# Milan Labus SD2B Assignment

## ***Q1.***

**introductory statement:** for this hypothesis test we want to test if the mean of the sample data is greater than the normal distribution of 3000g and we will do this in R by calculating the mean of the sample.

**exploratory analysis:**
The sample size n is 189

NULL hypothesis:
H0 mean > 3000

Alternative hypothesis:
HA mean <= 3000g

The hypothesis is disproved because the mean weight was not over 3000 so we reject HO

**Results:** the mean for this data set was 2944.587g so HA is disproved as it is not bigger than 3000g

**Conclusions:** The alternative hypothesis was wrong this dataset did not have a mean of over 3000g

## Q2.

1. Based on this sample find an estimate for birth weight for the population with a 95% level of confidence.

I can say with 95% confidence that the true mean lies between

[2873.304 and 3015.871] g

## Q3:

***Introductory Statement:*** In this analysis we are determining if smoking is associated with low birth weight. We have taken a sample of the population with smoking and non-smoking.

HO: The birth weight of babies from smoking mothers is not lower than the birth weight of babies from mothers who do smoke.

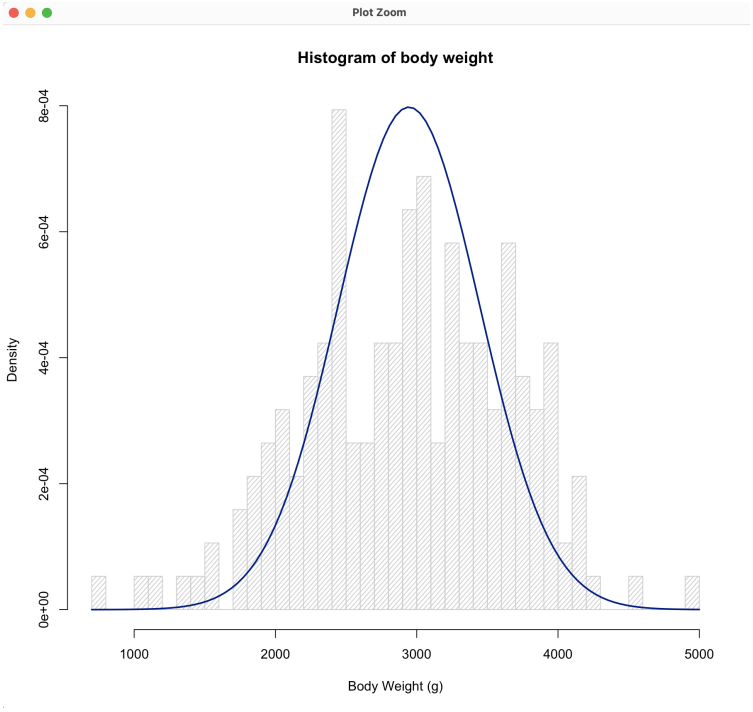
μ1>=μ2
HA: is the mean birth weight of children from smoking mothers is lower than the birth weight of babies from non smoking mothers. Showing that smoking is associated with low birth weight

