Analysis of Kansas Workers' Compensation Indemnity Claims 2014-2017

Workers' Compensation Division, Kansas Department of Labor

Beckett Malinowski* & David Sprick**

May 2023

Division of Workers' Compensation, Kansas Department of Labor

401 SW Topeka BLVD, Topeka, KS 66603

*Data Analyst, Data, Systems, and Statistics Unit

**Public Service Executive, Operations Section

Introduction

For the first time the Division is reporting a multi-year Kansas workers' compensation claims analysis for calendar years 2014-2017. Workers' compensation claims data is reported to the Division by insurers (self-insured and group-funded risk pools included) on all claims with indemnity payouts for claimants. Settlements with indemnity payments are also reported to the Division through its Electronic Data Interchange (EDI) or computer-to-computer data collection program. See the EDI program web site (https://www.dol.ks.gov/wc/oscar-and-edi) for more information on the method and data collection process between insurers and the Division. The sample size of claims collected over the 2014-2017 timeframe is 22,345 or n= 22,345. All claims included in the sample had to close out in CY2014-2017 to be included. For this quantitative data analysis, we will examine 10 variables routinely collected on closed workers' compensation claims.

Variables

A quantitative variable is a systematized understanding of similarities and differences among observed phenomena- as captured by your data. All of the 11 variables below used in our analysis are classified as continuous variables (measured at the interval level of measurement or ratio level) and defined below.

Claim Duration is defined as the date of injury until date claim is closed by the insurer.

Total Cost of the Claim is defined as the summation of all costs associated with the claim-medical, indemnity, legal etc.

Total Medical Payments or the sum of all medical payments (doctors, hospitals, medical mileage etc.) associated with the claim.

Total Indemnity Payments or the sum of all indemnity benefit type payments associated with the claim.

Total Lump Sum Payments or all payments paid out at once through a settlement.

Claimant Legal Payments Paid to date or the sum of all payments for legal services for the claimant associated with the claim.

Employer Legal Payments Paid to date or all the sum of payments for legal services for the employer associated with the claim.

Time Away from Work or the timeframe from the date of the injury until return-to-work date.

Time to Notify Insurer or the timeframe from date of injury until date insurer notified of injury.

Time to Medical Recovery or the timeframe from the date of injury until date of maximum medical improvement.

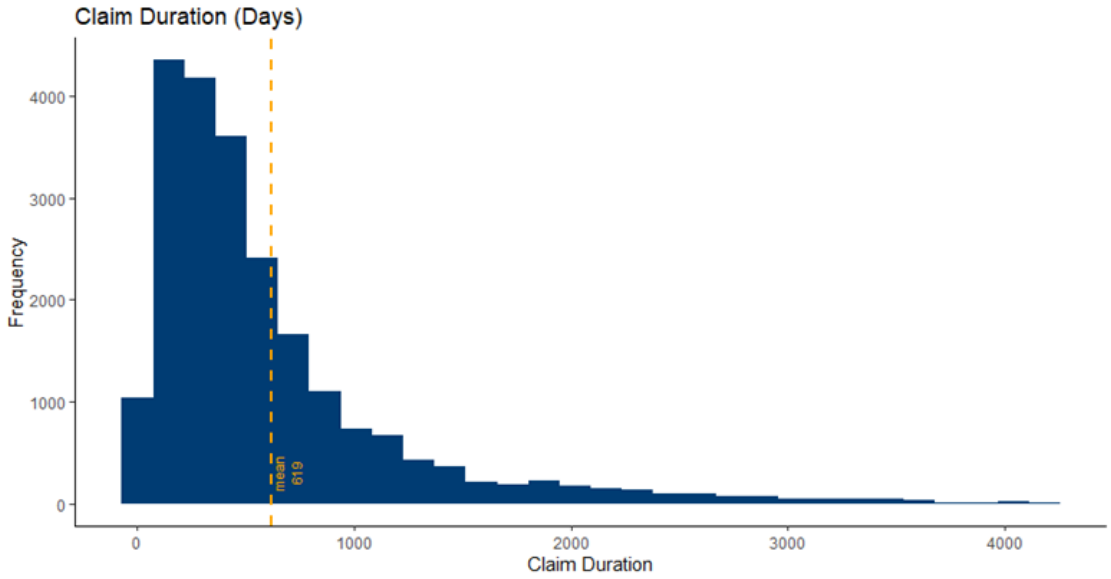
Time to First Payment or time from date of injury until time of the first indemnity payment made by the insurer.

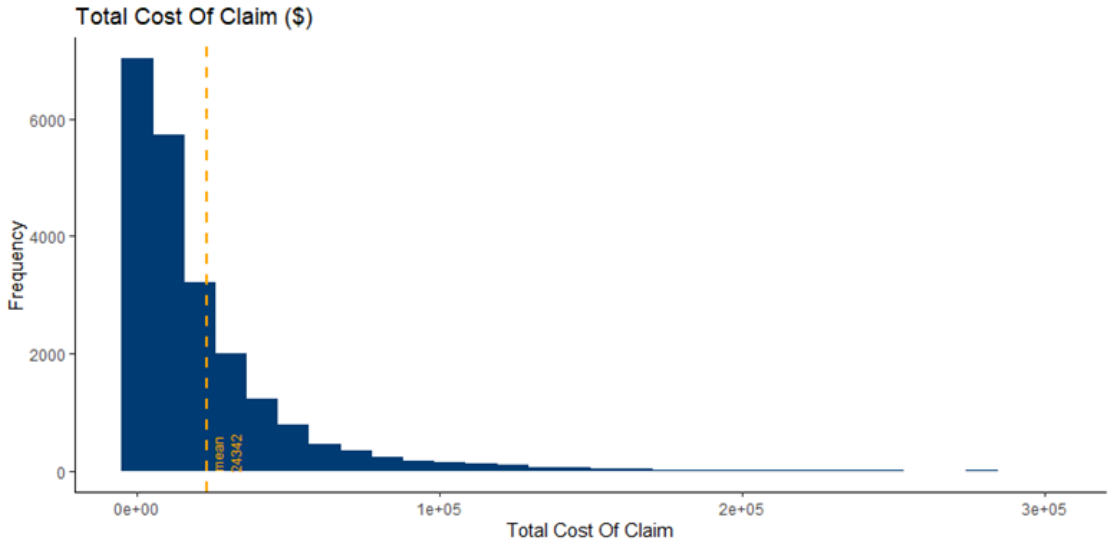
Exploratory Data Analysis Using Data Visualizations

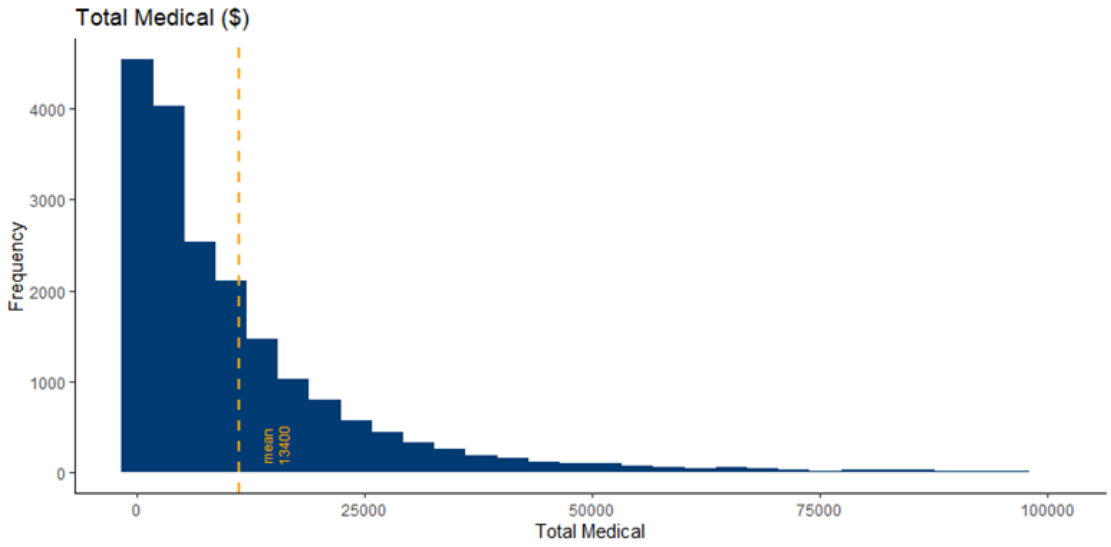
First, we will conduct an exploratory data analysis of the large sample of closed Kansas claims by summarizing its main characteristics through data visualizations. Below are histograms that reveal the shape of the distribution for each of the 11 continuous claims variables. If a variable is normally distributed, it would have a classic bell-shaped curve. None of our 11 variables are normally distributed. In fact, all are positively skewed distributions (see below). A positively skewed distribution will have a few observations on the right side of the distribution that will distort the mean. A negatively skewed distribution is the opposite- a few observations on the left side of the distribution distorting the mean. The overwhelming number of observations in our histograms of the variables bunch up nearer to the zero value on the x-axis whether measured in units of time or dollars. The mean scores are displayed in the histogram for reference.

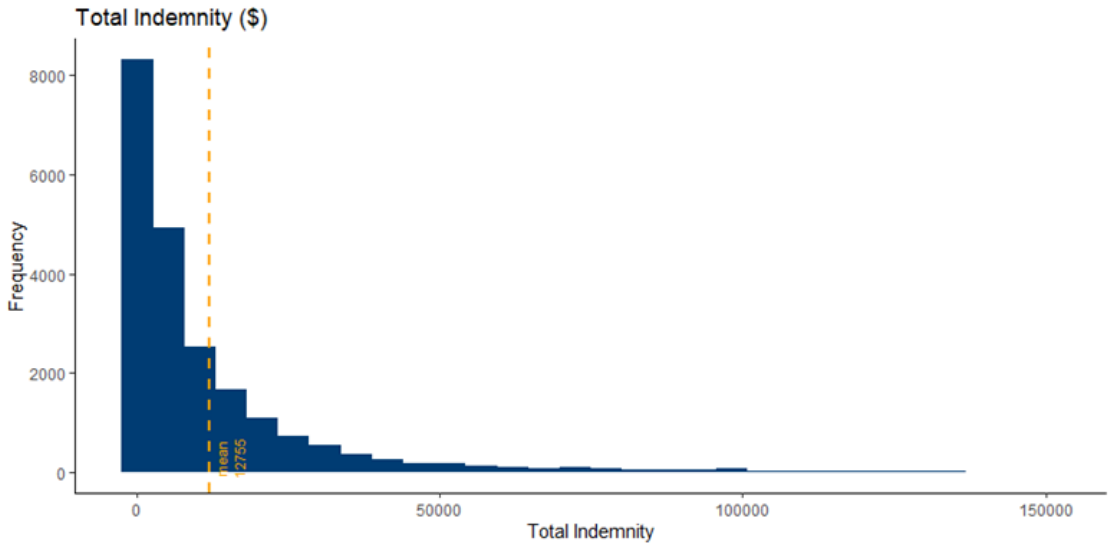
Descriptive Data analysis

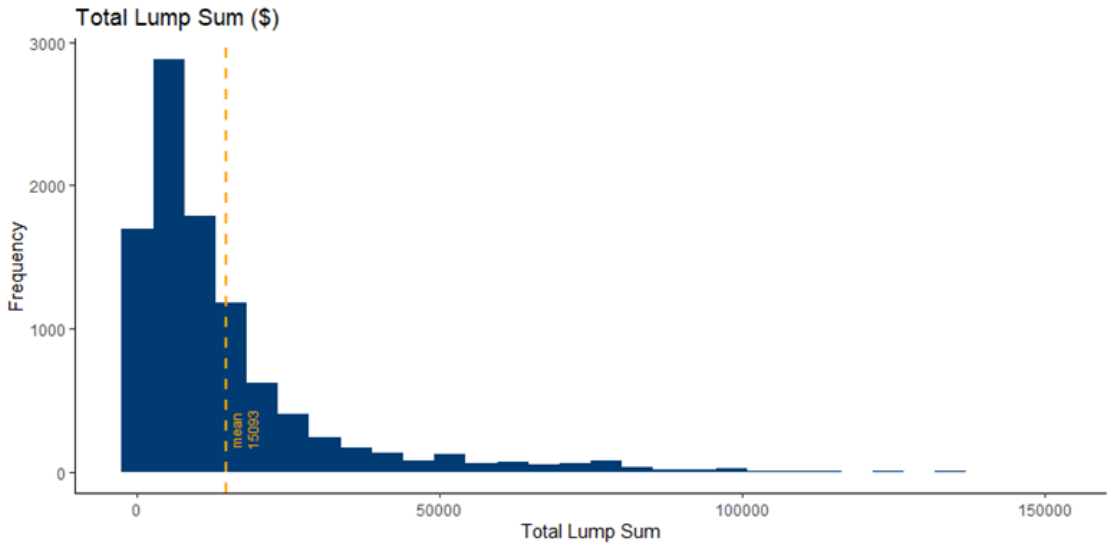
Figure 1.1 - Figure 1.11. Histogram per variable 2014-2017.

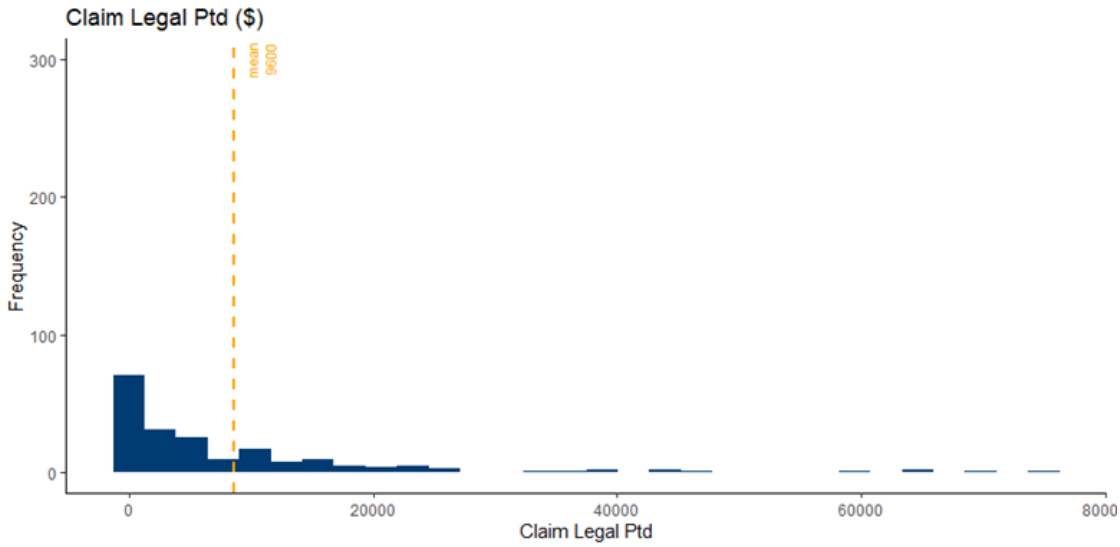


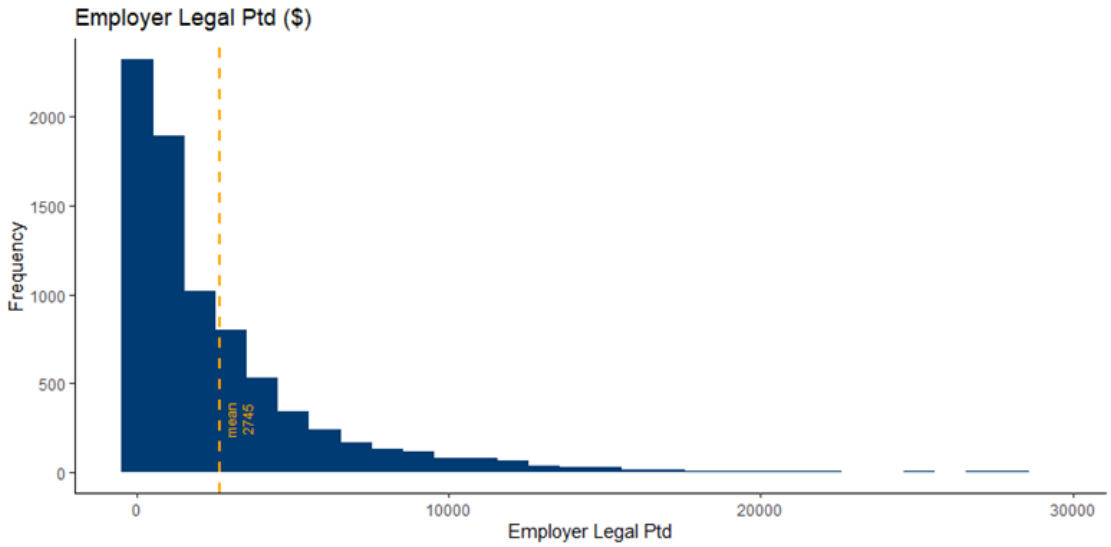


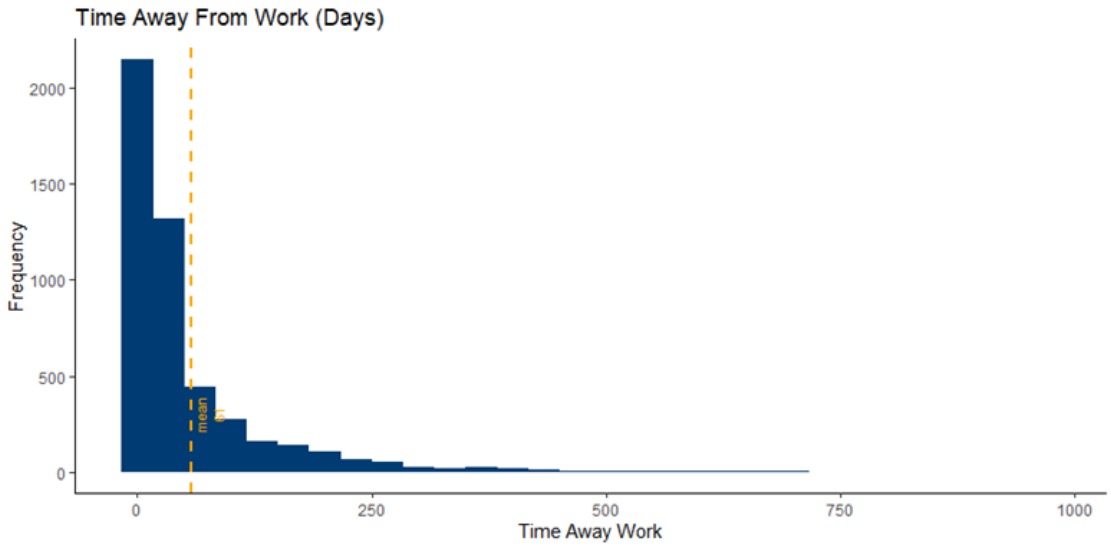

Total Medical ($)
Frequency
4000
3000
2000
1000
0
mean
13400
0
25000
50000
75000
100000
Total Medical

