JMP® ACADEMIC CASE STUDY

JMP036: Online Mortgage Application

Inferential Statistics: Estimation of Population Parameters

Produced by

Kevin Potcner, JMP Global Academic Team
kevin.potcner@jmp.com

Online Mortgage Application

Inferential Statistics: Estimation of Population Parameters

Key ideas:

This case study requires the use of various inferential statistical techniques to evaluate the potential improvement to the UI design of an online mortgage application process. These techniques include confidence interval for a population proportion; hypothesis test and confidence intervals for the population mean; evaluating the feasibility of using the normal probability distribution model; prediction intervals; interval estimates of cumulative percents; and percentiles. Comparisons of results are done across different groups, though no formal statistical comparative technique is used.

Background

An executive team for a mortgage lending company is concerned that the current online application process is a bit too difficult and complicated to complete, thus possibly frustrating potential customers. Since the current online application was implemented two years ago, approximately 10,000 applications have been submitted. This historic data has an average completion time of 72 minutes, with a standard deviation of 14 minutes. In addition, only 20% of the applicants were able to complete the application within an hour, and it required up to 90 minutes before at least 90% of the applicants completed the application.

The company's design team was asked to create a new UI design for the online application process that would be simpler than the current process, resulting in a more user-friendly experience for the customer and a reduced completion time.

Prior to rolling out the new design to all applicants, the UI design team wants to test it with a small focus group of 50 customers. Each customer was asked to provide a rating of the usability of the design. In addition, the time (in minutes) they took to complete the application process was recorded.

The Task

There are two primary questions the team would like to see if the data can answer.

1. Will at least 50% of applicants find the new design "Easier to Use" than expected?

2. Will the new UI design reduce the time it takes applicants to complete the process?
   a) Will the average completion time be less than 72 minutes?
   b) Will more than 20% of the applicants be able to complete the application within 60 minutes?
   c) Will 90% of the applicants be able to complete the application in less than 90 minutes?

The Data   Online Mortgage Application_1.jmp

Completion Time	Time (in minutes) it takes the applicant to complete the application process
Usability	Applicant's rating of ease of use as compared to their expectations (Easier, Same, More Difficult)

Analysis

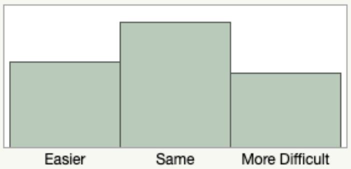
Inference for population proportions

We begin by summarizing the Usability rating by the participants in the focus group. Exhibit 1 displays a bar graph and frequency table of the results.

Exhibit 1 Distribution

Distributions

Usability



Frequencies

Level	Count	Prob
Easier	15	0.300
Same	22	0.440
More Difficult	13	0.260
Total	50	1.000

N Missing 0
3 Levels

To create, Analyze>Distribution. Select Usability as the Y variable.

Of the 50 applicants in the focus group, 15 (30%) found the new design “Easier to Use” than they expected, 22 (44%) rated it as “Same,” and 13 (26%) found it “More Difficult.”

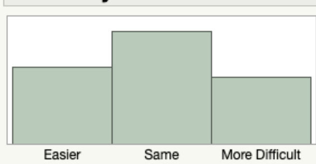
One of the goals was to create a design where at least 50% of the applicants found it was easier to use than they expected. Only 30% of the participants in the focus group found it Easier to Use than expected, so clearly there is no statistical evidence in the data to conclude that the goal would be met for a population of applicants.

To use these results to create an estimate of the proportion of a population of applicants who would find it Easier to Use, we construct 95% confidence intervals for each. The 95% confidence intervals have been added in Exhibit 2.

