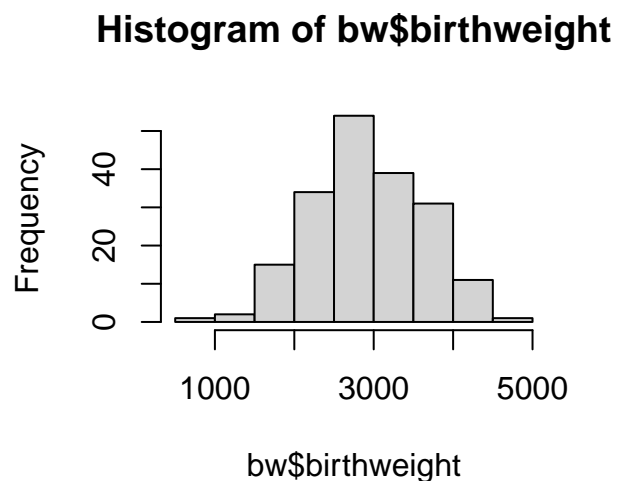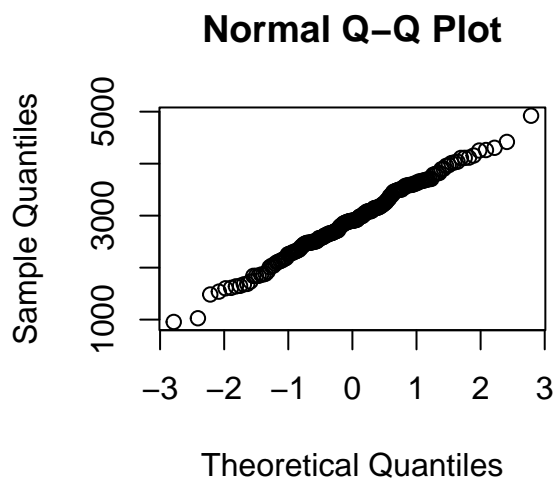# EDDA Assignment 1

Igor Mazurek, Nagul Ratna Pandian, Vidhya Narayanasamy

Group 66

Exercise 1:

a)Normality Check of the data:

```
qqnorm(bw$birthweight);hist(bw$birthweight);
```


The QQ plot and histogram leans towards the direction of normality.

```
shapiro.test(bw$birthweight)
```

```
##
##   Shapiro-Wilk normality test
##
## data:  bw$birthweight
## W = 0.99595, p-value = 0.8995
```

As a final check Shapiro-Wilk test was used. The p-value(0.8995) obtained was greater than 0.05 which implies that there is no strong evidence against normality. Hence we conclude the data is normally distributed.

Construct a bounded 96%-CI for mu:

```
n = 188; mu = mean(bw$birthweight); sig = sd(bw$birthweight)
za = qnorm(0.98); err = za*(sig/sqrt(n))
CI = c(lo_lim=mu-za*(sig/sqrt(n)), Up_lim=mu+za*(sig/sqrt(n))); CI
```

```
##    lo_lim   Up_lim
## 2808.817 3017.768
```

The 96% confidence Interval for mu is calculated as [2808.817, 3017.768]

Evaluate the sample size

```
n_new = (za^2*sig^2)/50^2 ; n_new
```

```
## [1] 820.8114
```

The sample size needed at length 100 for 96% CI is 821(rounded to the closest integer)

Compute a bootstrap 96%-CI for mean and compare it to the above CI

```
B=1000
Tstar=numeric(B)
for(i in 1:B){
Xstar=sample(bw$birthweight, replace=T)
Tstar[i]=mean(Xstar)
}
Tstar2=quantile(Tstar,0.02); Tstar98=quantile(Tstar,0.98)
sum(Tstar<Tstar2)
```

```
## [1] 20
```

```
c(2*mu-Tstar98,2*mu-Tstar2)
```

```
##      98%       2%
## 2806.712 3017.640
```