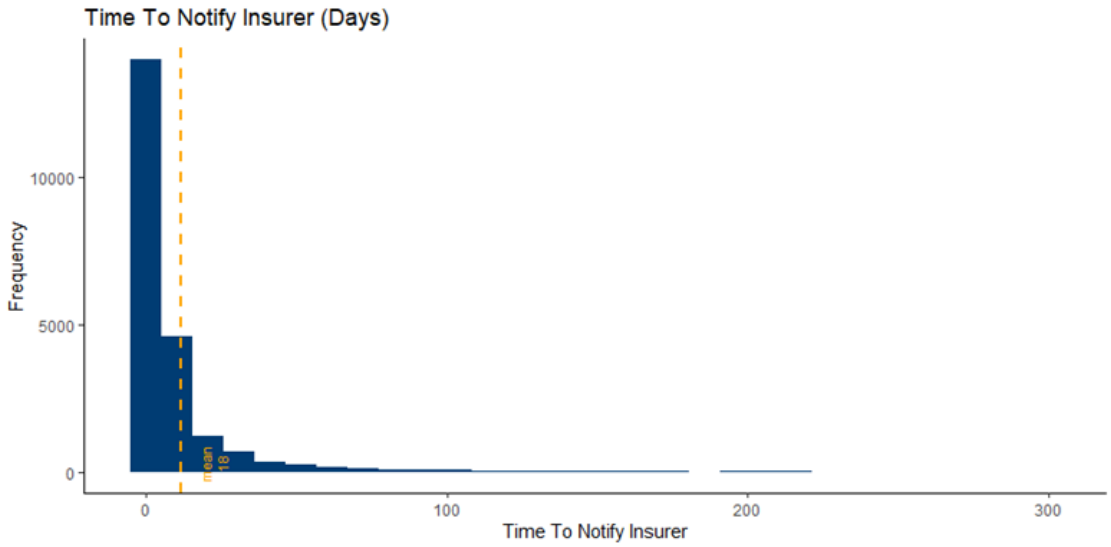

Total Indemnity ($)
Frequency
8000
6000
4000
2000
0
mean
12755
0
50000
100000
150000
Total Indemnity

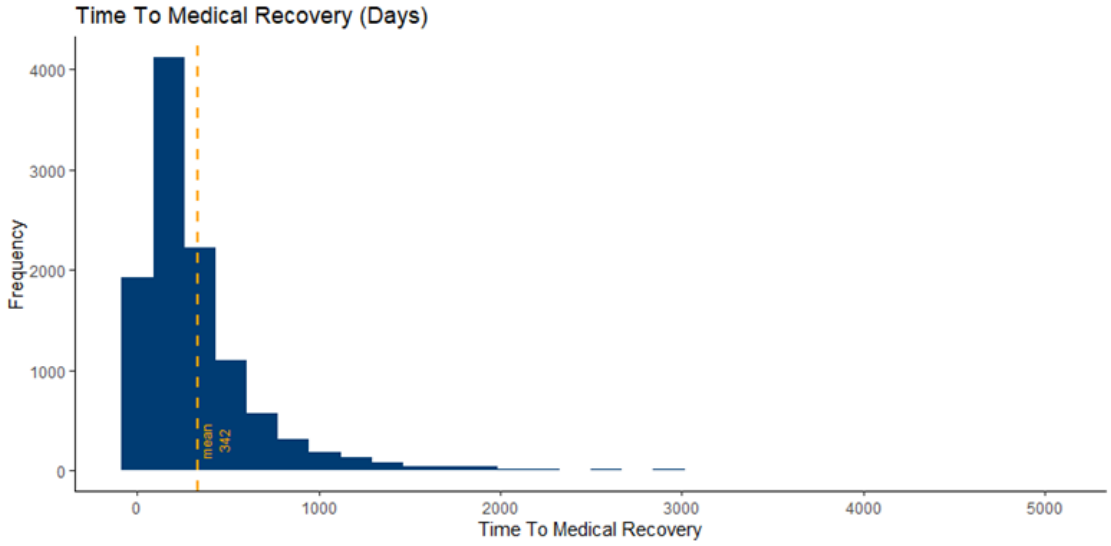

Total Lump Sum ($)
Frequency
3000
2000
1000
0
mean
15093
0
50000
100000
150000
Total Lump Sum

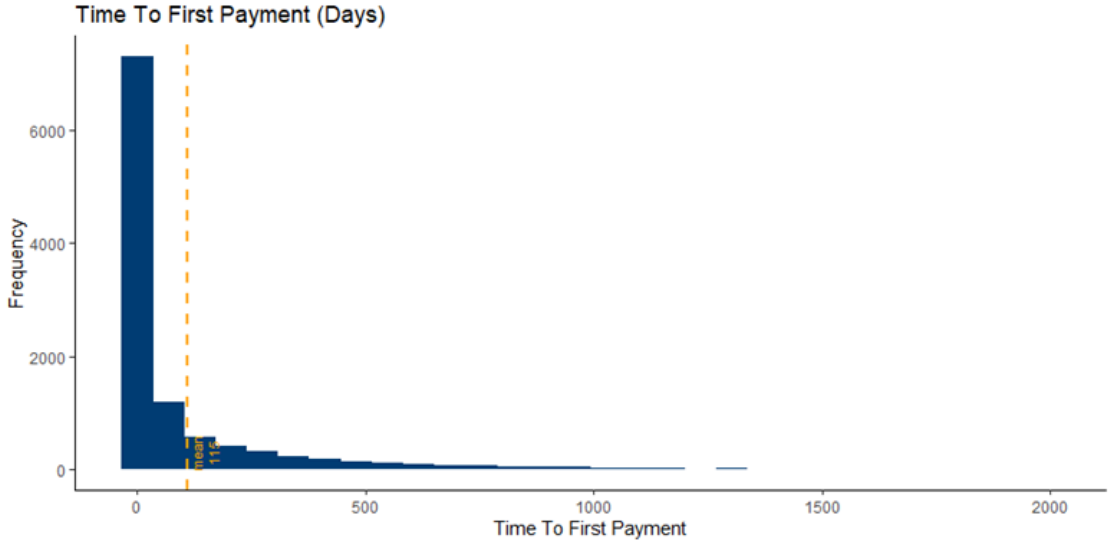

Claim Legal Ptd ($)
Frequency
Claim Legal Ptd
mean
9600
0
100
200
300
0
20000
40000
60000
8000


Employer Legal Ptd ($)
Frequency
Employer Legal Ptd
mean
2745
0
500
1000
1500
2000
0
10000
20000
30000


Time Away From Work (Days)
Frequency
Time Away Work
mean
0
500
1000
1500
2000
0
250
500
750
1000


Time To Notify Insurer (Days)
Frequency
10000
5000
0
mean
18
0
100
200
300
Time To Notify Insurer

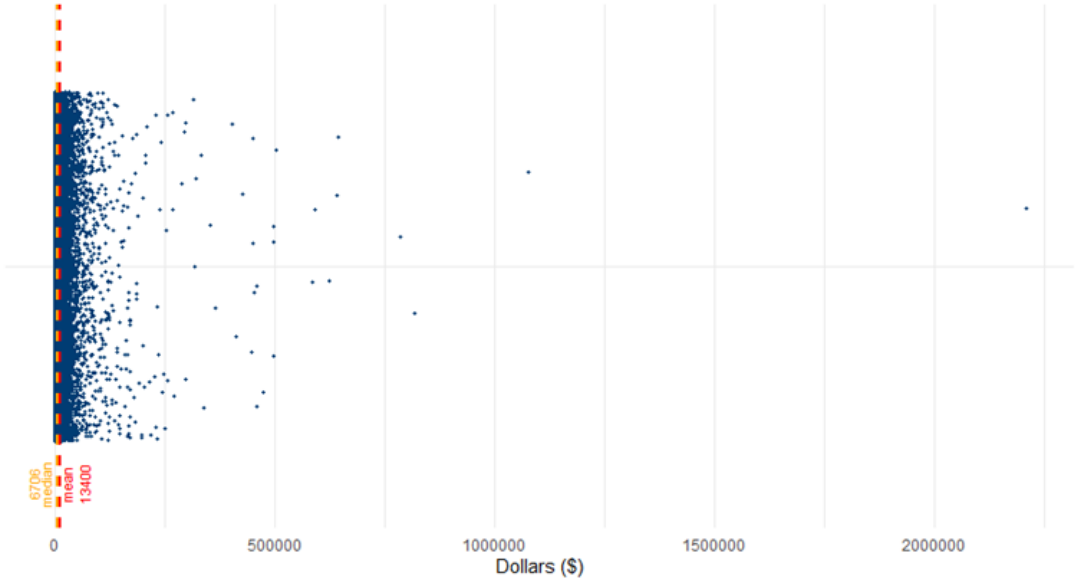

Time To Medical Recovery (Days)
Frequency
4000
3000
2000
1000
0
mean
342
0
1000
2000
3000
4000
5000
Time To Medical Recovery


Time To First Payment (Days)
Frequency
6000
4000
2000
0
mean
114
0
500
1000
1500
2000
Time To First Payment

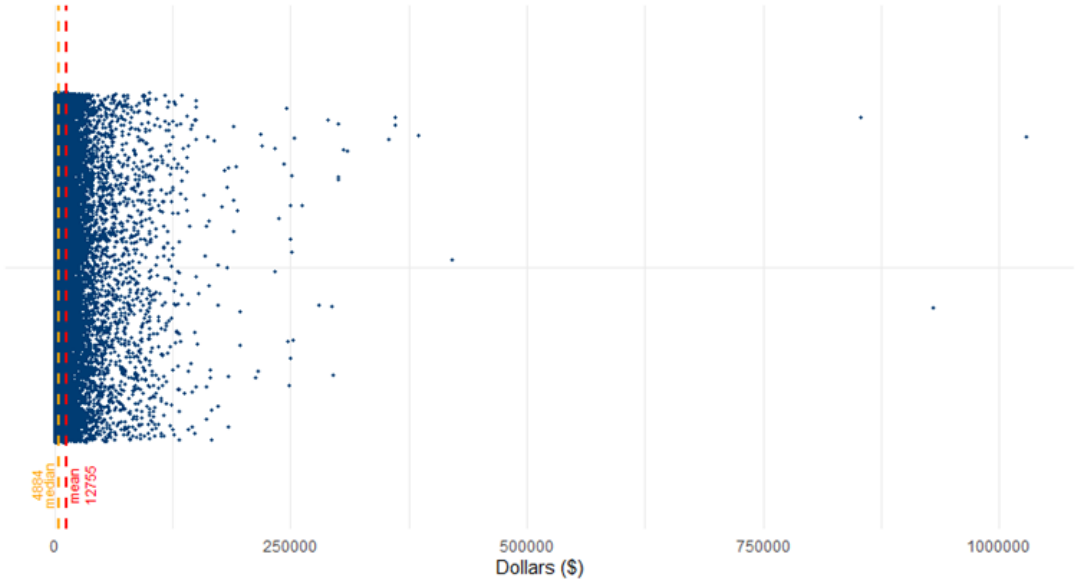
Below are jitter plots that display the data observations relevant to the measure used for the variables on the x axis. In our case it will be in days or dollars- depending on the variable under study. The claim duration jitter plot shows the overwhelming number of observations cluster about the median and median with some very large outlier values of duration beyond 10,000 days. The mean and median is shown in each jitter plot for reference for all 11 variables in the study below. These jitter plots and histograms were created using R/RStudio statistical software and the ggplot2 R package.

Figure 2.1 - Figure 2.11. Jitterplots per variable 2014-2017.

