

#Load dataset

```
Birthweights <- read.csv("~/Data1204/RData1204/Final Project/Birthweights.csv")
```

```
View(Birthweights)
```

```
library(lattice)
```

```
library(psych)
```

```
library(gmodels)
```

```
Birthweights
```

```
summary(Birthweights$bwt)
```

```
sd(Birthweights$bwt)
```

```
summary(Birthweights$gestation)
```

```
sd(Birthweights$gestation)
```

```
summary(Birthweights$age)
```

```
sd(Birthweights$age)
```

```
summary(Birthweights$height)
```

```
sd(Birthweights$height)
```

```
summary(Birthweights$weight)
```

```
sd(Birthweights$weight)
```

#Histogram bwt

```
library(ggplot2)
```

```
x=Birthweights$bwt
```

```
h<-hist(x, breaks=10, col="red", xlab="birthweight", main="Frequency Distribution of Birthweights")
```

```
xfit<-seq(min(x),max(x),length=50)
yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))
yfit<- yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col="blue", lwd=2)
```

```
str(Birthweights)
```

#Check the t-test is valid

```
library(ggplot2)
```

#Histogram

```
hist(Birthweights$bwt, freq = FALSE, col = 'yellow', main = "Birthweights Data", xlab = 'Scores')
lines(density(Birthweights$bwt), col = "red")
```

#Two Tail test

```
mean(Birthweights$bwt)
sd(Birthweights$bwt)
```

#Calculate z parameters

```
mu <- 3386.788
mu0 <- 3400 # Given mean
alpha <- 0.05 # Specify the significance level
sigma <- 519.6163 # population standard deviation
n <- nrow(Birthweights) # get the sample size
```

#Calculate z

```
z<-(mu-mu0)/(sigma/sqrt(n))
```

```
z
```

#Calculate p-value

```
2*pnorm(abs(z),lower.tail=FALSE)
```

#Linear Regression

#Build Linear Model

```
simple.fit<-lm(bwt~smoke, data=Birthweights)
```

```
LinearModel<-simple.fit
```

```
#Summary of Key Statistics of the Model
```

```
summary(LinearModel)
```

#Load Libraries

```
library(dplyr)
```

```
library(ggplot2)
```

```
ggplot(Birthweights,aes(x = factor(smoke), y = bwt)) + geom_point(colour = "green", size = 5, alpha = 0.5)
```

#Multivariate Regression

```
#Load Data
```

```
data(Birthweights)
```

```
Birthweights
```

#Create Dataset

```
input <- Birthweights[,c("bwt","gestation","region","age","height","weight", "smoke")]
```

```
print(head(input))
```

Create the relationship model.

```
model <- lm(bwt~gestation+region+age+height+weight+smoke, data = input)
```

Show the model.

```
print(model)
```

#Model Summary

```
summary(model)
```

#Predict

Create the relationship model with significant predictors

```
model1 <- lm(bwt~gestation+height+weight+smoke, data = input)
```

Show the model

```
print(model1)
```

```
newdata = data.frame(gestation=200, height = 200, weight = 100.0, smoke = 1)
```

```
predict(model1, newdata)
```