Data Analysis 2
Zoey Le
AST 550

There are many factors impacting on the weight of newborns based on the data from the National Center for Health Statistics National Vital Statistics Systems. This project is to identify the associations between four birth certificate variables and infant birth weight found on the copy of the Standard Certificate of Live Birth of November 2003. (1) Mother's education level, (2) Mother's race, (3) Total number of prenatal visits, and (4) Sex of Infant are picked for my report. Each variable is to find the relationship whether it has any impact on the weight of the newborns.

First of all, sex of the infant is the most fundamental factor to determine the weight of the newborns. Baby boys are naturally heavier than baby girls. The data shows that the means in weight are 29.93lbs and 30.72lbs* for baby girls and baby boys respectively. The numbers seem to have a slight difference between 2 genders; however, based on the Pearson's chi-square test, there is sufficient evidence saying that gender and weight are dependent on each other. On top of that, girls are more likely to be less than 2500 grams (5lbs) compared to boys; 4.8% and 4.5% for girls and boys respectively. Secondly, mother's race is also a factor impacting on the weight of newborns. There is more than 7% for Black and American Indian mothers to give birth to low-weight infants; more than twice higher than Asian mothers and 1.5 times higher than White mothers. Additionally, there is a wider spread (aka variance) in the data for Black and American Indian mothers. To be more specific, 273.38 for black and 270.89 for American Indian compared to 155.37 for Asian and 216.48 for white. These variances are telling us that black and American Indian mothers should be categorized into more levels of treatments during their pregnancy. Thirdly, hypothetically, mothers who visit to prenatal care services give birth to healthier infants in terms of their weight. Yet, the result turns out differently; the correlation between these two variables is 0.020. This number is closer to 0. In other words, there is no relationship between number of prenatal visits and the weight of the infant. Finally, mother' education level has a significant impact on the weight of the infant. Based on Pearson's chi-squared test, p-value nearly equals to 0. In other words, there is strong evidence that education levels of the mother psychologically have an impact on the weight of the infants. Mother who are in the groups of high school or GED completed, 8^{th} grade or less, and 9^{th} through 12^{th} with no diploma have higher chance to getting infants with low weight; 6.8% for 8^{th} grade or less, 6.6% for 9^{th} through 12^{th} with no diploma and 6.0 for high school or GED completed. Mothers who have a bachelor's degree or higher have the least chance of giving birth to low-weight babies.

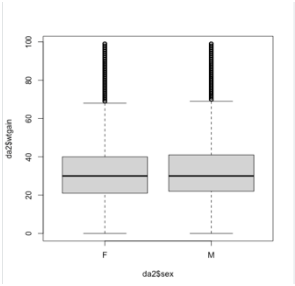
In the conclusion, number of prenatal visits doesn't determine the health of the infant. Particularly, spending more time with doctors doesn't mean the child will be healthier. Whereas the factors impacting on the child's weight are coming from the most natural status of the mother and the baby. Groups of baby girls, mothers from black and American India, and mothers giving birth at their early stage of education should be considered and treated more carefully.

APPENDIX:

Python code:

```
import pandas as pd
lb = pd.read_csv('linkedbirth.csv')
data = lb.loc[lb.index[0:],['mracerec','meduc','uprevis', 'sex','wtgain']]
data.to_csv('da2.csv')
```

[1] Infant gender:



fdata: weight for baby girls

> summary(fdata)

Min.	: 0.00
1st Qu.	:20.00
Median	:30.00
Mean	:29.93
3rd Qu.	:39.00
Max.	:98.00

mdata: weight for baby boys

>summary(mdata)

Min.	:0.00
1st Qu.	:21.00
Median	:30.00
Mean	:30.72
3rd Qu.	:40.00
Max.	:98.00

** the means are impacted by outliers*

gendertable

	Low weight	Normal
F	4.9%	95.1%
M	4.5%	95.5%

Chi-squared test:

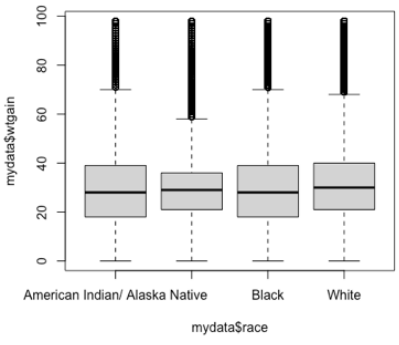
> gendertest <- chisq.test(gendertable)

Pearson's Chi-squared test with Yates' continuity correction

data: gendertable

X-squared = 306.94, df = 1, p-value < 2.2e-16

[2] Mother's race:



> summary(asiandata$wtgain)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	21.00	29.00	28.95	36.00	98.00

> summary(blackdata$wtgain)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	18.00	28.00	28.95	39.00	98.00

> summary(nativedata$wtgain)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	18.00	28.00	29.19	39.00	98.00

> summary(whitedata$wtgain)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	21.00	30.00	30.76	40.00	98.00

racetable

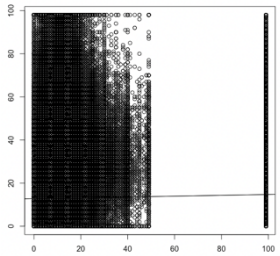
	low weight	normal
American Indian/Alaska	7.04%	92.96%
Asian/Pacific Islander	3.04%	96.96%
Black	7.54%	92.46%
White	4.22%	95.78%

Chi-squared test:

```
> testrace <- chisq.test(racetable)
> testrace
```

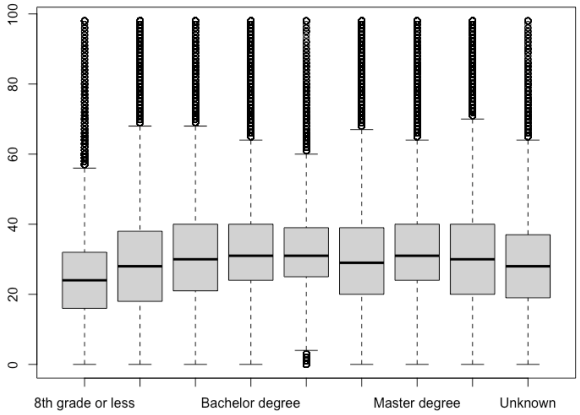
```
    Pearson's Chi-squared test

data:  racetable
X-squared = 14190, df = 3, p-value < 2.2e-16
```

[3] Number of prenatal visits:



```
> cor(x = da2.2$uprevis, y= da2.2$wtgain)
[1] 0.02073702
```

[4] Education Level:



edutable

	Low Weight	Normal
8th grade or less	6.8%	93.2%
9th through 12th with no diploma	6.6%	93.4%

Association Degree	4.6%	95.4%
Bachelor Degree	2.9%	97.1%
Doctorate of Prof Degree	1.9%	98.1%
High School or GED completed	6.0%	94.0%
Master Degree	2.4%	97.6%
Some College credit but not a degree	5.3%	94.7%
unknown	6.0%	94.0%

Chi-squared test:

```
> edutest <- chisq.test(edutable)

    Pearson's Chi-squared test

data:  edutable
X-squared = 17061, df = 8, p-value < 2.2e-16
```