Statistical Analysis of Correlation between TCB and TSB

**1.Introduction**

The aim of this study was to identify whether the relation between TSB and TCB and which factors could affect the correlation.

* 1. **Study Design**

When new babies were born, they have high risk to get Neonatal Hyperbilirubinemia. Because they have high level of bilirubin in their blood. Bilirubin is the main metabolite of heme in the body. It’s toxic and can cause irreversible damage to the brain and nervous system. Neonatal Hyperbilirubinemia may lead to seizures, cerebral palsy and nuclear jaundice.

TCB and TSB are two methods used to test bilirubin but it must be tested within the first 24 hours. TCB (Transcutaneous Bilirubin) can get from neonates’ blood. But TSB (Total Serum Bilirubin) this is painful for new born babies. People want to know if TCB can replace TSB.

In this project data was collected from Riverside County Regional Medical Center between January 1st and June 30th in 2014 (N=125). The goal of this project is response to answer the following questions. First one “Does TSB and TCB have correlation?” If there are relationship between two methods then second one “Does TSB depends on birth weight and gender?”

* 1. **Variables**

TCB is Y variable and TSB represent X variable. When TCB changes fellow with TSB then those two variables are correlated. If TSB and TCB are correlated there are two more factors which is gender and Birth weight. Gender are male and female. Birth weight is measure in the first 24 hours of new born. All new born are greater than 36 6/7 weeks and 42 weeks of gestation. Gender and birthweights are two factors that may affect the correlation between TCB and TSB. If there is correlation then doctors can use TCB replace TSB.

**2. Methodology**

Statistical analysis of this study was performed using statistical software R.3.6.3. Two statistical procedures were used in this analysis: descriptive analysis was used to plot the graph and summarize the data. Linear Regression was used to show correlation between TCB and TSB. Regression was also used to show how birthweight and gender affect correlation.

**2.1 Descriptive Analysis**

Descriptive Analysis is used to draw different plots. Scatterplot were used to determine the relationship between TCB and TSB. Residual plot used to analyze variance within factors. QQ plot used to analyze normality within treatments.

**2.2 Linear Regression**

For this project, we are interested in how TSB and TCB change. For question 1, we are only focus on what kind of relationship between TSB and TCB. For question 2, we are only focus on other factors may affect the relationship between TSB and TCB. Factors are birthweight and gender. Therefore, for each question, we use Regression to analyze the data. The model is listed below:

Y=β0+ β1X1 + β2X2+……+ βnXn +ε

Where:

Y =TCB method

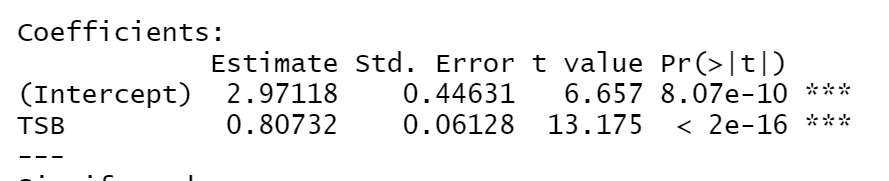
βi = the coefficient of Xi

ε = model’s error (residual)

