Problem Set 3

Michael Ghattas

Notes

Other students who I worked with on this assignment (if any): None

Introduction

These questions were rendered in R markdown through RStudio (https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf, http://rmarkdown.rstudio.com). Please generate your solutions in R markdown and upload a knitted pdf document to Gradescope. Please put your name in the "author" section in the header. The questions in this problem set use material from the slides on parameter estimation.

Context for Questions 1-3

We have seen that count k of successes from n Bernoulli trials is modeled as $Binomial(\text{size} = n, \text{probability} = p)$ then the maximum likelihood estimate of p equals $\frac{k}{n}$. Suppose this is repeated M times and k_1 successes are observed in n_1 Bernoulli trials, k_2 successes are observed in n_2 Bernoulli trials, and so on through k_M successes in n_M Bernoulli trials. The goal to find the maximum likelihood estimate of p if these are modeled as samples from $Binomial(\text{size} = n_1, \text{probability} = p)$, $Binomial(\text{size} = n_2, \text{probability} = p)$, and so on through $Binomial(\text{size} = n_M, \text{probability} = p)$.

Question 1

Consider the likelihood $L(k_1, k_2, ...k_n)$ function for $\{k_1, k_2, ...k_n\}$ as outcomes from M independent binomial distributions $Binomial(\text{size} = n_1, \text{probability} = p), Binomial(\text{size} = n_2, \text{probability} = p), ...Binomial(\text{size} = n_M, \text{probability} = p)$.

Question 1.1

Please give the likelihood function $L(k_1, k_2, ...k_n)$.

Your answer here:

$$L(k_1, k_2, \ldots, k_n) = \prod_{i=1}^{M} \binom{n_i}{k_i} p^{k_i} (1-p)^{n_i - k_i}$$

Question 1.2

Please give the log of the likelihood function as a sum of terms of the form $\log\left[\binom{n_i}{k_i}p^{k_i}(1-p)^{n_i-k_i}\right]$

Your answer here:

$$\log L(k_1, k_2, \ldots, k_n) = \sum_{i=1}^{M}\left[\log\binom{n_i}{k_i} + k_i\log(p) + (n_i - k_i)\log(1-p)\right]$$

Question 2

Question 2.1

Please give the derivative with respect to p of $\sum_{i=1}^{M}\left[\log\binom{n_i}{k_i} + k_i\log(p) + (n_i - k_i)\log(1-p)\right]$.

Your answer here:

$$\frac{d}{dp}\sum_{i=1}^{M}\left[\log\binom{n_i}{k_i} + k_i\log(p) + (n_i - k_i)\log(1-p)\right] = \sum_{i=1}^{M}\left[\frac{k_i}{p} - \frac{n_i - k_i}{1-p}\right]$$

Question 2.2

Please give the value of p that maximizes $\sum_{i=1}^{M}\left[\log\binom{n_i}{k_i} + k_i\log(p) + (n_i - k_i)\log(1-p)\right]$.

Your answer here:

$$\sum_{i=1}^{M}\left[\frac{k_i}{p} - \frac{n_i - k_i}{1-p}\right] = 0$$

Solving for p:

$$p = \frac{\sum_{i=1}^{M} k_i}{\sum_{i=1}^{M} n_i}$$

Question 3

If the M samples $\{k_1, k_2, ...k_n\}$ from M independent binomial distributions $Binomial(\text{size} = n_1, \text{probability} = p), Binomial(\text{size} = n_2, \text{probability} = p), ...Binomial(\text{size} = n_M, \text{probability} = p)$ are viewed as $\sum_{i=1}^{M} k_i$ successes in $\sum_{i=1}^{M} n_i$ independent Bernoulli trials with probability of success equal to p, what is the maximum likelihood estimate of p.

