475742	
P(T<=t) one-tail	0.3400354	
t Critical one-tail	1.303638589	
P(T<=t) two-tail	0.6800708	
t Critical two-tail	1.684875122	

picture 16

In here, we can see because the T critical one-tail is bigger than the T stat, the null hypothesis is not rejected.

I also did all of these steps for two tail test. You can see the result below:

Table C			
	Sample from 2015 Data	Sample from 2016 Data	
Sample Mean	1.67	1.63	
Sample Variance	0.04	0.13	
Sample Size	20.00	25.00	
Sample Means Difference	0.04	$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	
Sampling (standard) Error	0.01		
Degrees of Freedom (DF)	43.00	$df = n_1 + n_2 - 2$	
Hypothesis Testing (Population variances are known to be equal):			
	Parameter:	Inequality Type	Hypothesized Difference
Null Hypothesis H_o:	$\mu_1 - \mu_2$	" = "	0
Alternative Hypothesis H_a:	$\mu_1 - \mu_2$	" != "	0
Test Statistic:	2.6247		
P-value	0.1000		
Significance Level:	0.1000		
Ctitical Value(s):	1.6811		
Decision: Reject H_o?	yes	Explain why: t> t* so the null hypothesis is rejected	
	Answer "YES" or "NO"		
Decision in the context of the problem:			
There is not enough evidence to conclude that the two standard deviation are different.			

picture 17

Table D: Hypothesis Testing by Using the Data Analysis ToolPak:		
t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	*Variable 2*
Mean	1.667	1.632
Variance	0.039516842	0.128016667
Observations	20	25
Hypothesized Mean Difference	0	
df	39	
t Stat	0.415475742	
P(T<=t) one-tail	0.3400354	
t Critical one-tail	1.303638589	
P(T<=t) two-tail	0.6800708	
t Critical two-tail	1.684875122	

picture 18

In the past 8, I used sampling in data analytic package in order to randomly select 24 dependent pairs of 2015 and 2016 number of vehicles per household with each

pair of values corresponding to the same jurisdiction. Then I calculated the mean, variance, size sample and sampling standard error and degrees of freedom.
After that, I am going to test the hypothesis that the mean number of vehicles in 2015 was less than that of 2016 and the alpha=0.05.
You can see the formulas that I used below:

Mean of differences = $\bar{D}$
Standard Deviation of differences = s_D
Standard Error: $s_{\bar{D}} = \frac{s_D}{\sqrt{n}}$
Test Statistic: $t = \frac{\bar{D}-\mu_0}{s_{\bar{D}}}$
Level of Significance $= \alpha$

Critical T value (t*) = t with an area of α on its left.
Use Excel for critical t* value:
=T.INV(α,df)
P value = P($T \leq t$)
Use Excel for P value:
=T.DIST(t, df ,1)

Decision:
Method 1: Reject H_0 if sample $t \leq$ Critical t*
Method 2: Reject H_0 if P value $\leq \alpha$

picture 19

And the result is:

	Differences *d*		
Sample Mean	-0.02		
Sample Variance	0.10		
Sample Size	24		
Sampling (standard) Error	0.02		
Degrees of Freedom (DF)	23		
Hypothesis Testing:			
	Parameter:	**Inequality Type**	**Hypothesized Difference**
Null Hypothesis H_o:	μ_d	>=	0
Alternative Hypothesis H_a:	μ_d	<	0
Test Statistic:	-1.2165		
P-value	0.0500		
Significance Level:	0.0500		
Ctitical Value(s):	-1.7139		
Decision: Reject H_o?	No	Explain why: T*<T	
	Answer "YES" or "NO"		
Decision in the context of the problem:			
There is not enough evidence to conclude that the two standard deviation are different.			

picture 20

We can also do all of these steps by using t-test: paired two samples for means in a data analytics package. The result is:

Table B: Hypothesis Testing by Using the Data Analysis ToolPak:			
t-Test: Paired Two Sample for Means			
	Variable 1	*Variable 2*	
Mean	1.743333333	1.7675	
Variance	0.062249275	0.081323913	
Observations	24	24	
Pearson Correlation	0.325031813		
Hypothesized Mean Differen	0		
df	23		
t Stat	-0.37950636		
P(T<=t) one-tail	0.353896661		
t Critical one-tail	1.713871528		
P(T<=t) two-tail	0.707793321		
t Critical two-tail	2.06865761		

picture 21

Based on these results, because the t critical on the tail is bigger than t stat, the null hypothesis is not rejected.