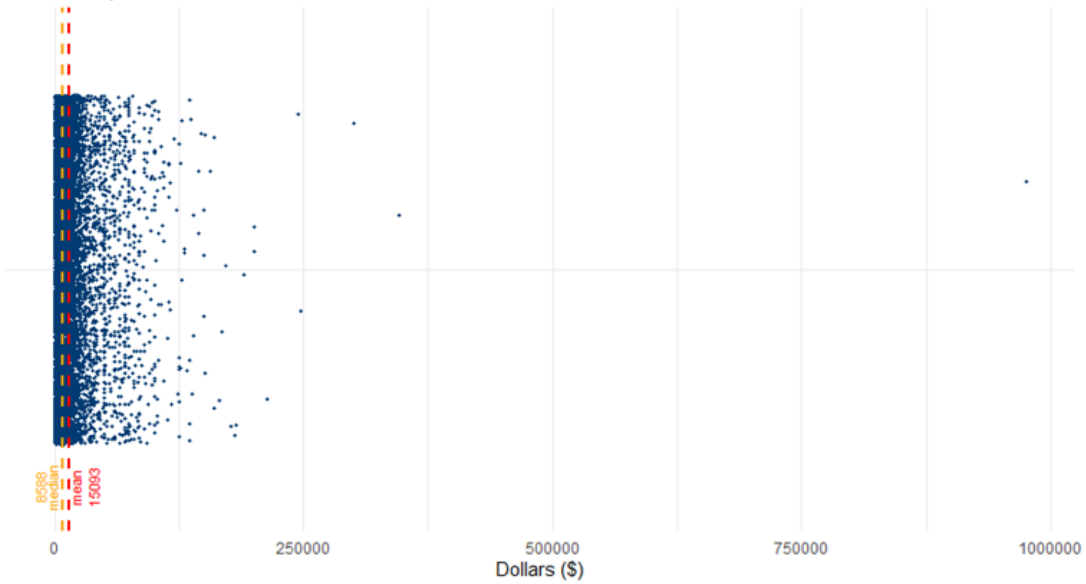
Total Medical



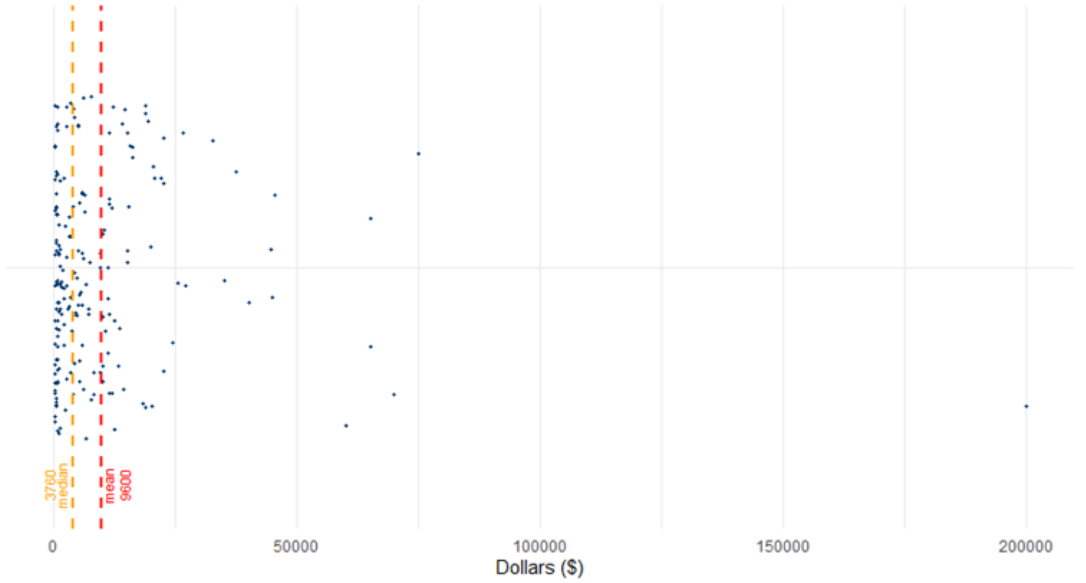
Total Indemnity



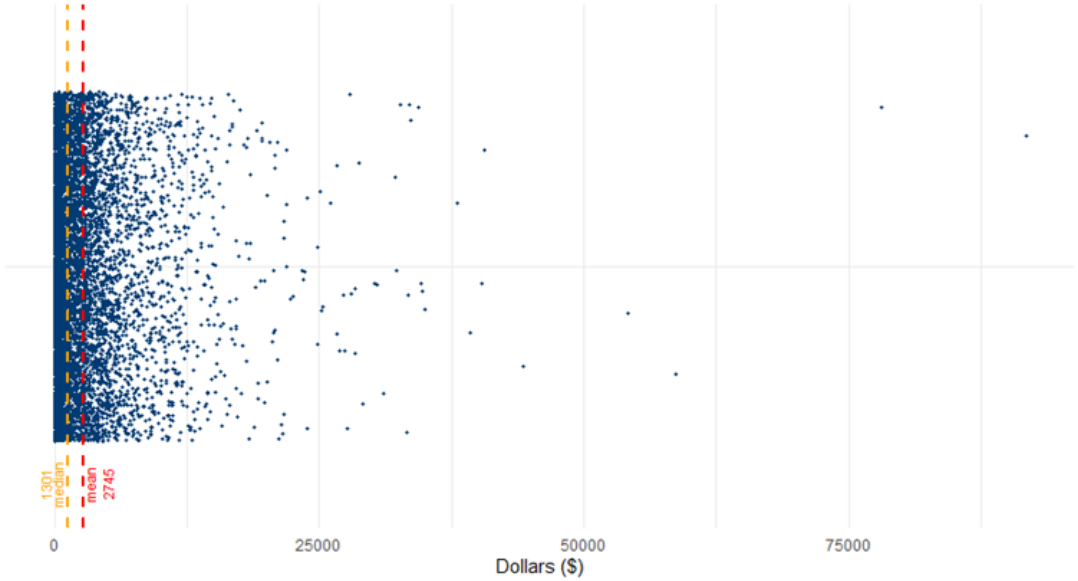
Total Lump Sum



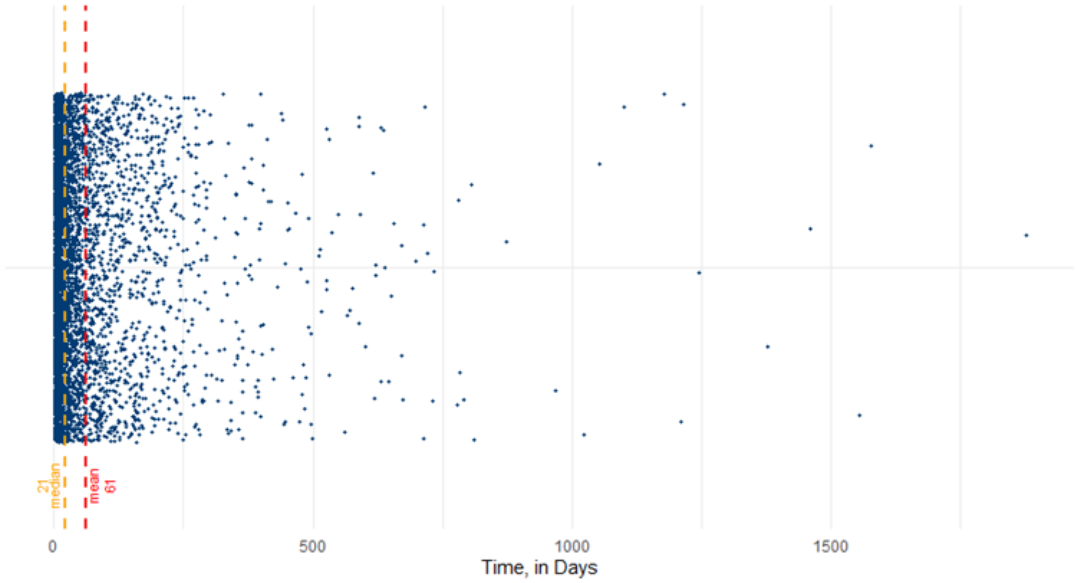
Claim Legal Ptd



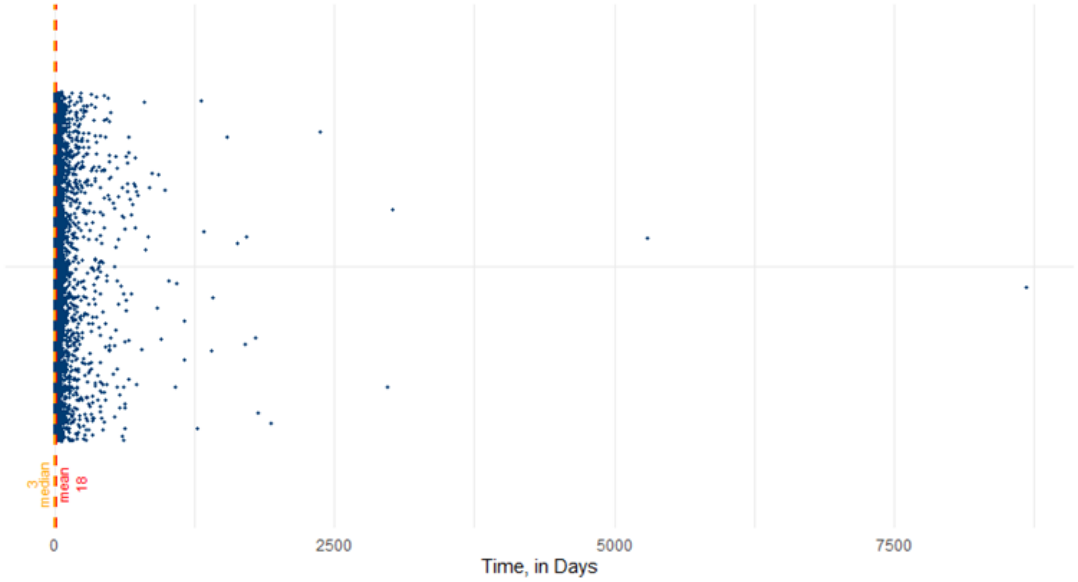
Employer Legal Ptd



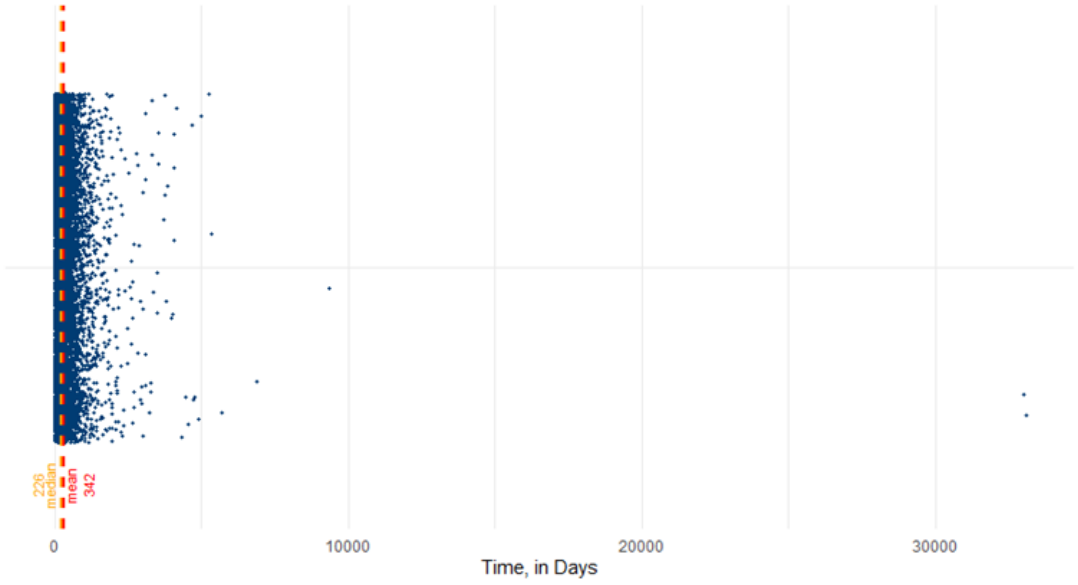
Time Away Work



Time to Notify Insurer

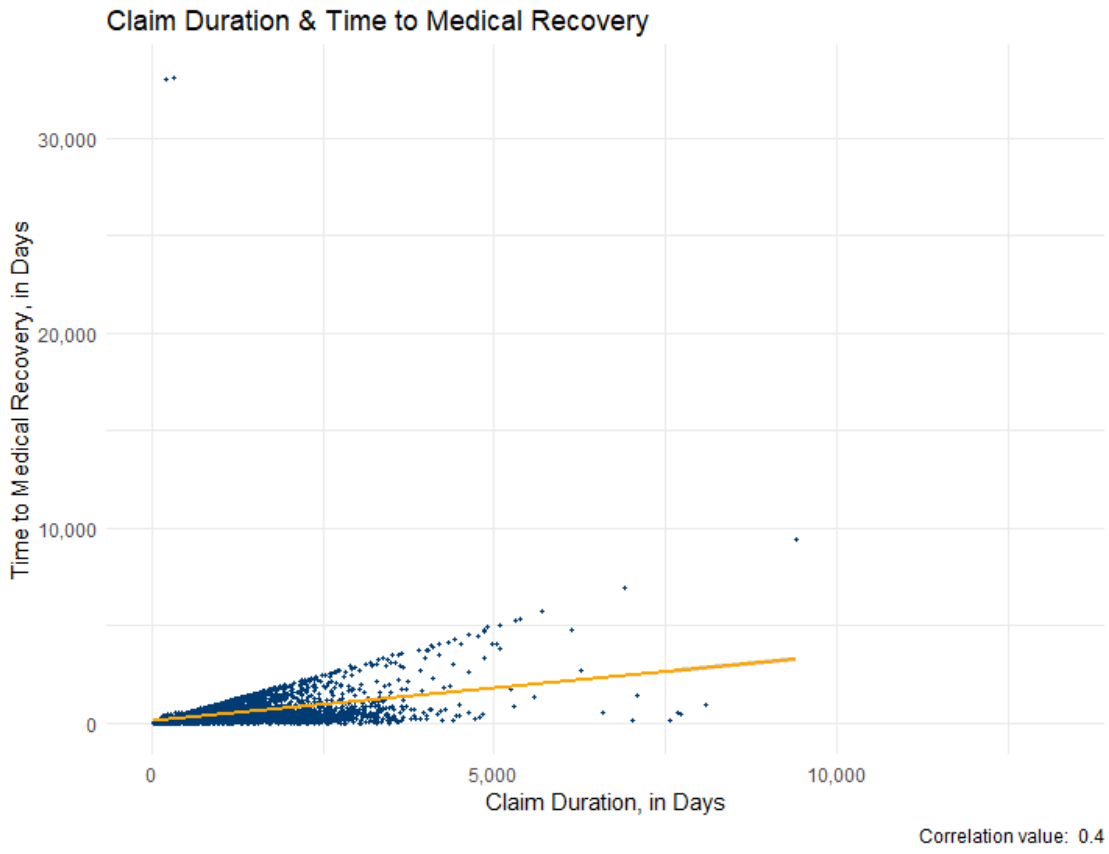


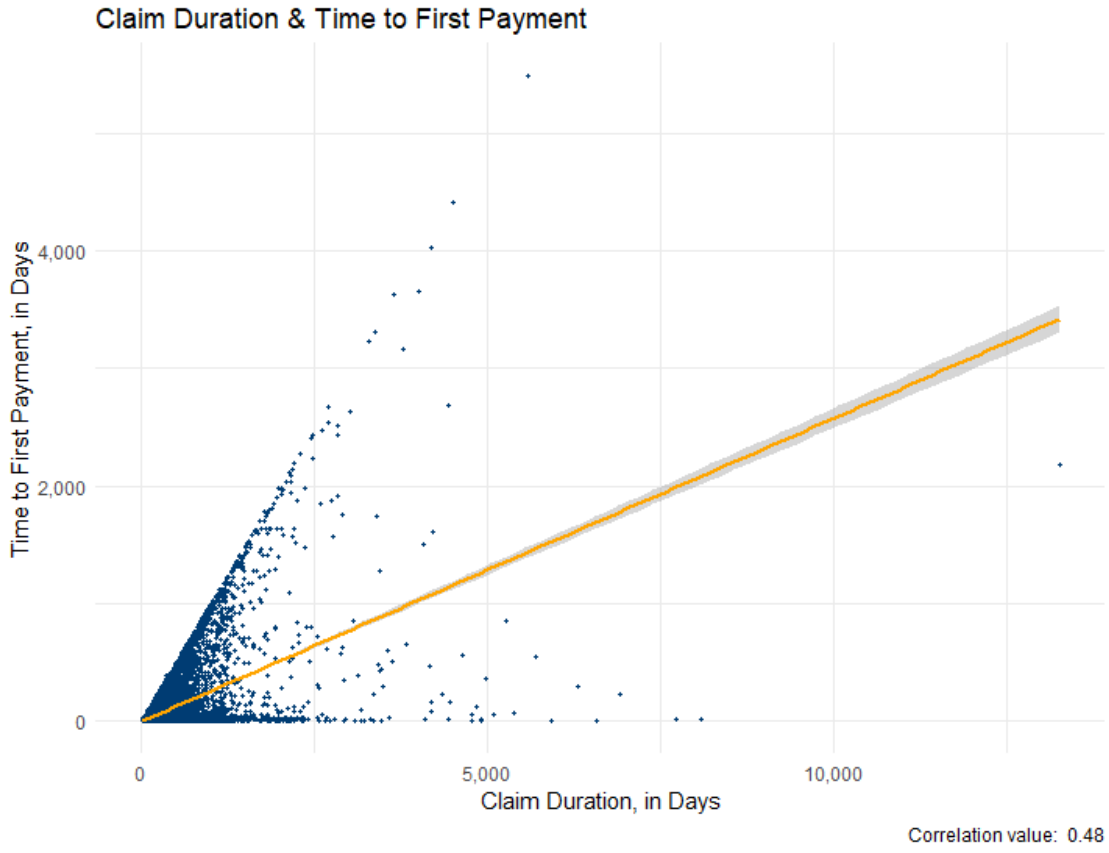
Time to Medical Recovery

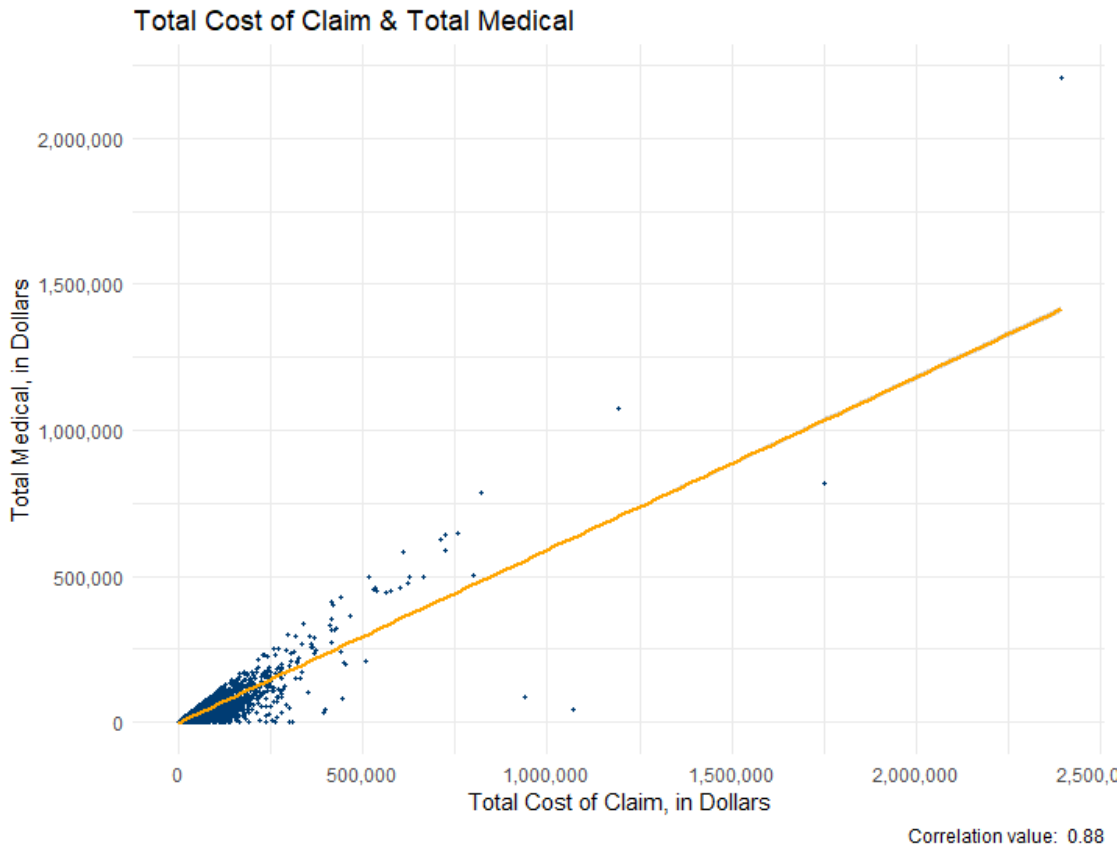


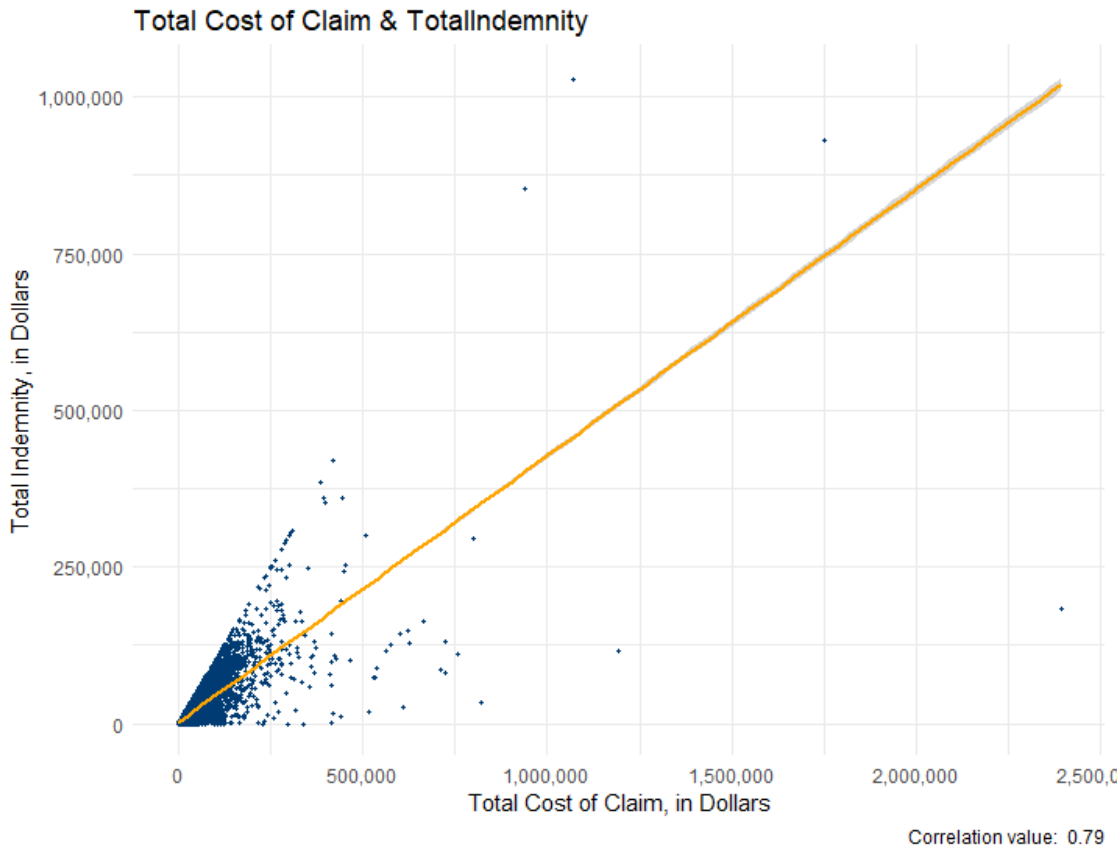
Below are scatter plots of two variables that are used for correlation analysis. These plots were created using R/RStudio statistical software and the ggplot2 R package as well. The first named variable in the scatterplot title is on the x-axis (horizontal) and the last-named variable is plotted on the y-axis (vertical). The correlation value- analyzed is next section- is displayed as is the regression line in yellow.

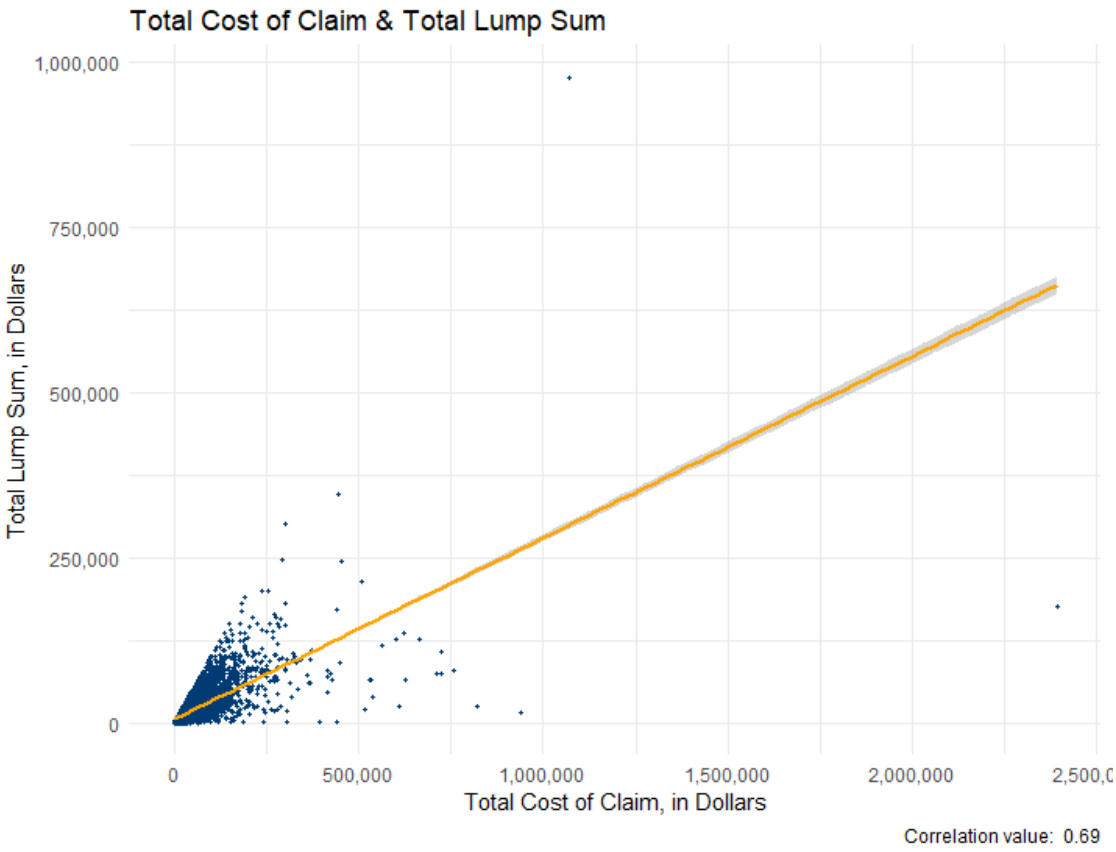
Figure 3.1 - Figure 3.9. Scatterplots per variable 2014-2017.

