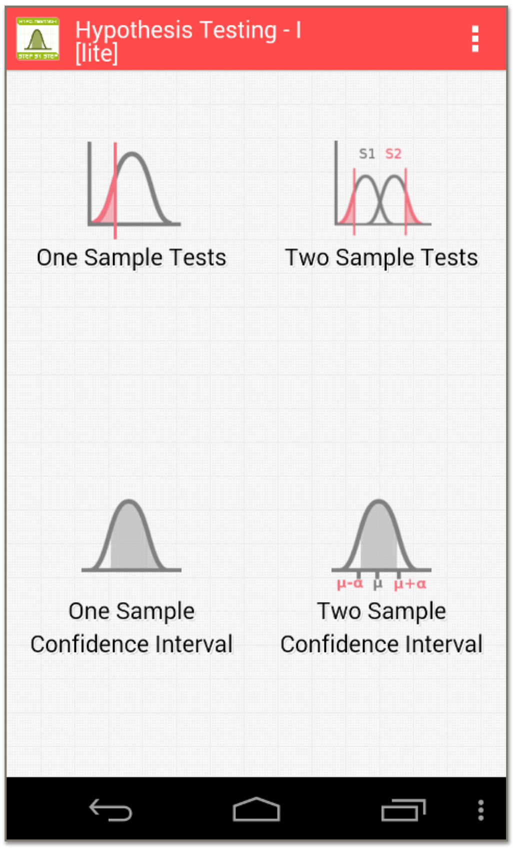
project week 4

One-sample Confidence Intervals & Hypothesis Testing



Maryam Heidari

Winter 2019

Project week 4

One-sample Confidence Intervals & Hypothesis Testing

Introduction

In this Assignment, we are going to learn confidence level and hypothesis testing for one sample. We use a one-sample statistic to estimate a population means by constructing confidence intervals by using an appropriate probability distribution, a population proportion by constructing confidence intervals by using an appropriate probability distribution, a population variance (or standard deviation) by constructing confidence intervals by using an appropriate probability distribution.

Discussion

In part one, I just calculate mean, variance and standard deviation of the population by using =average, =var and =SQRT functions. You can see the result in the below table:

Table A	
Population Mean:	**1.72**
Population Variance:	**0.09**
Population Standard deviation:	**0.30**

Table 1

In part two, after calculating mean, standard deviation, size, I used descriptive statistics in the data analytics package in order to calculate sampling standard error. Here because the sample size is bigger than 30, I used the z-test. Then calculate the z and margin error by using two function of =NORM.S.INV((1+c)/2) and (z*s)/sqrt(n) for different level of confidence. After I found the margin of error, I can calculate CI lower and upper limit. And at the end, I calculate the sample size for margin error of 0.05 for the different confidence level, by using the formula (z*s/E)^2 and I also used =ceiling in order to round the result to an upper number. you can find the result of these activities in the picture below:

Table A	
Sample Mean	1.80
Sample Standard deviation	0.23
Sample size	50
Sampling (standard) Error	0.03

Table B					
Confidence Level CI	z value corresponding to the CI level	Margin of Error	CI Lower Limit	CI Upper Limit	CI Width
92%	1.751	0.058	1.74	1.85	0.12
96%	2.054	0.068	1.73	1.86	0.14
98%	2.326	0.077	1.72	1.87	0.15
	* Note: You should decide whether a *z* or a *t* value should be used.				

Table C		
Confidence Level CI	Desired Margin of Error	Minimum Sample Size Needed
92%	0.05	68
96%	0.05	93
98%	0.05	119

Table D	
Population Mean:	1.72

Picture 1

Based on these result we can see only in confidence level of 98% the CI width contain the mean of the population and other confidence levels show use the range higher than the mean of the population, which is not suitable.

In part three, I used sampling in the data analytic package to make a sample of 18. And then, I calculated the mean, standard deviation, size, degree of freedom and sampling error for the sample by using the same function. In this part, because the sample size is smaller than 30, I used the t-test. Then calculate the t and margin error by using two functions of =T.INV((1+c)/2, df) and =(t*s)/sqrt(n). After calculating the margin of error we can calculate the CI lower and upper limit. You can see the results of these function in the picture below:

Table A	
Sample Mean	1.77
Sample Standard deviation	0.14
Sample Size	18
Degrees of Freedom (DF)	17
Sampling Error	0

Table B

Confidence Level CI	t value corresponding to the CI level	Margin of Error	CI Lower Limit	CI Upper Limit	CI Width
92%	1.862	0.060	1.71	1.83	0.12
96%	2.224	0.071	1.70	1.84	0.14
98%	2.567	0.082	1.69	1.85	0.16

* Note: You should decide whether a ***z*** or a ***t*** value should be used.

