**SCHOOL OF COMPUTING**

**COURSE: SECI2143**

PROBABILITY & STATISTICAL DATA ANALYSIS

# PROJECT 2



| **NAME** | **MATRIC ID** |
|---|---|
| NURIFFATISH SHIMA BINTI NORAZMI | A19EC0138 |

## **Introduction**

Birth weight is the weight of newborn baby. Birth weight that is less than 5.5lb is categorized as a low birth weight meanwhile birth weight that is more than 8.8lb is categorized as a high birth weight. A low birth weight baby can be born too early (premature), too small or both. There are several reasons a low birth weight can happen. It may happen because of genetic factors, mother's health, problems with the placenta or drug used by the mother. Several low birth weight babies can be further at risk for certain health problems. High birth weight babies are usually big because the parents are also big or the mother has diabetes throughout pregnancy. A high birth weight baby can be at a higher risk of birth injuries or problems with blood sugar.

The dataset that I used in this study was contributed by Ellen Marshall, University of Sheffield. This dataset named as Reduced Birth Weight and contains information of 42 newborn babies including their parents. It stated that the birth weight of the babies who mothers smoked have been adjusted a bit to magnify the differences between mothers who smoked and did not smoke in this dataset. Therefore, students can see more clearly the difference in a scatterplot with gestational age and scatter colour coded by smoking status. I am interested to test this dataset as I wanted to know the factors that affect the weight of newborn babies. So, the purpose of this study is to test on the mean of birth weight and to find the factors that will affect the birth weight. Specification of target population for this study is newborn babies including their parents.

# Hypothesis Testing

a) Hypothesis Testing Using a Single Sample**:** Test on Mean, variance unknown

As this dataset is from United Kingdom, I found that the average birth weight of United Kingdom babies is increasing, 7lb 8oz for boys and 7lb 4oz for girls. Therefore, I will use 7lb 6oz (7.4lb) as the mean population of birth weight of United Kingdom babies. I wanted to test whether the mean population is larger than 7.4lb at **0.05** level of significance.

$$H_0: \mu = 7.4$$

$$H_1: \mu > 7.4$$

meanbw : sample mean of birth weight

sdbw : sample standard deviation of birth weight

n : sample size

mu : claimed population mean of birth weight

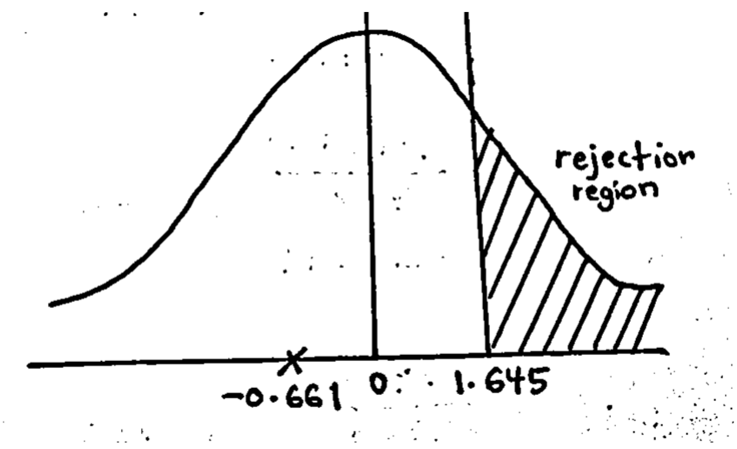
z : test statistic

z.alpha : critical value

```
> library(readxl)
> databw <- read_excel("~/psda proj 2/Birthweight reduced dataset (Shima_A19EC0138).xlsx")
> View(databw)
> meanbw<-mean(databw$Birthweight)
> sdbw<-sd(databw$Birthweight)
> n<-42
> mu<-7.4
> z<-(meanbw-mu)/(sdbw/sqrt(n))
> alpha<-0.05
> z.alpha<-qnorm(1-0.05)
```

```
> pnorm(z,lower.tail = FALSE)
[1] 0.7458317
```

| Values | |
|---|---|
| alpha | 0.05 |
| meanbw | 7.26428571428571 |
| mu | 7.4 |
| n | 42 |
| sdbw | 1.32973891231057 |
| z | -0.661429914279675 |
| z.alpha | 1.64485362695147 |