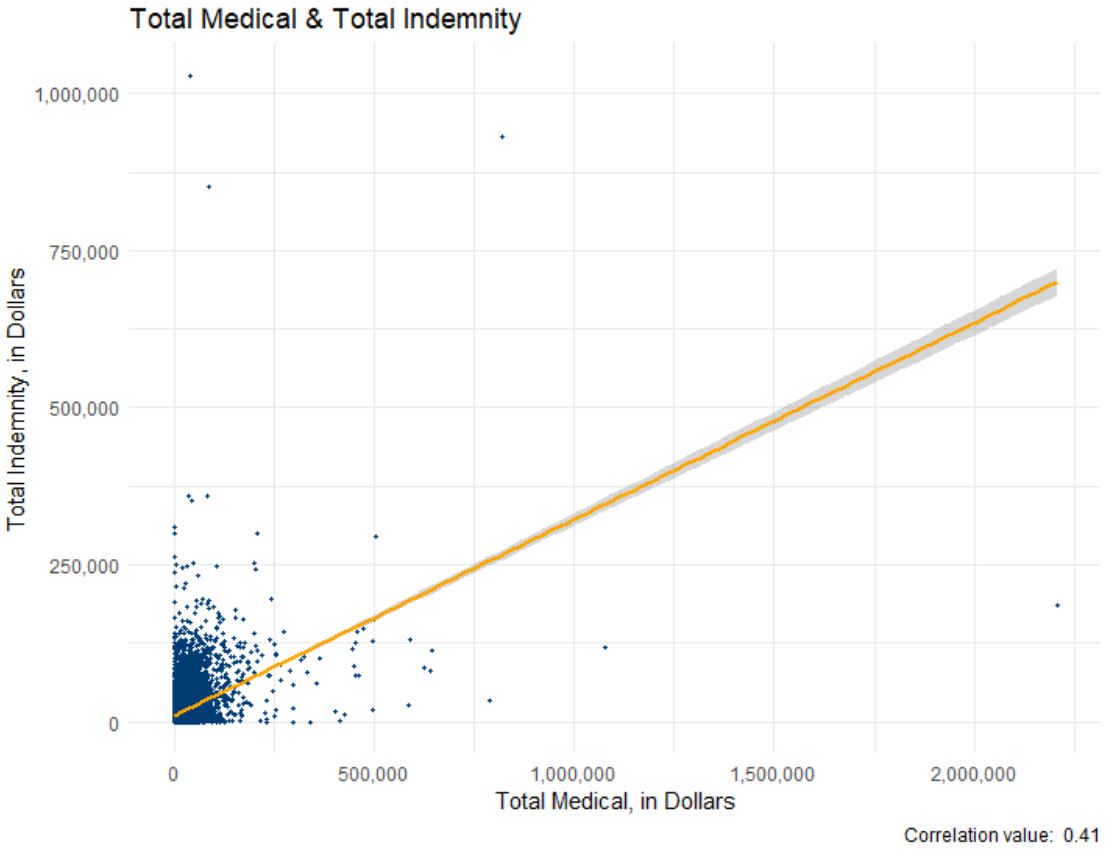

Claim Duration & Time to Medical Recovery
30,000
20,000
10,000
0
Time to Medical Recovery, in Days
0
5,000
10,000
Claim Duration, in Days
Correlation value: 0.4

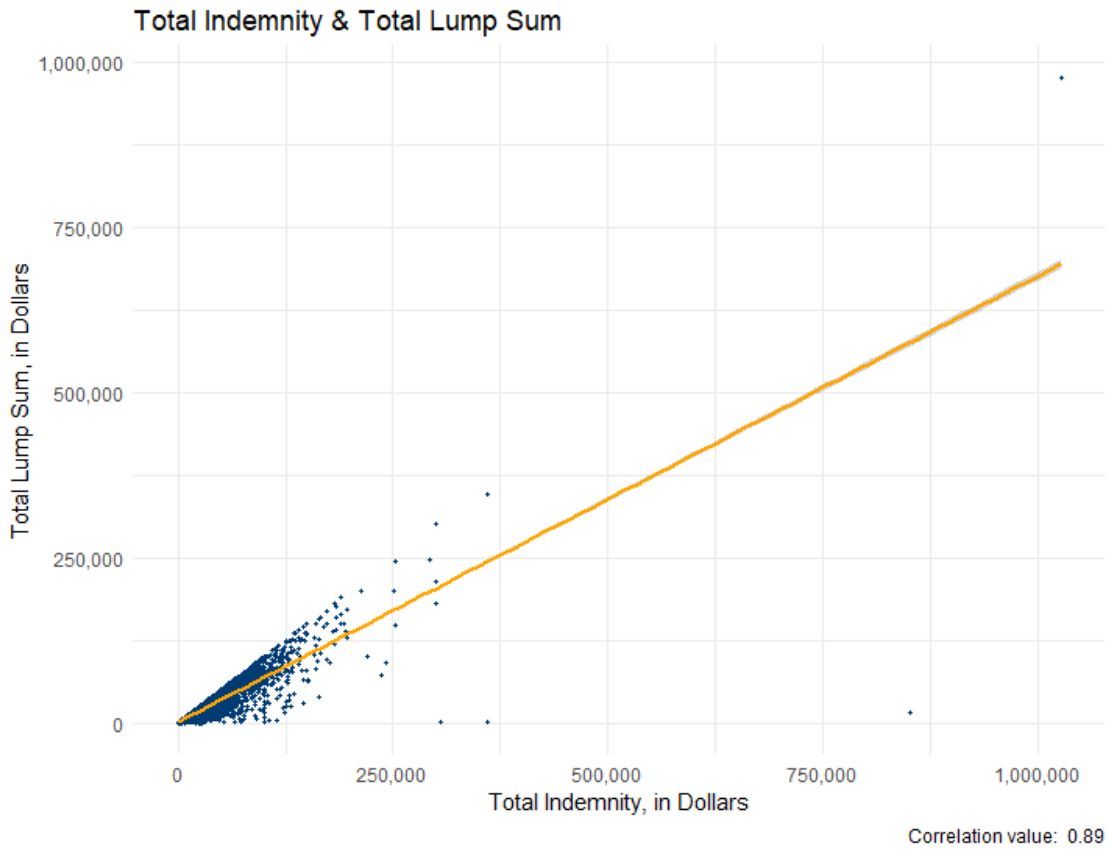

Claim Duration & Time to First Payment
4,000
2,000
0
Time to First Payment, in Days
0
5,000
10,000
Claim Duration, in Days
Correlation value: 0.48

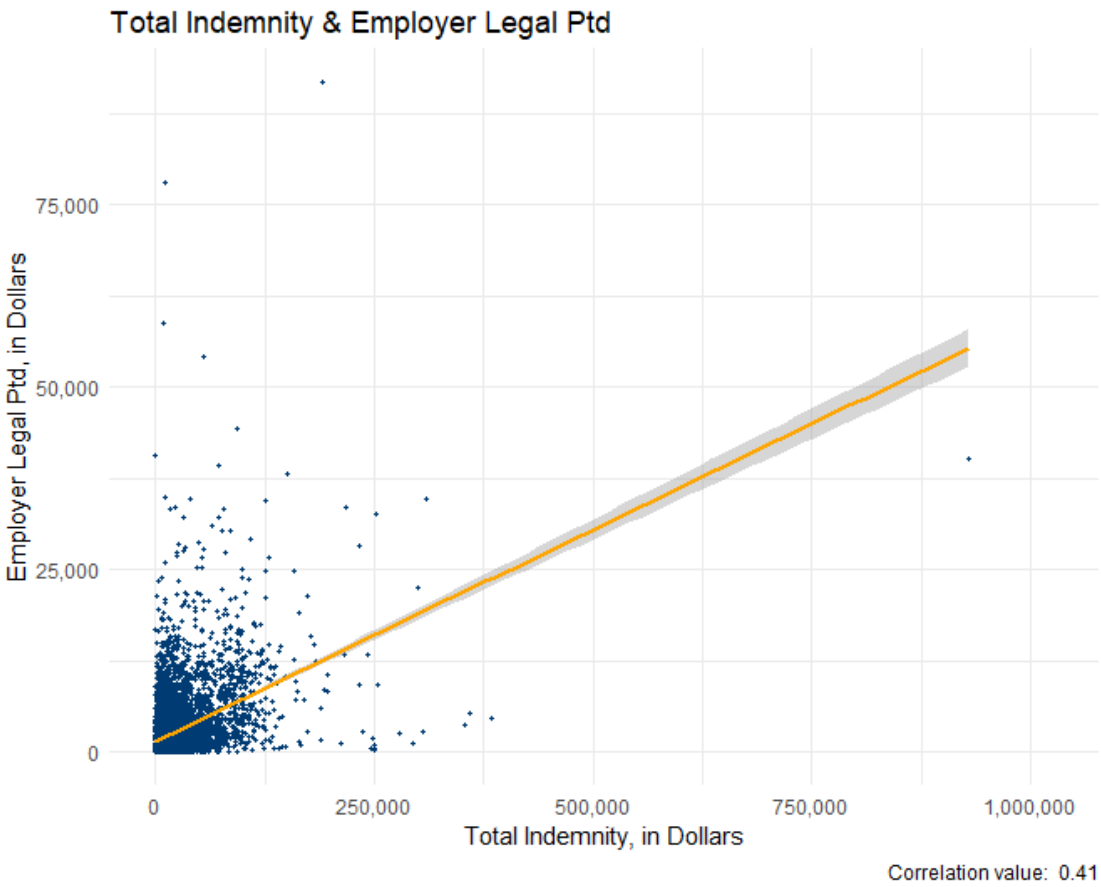

Total Cost of Claim & Total Medical
2,000,000
1,500,000
1,000,000
500,000
0
Total Medical, in Dollars
0
500,000
1,000,000
1,500,000
2,000,000
2,500,0
Total Cost of Claim, in Dollars
Correlation value: 0.88


Total Cost of Claim & TotalIndemnity
1,000,000
750,000
500,000
250,000
0
Total Indemnity, in Dollars
0
500,000
1,000,000
1,500,000
2,000,000
2,500,0
Total Cost of Claim, in Dollars
Correlation value: 0.79


Total Cost of Claim & Total Lump Sum
1,000,000
750,000
500,000
250,000
0
Total Lump Sum, in Dollars
0
500,000
1,000,000
1,500,000
2,000,000
2,500,0
Total Cost of Claim, in Dollars
Correlation value: 0.69


Total Medical & Total Indemnity
1,000,000
750,000
500,000
250,000
0
Total Indemnity, in Dollars
0
500,000
1,000,000
1,500,000
2,000,000
Total Medical, in Dollars
Correlation value: 0.41


Total Indemnity & Total Lump Sum
1,000,000
750,000
500,000
250,000
0
Total Lump Sum, in Dollars
0
250,000
500,000
750,000
1,000,000
Total Indemnity, in Dollars
Correlation value: 0.89


Total Indemnity & Employer Legal Ptd
Employer Legal Ptd, in Dollars
75,000
50,000
25,000
0
0
250,000
500,000
750,000
1,000,000
Total Indemnity, in Dollars
Correlation value: 0.41

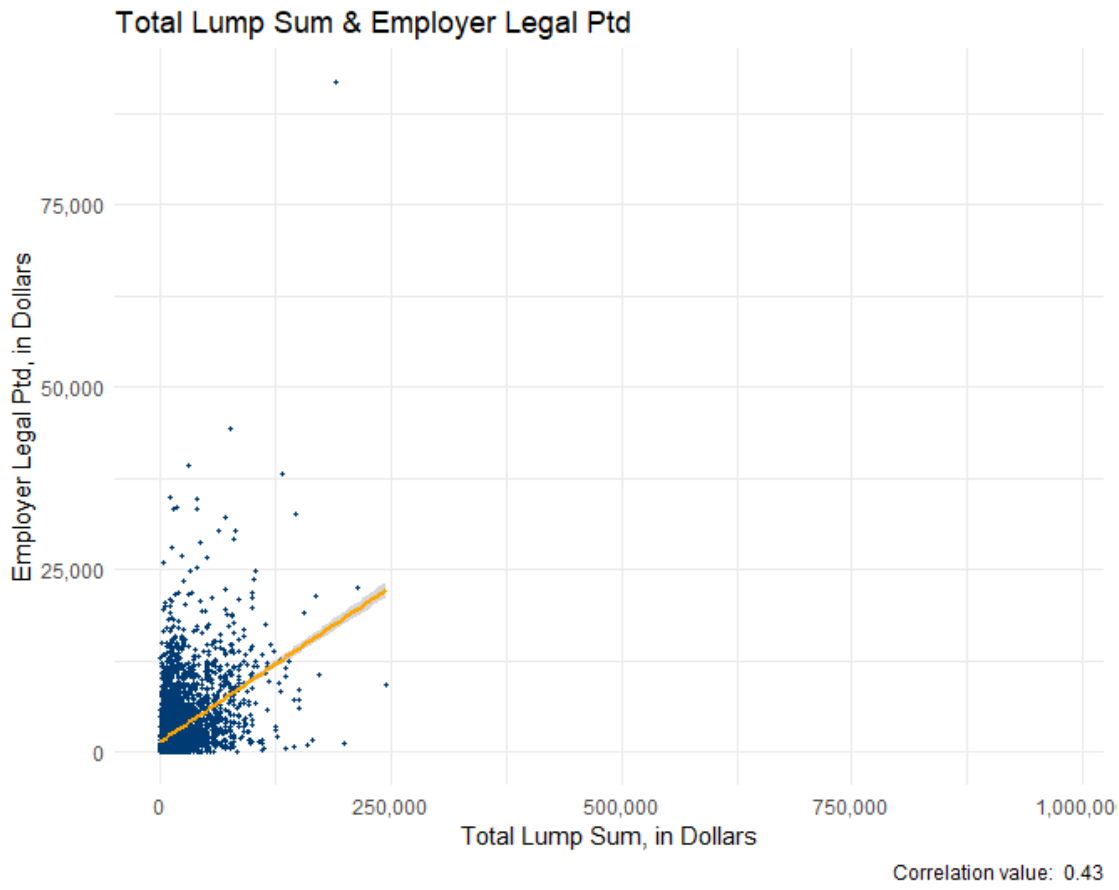


Table 1. Descriptive Statistics for Kansas Closed Indemnity Claims 2014-2017

	Claim Duration	Total Cost of Claim	Total Medical	Total Indemnity	Total Lump Sum	Claim Legal Ptd
mean	619	24,342	13,400	12,755	15,093	9,600
median	415	11,891	6,706	4,884	8,588	3,760
var	449,726	2,157,858,962	1,051,845,423	629,393,854	502,457,919	353,311,640
SD	671	46,453	32,432	25,088	22,416	18,797
kurtosis	23	463	1,237	290	359	52
skew	3	14	24	11	11	6
IQR	520	25,088	13,332	13,498	13,340	10,506
min	14	4	1	4	4	36
max	13,265	2,391,850	2,207,409	1,027,727	975,000	200,000

	Employer Legal Ptd	Time Away from Work	Time To Notify Insurer	Time To Medical Recovery	Time to First Payment
mean	2,745	61	18	342	115

	Claim Duration	Total Cost of Claim	Total Medical	Total Indemnity	Total Lump Sum	Claim Legal Ptd
median	1,301	21	3	226	18	
var	17,141,863	13,355	10,016	366,610	70,841	
SD	4,140	116	100	605	266	
kurtosis	60	50	2,980	1,582	53	
skew	5	6	42	31	6	
IQR	3,020	55	8	296	65	
min	6	1	0	0	0	
max	91,696	1,876	8,672	33,083	5,487	

Descriptive Statistics for Kansas Closed Indemnity Claims 2014-2017

Descriptive Statistics Analysis

Table 1- Descriptive Statistics for Kansas Closed Indemnity Claims 2014-2017 lists all descriptive statistics used in the analysis. In this section we will report but not analyze the variance (var) as it is reported in the variable squared (e.g., dollars squared). Instead, we will instead analyze the square root of the variance or the standard deviation- a standard measure of dispersion as it measures the average distance of each observation from the mean. Also, the minimum and maximum values for each variable are straightforward requiring no interpretation only identification. The reader can consult Table 1 for those values of each variable if that is of interest.

As the preceding section the histograms and jitter plots showed most of our workers' compensation claims variables are positively skewed with the mean values higher than the median. In normal distribution the mean and median values are the same.

For the variables measuring time, mean claim duration is 619 days, the median is 415 days. Median is an average but is the midpoint, the 50th percentile with half the values higher and the other half lower. All other time variables all had higher mean values than median values as well. Mean time away from work was 61 days, the median value was only 21 days. Mean time to notify the insurer was 18 days, median is 3 days. Mean time to medical recovery was 342 days or nearly a year while the median was 226. Mean time to first payment was 115 days and the median value was 18 days. When reporting the "average" as a measure of central tendency with positive skewed distribution you should use the median value instead of the mean. A few larger values will "pull" the mean value higher than the median. The median is considered a robust statistic in that outlier values do not affect its calculation. Report the mean but use the median as the average.

Mean total cost of claim is ($24,342), mean total medical ($13,400), mean total indemnity ($12,755), mean total lump sum settlement ($15,093), mean claimant legal paid to date ($9,600), and mean employer legal paid to date ($2,745) are all higher values than the associated median values for each cost variable. We report the mean values but will call the median values for all these variables the "average." Average or median total cost of the claim is $11,891, median total medical is $6,706, median total indemnity $4,884, total lump

sum settlement had a median value of $8,588, median claimant legal paid to date was $3,760, and finally, employer legal paid to date median was $1,301.

As discussed earlier the standard deviation is a key measure of dispersion. It measures the average distance of each observation from the mean and is interpreted in the language of the variable. In Table 1 the standard deviation for claim duration is 671, which is really 671 days. The lower that value the more the data cluster about the mean, the higher the standard deviation the more dispersed the data from the mean. Claim duration has a lot of variation in our sample of claims. The rest of the reported standard deviations are in Table 1. The IQR or interquartile range measures where in the sample the middle fifty percent of the distribution is located. Again, for the variable claim duration 50% of the sample lies within a spread of 520 days. A good data visualization for the IQR is the box plot or sometime called the box and whisker plot. IQR values are easy to interpret and located in Table 1.

The quantified skewness values supplement the earlier data visualizations and the R statistical program used in our analysis calculates a positively skewed distribution as a positive number, a negatively skewed distribution as a negative number, and a normally distributed variable as zero. All 11 variables have skewness values that range from positive 3 to positive 42. See Table 1. Kurtosis is a quantified measure of how peaked or flat a distribution is. We are interested in how "fat" or "thin" the tails of the distribution are relative to a normal distribution. Normally distributed sample have a zero (0) kurtosis value, a negative kurtosis means a flatter data distribution, while a positive value indicates a more peaked distribution. The claim duration variable has a kurtosis value of 23 or a peaked data distribution with less data in the tails (thin) while the total cost of the claim kurtosis value is 463. That distribution is even more peaked with very thin tails indicating some very high and very low outlier values in the sample. The rest of the kurtosis values are in Table 1.

Correlation Analysis

A correlation is a measure of association between two variables. It is an analysis of whether the two variables "co-vary" (covariation) together or not. We use the Pearson's correlation coefficient to measure this covariation below and then visualize this relationship with the use of scatterplots.

The correlation coefficient varies from -1.0 to 1.0. A negative 1.0 value means a perfectly inverse or negative relationship, as one variable rises in value the other decreases. A positive 1.0 correlation coefficient means a perfectly positive relationship, as one variable rises in values so does the other. The social world is a little messier, so we need to aid your interpretation with a quick explanation on how to describe correlations between our workers' compensation claims variables. A correlation value of zero (0) means no correlation between the variables.

• Values ranging from 0 through 0.25 are describe as weak positive correlation, the 0.25 to 0.75 is a moderate positive correlation and 0.75 to 1 is a strong positive correlation.

• Values ranging from 0 through -0.25 are describe as weak negative correlation, coefficients from -0.25 to -0.75 is a moderate negative correlation and -0.75 to -1 is a strong negative correlation. This is sometimes also referred to as an inverse relationship.

• There is no such thing as a "direct correlation" based on the interpretation above.