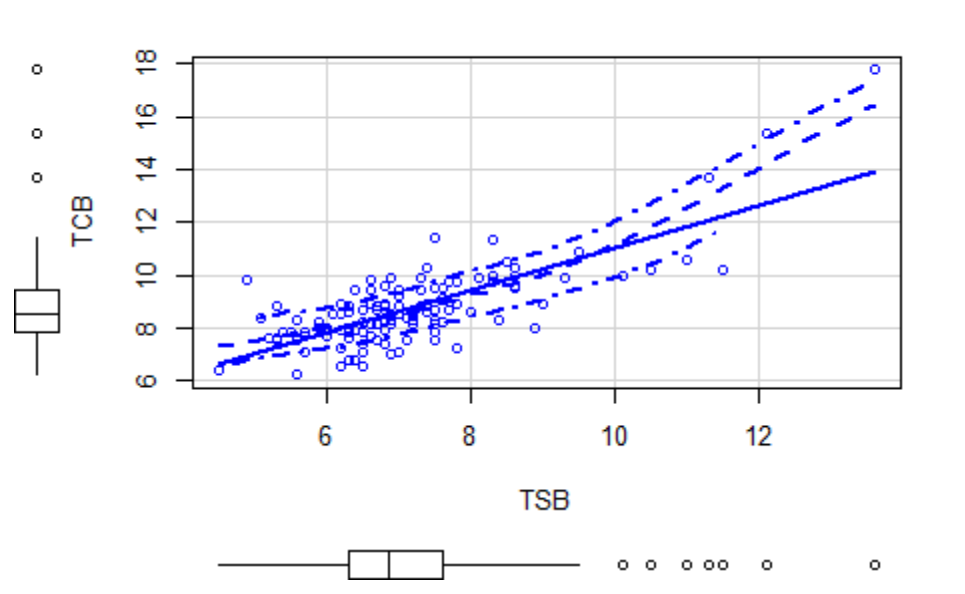
**3. Results**

**3.1 Question 1 TSB affect**

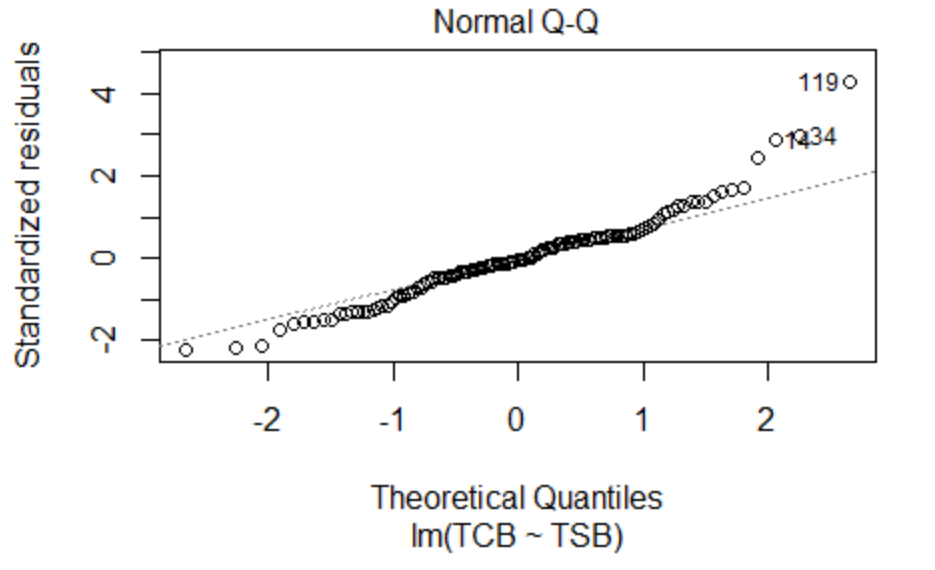
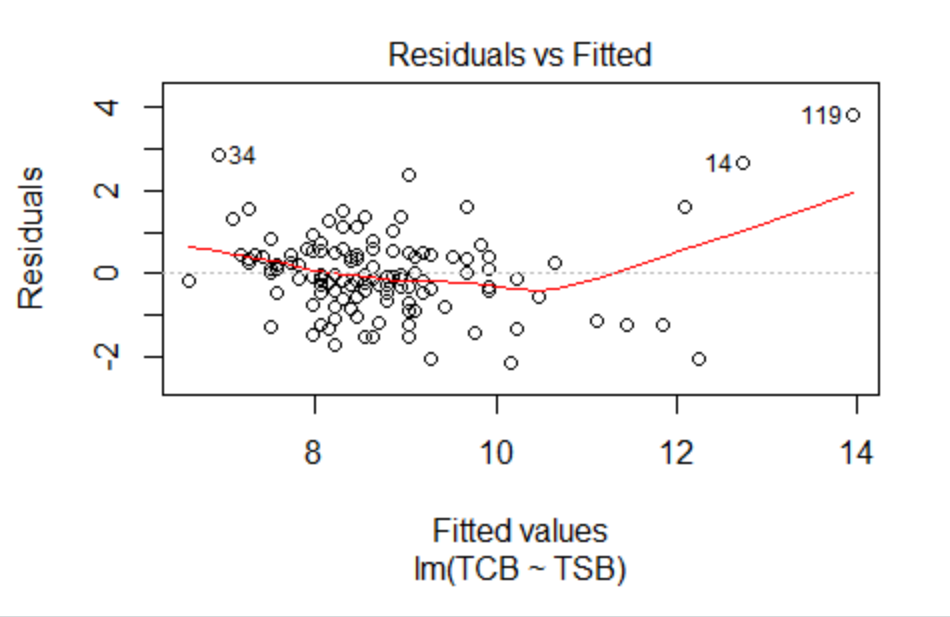
In the beginning the model I built is Y= β0+ β1X1. X is value of TSB. Based on the coefficient value of TSB I can easily see the overall relationship between TSB and TCB.