Table C	
Population Mean:	1.72

Picture 2

Based on these result we can see for all of the confidence levels, the CI width contains the mean of the population.

In part 4, I used sampling in data analytic package in order to make random sample size 130. And then I used COUNTIF function in order to count how many of my sample is bigger than 2. In order to calculate the proportion of success, I divided the number of success by the total number of sample. When I subtract one from the proportion of success, we can find the proportion of failure. After that, I used =Norm.S.INV ((1+c)/2) and = z*sqrt ($\hat{p}$ q/n) in order to calculate z and margin of error for the different confidence level. After I find the margin of error, I calculated the upper and lower CI limit for each confidence level. In the end, by using the E=0.08 and z, I calculate the size of the sample for a different confidence level.

Table A	
Sample Proportion of Success	0.17
Sample Proportion of Failure	0.83
Sample Size	130
Sampling Error	0.03

Table B

Confidence Level CI	z value corresponding to the CI level	Margin of Error	CI Lower Limit	CI Upper Limit	CI Width
90%	1.64	0.05	0.12	0.22	0.11
95%	1.96	0.06	0.10	0.23	0.13
99%	2.58	0.08	0.08	0.25	0.17

Table C

Confidence Level CI	Desired Margin of Error	Minimum Sample Size Needed
90%	0.08	60
95%	0.08	85
99%	0.08	146

Table D	
Population Proportion of Success	84.54%
Population Proportion of Failure	15.46%

You can find the results of all of these steps in the picture below:

picture 3

As you can see, none of the CI widths of different confidence level contain the population proportion.

In part 5, I used sampling in data analytic package in order to make random sample size 20. And then used the chi-squared distribution formula in this part which is:

$$\textbf{CI Lower Limit for } \sigma\textbf{:} \quad \sqrt{\frac{(n-1)\,s^2}{\chi^2_R}}$$

$$\textbf{CI Upper Limit for } \sigma\textbf{:} \quad \sqrt{\frac{(n-1)\,s^2}{\chi^2_L}}$$

$$\chi_R{}^2 := CHISQ.INV\left(\frac{1+c}{2}, DF\right)$$

$$\chi_L{}^2 := CHISQ.INV\left(\frac{1-c}{2}, DF\right)$$

$$DF = n - 1;$$

picture 4

Table A

Sample Variance	0.05
Sample Standard deviation	0.23
Sample Size	20
Degrees of Freedom (DF)	19

Table B

Confidence Level CI	χ^2_{Left}	χ^2_{Right}	CI Lower for Variance	CI Upper for Variance	CI Lower for Standard deviation	CI Upper for Standard deviation
92%	9.698	31.037	0.034	0.107	0.183	0.328
96%	8.567	33.687	0.031	0.122	0.176	0.349
98%	7.633	36.191	0.029	0.137	0.170	0.370

* Note: You should decide whether a ***z*** or a ***t*** value should be used.

Table C

Population Standard Deviation:	0.30

you can see the result in the picture below:

picture 5

In part 6, I test the hypothesis that the population average number of vehicles per household is more than 1.5 vehicles. For this part, I assume the null hypothesis that the μ<= 1.5 and calculated the z test statistic and critical value and p-value for z-

Table A				
Sample Mean	1.80		Population Mean:	1.72
Sample Standard deviation	0.23			
Sample size	50			
Sampling (standard) Error	0.03			
Hypothesis Testing:				
	Parameter:	Inequality Type	Hypothesized mean	
Null Hypothesis H_o:	μ	<=	1.5	
Alternative Hypothesis H_a:	μ	>	1.5	
Test Statistic:	8.9343			
P-value	0.0000000			
Significance Level:	0.0300			
Ctitical Value(s):	1.8808			
Decision: Reject H_o?	yes	Explain why: because significance level is more than p-value, the null hypothesis is rejected		
	Answer "YES" or "NO"			
Decision in the context of the problem:				
There is sufficient evidence supporting the claim that the population average number of vehicles per household is more than 1.5 vehicles				

test. You can see the result below:

picture 6

As you can see, because p-value is smaller than the significance level (0.03), the null hypothesis is rejected, and we can say, there is sufficient evidence supporting the claim that the population average number of vehicles per household is more than 1.5 vehicles.