Conclusion:

Since $Z = -0.661 < Z_{0.05} = 1.645$ and P-value = 0.746 > 0.05, we failed to reject the null hypothesis at 0.05 level of significance. There is not enough evidence to show that the mean population birth weight of United Kingdom babies is more than 7.4lb. Since nowadays there are a lot of instant foods, many mothers whom do not care for their food intake may eat instant food such as eating instant noodles that lacks of all important nutrients that mothers need during pregnancy. It can cause health problems to mothers. A low birth weight baby may be born because of these health problems. If there are many mothers whom do not care of their food intake, a lot of low birth weight babies may be born and might affect the average of birth weight.

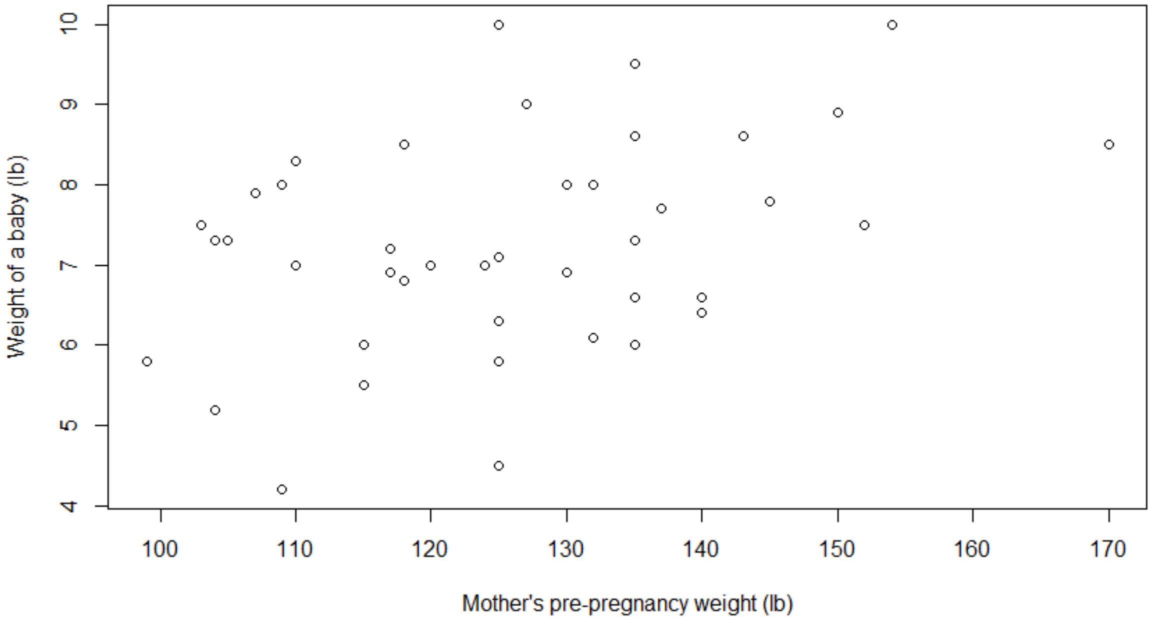
b) Correlation Analysis: The relationship between mother's pre-pregnancy weight (lb) and weight of a baby (lb)

mppwt : Mother's pre-pregnancy weight (lb)

Birthweight : Weight of a baby (lb)

```
> x<-databw$mppwt
> y<-databw$Birthweight
> cor(x,y)
[1] 0.3895806
```

```
> plot(x,y, xlab="Mother's pre-pregnancy weight (lb)", ylab="Weight of a baby (lb)")
```




It can be seen that the mother's pre-pregnancy weight increases as the weight of a baby increases. A scatter plot and correlation analysis of the data indicates that there is positive relationship between mother's pre-pregnancy weight and weight of a baby. As the correlation efficient (r) is 0.3895806, the strength of relationship between these two variables is weak.