Exhibit 2 Distribution

Distributions

Usability



Frequencies

Level	Count	Prob
Easier	15	0.300
Same	22	0.440
More Difficult	13	0.260
Total	50	1.000

N Missing 0
3 Levels

Confidence Intervals

Level	Count	Prob	Lower CI	Upper CI	1-Alpha
Easier	15	0.300	0.191	0.438	0.950
Same	22	0.440	0.312	0.577	0.950
More Difficult	13	0.260	0.159	0.396	0.950
Total	50				

Note: Computed using score confidence intervals.

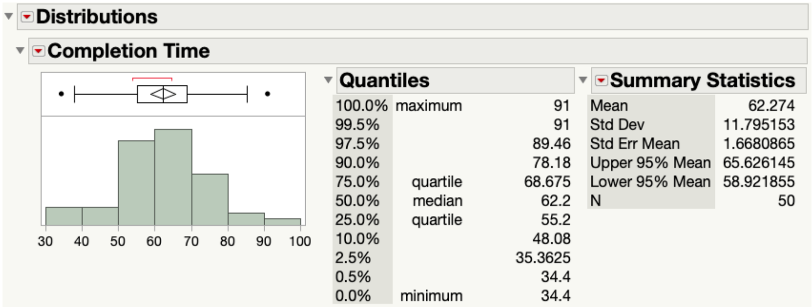
To create, Analyze>Distribution. Select Usability as the Y variable.

Using these confidence intervals, we can conclude with 95% confidence that it is estimated that between 19% and 44% of a population of applicants would find the design easy to use. You may notice that the imprecision in estimating this population parameter is quite large as measured via the width of the confidence interval (44 – 19 = 25). The lower bound of the confidence would need to be larger than 50% in order for there to be the statistical evidence needed to conclude that at least 50% of the proportion of a population of applicants would find it easier to use.

Graphical and numerical summarization of a continuous variable

We begin by summarizing the data using the Distribution Platform. Exhibit 3 displays that output for all 50 observations.

Exhibit 3 Distribution



Quantiles

100.0%	maximum	91
99.5%		91
97.5%		89.46
90.0%		78.18
75.0%	quartile	68.675
50.0%	median	62.2
25.0%	quartile	55.2
10.0%		48.08
2.5%		35.3625
0.5%		34.4
0.0%	minimum	34.4

Summary Statistics

Mean	62.274
Std Dev	11.795153
Std Err Mean	1.6680865
Upper 95% Mean	65.626145
Lower 95% Mean	58.921855
N	50

To create, Analyze>Distribution. Select Completion Time as the Y variable.

We can see that the distribution for these data of completion time for the new UI design is symmetric with the familiar normal distribution shape. The sample mean is 62.27, the sample standard deviation is 11.80, and the 90th percentile is 78.18. Notice that the sample standard deviation is very close to the standard deviation of 12 from the approximately 10,000 applicants that used the current design, suggesting that the variation has not changed. Also notice that the sample average of these 50 applications is lower than 72 minutes (the average completion time for all those that used the current design).