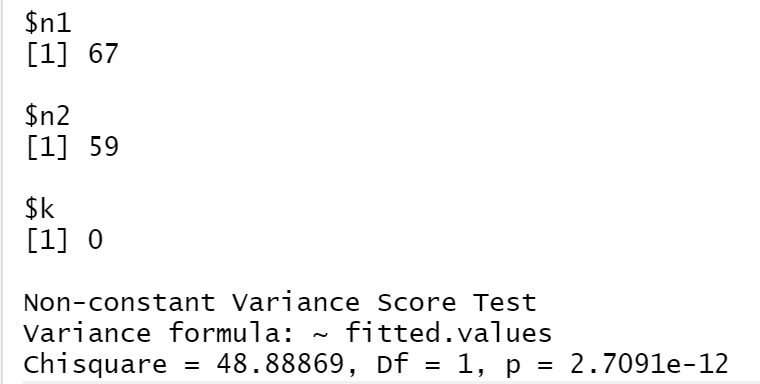
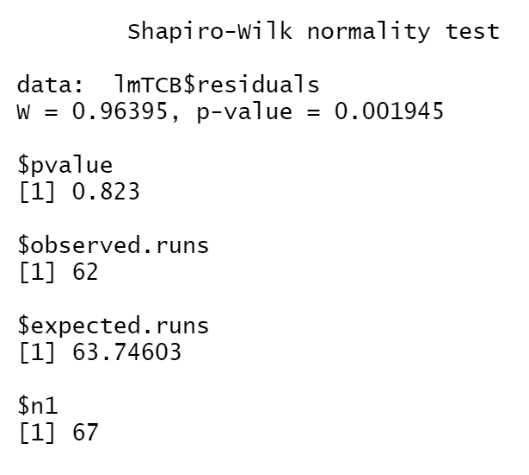


From the result we can see that P value of TSB is almost 0 which is smaller than alpha=0.05. This can be indicated that TSB have significant meaning for TCB. The coefficient of TSB is 0.807 which is 80.7% correlated with TCB. This is a high correlation value which means TSB and TCB have strong relationship. We also need to check scatter plot, if in the scatter plot there is linear trend also can prove TSB and TCB are correlated. Residual plot and QQ plot can show if this residual model has constant variance and normality.



From the scatter plot we can see that most of points are on the linear and only few data points are out of the line. This scatter plot can also indicate that in overall model TSB and TCB have relationship.