The bootstrap 96% confidence interval for the population mean is [2805.296, 3011.012] around its
mean mu= 2913.293. The bootstrap CI is less than the previous CI.

b)

```
t.test(bw$birthweight, alternative = "greater", mu=2800)
```

```
##
##  One Sample t-test
##
```

```
## data:  bw$birthweight
## t = 2.2271, df = 187, p-value = 0.01357
## alternative hypothesis: true mean is greater than 2800
## 95 percent confidence interval:
##  2829.202      Inf
## sample estimates:
## mean of x
##  2913.293
```

The null hypothesis is that mean is less than 2800. We reject the null hypothesis(mu is less than or equal to mu0) as p<0.05. The one sample t-test we used here is right sided. So we are dealing with one sided hypothesis, The confidence interval is connected to the problem of the hypothesis testing and it has to be one sided as well. Hence we get inf as the upper bound of the CI.

```
binom.test(sum(bw$birthweight > 2800), n, p=0.5, alternative ="greater")
```

```
##
##   Exact binomial test
##
## data:  sum(bw$birthweight > 2800) and n
## number of successes = 107, number of trials = 188, p-value = 0.03399
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.5065781 1.0000000
## sample estimates:
## probability of success
##               0.5691489
```

```
wilcox.test(bw$birthweight, mu=2800, conf.int=T, conf.level = 0.95)
```

```
##
##   Wilcoxon signed rank test with continuity correction
##
## data:  bw$birthweight
## V = 10498, p-value = 0.03064
## alternative hypothesis: true location is not equal to 2800
## 95 percent confidence interval:
##  2811.5 3022.5
## sample estimates:
## (pseudo)median
##           2914
```

We used the sign test and wilcoxon sign test as they can be used for one sample normally distributed data. Similar to the t-test the null hypothesis is rejected in the sign tests as well, as the p values for both the sign tests are less than 0.05

C) Powers of the t-test and sign test at mu > 2800:

```
B = 1000
ttest <- numeric(B); sign <- numeric(B)
for(j in 1:B) {
  sample_values = rnorm(n, 2800, sig)
  ttest[j] <- t.test(sample_values, mu=2800, alternative='greater')[[3]]
  sign[j] <- binom.test(sum(sample_values>2800), n, p=0.5, alt='g')[[3]]
}
p_value = 0.05
print("Power of the t-test"); print(sum(ttest < p_value)/B)
```

```
## [1] "Power of the t-test"
```

```
## [1] 0.05
```

```
print("Power of the sign test"); print(sum(sign < p_value)/B)
```

```
## [1] "Power of the sign test"
```

```
## [1] 0.048
```

Power is the probability of rejecting H0 correctly and it depends on the amount of data and the probability(type2error). In this case the t-test rejects H0 more effectively than the sign test.

d)

```
t1 = sum(bw$birthweight< 2600)
p_est = mean(bw$birthweight < 2600)
phat_l = 0.25
me = p_est - phat_l ;
phat_r = me + p_est;
conf_int = c(phat_l=0.25, phat_r=me + p_est);conf_int
```

```
##     phat_l     phat_r
## 0.2500000 0.4095745
```

```
z_clt = me/(sqrt(p_est*(1-p_est)/n));
conf_level = pnorm(z_clt); conf_level
```

```
## [1] 0.9900163
```

The right end $\hat{p}_r$ of the confidence interval is 0.4096. The whole confidence interval is $[0.250, 0.4096]$ and the confidence level is 0.990

e)

```
prop.test(c(34,61), c(62,126))
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(34, 61) out of c(62, 126)
## X-squared = 0.45343, df = 1, p-value = 0.5007
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.09929476  0.22781499
## sample estimates:
##    prop 1    prop 2
## 0.5483871 0.4841270
```

As the p value is greater than 0.05 the null hypothesis is not rejected which implies that the expert's claim (mean weight is different for male and female babies) is not true.