μ1<μ2

***Exploratory Analysis:***

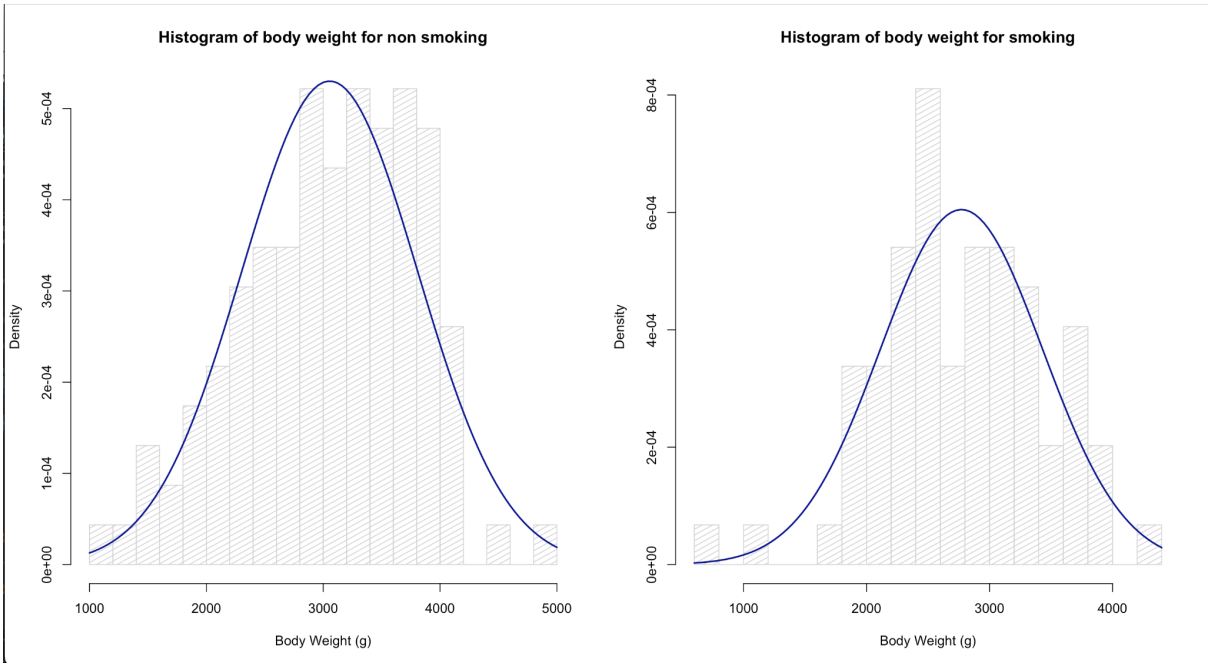
Significance level = 0.05

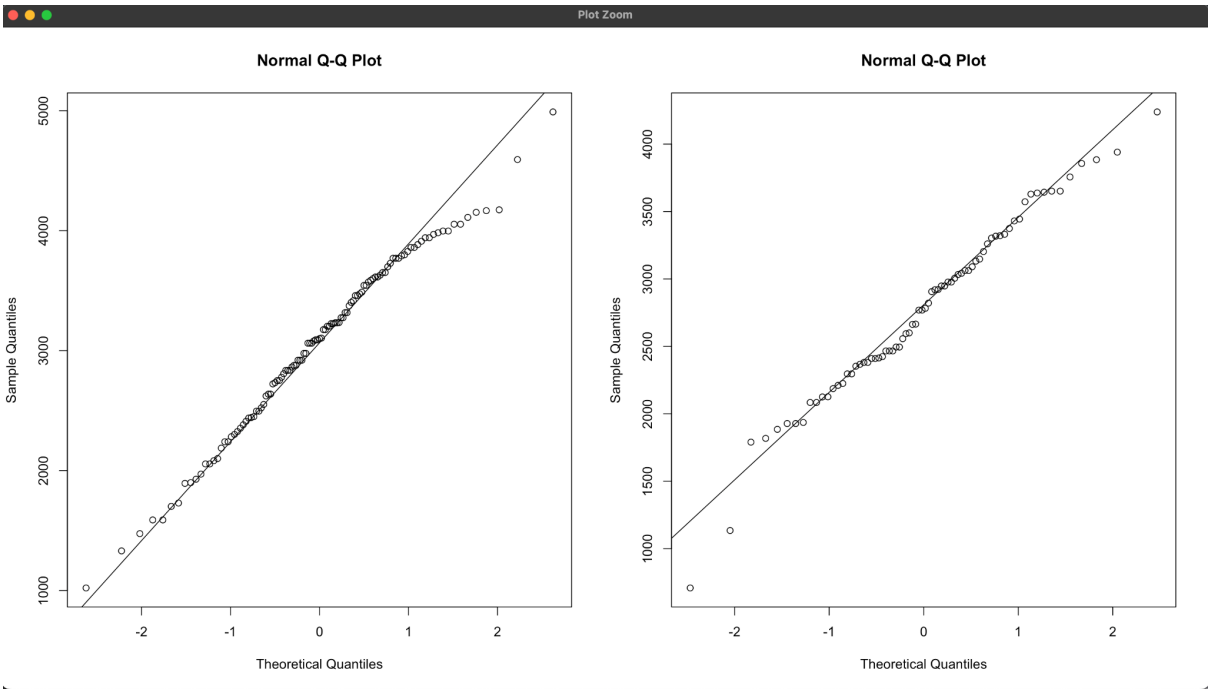
**First we Testing whether the entire sample is normally distributed:**



The sample is normally Distributed.