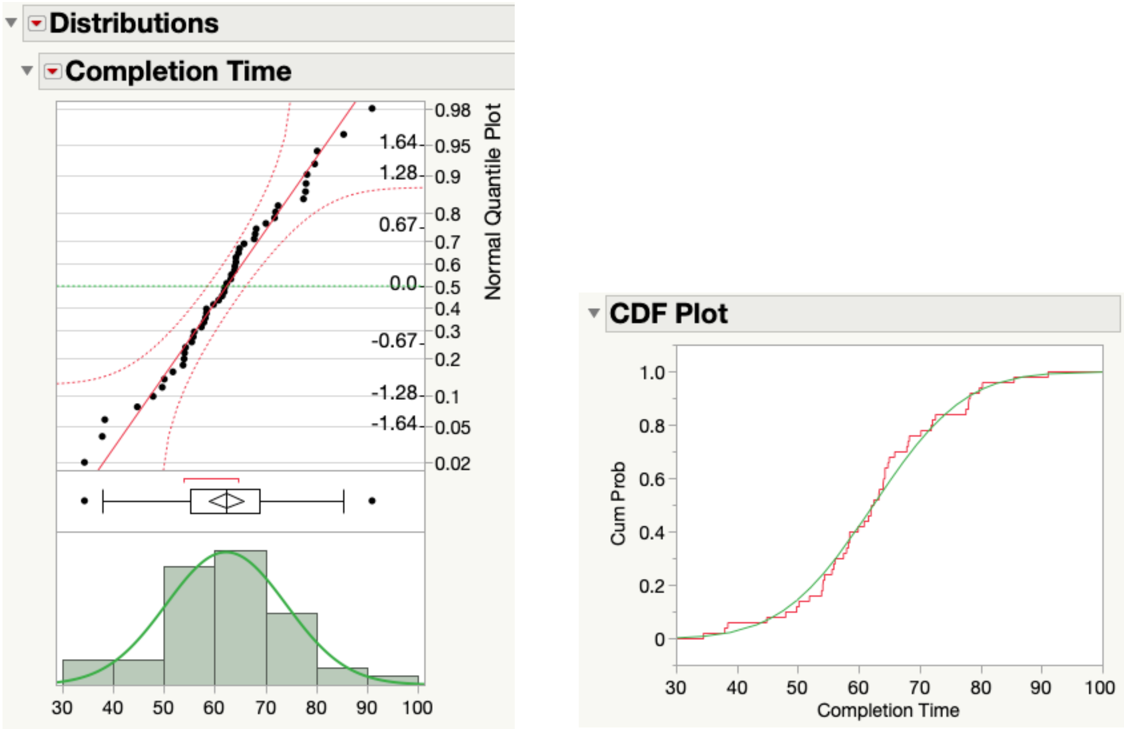
We should, however, reserve concluding just yet that we have produced the necessary statistically significant evidence to indicate that an improvement has been made to a population of applicants in general. We will need to conduct formal inference-based statistical analyses to determine that.

Modeling a variable with a probability distribution

Many statistical procedures are based upon certain assumptions. One common assumption is related to a probability distribution that is an appropriate model for the variable being studied.

Here we will evaluate if the normal distribution is a good model to use. An examination of the histogram in Exhibit 3 shows that this is a reasonable assumption. We can expand upon that evaluation by fitting a normal distribution to these data. Exhibit 4 shows the histogram with a normal distribution added along with a normal quantile plot and a cumulative distribution function plot (CDF).

Exhibit 4 Fitting a Distribution to the Data



To create, choose Continuous Fit>Fit Normal, as well as Normal Quantile Plot and CDF Plot under the red triangle next to the variable name Completion Time.

Comparing the histogram to the normal distribution curve suggests a good fit to the data. The normal quantile plot is another way to evaluate how well the probability model fits the data. This graph plots the quantile scores (Y axis) versus the actual data values (X axis). The Y axis is scaled so that the quantiles for a normal distribution correspond to the straight line shown. If the points roughly correspond to that straight line and are within the confidence bounds shown, as we see here, the assumption of normality of the underlying distribution is appropriate.

The CDF is another way to view how the data compares to a normal distribution. The green line represents the cumulative proportion of an assumed normal distribution as Completion Time goes from low to high. The red step function is the cumulative percentiles of the data. In the same way the quantile plot is evaluated, we look to see if the percentiles of the data follow the percentiles of the normal distribution. The data follows that normal CDF quite well.

Sometimes reaching this conclusion is not as easy as it is here, since the data points might vary quite a bit from the straight line in the normal quantile plot or there might not be enough data for a histogram to reveal a clear shape. In these cases, one can conduct a formal statistical test to assess how well the data match the normal probability distribution.

This can be stated as:

H_O : The random variable Completion Time can be well modeled with a normal distribution
H_A : The normal distribution is not a good model for the random variable Completion Time

Exhibit 5 shows the results of this statistical test.

