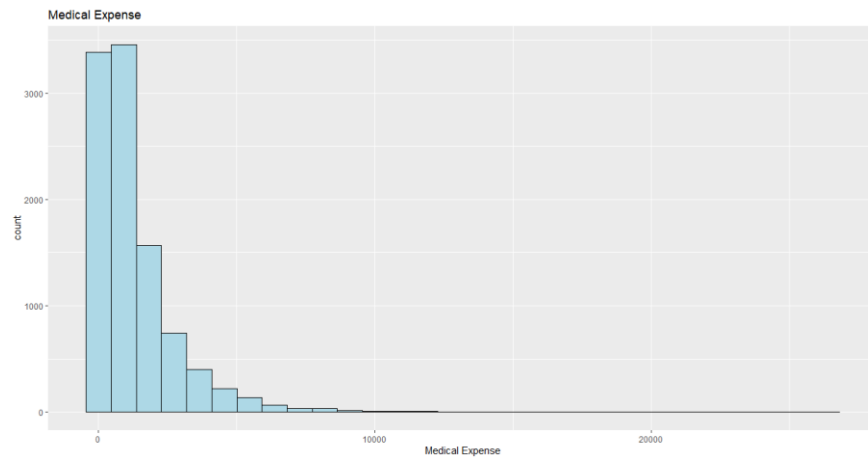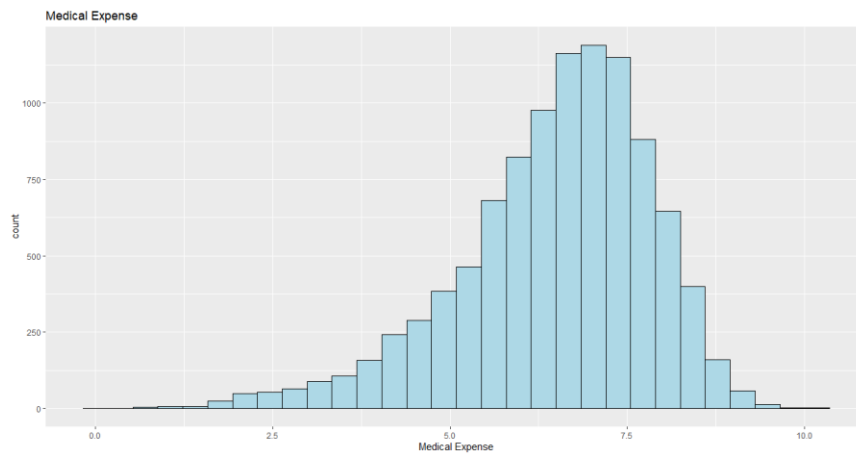# Effect of Health Insurance on Medical Expenses

Histogram of medexpense

The data is right skewed and is not normally distributed. Therefore, is not suitable for OLS regression model. Log transformation can be applied to make the variable closer to normal.

The figure on below shows the distribution of transformed variable (log of medical expense).

Analysis of predictor variables

Predictor Name	Sign of the Hypothesized effect	Rationale for that Predicted Effect
healthins	Negative	If patient has health insurance, the out of expense decreases as the insurance copays
age	Positive	It is expected that older people have more ailments, therefore as age increases out of pocket expense increases
female	Positive	Women tend to have more medical expenses as they are more affected
blackhisp	No effect	Race may not have an effect on medical expense, as a race doesn't have any special illnesses
income	Maybe positive or negative	Higher income people tend to buy premium insurances which cover a lot, however they tend to attend premium hospitals which nullifies effect, lower income people may not have insurance and end up paying high.
illnesses	Positive	More illnesses a person has more visits to hospitals increasing expenses
ssiratio	No effect	The social security benefits a person gets may not have an effect on the medical expense, as a person getting more money as benefits need not spend more money on healthcare
lowincome	Positive	Low income people tend to have no insurance or a nominal plan. Therefore any illness could mean they have to spend a lot of money on treating
firmsize	No effect	Size of the firm may not have any effect on the medical expense as healthcare is personal and firm size may not have affect on health
firmlocation	No effect	The location a person works may not have an effect on how is health is affected, therefore medical expenses may not have any effect
educyr	Negative	High education could mean more awareness managing expenses and may take better health insurance plans, reducing expenses
private	Positive	Private insurance requires more copays and deductibles than public therefore more out of pocket expenses
hisp	No effect	Race may not have an effect on medical expense, as a race doesn't have any special illnesses
black	No effect	Race may not have an effect on medical expense, as a race doesn't have any special illnesses
married	No effect	May not have any effect as being married or unmarried doesn't have affect on illnesses faced
verygood	Negative	Healthy people tend to visit hospitals less therefore less medical expenses
good	Negative	Healthy people tend to visit hospitals less therefore less medical expenses
fair	Maybe positive or negative	May visit hospital more therefore more expenses
poor	Positive	Unhealthy people tend to visit hospitals more therefore less medical expenses
poverty	Positive	Lower income people may not have insurance and end up paying high.
midincome	Negative	Middle income people may have insurance and end up paying less and also they don't visit expensive hospitals
msa	Positive	Expenses are possibly high for rural populations than for urban populations, possibly due to a greater prevalence of poor health status in rural populations
prioritylist	Positive	If the patient is elderly or fragile, hospitalizations tend to be more increasing expenses