In past 7, I did all of the steps in part 5, but for T-test. You can see the formula and the result in the pictures below:

T Distrubution

=T.DIST(*t* , DF , 1)

= T.INV(*Probability* , *DF*)

CI: $(\bar{x} - E, \bar{x} + E)$

$$E = t_c \frac{s}{\sqrt{n}}$$

$$DF = n - 1;$$

Test Statistic t	$t = \frac{(\bar{x} - \mu_o)\sqrt{n}}{s}$

Table A				
Sample Mean	1.77		Population Mean:	1.72
Sample Standard deviation	0.14			
Sample Size	18			
Degrees of Freedom (DF)	17.00			
Sampling (standard) Error	0.03			
Hypothesis Testing:				
	Parameter:	Inequality Type	Hypothesized mean	
Null Hypothesis H_o:	μ	>=	2	
Alternative Hypothesis H_a:	μ	<	2	
Test Statistic:	-7.1094			
P-value	0.0000			
Significance Level:	0.0500			
Ctitical Value(s):	-2.1098			
Decision: Reject H_o?	Yes	Explain why: The p-value is smaller then signigicance level, so the null hypothesis is rejected		
	Answer "YES" or "NO"			
Decision in the context of the problem:				
There is sufficient evidence supporting the claim that the population average number of vehicles per household is less than 2 vehicles.				

picture 7

picture 8

Because the p-value is smaller than the significance level (0.05), the null hypothesis is rejected. So we can say there is sufficient evidence supporting the claim that the population average number of vehicles per household is less than 2 vehicles.

In part 8, In here I test the null hypothesis for the proportion of success. You can see the formula and the result is below:

CI: $(\hat{p} - E, \hat{p} + E)$

$$E = z_c\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

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

Determining the sample size: $n = \hat{p}\,\hat{q}\left(\frac{z_c}{E}\right)^2.$

Test Statistic z	$z = \dfrac{\hat{p} - p_o}{\sqrt{\dfrac{p_o q_o}{n}}}$

Table A			
Sample Proportion of Success	0.1692		
Sample Proportion of Failure	0.8308		
Sample Size	130		
Hypothesis Testing:			
	Parameter:	Inequality Type	Hypothesized Proportion
Null Hypothesis H_o:	*p*	>=	0.2
Alternative Hypothesis H_a:	*p*	<	0.2
Sampling (standard) Error	0.0306		
Test Statistic:	-0.8771		
P-value	0.1902		
Significance Level:	0.0100		
Ctitical Value(s):	0.0125		
Decision: Reject H_o?	No	Explain why: The p-value is more then signigicance level, so the null hypothesis is not rejected	
	Answer "YES" or "NO"		
Decision in the context of the problem:			
There is not sufficient evidence supporting the claim that so less than 20% of the population own more than two vehicles.			

Column1	
Mean	1.691154
Standard Error	0.030617
Median	1.735
Mode	1.91
Standard Deviation	0.349082
Sample Variance	0.121858
Kurtosis	0.440257
Skewness	-0.81616
Range	1.66
Minimum	0.63
Maximum	2.29
Sum	219.85
Count	130
	3E+235

picture 9

picture 10

Because the p-value is smaller than the significance level (0.01), the null hypothesis is not rejected. So we can say there is not sufficient evidence supporting the claim that the population average number of vehicles per household is less than 2 vehicles.

In part 9, I test the null hypothesis for variances. You can see the formula and results below:

Chi-squared Distribution

=CHISQ.DIST(χ^2 , *DF* , 1)

= CHISQ.INV(*Probability* , *DF*)

CI Lower Limit for σ: $\sqrt{\frac{(n-1)\,s^2}{\chi^2_R}}$

CI Upper Limit for σ: $\sqrt{\frac{(n-1)\,s^2}{\chi^2_L}}$

Test Statistic χ^2 $\quad \chi^2 = \frac{(n-1)s^2}{\sigma_o^2}$

Table A			
Sample Variance	0.05		
Sample Standard deviation	0.23		
Sample Size	20		
Degrees of Freedom (DF)	19.00		
Hypothesis Testing:			
	Parameter:	Inequality Type	Hypothesized Variance
Null Hypothesis H_o in terms of Variance:	σ^2	"="	0.3
Alternative Hypothesis H_a in terms of Variance:	σ^2	"!="	0.3
Test Statistic:	3.4740		
P-value	0.0001		
Significance Level:	0.1000		
Ctitical Value(s):	17.5894		
Decision: Reject H_o?	yes	Explain why: The p-value is smaller then signigicance level, so the null hypothesis is rejected	
	Answer "YES" or "NO"		
Decision in the context of the problem:			
the standard deviation of the number of vehicles per household is different from 0.3			

picture 11

picture 12

Because the p-value is smaller than the significance level (0.1), the null hypothesis is rejected, so we can say the standard deviation of the number of vehicles per household is different from 0.3.

conclusion

In this assignment, I learn about z-test and t-test for a small and big sample. And I learn about null hypothesis and how to analyze the results of it.