Table 2. Correlation Table

	Claim Duration	Total Cost of Claim	Total Medical	Total Indemnity	Total Lump Sum	Claim Legal Ptd	Employer Legal Ptd	Time Away from Work	Time To Notify Insurer	Time To Medical Recovery	Time to First Payment
Claim Duration	1.00	0.32	0.20	0.33	0.28	-0.11	0.31	0.29	0.20	0.40	0.48
Total Cost of Claim	0.32	1.00	0.88	0.79	0.69	0.07	0.34	0.36	0.01	0.20	0.06
Total Medical	0.20	0.88	1.00	0.41	0.33	0.15	0.18	0.25	0.00	0.18	0.00
Total Indemnity	0.33	0.79	0.41	1.00	0.89	0.01	0.41	0.39	0.03	0.24	0.13
Total Lump Sum	0.28	0.69	0.33	0.89	1.00	0.26	0.43	0.29	0.01	0.31	0.06
Claim Legal Ptd	-0.11	0.07	0.15	0.01	0.26	1.00	-0.04	-0.06	-0.05	0.07	-0.19
Employer Legal Ptd	0.31	0.34	0.18	0.41	0.43	-0.04	1.00	0.14	0.02	0.34	0.19
Time Away from Work	0.29	0.36	0.25	0.39	0.29	-0.06	0.14	1.00	0.04	0.16	0.07
Time To Notify Insurer	0.20	0.01	0.00	0.03	0.01	-0.05	0.02	0.04	1.00	0.16	0.24
Time To Medical Recovery	0.40	0.20	0.18	0.24	0.31	0.07	0.34	0.16	0.16	1.00	0.22
Time to First Payment	0.48	0.06	0.00	0.13	0.06	-0.19	0.19	0.07	0.24	0.22	1.00

All correlations between the 11 claims variables are shown in the Correlation Table 2. We only discuss correlations with coefficients greater than or equal 0.40 (moderate positive) or less than or equal to -0.40 (moderate negative). For the variable claim duration, it has a moderate positive correlation with time to medical recovery (0.40), and time to first payment (0.48). Total cost of the claim has strong positive correlations with the variables total medical (0.88) and total indemnity (0.79) and a moderate positive correlation with total lump sum (0.69). Total medical paid has a strong positive correlation with total cost of the claim (0.88) and a moderate positive correlation with total indemnity (0.41). Total Indemnity has a strong positive correlation with total cost of the claim (0.79) and total lump sum (0.89) and a moderate positive correlation with total medical (0.41) and employer legal costs paid to date (0.41). The only other strong positive correlations were with total lump sum paid with the variable total indemnity (0.89).

Bootstrap Confidence Intervals: The Concept

Confidence intervals are an interval estimate which means the range of plausible values where we may find the true population value. For example, while we calculated sample mean estimates (called point estimates) such as 619 days of duration for a claim, we know this has some error because we sampled from the population. We want to use either math or computers to make an interval estimate (a confidence interval) where we think the population mean lies.

Statisticians mostly use mathematical models (e.g., probability theory) to make this inference with a confidence interval. However, we can now also use computational inference methods to calculate confidence intervals as well- such as bootstrapping. A foundation principle of inference is we use the data we have (the sample) to make inferences about the data we don't have.

Bootstrapping is method of computational inference created by the statistician Bradley Efron in the late 1970s. Much like probability theory, bootstrapping also uses the sample to make inferences to the population of data we did not collect. It just uses a different algorithm and our modern computation power. The mathematical model (e.g., probability theory) assumes if we could, we would measure the variability of the statistics by repeatedly taking sample data from the population and compute the sample statistic each time. Then we could do it again. And again. And so, on until we have a good sense of the variability of our original estimate. The mathematical model works off this assumption with the idea from probability called the sampling distribution of sampling means.

Bootstrapping is using this very same idea but instead of a hypothetical repeated sampling based on a theory, it uses resampling from the original data set. The method is we repeatedly sample the original data, with replacement, using the same sample size as the original data. The resampling procedure is a simulation, in this analysis we do 1000 runs or resamples from the original sample then calculate the statistic of interest and plotting the resulting bootstrap distribution.

The interpretation of the computationally produced confidence interval works the same as a regular confidence interval. We calculate an upper and lower limit to produce an interval

estimate for means, medians, correlation coefficients below. We could bootstrap standard deviations or even regression coefficients but for this analysis we just produce estimates for mean values, medians, and correlation coefficients. We used a R package called 'infer' to bootstrap all confidence intervals and produce the nice bootstrap distributions below.

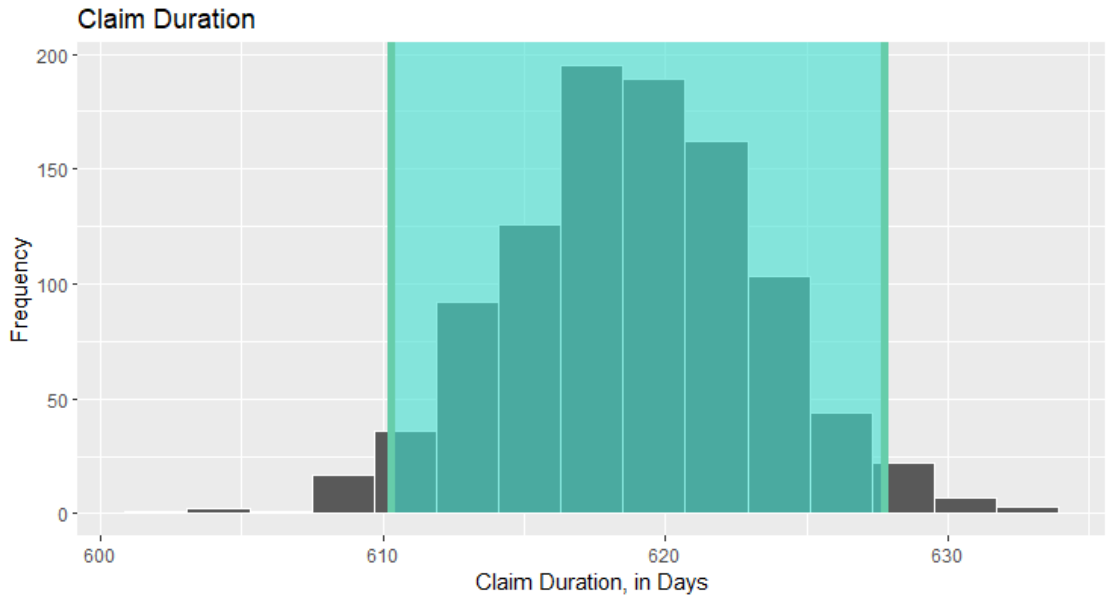
Bootstrap confidence interval are created using one of two methods. First, there are bootstrap percentile interval- obtain the endpoints representing the middle (95%) of the bootstrapped statistics. The endpoints will be the confidence interval. This is the method we will use in this analysis. We will visualize it and compute the exact upper and lower limits. The other method is to calculate a bootstrap standard error (SE) interval. The confidence interval will be given by the original observed statistic plus or minus some multiple (e.g., 2) of standard errors. This is more abstract method so we will not use it, just be aware of it.

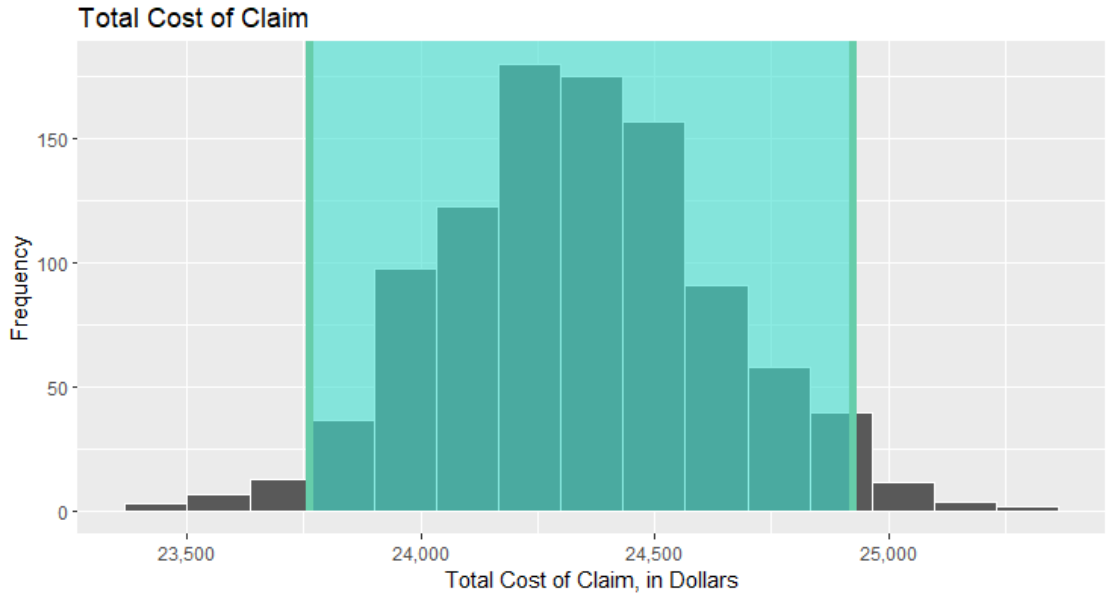
Bootstrap Confidence Intervals: Sample Means, Medians, and Correlation Coefficients

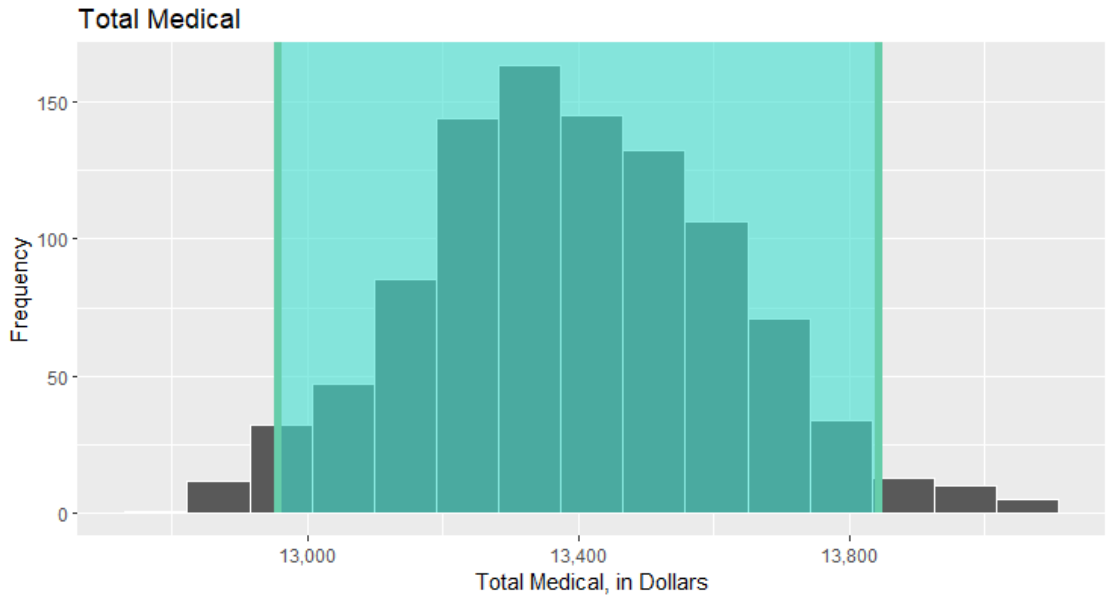
First, we will display the bootstrap distributions for all variable means, medians, and selected correlation coefficients. These visualizations were created using R/R Studio statistical software using the infer R package. For all bootstrap resampling simulations, we did 1,000 runs (resamples) and then plot out the resulting means, medians, and/or correlation coefficient bootstrap distributions with the estimated confidence intervals highlighted in aqua blue in the graph.

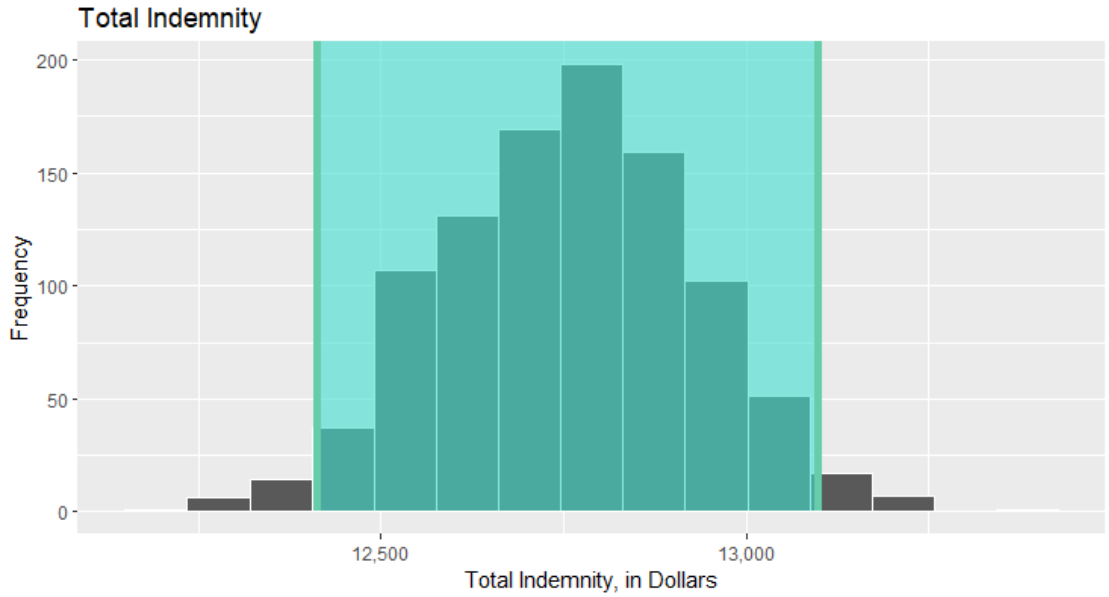
The bars or bins are the number of resamples (count) of mean (or median or correlation coefficients) from the 1,000 simulations and the resulting histogram is the bootstrap distribution. The aqua blue shading in the confidence interval, with the lower limit on the left in bold, the upper limit on the right in bold (the endpoints). The endpoints are the bootstrap confidence interval- representing the middle 95% of the bootstrap statistics.

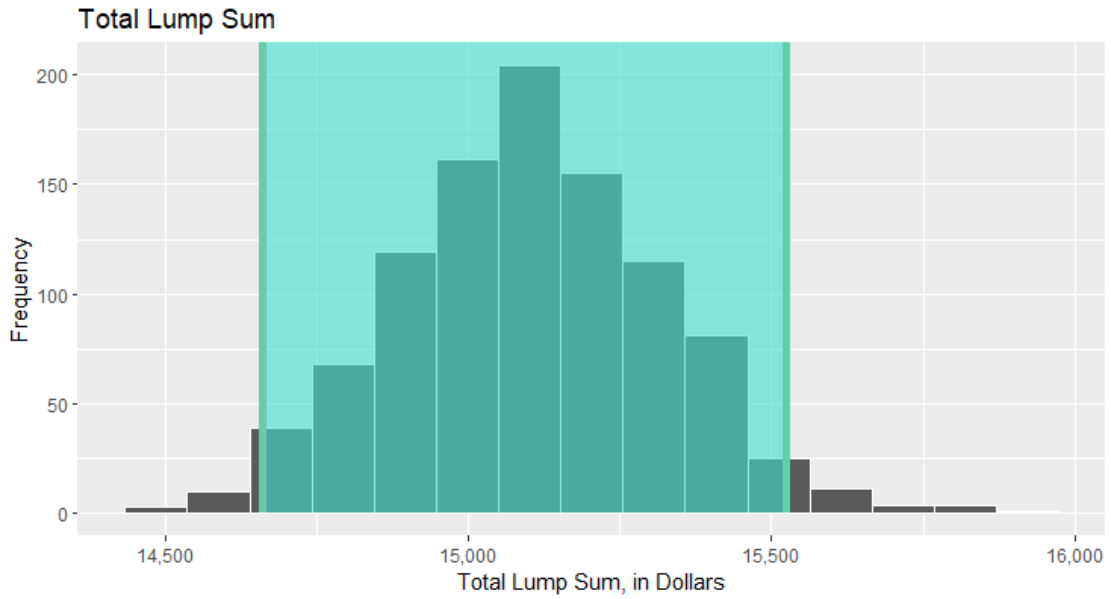
Figure 4.1- Figure 4.11. Bootstrap means per variables 2014-2017.

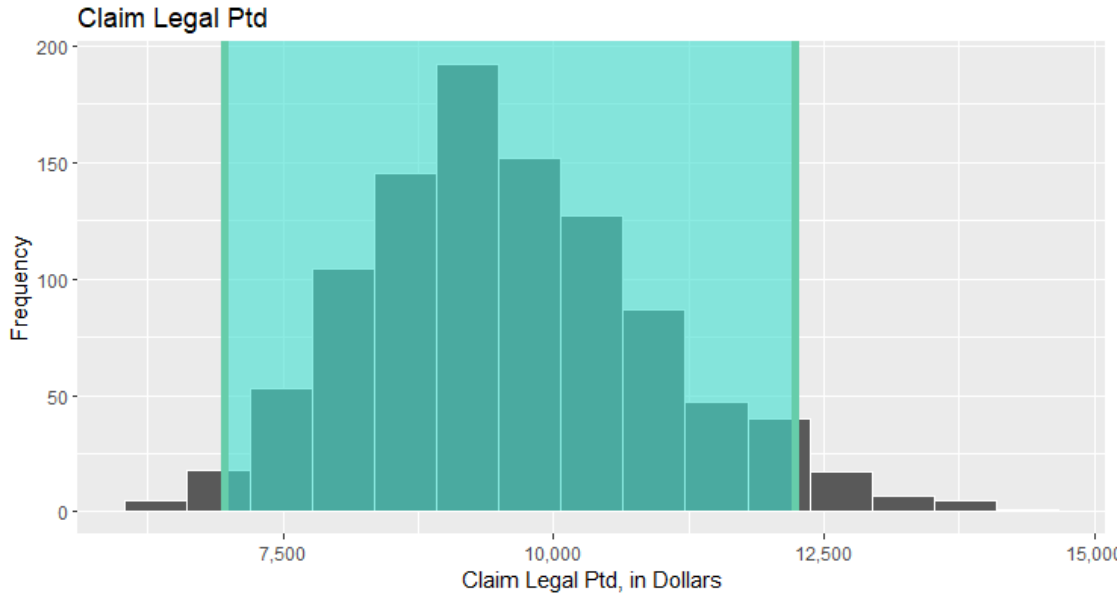

Claim Duration
200
150
100
50
0
Frequency
600
610
620
630
Claim Duration, in Days

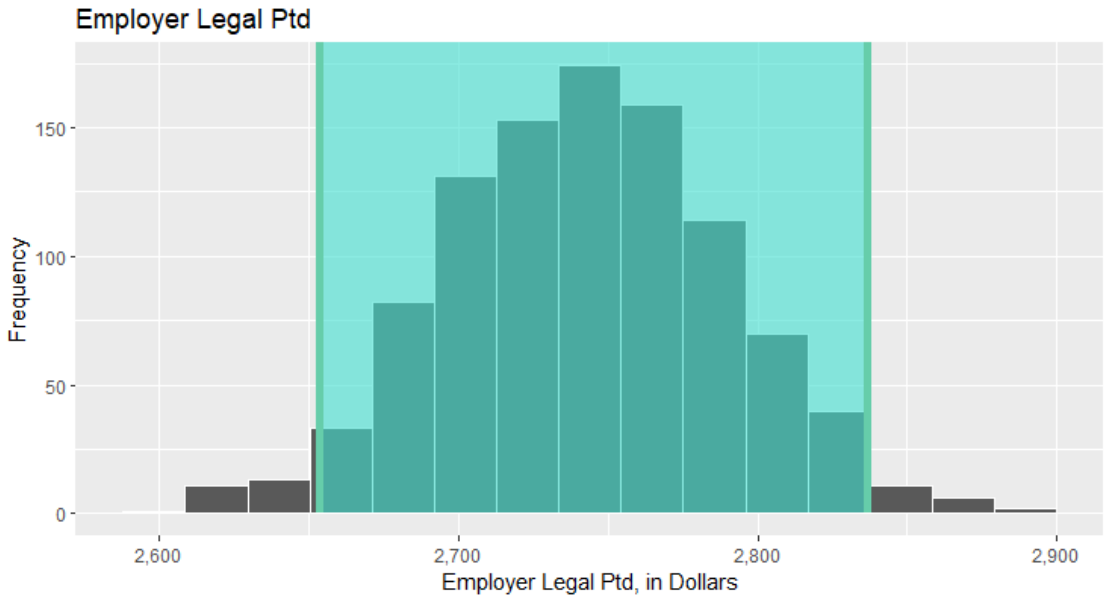

Total Cost of Claim
150
100
50
0
Frequency
23,500
24,000
24,500
25,000
Total Cost of Claim, in Dollars