Your answer here:

$$\hat{p} = \frac{\sum_{i=1}^{M} k_i}{\sum_{i=1}^{M} n_i}$$

Question 4

Context

The code below generates a sample, `samp1`, of size 10,000 from the $Binomial(\text{size} = 20, \text{probability} = 0.5)$ distribution and a sample, `samp2` of size 10,000 from the $Binomial(\text{size} = 50, \text{probability} = 0.3)$ distribution.

```r
set.seed(12345)
samp1<-rbinom(10000,20,.5)
dat1<-data.frame(x=samp1)
samp2<-rbinom(10000,50,.25)
dat2<-data.frame(x=samp2)
```

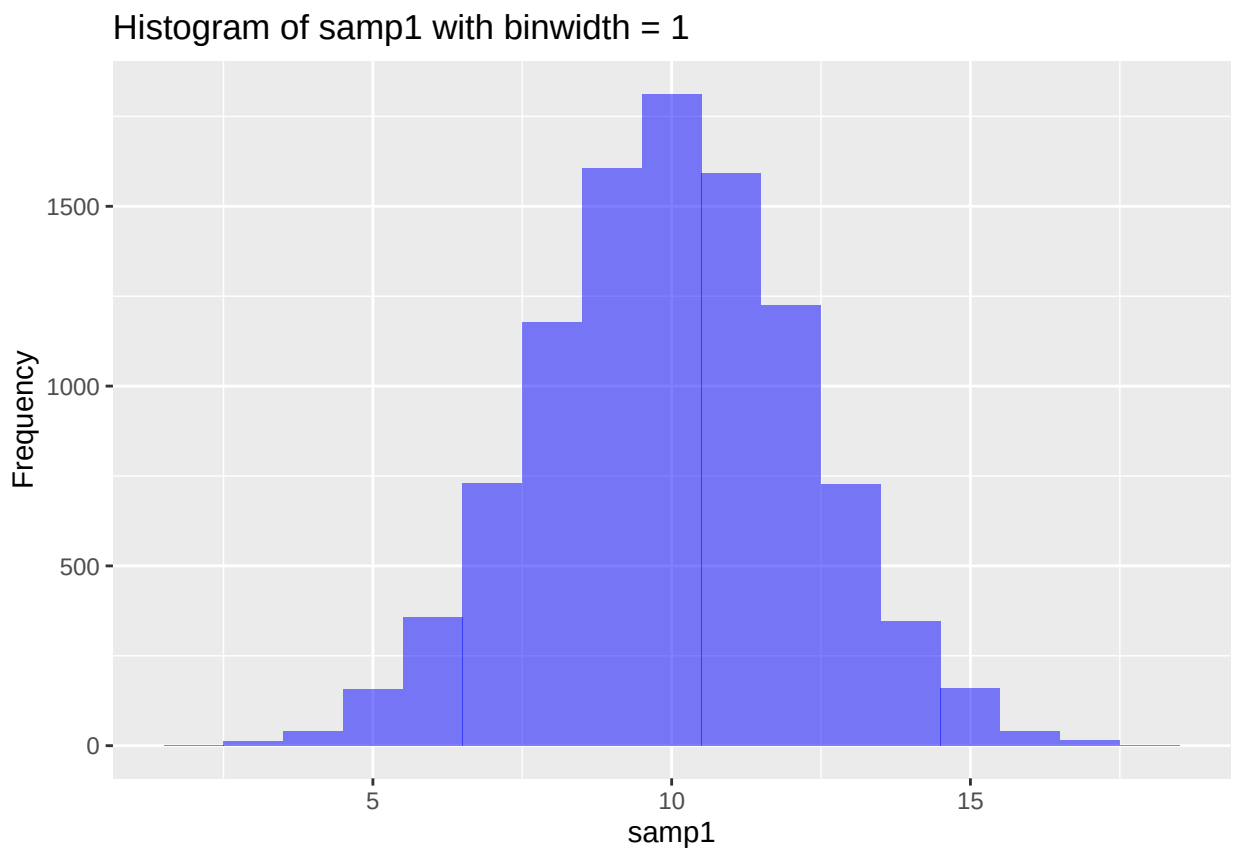
Question 4.1

Please display separate histograms of `samp1` and `samp2` with binwidth equal to 1.