To test whether there is an evidence of a linear relationship between mother's pre-pregnancy weight and weight of a baby at the 0.05 level of significance, I used cor.test() function to get the value of test statistic and P-value.

$$H_0: \rho = 0 \text{ (No linear correlation)}$$

$$H_1: \rho \neq 0 \text{ (linear correlation exists)}$$

```
> cor.test(x,y)

        Pearson's product-moment correlation

data:  x and y
t = 2.6753, df = 40, p-value = 0.01077
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.09715273 0.62008979
sample estimates:
      cor
0.3895806
```

test statistic, t = 2.6753

Since P-value = 0.01077 < 0.05, we reject the null hypothesis. There is sufficient evidence to show that a linear correlation exists between mother's pre-pregnancy weight and weight of a baby at the 0.05 level of significance. A research published in the journal Acta Obstetricia et Gynecologica Scandinavica (AOGS) shows that both pre-pregnancy weight (body mass index, BMI) and weight gain during pregnancy are important predictors of birth weight for a newborn baby. Results of the population-based pregnancy study performed by researchers led by Unni Mette Stamnes Koepp of the Department of Pediatrics at Soerlandet Hospital and the University of Oslo, Norway showed that birth weight increased as the maternal pre-pregnant BMI increased. Koep said, "encouraging women to attain a healthy weight before conception and keep a moderate weight gain during pregnancy is important to avoid high or excessive birthweight in offspring". Therefore, every mother needs to take of their body weight before pregnancy.

c) Regression Analysis: The relationship between gestation (weeks) and birth weight (lb)