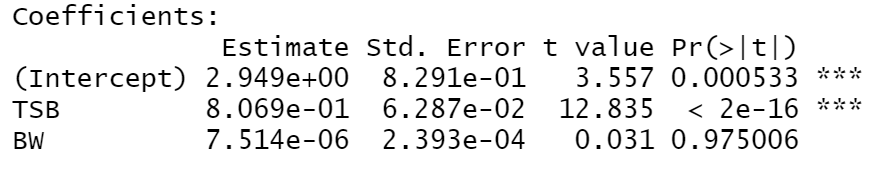




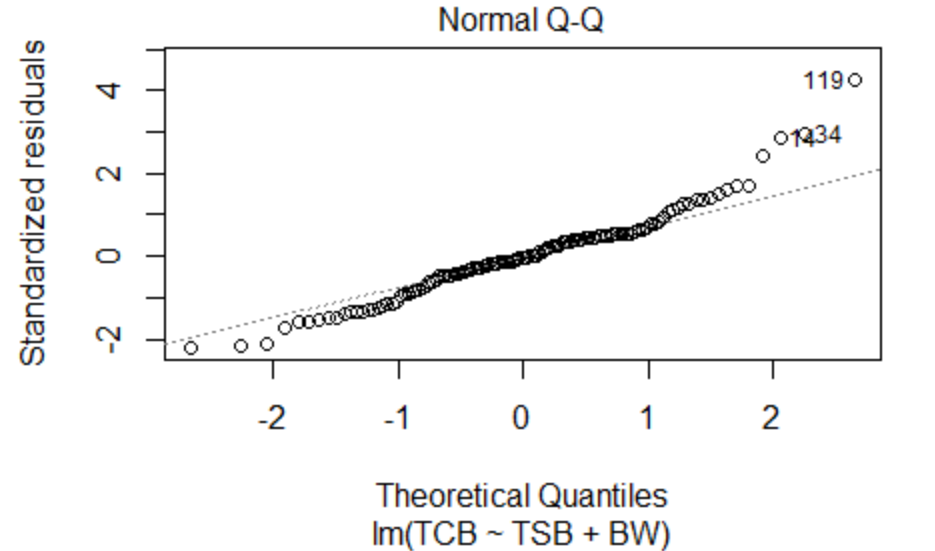
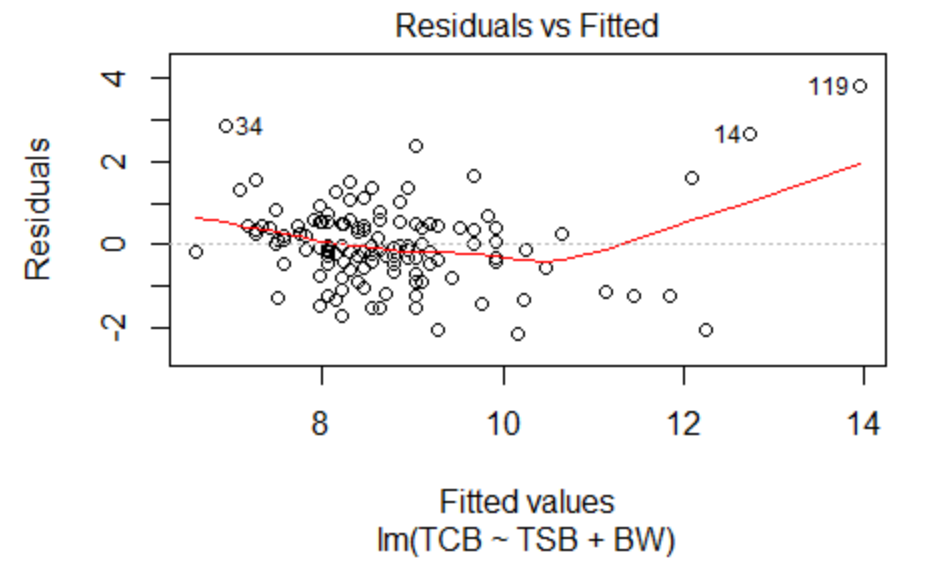
According to residuals vs fitted model most observations have constant variance but there is couple outlier in the end. From Normal QQ plot most data are rely on the linear line so it’s normality. From Shapiro Wilk test normality test satisfied. But based on the constant variance test P value < alpha constant variance satisfies. Based on run test P > alpha that shows TCB and TSB are not independent.