Exhibit 5 Goodness of Fit Test

Fitted Normal Distribution

Parameter		Estimate	Std Error	Lower 95%	Upper 95%
Location	μ	62.274	1.6680865	59.00461	65.54339
Dispersion	σ	11.795153	0.3487112	11.131116	12.257075

Measures	
-2*LogLikelihood	387.66272
AICc	391.91804
BIC	395.48677

Goodness-of-Fit Test

	W	Prob<W
Shapiro-Wilk	0.9842587	0.7393

	A2	Simulated p-Value
Anderson-Darling	0.3458201	0.4860

To create, choose Goodness of Fit under the red triangle next to the Fitted Normal Distribution output.

Two test statistics are shown (Shapiro-Wilk and Anderson-Darling). These two tests are based on different mathematical and statistical methodologies but address the same question. Both of them have large p-values (0.7393 for the Shapiro-Wilk test and 0.4860 for the Anderson-Darling test), indicating that there is no statistical evidence against the hypothesis that the normal distribution is a good probability model to use. This was as expected as our previous graphical-based examination indicated the normal distribution fits the data well.

Hypothesis test for the population mean

Inferential statistical techniques are focused on generalizing to a larger population based upon the analysis of a sample. We can think of the 50 people in the focus group as just a sample from a large population of applicants. The primary question is not if the sample mean of these 50 people is different than 72 minutes, but if we can conclude, based on the sample information, that the average completion of a large population of applicants using the new design is less than 72 minutes.

We could formally state this question of interest regarding the mean as follows:

$$H_O : \mu = 72$$
$$H_A : \mu < 72$$

The null hypothesis H_O states that average Completion Time for a population of applicants using the new design is equal to 72 (the average completion for the current design). The alternative hypothesis H_A states that the population mean is less than 72.

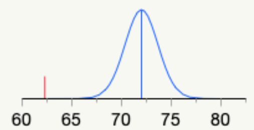
There are two primary ways this can be done: hypothesis test and confidence interval. We will illustrate the hypothesis test first. Exhibit 6 displays the results of this test known as the t-test.

Exhibit 6 Hypothesis Test for the Population Mean

Test Mean

Hypothesized Value	72
Actual Estimate	62.274
DF	49
Std Dev	11.7952

	t Test
Test Statistic	-5.8306
Prob > |t|	<.0001*
Prob > t	1.0000
Prob < t	<.0001*



To create, choose Test Mean from the red triangle next to variable name Completion Time in the Distribution output that was done to produce results shown in Exhibit 3. Enter 72 for the Specify Hypothesized Mean field.

Since we are performing a lower one-tailed hypothesis test, the p-value to examine for the test is Prob < t. The p-values for this one-tailed test is < 0.0001. That is, there is less than a 0.01% chance that we would obtain a sample mean of 62.3 or less from a sample of 50 applicants if, in fact, the population mean were equal to 72. As you can see, this is highly unlikely. Therefore, we would conclude that we have statistically significant evidence to reject the null hypothesis that the population mean is equal to 72 in favor of the alternative hypothesis that the population mean is less than 72. How different? We will evaluate that next.

Confidence interval for the population mean

A confidence interval estimate for a parameter is a range of plausible values for that parameter. The confidence level is a measure of certainty that this interval estimate (calculated from these data) contain the unknown value of the parameter, in this case the average completion time of a population of applicants using the new design.

Exhibit 7 below displays a 95% confidence interval for the population mean based on these 50 applicants.

Exhibit 7 Confidence Interval for the Population Mean

Summary Statistics

Mean	62.274
Std Dev	11.795153
Std Err Mean	1.6680865
Upper 95% Mean	65.626145
Lower 95% Mean	58.921855
N	50

The confidence interval is part of the output from the distribution analysis used to produce Exhibit 3.