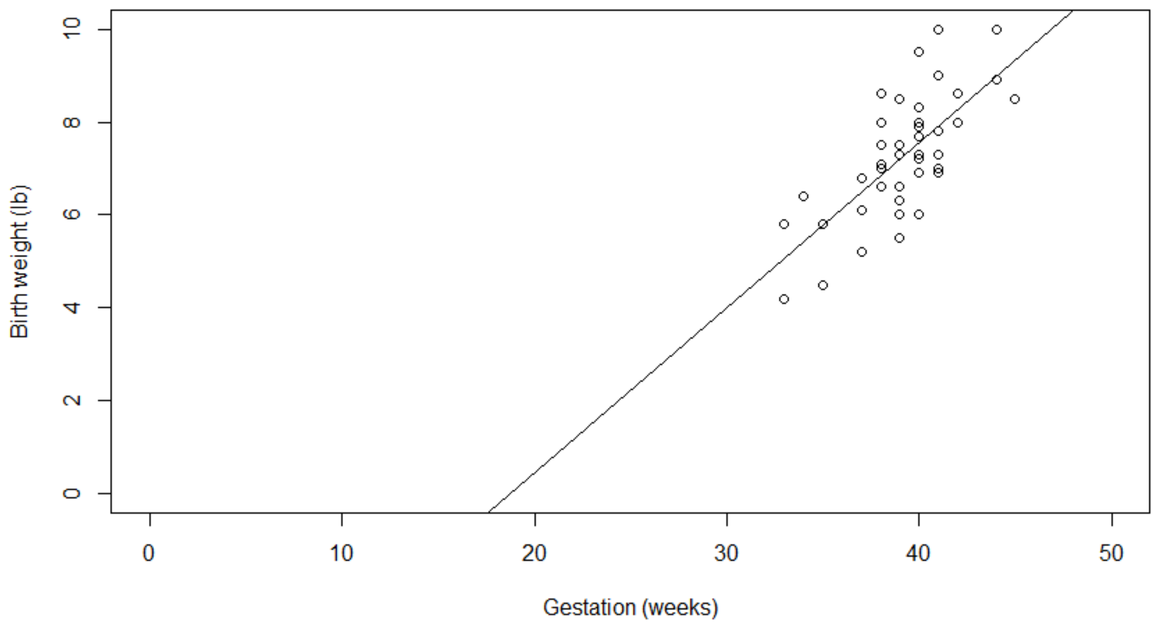
Dependent variable: Birth weight

Independent variable: Gestation

$H_0$: $\beta_1 = 0$ (no linear relationship)
$H_1$: $\beta_1 \neq 0$ (linear relationship does exist)

```
> x<-databw$Gestation
> y<-databw$Birthweight
> plot(x,y, xlim=c(0,50),ylim=c(0,10), xlab="Gestation (weeks)", ylab="Birth weight (lb)")
> model<-lm(y~x)
> model

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
    -6.6602       0.3553
```

```
> summary(model)

Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max
-1.69661 -0.78430 -0.02426  0.60604  2.09279

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.66019    2.21162  -3.011  0.00449 **
x            0.35530    0.05631   6.310 1.73e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.953 on 40 degrees of freedom
Multiple R-squared:  0.4988,	Adjusted R-squared:  0.4863
F-statistic: 39.82 on 1 and 40 DF,  p-value: 1.733e-07

> abline(model)
```




Based on the functions I used in the coding, I got ŷ = -6.6602 + 0.3553x. Since the value of P-value = 1.733e$^{-07}$ < 0.05, we reject the null hypothesis. There is sufficient evidence that linear relationship does exist between gestation and birth weight at the 0.05 level of significance. Based on the summary, $R^2$ = 0.4988 shows that 50% of the variation in birth weight is explained by variation in gestation. Therefore, I agree that gestational age is a good predictor of birth weight as mentioned in an article titled References of Birth Weights for Gestational Age and Sex from a Large Cohort of Singleton Births in Cameroon.

d) Goodness-of-fit test

$H_0$: $p_{low} = 0.2$, $p_{normal} = 0.8$

$H_1$: At least one of the proportions is different from the claimed value.

```
> freq.lbw<-table(databw[,17])
> prob<-c(0.2,0.8)
> chisq.test(freq.lbw, p=prob, correct=FALSE)

        Chi-squared test for given probabilities

data:  freq.lbw
X-squared = 0.85714, df = 1, p-value = 0.3545

> alpha<-0.05
> x2.alpha<-qchisq(alpha,df=1,lower.tail = FALSE)
> x2.alpha
[1] 3.841459
```

```
> freq.lbw

   Low Normal
     6     36
```

| Low birth weight | p | Observed frequency | Expected frequency |
|---|---|---|---|
| Low | 0.2 | 6 | 8.4 |
| Normal | 0.8 | 36 | 33.6 |

Based on the table, observed frequency and expected frequency are close. I used the chisq.test() function to get the value of $\chi^2$ and P-value. The value of $\chi^2 = 0.85714$ is small and the value of P-value is large. Since P-value = 0.3545 > 0.05, we failed to reject the null hypothesis at the 0.05 level of significance. There is not enough evidence to show that at least one of the proportions is different from the claimed value. We can say that this is a good fit with assumed distribution. Therefore, we can assume that there are many mothers whom take care of their health by taking care of their food intake to make sure their babies are born with a normal birth weight.

## Discussion and Conclusion

Based on the hypothesis testing using a single sample that I have done to test on the mean of birth weight with variance unknown, the null hypothesis is failed to reject. There is not enough evidence to show that the mean population of birth weight of United Kingdom is larger than 7.4lb. This probably might happen because of nowadays there are a lot of instant foods. Many mothers do not care for their health so they also do not care for their food intake. They may eat instant food such as eating instant noodles that lacks of all important nutrients like protein, vitamins, minerals, vegetables and fibre that mothers need during pregnancy. This can cause health problems to mothers. A low birth weight baby may be born because of these health problems in the mother. If mothers whom do not care of their food intake increases, a lot of low birth weight babies may be born and might affect the average birth weight of babies in the next years. Therefore, it may decrease the mean birth weight from 7.4lb in the next years.