Now we test the normal distribution of smoking and non smoking:
Check for normality:




They both show normal distribution

Shapiro-Wilk normality test
HO=Data normally distributed
HA=Data not normally distributed

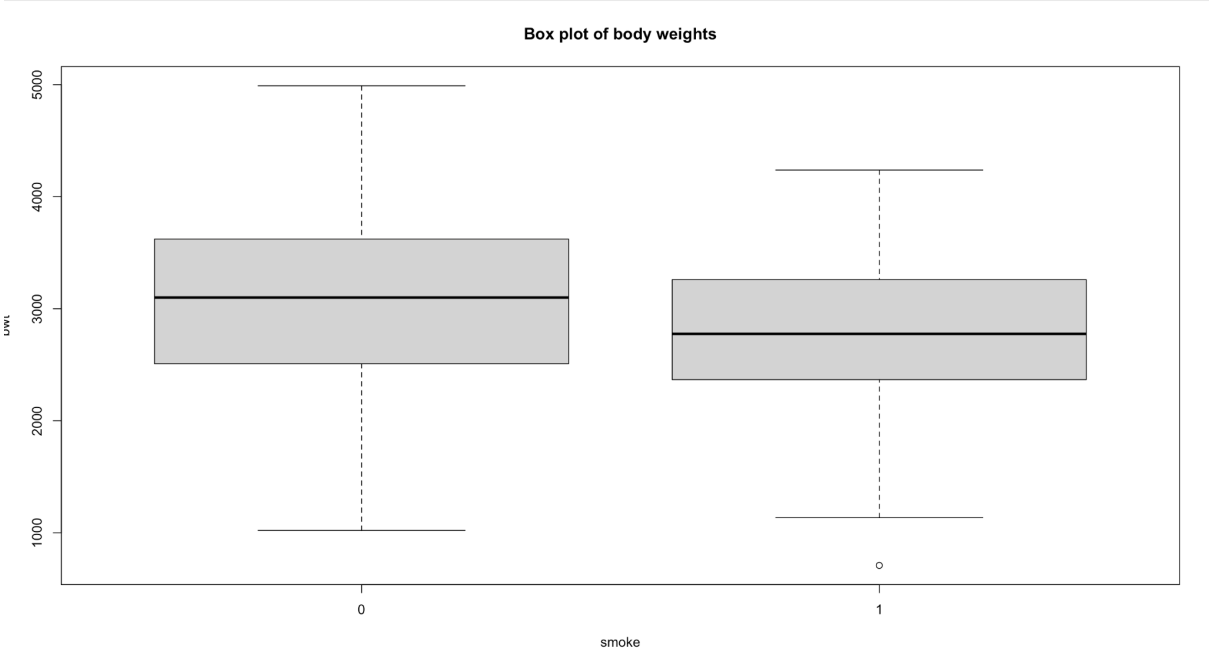
Non smoking: p-value = 0.3337
Smoking: p-value = 0.4195
The P-values are not less than 0.05(significance level) so we reject the alternative hypothesis meaning the data is normally distributed.

***Results:***

Assess equality of variance:

HO: Populations have equal variance

HA: Populations do not have equal variance




The p-value of f-test is 0.2254 which is not less than 0.05 (significance level) so we accept H0 meaning the populations have equal variance

Performing the two sample t-test:

Test value = 2.6529

p-value = 0.008667

| mean in group 0 (non smoking) | mean in group 1 (smoking) |
|---|---|
| 3055.696 | 2771.919 |

95 percent confidence interval:

72.75612 - 494.79735

**Conclusion:**

Based on these results we can reject HO so smoking is associated with low birth rate. The boxplot shows that the non smoking weights are higher and so does the mean. We reject H0 because the pvalue is less than 0.05(significance), meaning the test is statistically relevant and we reject Ho. 0 is not included in the confidence interval between the two means so they are not equal and the difference in weight between the two groups lies between 73 and 495 with 95% confidence showing that in fact smoking does associate with a lower birth rate.

**Q4.**

***Introductory Statement:*** In this analysis we are determining if hypertension is associated with low birth weight. We have taken a sample of the population with hypertension and non-hypertension.

Ho: The birth weight of babies from hypertension is not lower than the birth weight of babies from hypertension.