The 95% confidence interval is [58.9 , 65.6], which is interpreted as being 95% confident that the average completion time of a population of applicants using the new design is between 59 and 66 minutes. Note that this range of plausible values does not include 72, the average completion time of the approximately 10,000 applicants that used the current design. Thus, we have statistical evidence at the 95% level that average completion time has been reduced. The confidence interval answers both the question about whether there is or isn't statistically significant evidence to believe the alternative hypothesis, and where we estimate that population mean to be. The confidence interval also provides us with a measure of the uncertainty in the estimate, which is quantified through the width of the confidence interval: 65.6 – 58.9 = 6.7. Sometimes confidence intervals are written as point estimate +/- margin of error. For our analysis, this would be 62.3 +/- 3.3. The confidence interval directly answers the question of how much different from 72 we believe the average completion time for the new design is. For our data, that would be 72 – 66 = 6 to 72 – 59 = 13 minutes. In other words, we estimate, with 95% confidence, that the new design will take between 6 and 13 minutes less time to complete on average.

Technical note: You may have noticed that there was also a confidence interval provided for the population mean when we were evaluating the normal distribution fit in Exhibit 5. That confidence interval is [59.0 , 65.5], which is extremely close but not exactly the same as the confidence interval shown in Exhibit 7. The estimation procedure used to produce the results in Exhibit 5 is based upon a technique called maximum likelihood estimation, which is not the same estimation technique used to produce the results in Exhibit 7.

The inferential procedures used to construct a confidence interval and to perform a hypothesis test for the population mean are based on the assumption that the underlying distribution of the population is normally distributed. We had evaluated this earlier and concluded that the normal distribution was an appropriate model to use. It is important to note however that the inferential procedures for a population mean are not very sensitive to this assumption, especially with the sample size of 50 that we have in our study. It would take very extreme departures from this assumption to cause any concern.

Prediction interval

Another useful interval estimate is the prediction interval. A prediction interval provides a lower and upper bound estimate of the completion time for an individual applicant. Exhibit 8 displays a prediction interval.

Exhibit 8 Prediction Interval

Prediction Interval

Parameter	Future N	Lower PI	Upper PI	1-Alpha
Individual	1	38.33489	86.21311	0.950

To create, choose Prediction Interval from the red triangle next to variable name Completion Time in the Distribution output that was done to produce results shown in Exhibit 1. Use 0.95 for the Confidence Level. Enter 1 for Number of Future Samples. Choose Two-Sided.

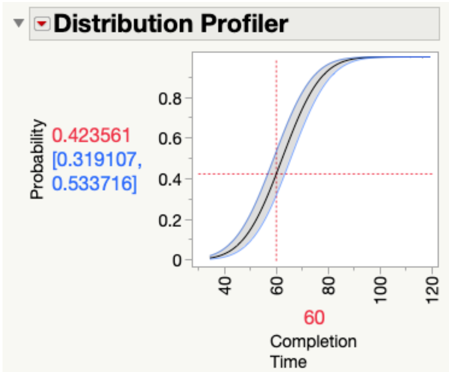
The prediction interval is interpreted as having 95% confidence that an individual applicant will complete the application somewhere between 38.3 and 86.2 minutes. Note that the width of this prediction interval (82.6 – 38.3 = 44.3) reflects the level of uncertainty in predicting the outcome of an individual applicant.

Unlike the confidence interval, the prediction interval is very sensitive to the underlying probability distribution being used to model the random variable. We had evaluated this distributional assumption earlier and found that the normal distribution was a good model to use so we have no cause for concern.

Estimating cumulative percents

Recall that only 20% of the all the applicants that had used the current design were able to complete the application within an hour. To evaluate if there is any evidence for an improvement using this new design, we will use the results of the focus group to estimate the proportion of a population that would complete the application within 60 minutes. From the normal distribution fit output shown in Exhibits 4 and 5, an interactive Profiler tool can be accessed that will allow us to generate estimates of cumulative percents along with corresponding confidence intervals. Exhibit 9 displays that Profiler.

Exhibit 9 Profiler



To create, choose Profiler under the red triangle next to the Fitted Normal Distribution output.