Based on the correlation analysis on the relationship between mother's pre-pregnancy weight and weight of a baby that I have done, the mother's pre-pregnancy weight increases as the weight of a baby increases. Based on the scatter plot and correlation analysis of the data, there is positive relationship between mother's pre-pregnancy weight and weight of a baby. Based on the correlation efficient that I got, the strength of relationship between these two variables is weak. The P-value that I got lead to the rejection of the null hypothesis. There is enough evidence to prove that a linear correlation exists between mother's pre-pregnancy weight and weight of a baby at the 0.05 level of significance. This analysis result can prove the research published in the journal Acta Obstetricia et Gynecologica Scandinavica (AOGS) which shows that both pre-pregnancy weight (body mass index, BMI) and weight gain during pregnancy are important predictors of birth weight for a newborn baby.

Based on the regression analysis on the relationship between gestation and birth weight that I have done, the P-value that I got lead to the rejection of the null hypothesis. There is enough evidence to show that linear relationship does exist between gestation and birth weight at the 0.05 level of significance. The value of $R^2$ that I obtained shows that 50% of the variation in birth weight is explained by variation in gestation. This analysis result can prove the statement that I found from an article which stated gestational age is a good predictor of birth weight.

Based on the goodness-of-fit test that I have done on frequency of low and normal birth weight with unequal frequencies, the observed frequency and expected frequency are close. Based on the results of the coding, the value of $\chi^2$ that I obtained is small and the P-value is large. As the null hypothesis is failed to reject, I can say that this is a good fit with assumed distribution. As the proportion of normal birth weight is larger than low birth weight, we can assume that there are many mothers whom take care of their health by taking care of their food intake to make sure their babies are born with a normal birth weight.

As conclusion, the birth weight of a baby is important to be taken to check whether the baby is in a good condition or not. A baby's weight should be 5.5lb and above to be considered as normal birth weight. According to a variety of studies, a baby with birth weight low than 5.5lb is more exposed to heart disease and diabetes, slowed growth, low I.Q. and other problems. Therefore, we need to know the factors that will affect birth weight of babies. Based on the correlation and regression analysis results, I found that mother's pre-pregnancy weight and gestation are some of the factors that will affect weight of a newborn baby.

## References


1. Heeren, T., & Milton, J. (n.d.). Basic Statistical Analysis Using the R Statistical Package. Retrieved from https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R-Manual/R-Manual_print.html
2. BBC News. (2020). An Average Baby?. Retrieved from https://www.bbc.com/news/uk-23403391
3. MedlinePlus. (April 2020). Birth Weight. Retrieved from https://medlineplus.gov/birthweight.html#:~:text=Birth%20weight%20is%20the%20first,(premature)%2C%20or%20both.
4. ScienceDaily. (December 2011). Mothers' weight before and during pregnancy affects baby's weight. Retrieved from https://www.sciencedaily.com/releases/2011/12/111213110521.htm
5. PRB. (June 2007). Is Low Birth Weight a Cause of Problems, or a Symptom of Them?. Retrieved from https://www.prb.org/lowbirthweight/#:~:text=Babies%20whose%20birth%20weight%20is,to%20a%20variety%20of%20studies.
6. Mathematics and Statistics Help (MASH). Datasets for Teaching. Retrieved from https://www.sheffield.ac.uk/mash/statistics/datasets
7. Hindawi. Obstetrics and Gynecology International. (2014). References of Birth Weights for Gestational Age and Sex from a Large Cohort of Singleton Births in Cameroon. Retrieved from https://www.hindawi.com/journals/ogi/2014/361451/