$\mu 1 >= \mu 2$

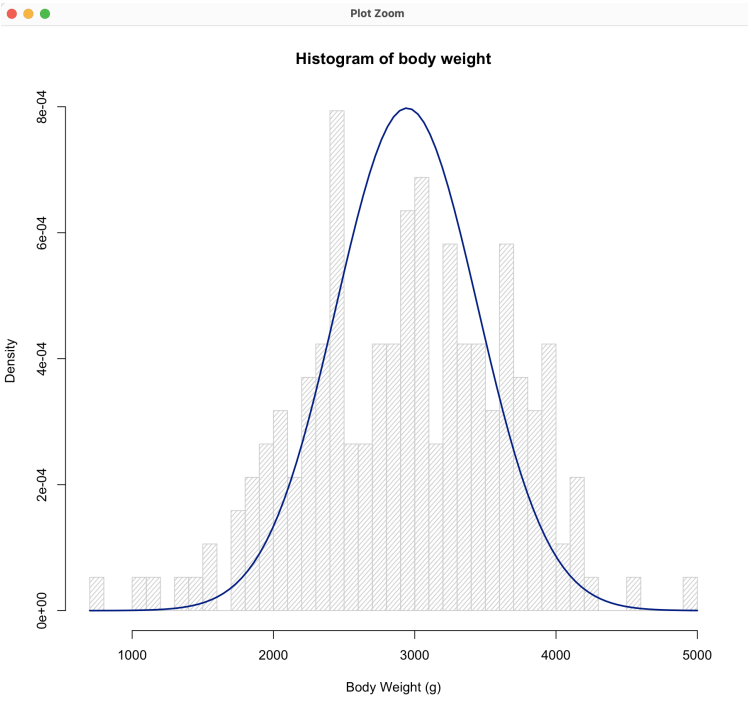
HA: is the mean birth weight of children with history of hypertension  is lower than the birth weight of babies from non hypertension show that it is associated with lower birth weight

$\mu_1 < \mu_2$

***Exploratory Analysis:***

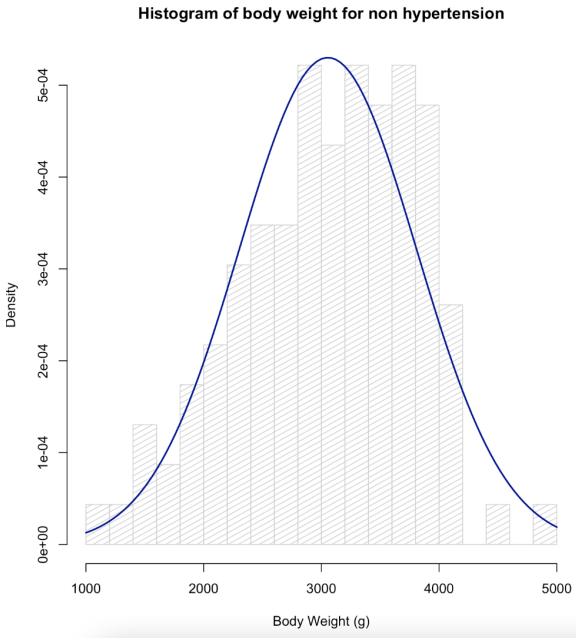
Significance level = 0.05

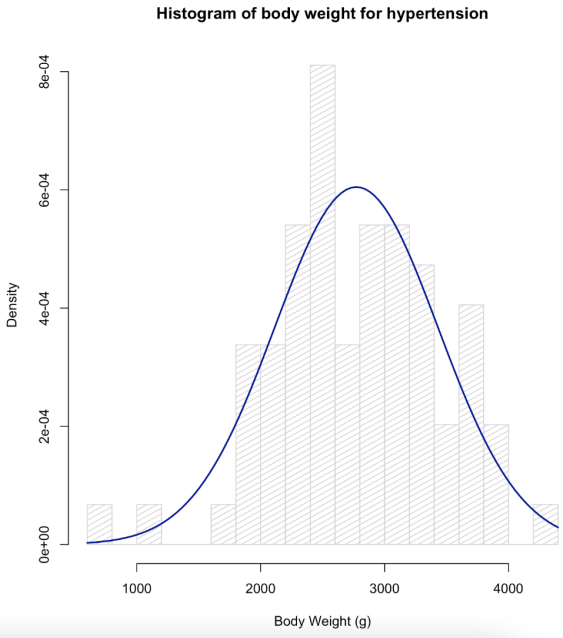
**First we Testing whether the entire sample is normally distributed:**