The Profiler allows us to pick a Completion Time by either dragging the red dotted line or typing in the desired Completion Time (60 minutes was chosen here). The point estimate and the corresponding confidence interval are shown on the left, which is interpreted as 95% confidence that between 32% to 53% of the applicants using the new design would complete the application in 60 minutes or less. Notice that the lower bound of 32% exceeds the 20% completion rate from the historic data. Thus, we have statistical evidence to indicate an improvement in the new design as measured by the proportion of applicants that can complete the application within 60 minutes.

Estimating quantiles

Another question to evaluate if there has been an improvement using the new design is to see if the Completion Time for 90% of the applicants is less than the historic data (i.e., 90 minutes). This measure is called the 90th quantile. In Exhibit 1, we had produced sample statistics for some of the quantiles, which are the ones displayed by default: 0.5th, 2.5th, 10th, 25th, 50th, 75th, 90th, 97.5th, and 99.5th quantiles. There we saw that the 90th quantile is 78.2 minutes. Though this point estimate is less than 90, we need to generate a confidence interval to incorporate the statistical uncertainty. Exhibit 10 shows point estimates and 95% confidence intervals for the 90th quantile.

Exhibit 10 Distribution with Custom Quantiles

Custom Quantiles

Quantiles

Quantile	Estimate	Lower 95%	Upper 95%	Actual Coverage
90%	78.18	72	91	97.03

Smoothed Empirical Likelihood Quantiles

Quantile	Estimate	Lower 95%	Upper 95%
90%	78.5711	72.5559	85.166

To create, choose Display Options>Custom Quantiles under the red triangle next to the Completion Time title. Enter the values 0.90.

Technical Note: Two methods of estimating quantiles are displayed. The first is based on calculations from the data via a non-parametric method (i.e., not assuming a particular probably distribution) and the second is based on a function (kernel density) fit to the data.

Using the smooth empirical likelihood estimation method, we would interpret the 90% quantile estimate as: It is estimated, with 95% confidence, that 90% of the applicants would complete the application in at least 73 minutes and at most 85 minutes. Notice that upper bound of the confidence interval is less than 90 minutes, the 90th quantile of the historic data. Thus, we have produced the statistical evidence to indicate that there has been a reduction in this characteristic of interest.

Summary

Statistical insights

Our analyses have not demonstrated any statistical evidence to conclude that at least 50% of a population of applicants using this new design would find it Easier to Use. In fact, it was estimated to only be 19% to 44%.

We have shown, however, that there has been an improvement in the average time it would take for applicants to complete the application process from 72 minutes to somewhere between 59 and 66 minutes. This is an improvement of 6 to 13 minutes.

We also estimated the range of time by which an individual applicant would complete the new application (38 to 86 minutes).

Through estimates of the cumulative percent for 60 minutes and the 90th quantile, we also demonstrated statistical evidence of an improvement. Specifically, we estimated that between 32% and 53% of the applicants using the new design would complete the application in 60 minutes or less – an improvement over the 20% using the current design. We also estimated that 90% of the applicants would have the application completed by at best 73 minutes and at most 85 minutes – a reduction from the 90 minutes from the historic data.

Implications and further study

Though our data has found the evidence to indicate a statistically significant reduction in the average completion time from 72 to somewhere between 59 and 66 minutes, this difference may not be practically important. This practical importance is decided not by the statistical analysis but by objectives and goals set by the company. For example, if management had wanted the improvement to be at least 15 minutes on average, then a confidence interval would need to have an upper bound less than 72 – 15 = 57, such as [49, 56].

Similarly, though we demonstrated statistically significant evidence to show an improvement in the cumulative percent at 60 minutes and the 90th quantile, this may not achieve the amount of improvement desired.