Linear Regression Models

lm1 =lm(logmedexpense~healthins+age+female+income+illnesses+lowincome+educyr
+private+verygood+good+fair +poor+poverty+midincome+msa+prioritylist,hi)

lm2 = lm(logmedexpense~healthins+age+female+illnesses+lowincome+educyr
+private+verygood+good+fair+poor+poverty+midincome+msa+prioritylist,hi)

lm3 <- lmrob(logmedexpense~healthins+age+female+illnesses+lowincome+educyr
+private+verygood+good+fair+poor+poverty+midincome+msa+prioritylist, data=hi)
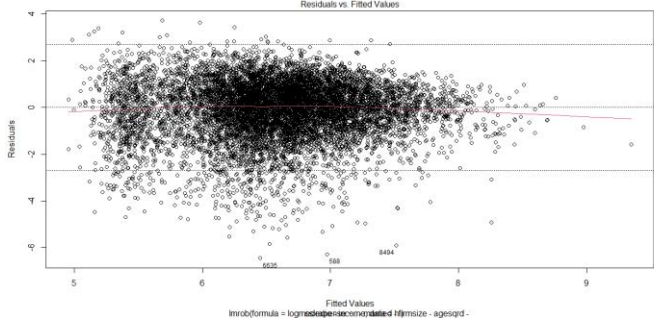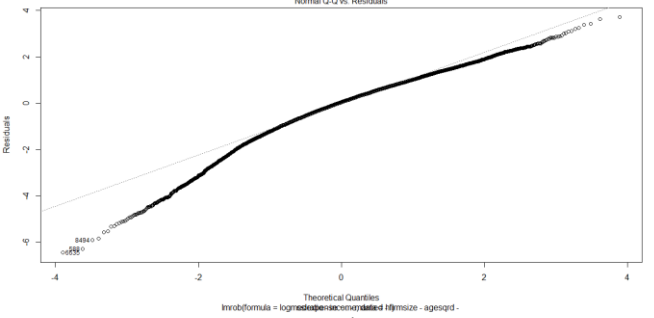
<div align="center">Descriptive Statistics</div>