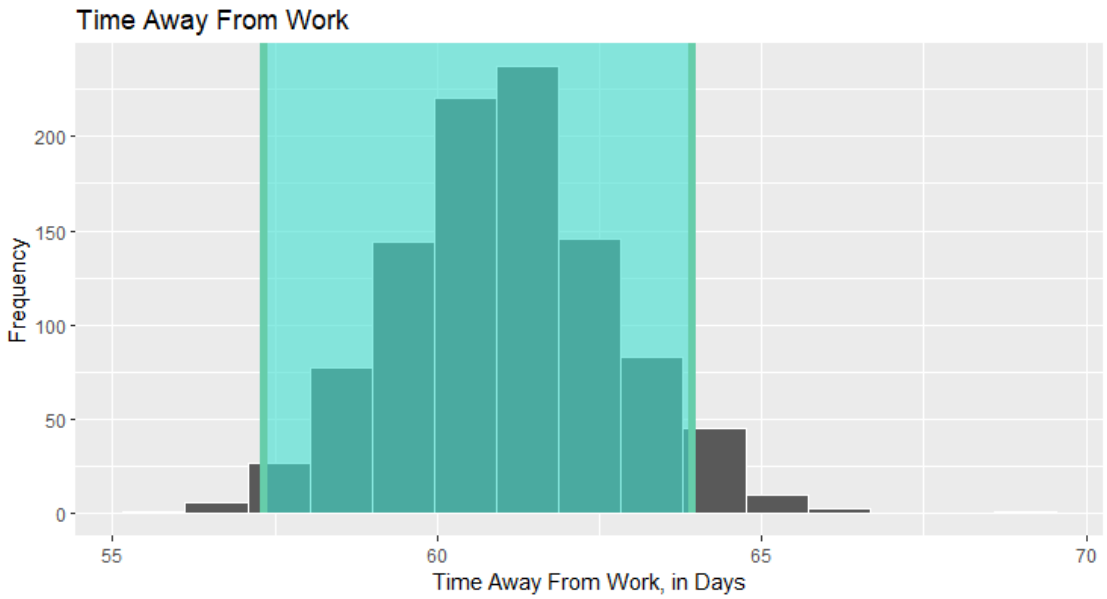

Total Medical
150
100
50
0
Frequency
13,000
13,400
13,800
Total Medical, in Dollars

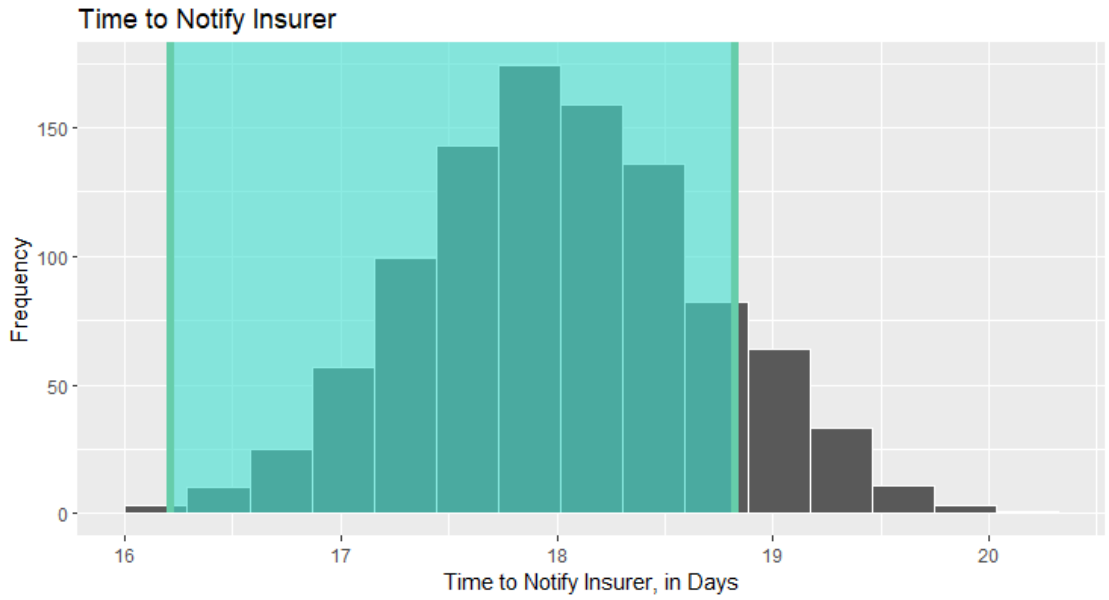

Total Indemnity
200
150
100
50
0
Frequency
12,500
13,000
Total Indemnity, in Dollars

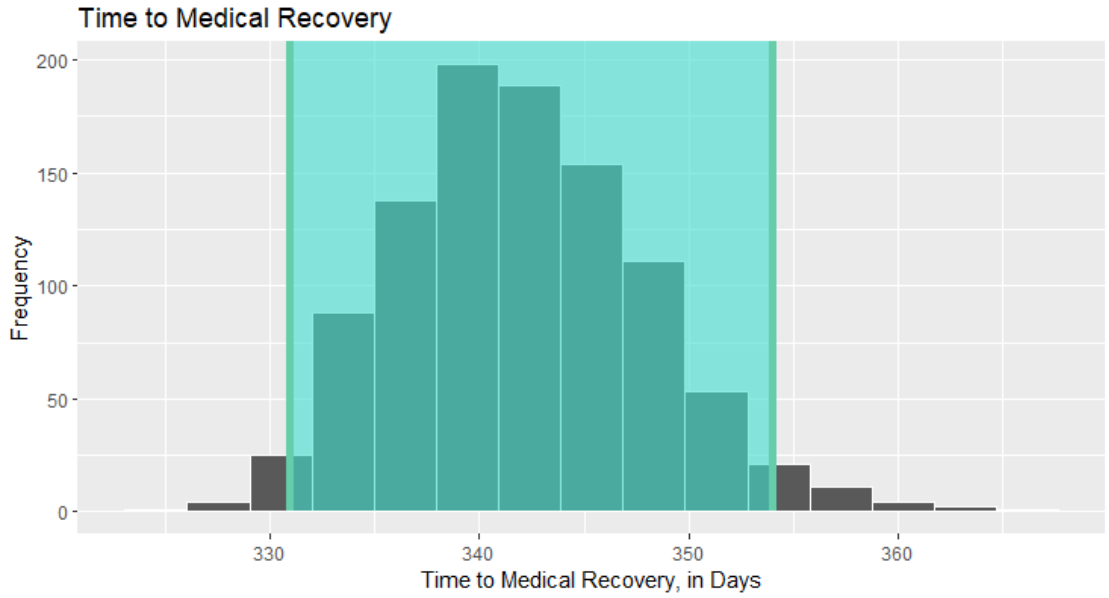

Total Lump Sum
200
150
100
50
0
Frequency
14,500
15,000
15,500
16,000
Total Lump Sum, in Dollars

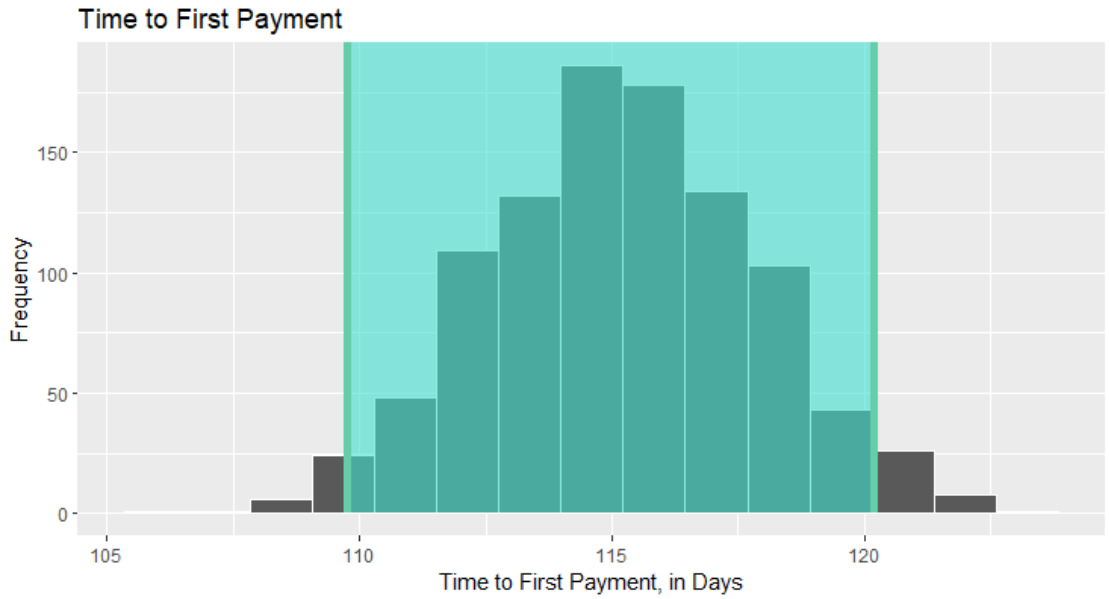

Claim Legal Ptd
200
150
100
50
0
Frequency
7,500
10,000
12,500
Claim Legal Ptd, in Dollars


Employer Legal Ptd
Frequency
150
100
50
0
2,600
2,700
2,800
2,900
Employer Legal Ptd, in Dollars


Time Away From Work
Frequency
200
150
100
50
0
55
60
65
70
Time Away From Work, in Days


Time to Notify Insurer
Frequency
150
100
50
0
16
17
18
19
20
Time to Notify Insurer, in Days


Time to Medical Recovery
Frequency
200
150
100
50
0
330
340
350
360
Time to Medical Recovery, in Days

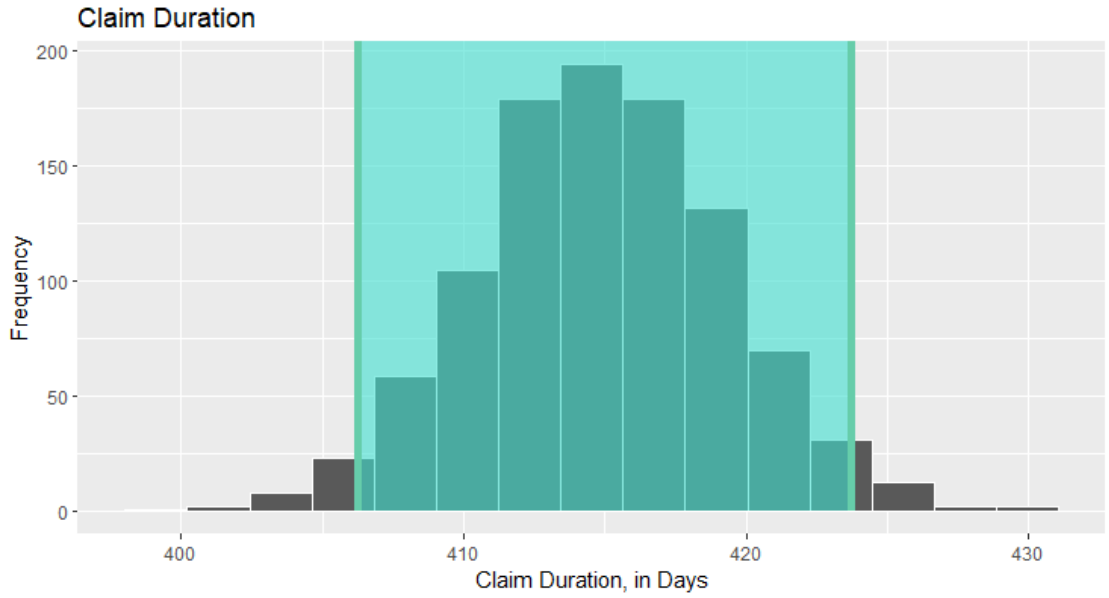


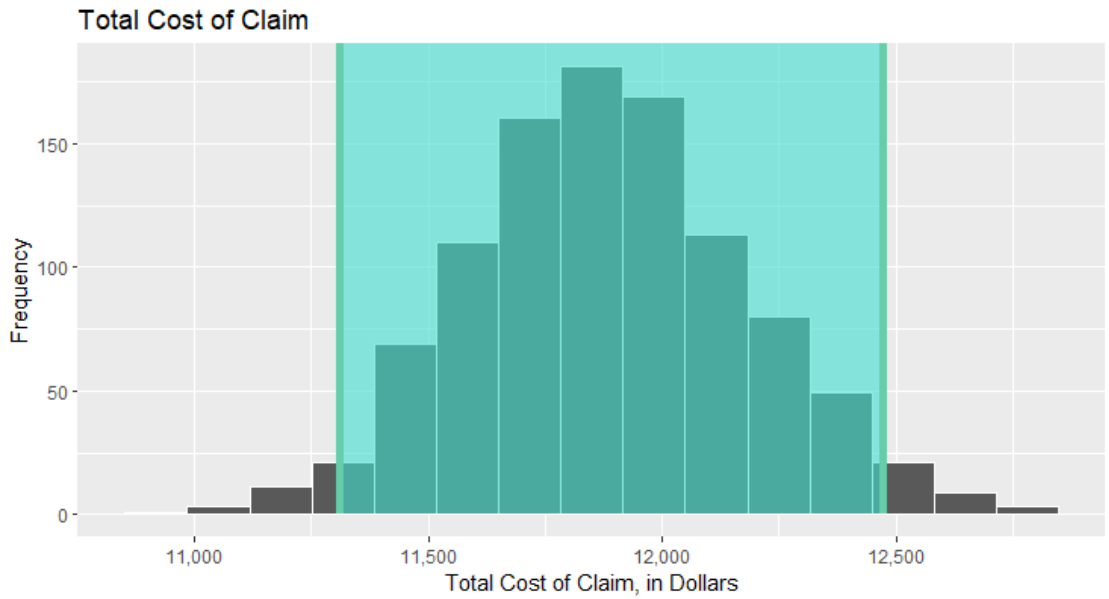
Below in Table 3.1 is a listing of all bootstrap confidence intervals for each variable mean. An example interpretation would be mean claim duration from the sample was 619 days. The bootstrap confidence interval estimate has a lower limit of 611 days and an upper limit of 628 days. The inference is the population mean for claim duration lies within 611 to 628 days, with 95% level of confidence. The rest of the bootstrap confidence intervals for the other variables are below.

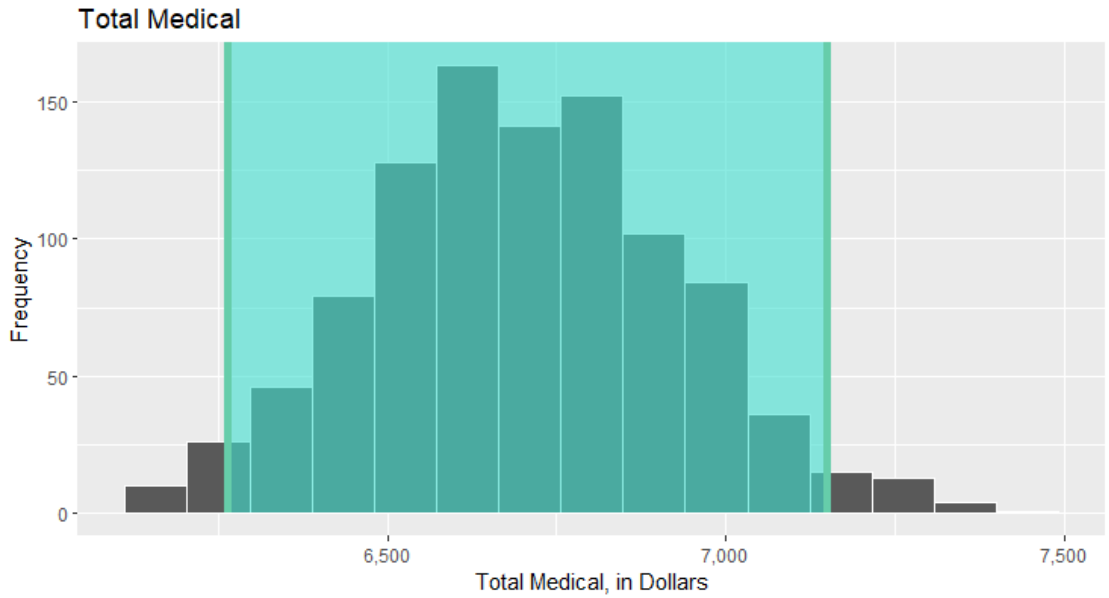
Table 3.1. Bootstrap Mean Confidence Intervals

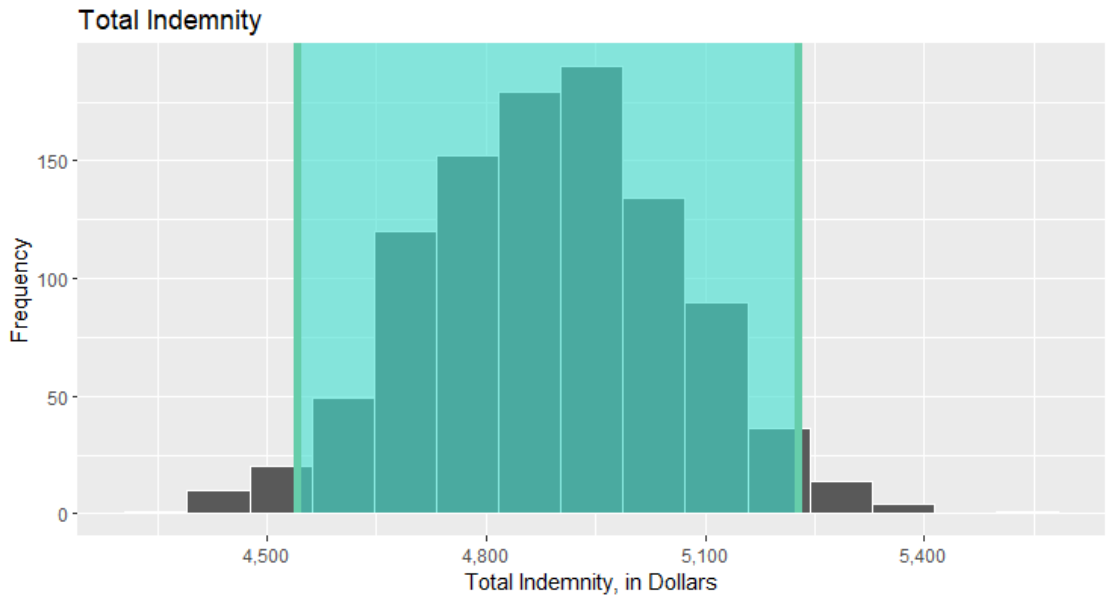
Variable	Lower CI	Observed Mean	Upper CI
Claim Duration	611	619	628
Total Cost of Claim	23754	24342	24951
Total Medical	12958	13400	13903
Total Indemnity	12445	12755	13088
Total Lump Sum	14686	15093	15544
Claim Legal Ptd	7135	9600	12618
Employer Legal Ptd	2654	2745	2844
Time Away from Work	58	61	64
Time To Notify Insurer	16	18	19
Time To Medical Recovery	332	342	355
Time To First Payment	110	115	120

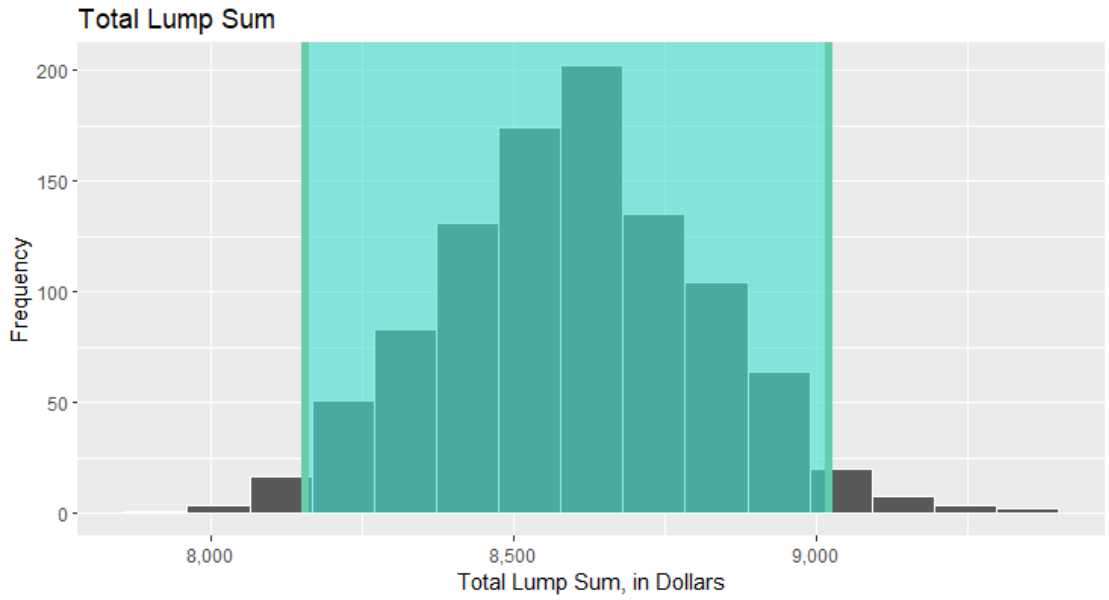
Figure 5.1-Figure 5.11. Bootstrap median samples per variables 2014-2017.