The overall residual model is Y=2.97118+0.807X. In this model TSB (X) and TCB (Y) have strong correlation.

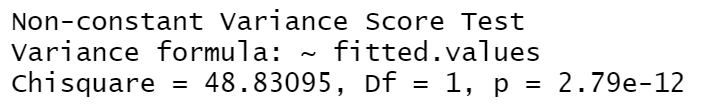
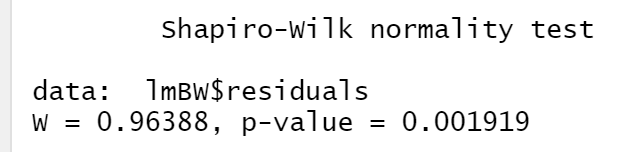
**3.2 Birthweight affect**

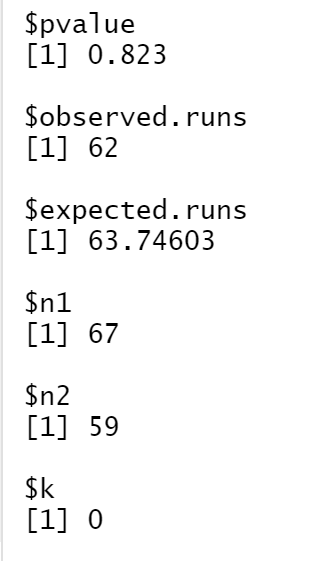
There are two factors may affect the correlation. First one is birthweight. Birthweight is my new factor the new residual model become Y= β0+ β1X1 + β2X2. X1 represent for TSB and X2 represent Birthweight. 

From this result the new model include BW is Y=2.949+0.8069\*TSB+0.000007514\*BW. We can see that P value for BW is 0.975 which is higher than alpha that indicate BW does not have significant meaning in the model. We need get rid of β2X2 from our model. The coefficient of TSB is 0.8069 almost same as overall residual model. Even though BW does not have significant meaning I still draw the residual plot, normality plot, and independent test. In order to see how normality, constant variance and independent change.

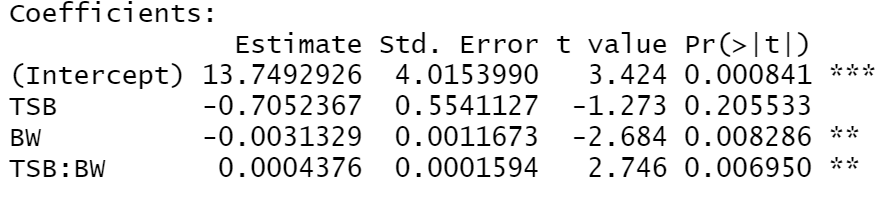


Compared with overall residual model residual vs fitted model looks same. Most data points are having constant variance but there are couple outliers. The normality plot has some outliers in the beginning and end some of points are little bit away from linear model. But overall looks like normality test satisfied.





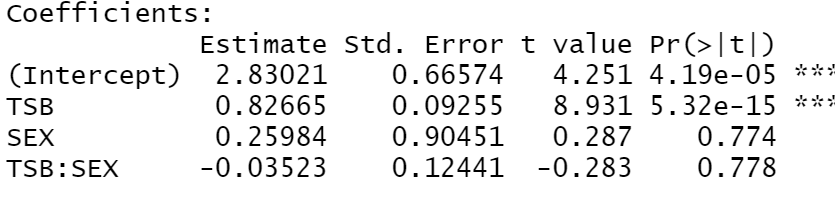
According to Shapiro test, constant variance test and run test can prove that both normality and constant variance test satisfied. All data are independent. For this model correlation value is also high but birthweight does not have significant meaning in the model. From there we can see that birthweight does not affect the correlation of the model. After that I tried to use beta as the slope of TSB to see how’s that change the value of TCB.