In the past 9, I am going to test the p1 (the proportion of the 2015 households with more than 2 vehicles) and p2 (the proportion of the 2016 households with more than 2 vehicles) are different and alpha= 0.05.
The formulas that I used are:

Z-Test Hypothesis Testing for $p_1 - p_2$:

Left Tailed problems

Sample sizes: n_1 and n_2

Samples are randomly selected and independent of each other.

Sample Proportions: $\hat{p}_1 = \frac{x_1}{n_1}$ and $\hat{p}_2 = \frac{x_2}{n_2}$

Requirements: $n_1\hat{p}_1 \geq 10$, $n_1\hat{q}_1 \geq 10$, $n_2\hat{p}_2 \geq 10$, and $n_2\hat{q}_2 \geq 10$

Pooled Estimate of proportions:

$$\bar{p} = \frac{n_1\,\hat{p}_1 + n_2\,\hat{p}_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\bar{q} = 1 - \bar{p}$$

Null Hypothesis: H_0: $p_1 - p_2 \geq p_0$

Alternative Hypothesis: H_A: $p_1 - p_2 < p_0$

Standard Error $\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\bar{p}\,\bar{q}\,\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$

Test Statistic $z = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sigma_{\hat{p}_1-\hat{p}_2}}$

Level of significance $= \alpha$

Critical *Z* value (z^*) = *z* with an area of α on its left.

Use Excel for ***critical*** z^* ***value***:

$= \mathrm{NORM.S.INV}(\alpha)$

P value = $P(Z \leq z)$

Use Excel for ***P value***:

$= \mathrm{NORM.S.DIST}(z, 1)$

Decision:

Method 1: ***Reject Ho if*** $\textit{sample } z \leq \textit{Critical } z^*$

Method 2: ***Reject Ho if*** $\textit{Pvalue} \leq \alpha$

picture 22

The result is:

	Table A		
	Sample from 2015 Data	**Sample from 2016 Data**	
Sample Size	**120.0000**	**125.0000**	
Sample Number of Success *	**19**	**18**	
Sample Proportion of Success *	**0.1583**	**0.0000**	
Sample Proportion of Failure	**0.8417**	**0.0000**	
* Note: A proportion of Success is the proportion of households with more than two vehicles.			
Difference of Sample Proportions	**0.0143**		
***p-bar*: Pooled estimate for proportions**	**0.081306306**		
q-bar* = 1 - *p-bar	**0.918693694**	$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\bar{p}\,\bar{q}\,\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$	
Sampling (standard) Error	**0.0899**		
Hypothesis Testing:			
	Parameter:	**Inequality Type**	**Hypothesized Proportions Difference**
Null Hypothesis H_o:	$p_1 - p_2$	**" = "**	**0**
Alternative Hypothesis H_a:	$p_1 - p_2$	**" != "**	**0**
Test Statistic:	**1.7613**		
P-value	**0.0500**		
Significance Level:	**0.0500**		
Ctitical Value(s):	**-1.6449**		
Decision: Reject H_o?	**No**	**Explain why: z>z***	
	Answer "YES" or "NO"		
Decision in the context of the problem:			
There is not enough evidence to conclude that the two proportion of succes are different.			

picture 23

Based on picture 23, there is not enough evidence to conclude that the two

Conclusion

In this assignment, I learn about two sample confidence intervals and hypothesis testing for a large and small sample, dependent or independent, and how to compare their means, standard deviation or proportion of success and draw a conclusion out of the null hypothesis testing.