	Dependent variable:		
	logmedexpense		
	OLS		MM-type linear
	(1)	(2)	(3)
healthins	0.078** (0.032)	0.086*** (0.032)	0.083*** (0.030)
age	-0.004** (0.002)	-0.005** (0.002)	-0.005*** (0.002)
female	0.079*** (0.025)	0.077*** (0.025)	0.061** (0.024)
income	0.0004 (0.001)		
illnesses	0.354*** (0.010)	0.354*** (0.010)	0.330*** (0.009)
lowincome	-0.008 (0.042)	-0.011 (0.037)	0.009 (0.035)
educyr	0.013*** (0.004)	0.010** (0.004)	0.010*** (0.004)
private	0.009 (0.032)	-0.010 (0.032)	-0.010 (0.030)
hisp		-0.156*** (0.050)	-0.174*** (0.050)
black		-0.153*** (0.042)	-0.157*** (0.042)
verygood	0.157*** (0.039)	0.158*** (0.039)	0.149*** (0.040)
good	0.248*** (0.038)	0.250*** (0.038)	0.246*** (0.039)
fair	0.409*** (0.043)	0.418*** (0.043)	0.400*** (0.043)
poor	0.502*** (0.058)	0.502*** (0.058)	0.520*** (0.053)
poverty	-0.039 (0.045)	-0.030 (0.040)	0.014 (0.038)
midincome	0.030 (0.035)	0.026 (0.031)	0.036 (0.029)
msa	-0.055** (0.028)	-0.037 (0.028)	-0.032 (0.026)
prioritylist	0.574*** (0.038)	0.574*** (0.038)	0.582*** (0.045)
Constant	5.157*** (0.159)	5.302*** (0.160)	5.462*** (0.158)
Observations	10,089	10,089	10,089
R^2	0.205	0.206	0.221
Adjusted R^2	0.204	0.205	0.220
Residual Std. Error	1.216 (df = 10072)	1.214 (df = 10071)	1.078 (df = 10071)
F Statistic	162.149*** (df = 16; 10072)	154.112*** (df = 17; 10071)	

Note: *p<0.1; **p<0.05; ***p<0.01

Checking *assumptions of OLS regression*

Assumptions	Result	Test
Linearity	No, the residuals are not linear	 Residuals vs. Fitted Values

Normality	No, the residuals are not normally distributed.	
		Normal Q-Q vs. Residuals
		
		`Two-sample Kolmogorov-Smirnov test`
		`data:  rnorm(10000) and lm3$res`
		`D = 0.049973, p-value = 2.554e-11`
		`alternative hypothesis: two-sided`

Homoscedasticity	No, the residuals are heteroscedastic	`Bartlett test of homogeneity of variances`
		
		`data:  list(lm3$res, lm3$fit)`
		`Bartlett's K-squared = 4814.1, df = 1, p-value`
		`< 2.2e-16`

Multicollinearity	No independent variables are multicollinear	``` healthins age female illnesses 1.652263 1.087206 1.031141 1.227621 lowincome educyr private hisp 1.411635 1.271651 1.625294 1.103268 black verygood good fair 1.046284 2.424101 2.689111 2.355888 poor poverty midincome msa 1.665330 1.464569 1.426601 1.049796 prioritylist 1.115050 ```

Autocorrelation	No, autocorrelation is observed	`Durbin-Watson test`
		`data:  lm3`
		`DW = 1.7917, p-value < 2.2e-16`
		`alternative hypothesis: true autocorrelation`
		`is greater than 0`

Based on my analysis, the OLS assumptions linearity, normality, homoscedasticity have failed. The model may not be the best model, however, could be an approximation.

Insights

- *Do people with health insurance have higher or lower medical expense than people without health insurance, when other variables are controlled? By how much? Why do you think this happens?*
  People with health insurance have higher medical expense. There is an increase of 8.3% in medical expense for people with health insurance. It may be because people with health insurance took health insurance because they are more health conscious. They may visit hospital more as they can dependent on health insurance. It could also be people without health insurance do not take tests ordered after visit as they are expensive.

- *Do people with private insurance pay more or less than people with public insurance? By how much?*
  People with private insurance pay less than public insurance by 1%.

- *Do people with more illnesses have higher or lower medical expense than people with less illnesses? By how much?*
  People with more illnesses have higher medical expense than people with less illnesses by 33%.

- Do males have higher medical expense than females? By how much?
  Males have lower medical expense than females by 6.1%.

- Do older people have higher medical expense than younger people? By how much?
  Older people have higher medical expense than relatively younger people by .5%. However, it has to be remembered that the age of all patients is between 65 and 91, so demographically they are old.

- Do minority groups (Blacks/Hispanics) have higher or lower medical expenses than the non-minority population? By how much?
  Minority groups have lower medical expenses than non-minority population, blacks have 17.4% less and Hispanics 15.7% less.

- How do people's income level relate to their medical expense, when controlled for other factors? By how much?
  People in poverty tend to spend 3.9% less, people with low income spend .8% less and people with middle income spend 3% more.