Based on this result we can see the new model is Y=13.749-0.00313\*BW+ (-0.7052+0.0004376\*BW) \*TSB. The coefficient of TSB is ((-0.7052+0.0004376\*BW) that means the change of TSB depend on BW. Also the P value is 0.008 which is smaller than alpha=0.05 so BW have significant meaning in the model

Another factor I need to test on is gender. I separate the while data into two parts, male and female. Use two data to check how does gender affect the correlation between TSB and TCB. The following result will be based on gender.

**3.3 Gender affect**

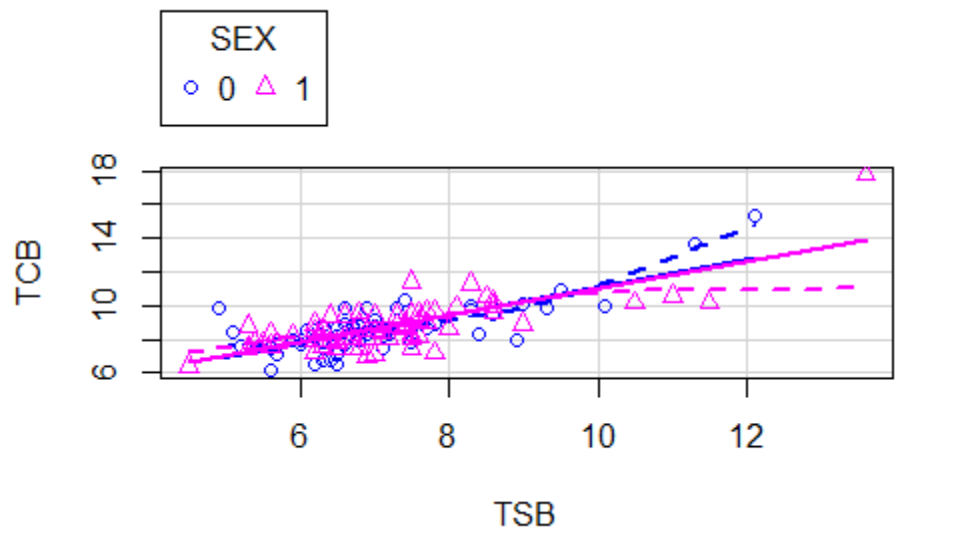


From the result we can see that P value of TSB is almost 0 which is smaller than alpha=0.05. This can be indicated that TSB have significant meaning for TCB. The coefficient of TSB is 0.827 which is 82.7% correlated with TCB. The overall correlation is 80.7%. Which is almost same as correlation based on gender? This is a high correlation value which means TSB and TCB have strong relationship. We also need to check scatter plot, if in the scatter plot there is linear trend also can prove TSB and TCB are correlated. Residual plot and QQ plot can show if this residual model has constant variance and normality.

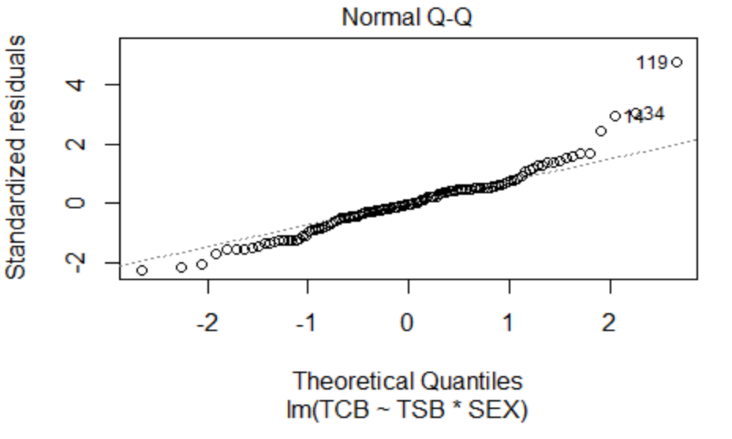
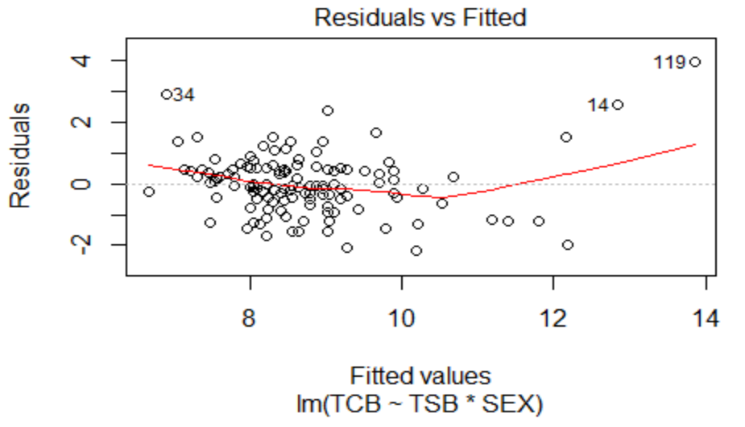
Female: TCB=2.83+0.82665\*TSB