Recall that a sample of 50 people was used in the focus group to evaluate the new design. A key assumption in all of our analyses is that these 50 people are representative of the population the inference is being made to (i.e., all future applicants using the new design). This is not assessed through statistical analysis tools but instead by evaluating how the sample of applicants were chosen, their characteristics, etc., in order to decide if it's reasonable to assume this focus group is indeed a representative sample from the desired population.

We did not incorporate into our analyses important characteristics of these participants, for example, whether or not the applicant had ever completed a mortgage application in the past. It would be very informative to see how this prior experience impacts the Usability Rating and Completion Time. To explore this, a binary variable (Yes, No) indicating whether each participant in the focus group had prior experience in completing a mortgage application has been added to the data. The exercises will ask you to incorporate this new information to perform a similar set of analyses as was demonstrated.

Exercises

Use the data in file Online Mortgage Application_2.jmp data set to answer the following questions:

These data contain the variables:

Usability	Applicant's rating of ease of use as compared to their expectations (Easier, Same, More Difficult)
Completion Time	Time (in minutes) it takes the applicant to complete the application process
Prior Experience	Binary variable indicating if the participant had prior experience in completing a mortgage application (Yes, No).

1. The results of the focus study estimated that between 19% and 44% of a population of applicants would find the design Easy to Use, which was a disappointing result as the goal was at least 50%. The management followed up with asking us if it can be estimated that 50% or more of applicants would at least find it Not More Difficult to Use. Create a 95% confidence interval for the proportion of a population of applicants who would not find it more difficult. Was there a satisfactory result for this inquiry?

   Hint: This will require recoding the variable Usability to contain only two values: More Difficult, Not More Difficult.

2. Our analyses did not incorporate the participant's experience with having completed a mortgage application in the past. Create 95% confidence intervals for the proportion of a population of applicants who would find the application Not More Difficult to Use for those that have completed a mortgage application prior and for those that have not. Create a graph that displays the confidence intervals together comparing them to 50%. Does there appear to be an impact on the Usability Rating based upon the applicant's prior experience completing an application? If so, explain that difference and how it compares to the value 50%.

   Hint: This will require creating a new data table that has the values for the point estimates and the lower and upper bounds of each confidence interval in one column and another column identifying the two experience levels.

3. Our analyses on Completion Time had also not incorporated the participant's prior experience with completing a mortgage application. Calculate 95% confidence intervals for the population mean Completion Time for each level of Prior Experience (Yes, No). Use Graph Builder to create a graph that displays Completion Time separately for each of the two levels of Prior Experience. Display 95% confidence intervals for each, comparing them to 72 minutes (the average for the current design). Does there appear to be an impact on Completion Time based upon the applicant's prior experience completing an application? If so, explain that difference and how the estimates of the population mean Completion Time using the new design compares to 72 minutes.

4. Evaluate if the normal probability distribution is a good model to use for the Completion Time variable for each of the two groups of Prior Experience. Is there any evidence to suggest that the normal distribution may not be a good model to use?

5. Estimate (with 95% confidence) the range in time that an individual applicant will complete the application process, both for someone who has completed a mortgage application prior and someone who has not. Provide an interpretation for these interval estimates.

6. Estimate (with 95% confidence) the proportion of applicants of each level of prior experience who will complete the application within 60 minutes. Provide an interpretation for these interval estimates.

7. Estimate (with 95% confidence) the time at which 90% of applicants will be able to complete the application process for both the applicants who have completed a mortgage application prior and for those who have not. Provide an interpretation for these interval estimates.

8. The design team is curious to know if those who took a long time to complete the application also tended to provide a low Usability rating. Create a graphical display that shows the Completion Times for each of the three Usability Ratings (Easier, Same, More Difficult). Do applicants who take longer to complete the application tend to also rate the usability of the design less favorably? Augment the graph to include a reference line at 60, 72, and 90 minutes. Using different colors, delineate those applicants that have prior experience in a mortgage application and those that do not. Interpret the findings.

9. Create an Executive Summary that contains one or two visualizations and five to seven bullet points summarizing the key findings.