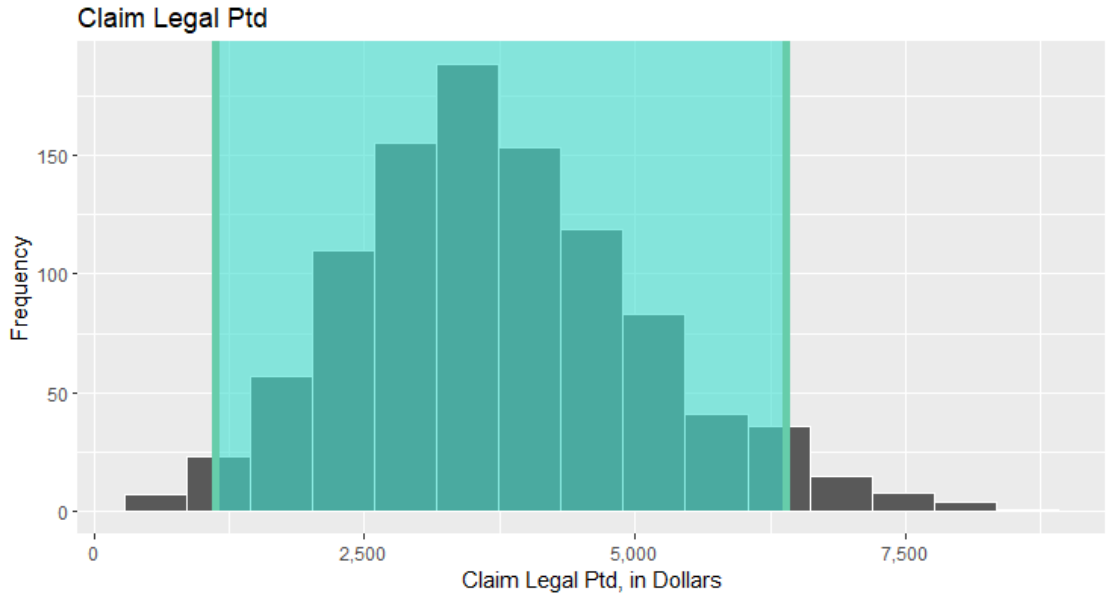


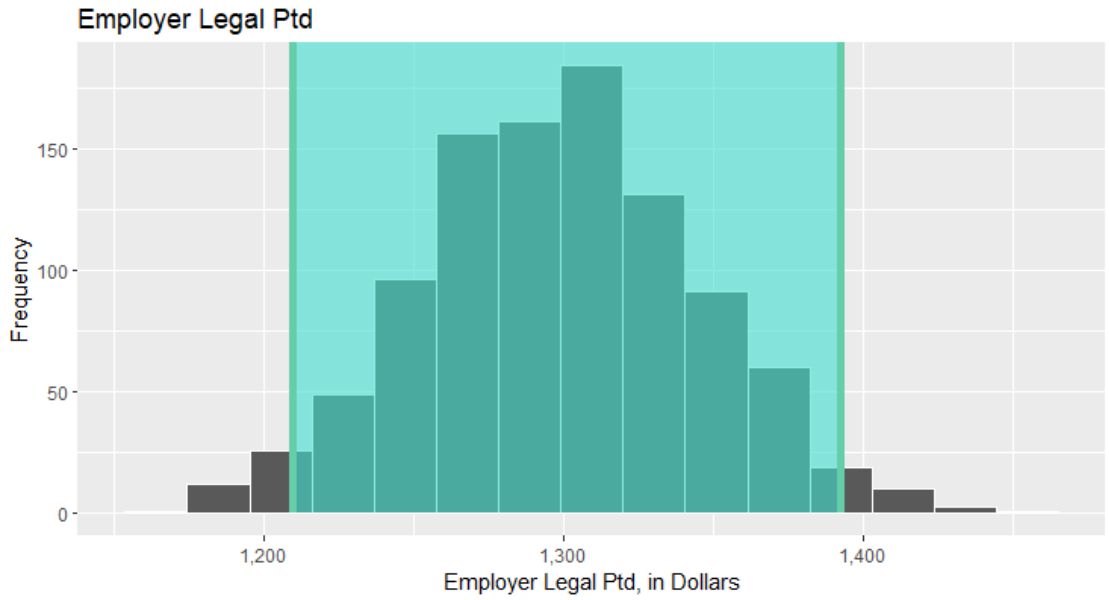


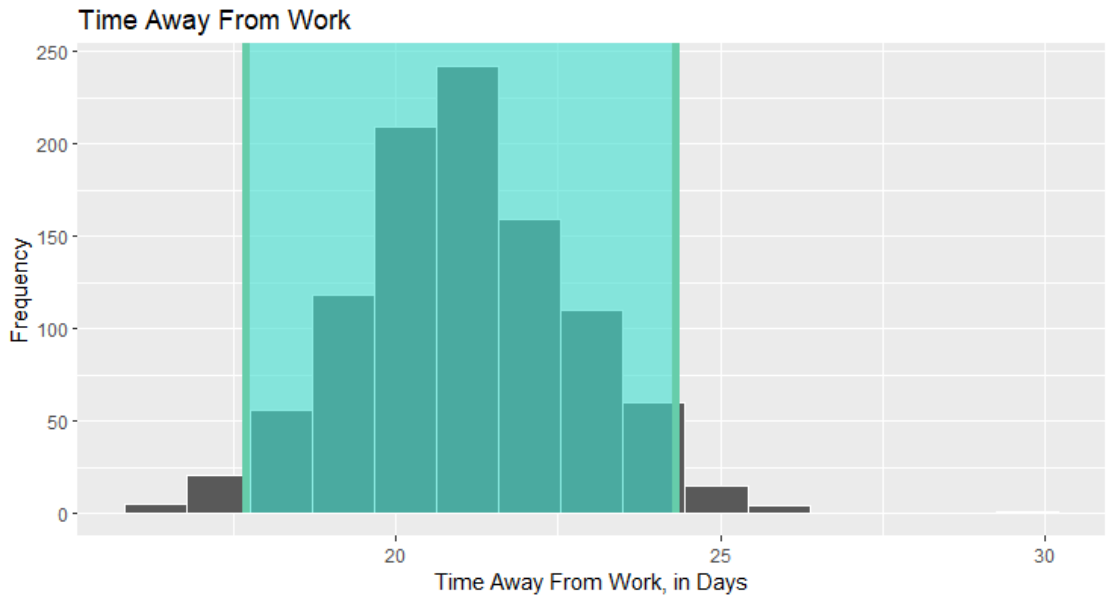

Total Medical
150
100
50
0
Frequency
6,500
7,000
7,500
Total Medical, in Dollars

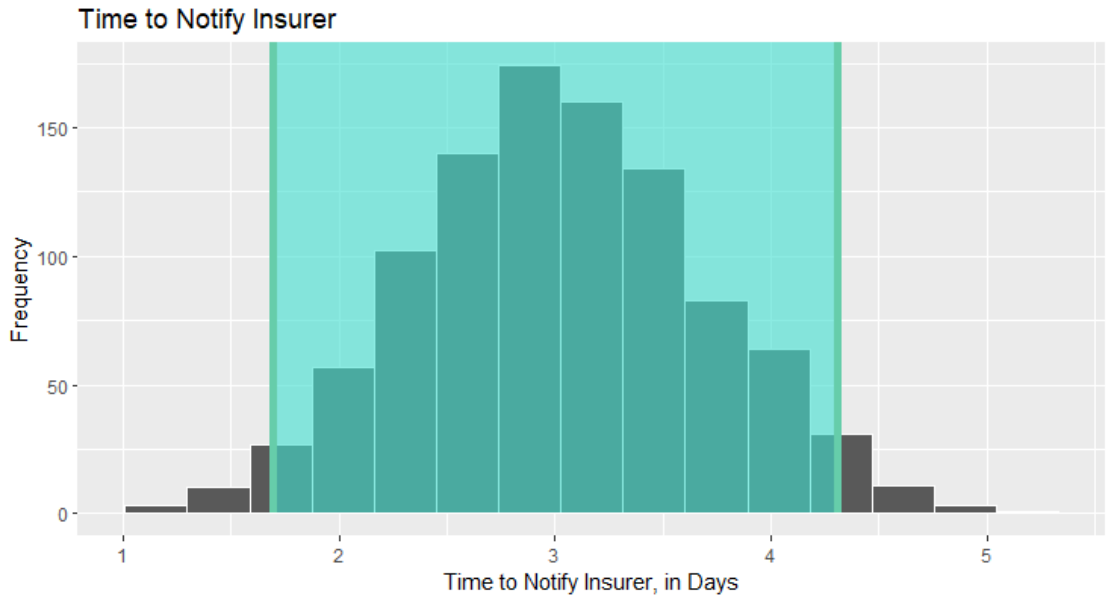

Total Indemnity
150
100
50
0
Frequency
4,500
4,800
5,100
5,400
Total Indemnity, in Dollars

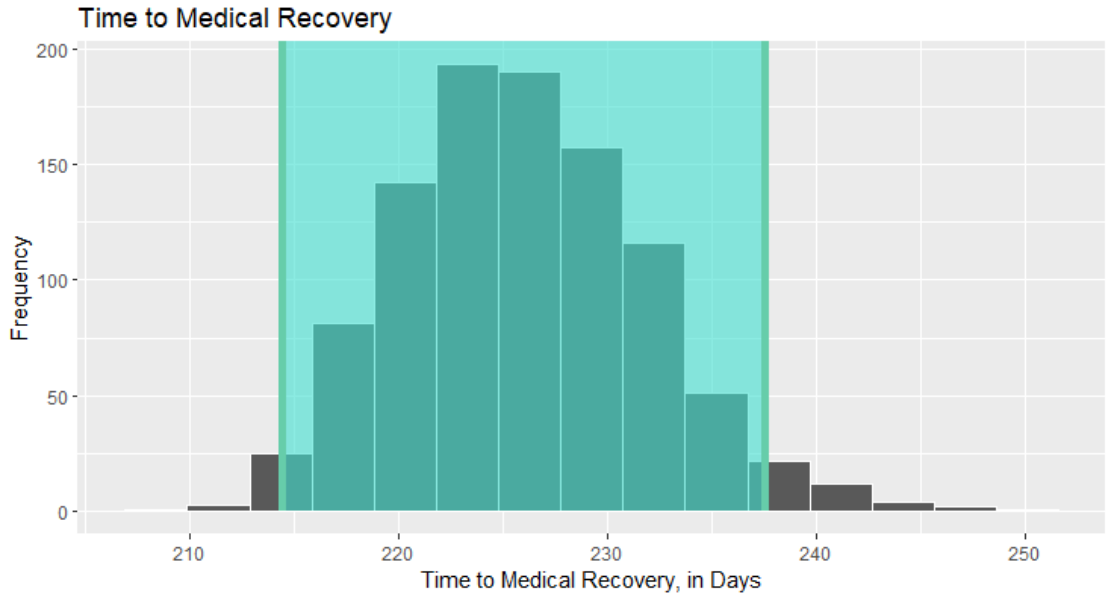

Total Lump Sum
200
150
100
50
0
Frequency
8,000
8,500
9,000
Total Lump Sum, in Dollars

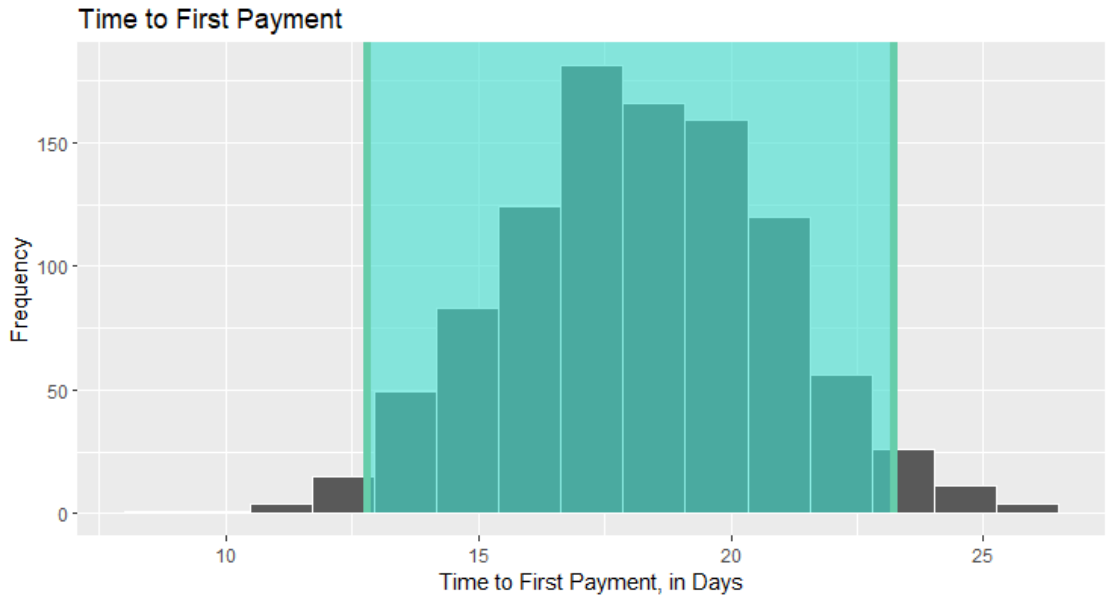

Claim Legal Ptd
150
100
50
0
Frequency
0
2,500
5,000
7,500
Claim Legal Ptd, in Dollars


Employer Legal Ptd
Frequency
150
100
50
0
1,200
1,300
1,400
Employer Legal Ptd, in Dollars


Time Away From Work
Frequency
250
200
150
100
50
0
20
25
30
Time Away From Work, in Days


Time to Notify Insurer
Frequency
150
100
50
0
1
2
3
4
5
Time to Notify Insurer, in Days


Time to Medical Recovery
Frequency
200
150
100
50
0
210
220
230
240
250
Time to Medical Recovery, in Days

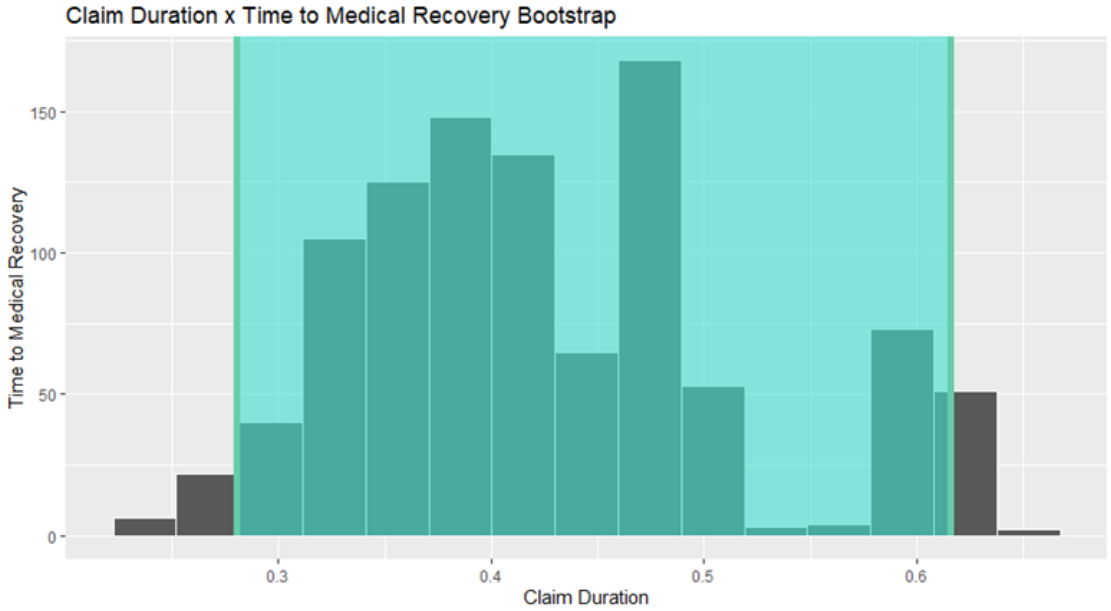


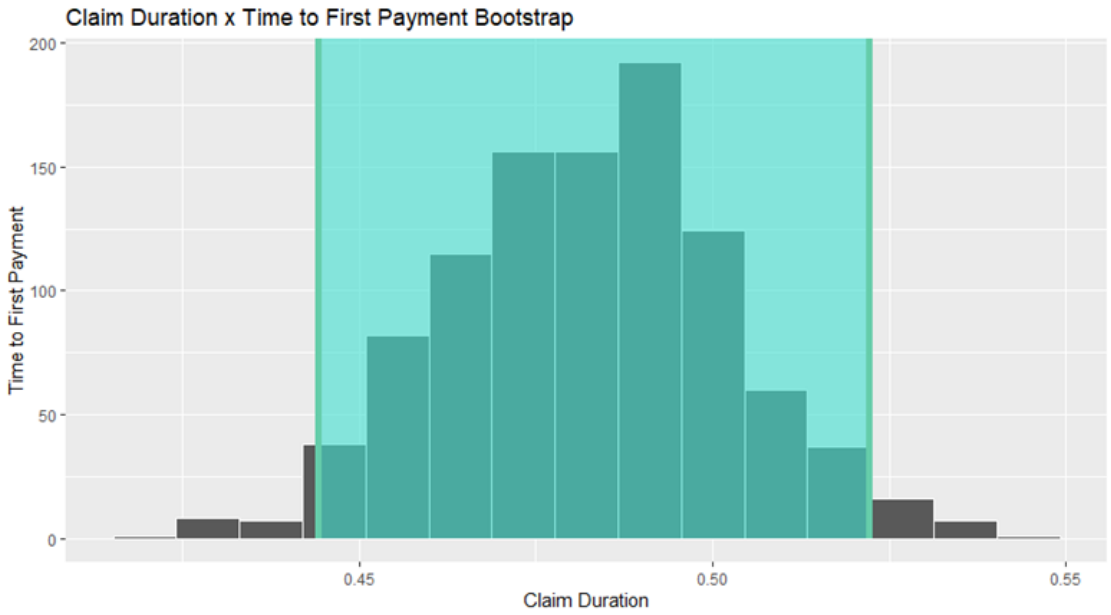
Below in Table 3.2 is a listing of all bootstrap confidence intervals for each variable median. The median is the midpoint in the distribution (50th percentile) and we can estimate the population median using the bootstrap confidence interval as well. An example interpretation would be median total cost of the claim variable from the sample was $11,819 days. The bootstrap confidence interval estimate has a lower limit of $11,618 and an upper limit of $12,160. The inference is the population median for claim duration lies within $11,618 to $12,160 with 95% level of confidence. The rest of the bootstrap confidence intervals for the other variables are below.