Male: TCB= (2.83+0.26)+(0.8265-0.035)\*TSB

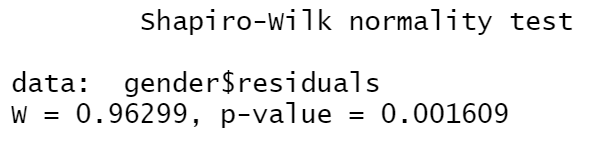
The P value for sex is 0.774 which is larger than alpha which mean that there is no significant difference between male and female.

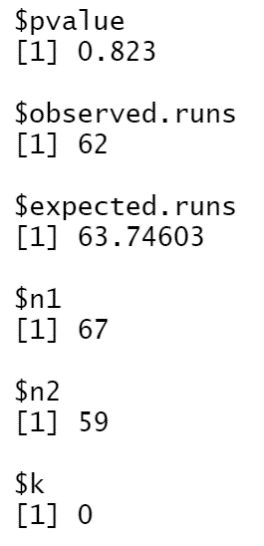


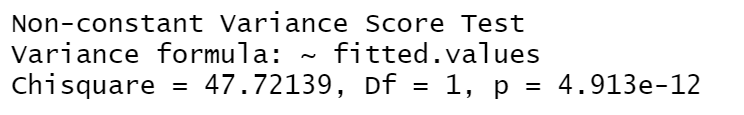
From the scatter plot we can see that most of points are on the linear and only few data points are out of the line. Blue line represent for female and pink line is for male. This scatter plot can also indicate that in overall model TSB and TCB have relationship.



Most data points are having constant variance but there are couple outliers between 8-10. The normality plot has some outliers in the end. But overall looks like normality test satisfied. Use Shapiro test, constant variance test and run test double check if this residual has normality constant variance and independent



. 



According to Shapiro test, constant variance test and run test can prove that both normality and constant variance test satisfied. All data are independent. For this model correlation value is also high which is 0.8265. The overall model residual is 0.807. Compare with overall model, residual model include gender is a little bit lower than overall model. I can conclude that gender (male) does not affect the correlation between TSB and TCB.

**4. Conclusion**

In this study we tried to find out that can we use TSB replace TCB and what kind of factors will be affect?

1. “Is there any relationship between TSB and TCB?”

Yes, based on the overall model TCB=2.97118+0.807\*TSB. TCB and TSB have pretty strong relationship. Also, data fellow with normal distribution with constant variance. We can use TSB to replace TCB.

1. What factors will affect the correlation?

**Birthweight**: Birthweight as a factor in regression have high P value which does not have significant meaning in the regression line. But birthweight as the slope in regression it can change the value of TSB.

**Gender:** Based on male the regression line shows correlation based on gender is 82.65% does not have significant different with overall regression. The P value for sex is higher than alpha which means gender does not affect the correlation.