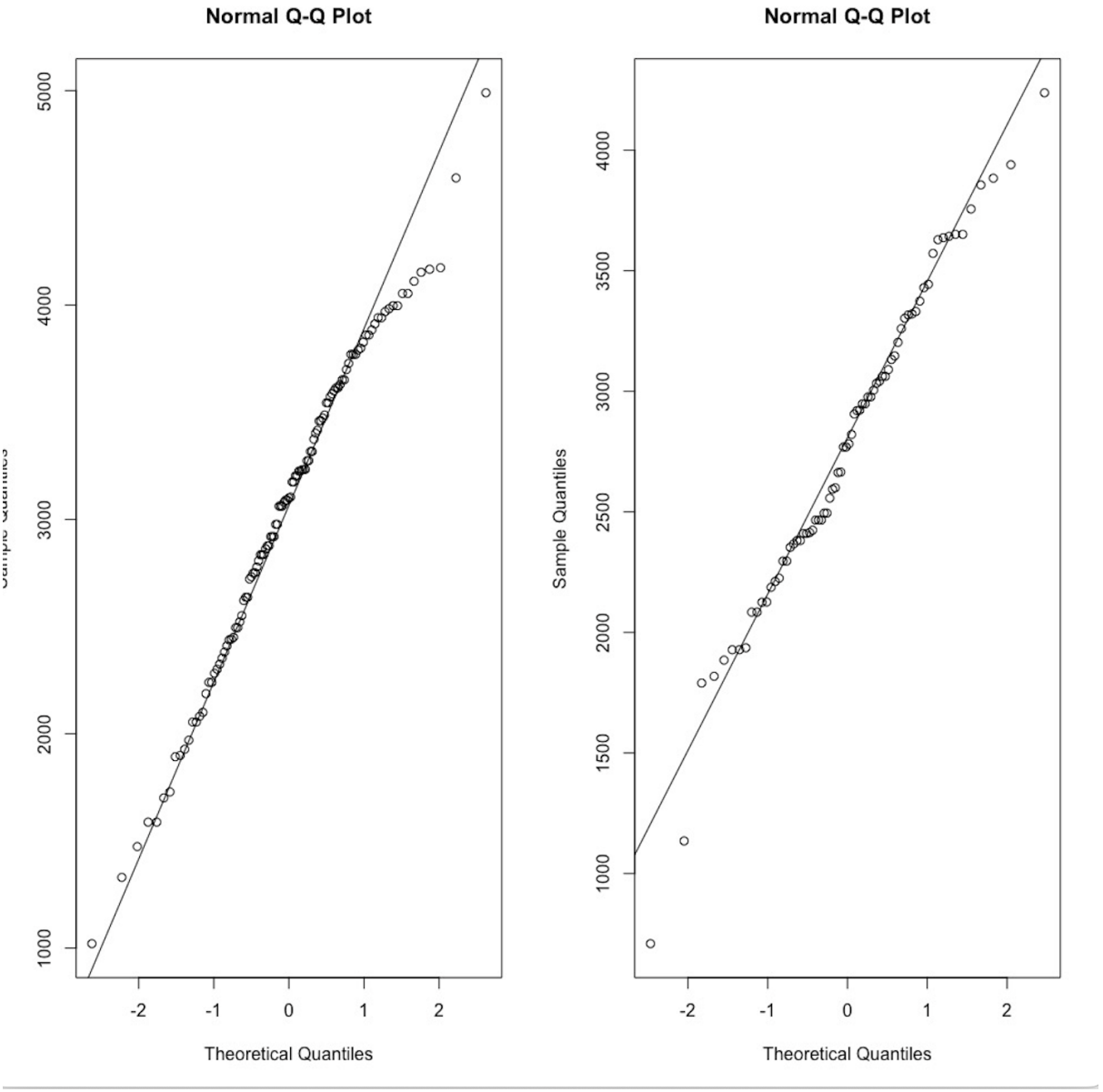
**Now we check for normality for hypertension and non hypertension**

Check no1 Histogram:




**Check no.2 for normality qqplot hypertension and non hypertension**




**Check for normality no.3**
Shapiro-Wilk normality test
HO=Data normally distributed
HA=Data not normally distributed

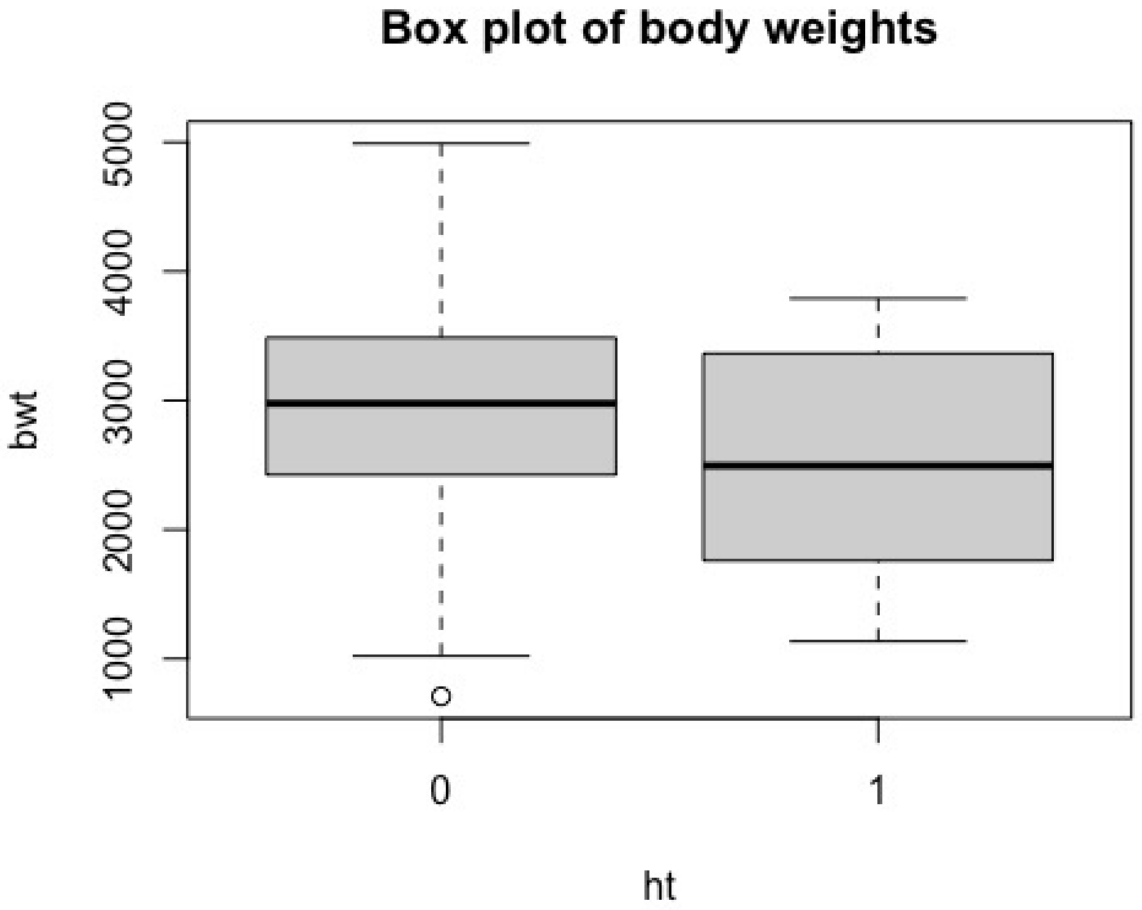
Non hypertension: p-value = 0.5019
Hypertension: p-value = 0.4403

The P-values are not less than 0.05(significance level) so we reject the alternative hypothesis meaning the data is normally distributed.
**Results:**

Assess equality of variance:

HO: Populations have equal variance

HA: Populations do not have equal variance




The p-value of f-test is 0.1662 which is not less than 0.05 (significance level) so we accept H0 meaning the populations have equal variance

Performing the two sample t-test:

Test value = 2.0179

p-value = 0.04503

95 percent confidence interval:
9.743799 - 861.052811

| mean in group 0 (non hypertension) | mean in group 1 (hypertension) |
| --- | --- |
| 2972.232 | 2536.833 |

**Conclusion:**

Based on these results we can reject HO meaning hypertension is associated with low birth rate. The boxplot shows that the non hypertension weights are higher and so does the mean. The p-value is less than 0.05(significance) meaning the results are significant and we reject HO. 0 is not included in the confidence interval between the two means so they are not equal and the difference in weight between the two groups lies between 9.743799 and 861.052811 with 95% confidence showing that yes in fact hypertension does associate with lower birth weights.