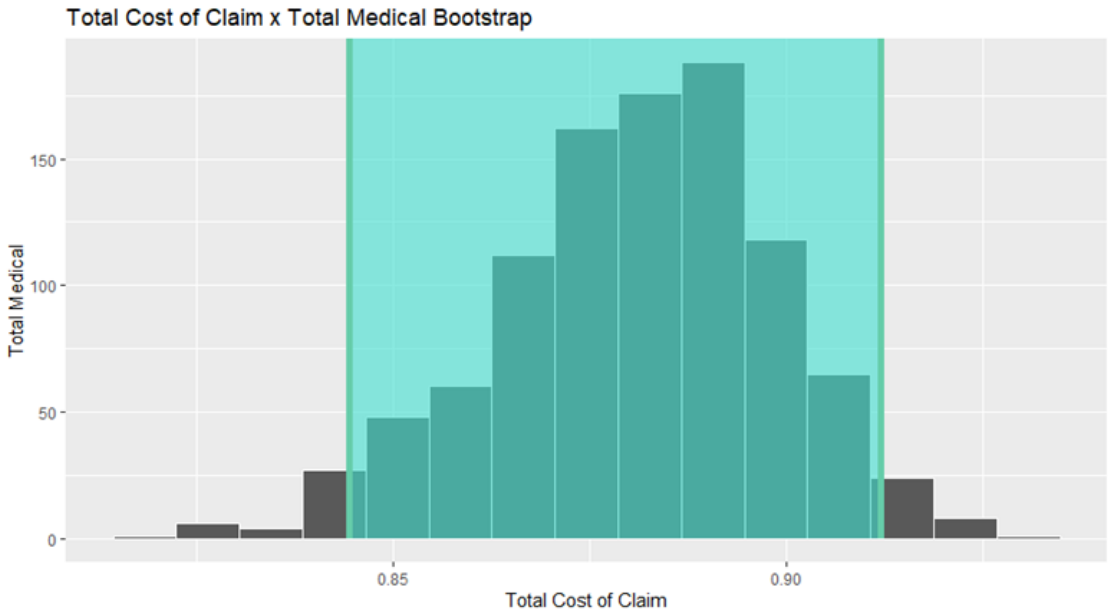
Table 3.2. Bootstrap Median Confidence Intervals

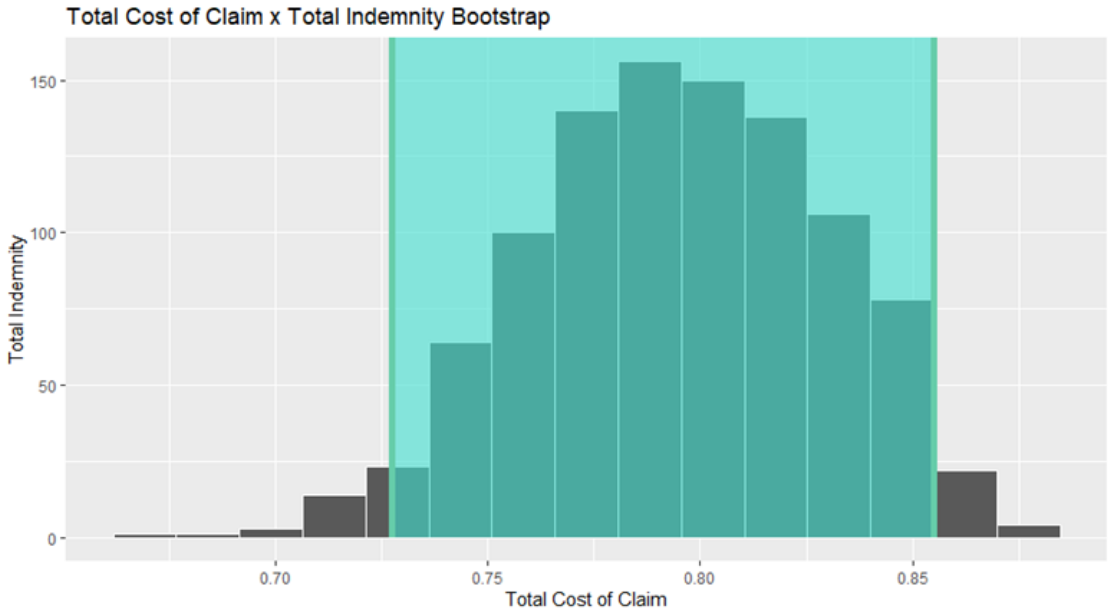
Variable	Lower CI	Observed Median	Upper CI
Claim Duration	410	415	421
Total Cost of Claim	11618	11819	12160
Total Medical	6522	6706	6890
Total Indemnity	4700	4884	5000
Total Lump Sum	8376	8588	9000
Claim Legal Ptd	2500	3760	5143
Employer Legal Ptd	1241	1301	1385
Time Away from Work	20	21	22
Time To Notify Insurer	3	3	3
Time To Medical Recovery	222	226	230
Time To First Payment	18	18	19

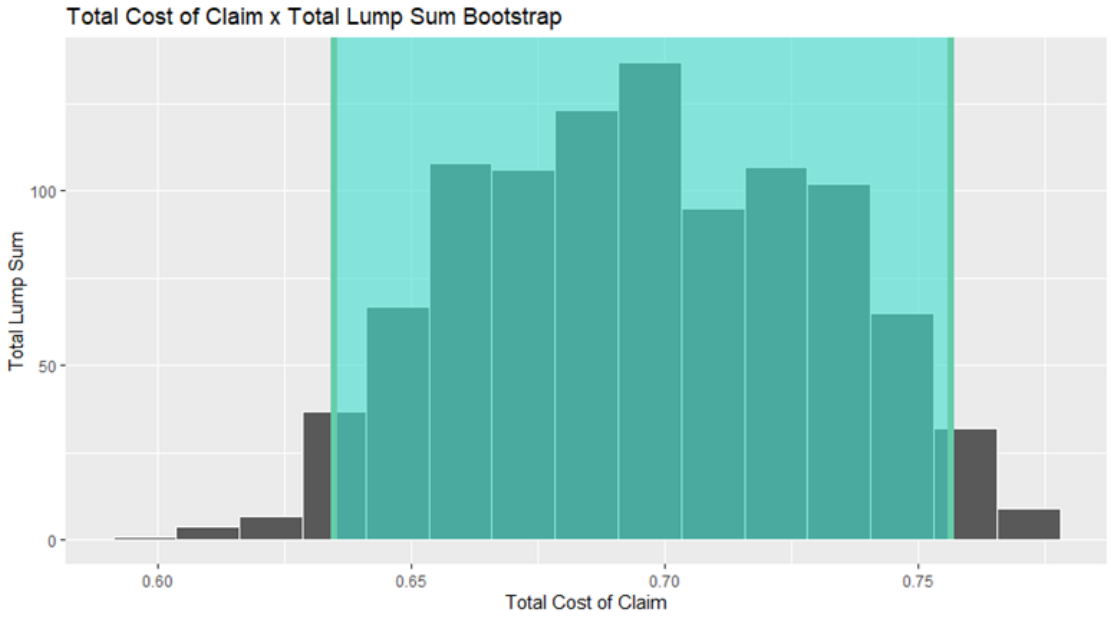
Figure 6.1-Figure 6.9. Bootstrap correlation coefficients for all variables with a sample correlation coefficient of greater than or equal to 0.4 and less than or equal to -0.4.

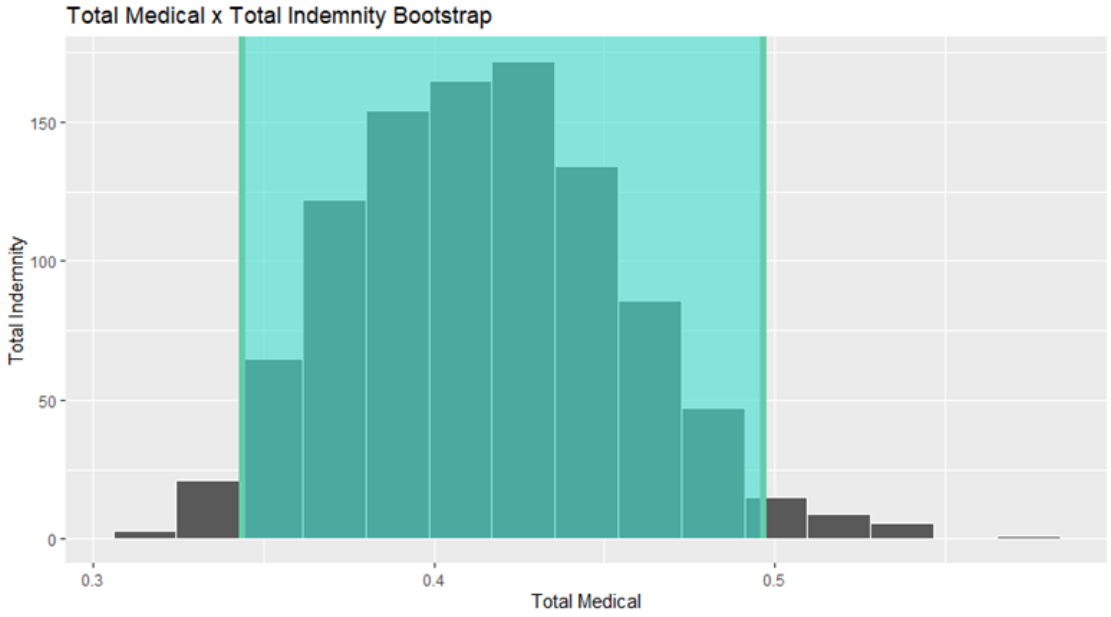


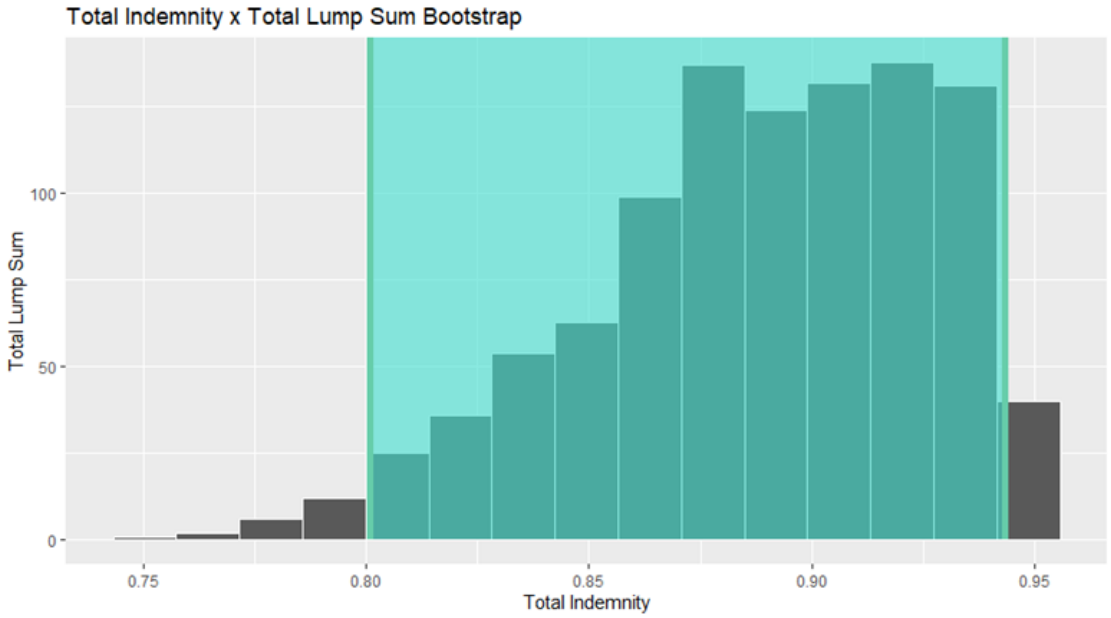


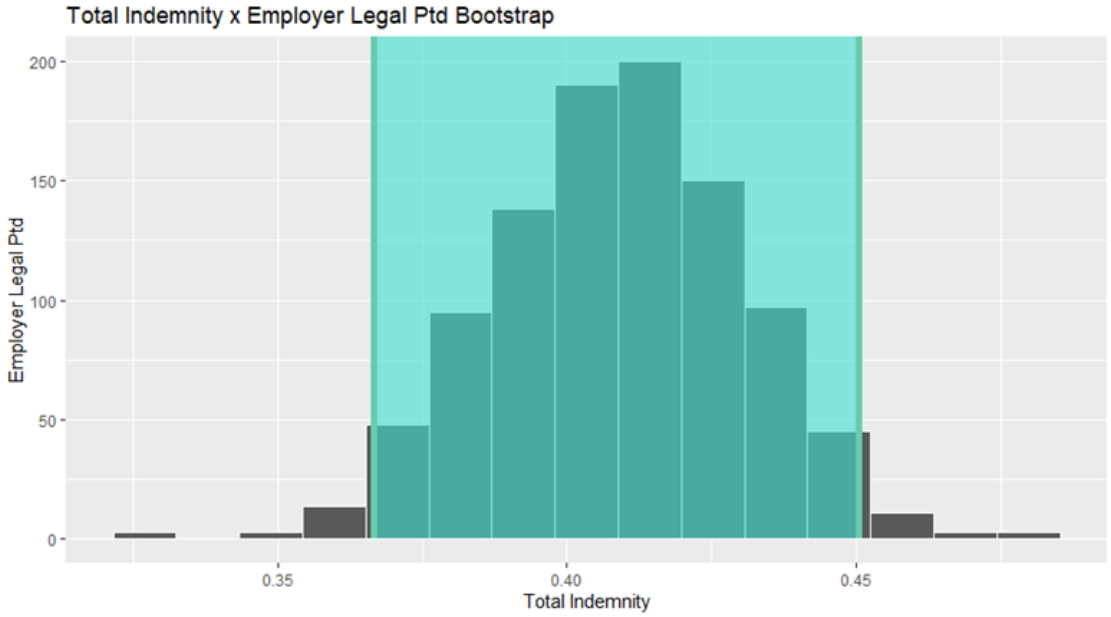

Total Cost of Claim x Total Medical Bootstrap
Total Medical
150
100
50
0
0.85
0.90
Total Cost of Claim


Total Cost of Claim x Total Indemnity Bootstrap
Total Indemnity
150
100
50
0
0.70
0.75
0.80
0.85
Total Cost of Claim


Total Cost of Claim x Total Lump Sum Bootstrap
Total Lump Sum
100
50
0
0.60
0.65
0.70
0.75
Total Cost of Claim


Total Medical x Total Indemnity Bootstrap
Total Indemnity
150
100
50
0
0.3
0.4
0.5
Total Medical


Total Indemnity x Total Lump Sum Bootstrap
Total Lump Sum
100
50
0
0.75
0.80
0.85
0.90
0.95
Total Indemnity


Total Indemnity x Employer Legal Ptd Bootstrap
Employer Legal Ptd
200
150
100
50
0
0.35
0.40
0.45
Total Indemnity

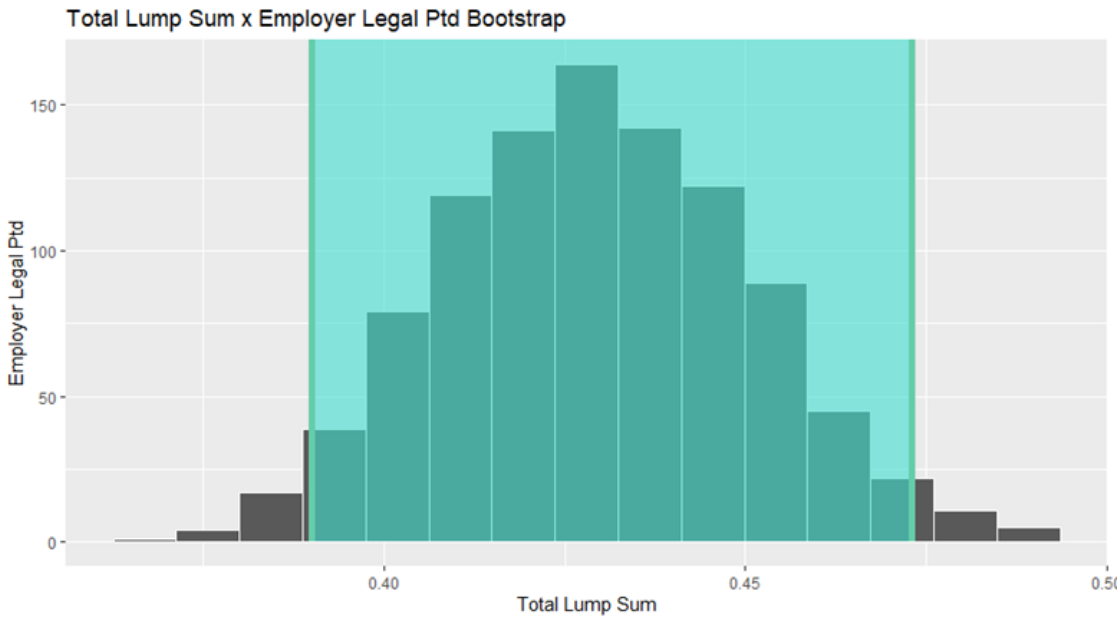


Table 3.2. Bootstrap Confidence Intervals for Correlation Coefficients

Variable 1	Variable 2	Lower CI	Correlation	Upper CI
Claim Duration	Time to Medical Recovery	0.279	0.4	0.617
Claim Duration	Time to First Payment	0.442	0.48	0.519
Total Cost of Claim	Total Medical	0.842	0.88	0.91
Total Cost of Claim	Total Indemnity	0.722	0.79	0.856
Total Cost of Claim	Total Lump Sum	0.629	0.69	0.755
Total Medical	Total Indemnity	0.339	0.41	0.496
Total Indemnity	Total Lump Sum	0.803	0.89	0.942
Total Indemnity	Employer Legal Ptd	0.363	0.41	0.449
Total Lump Sum	Employer Legal Ptd	0.384	0.43	0.473

Table 3.3 above lists the two variables that were analyzed for a correlation, the correlation coefficient, and the lower and upper confidence interval (CI) for the estimated population correlation coefficient using the bootstrap method of inference.