Your code and plots here:

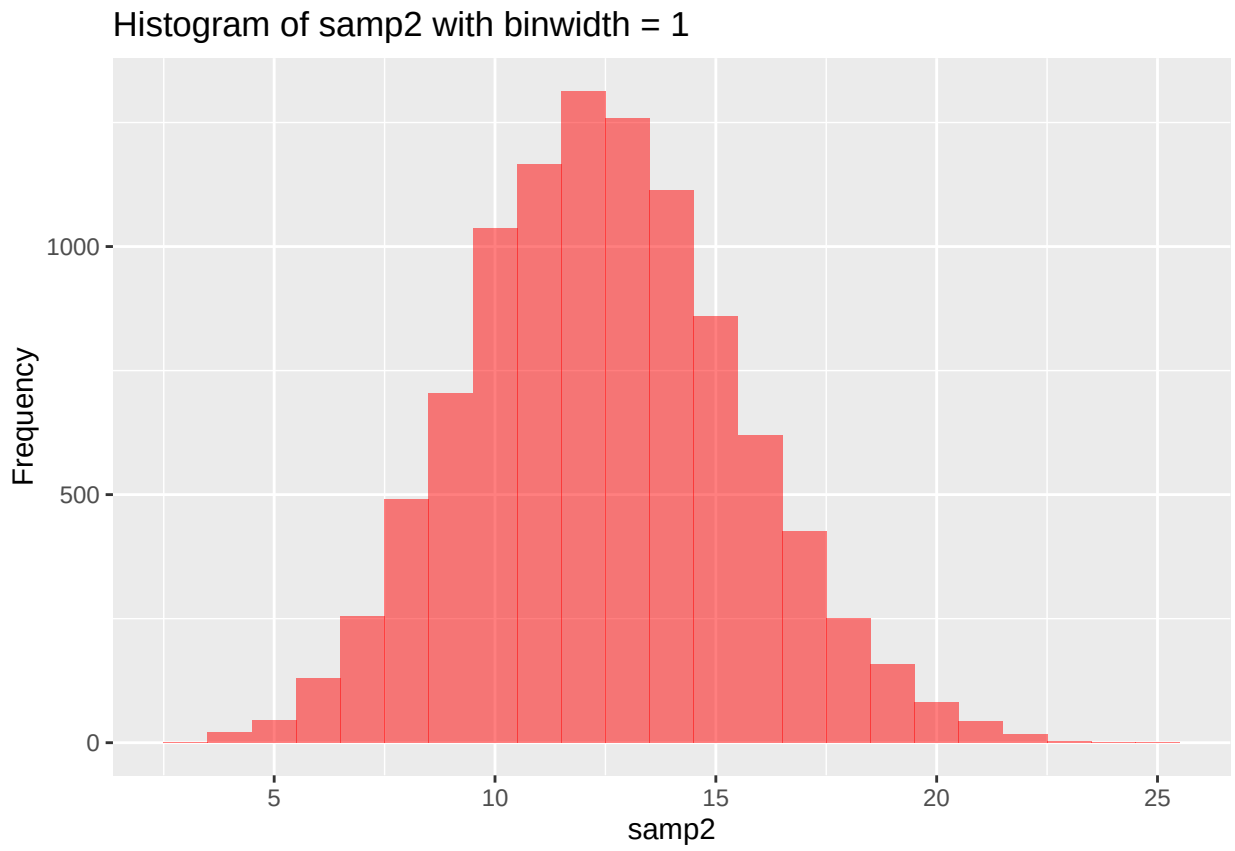
```r
# Plot histograms
library(ggplot2)

# Histogram for samp1
ggplot(dat1, aes(x = x)) +
  geom_histogram(binwidth = 1, fill = "blue", alpha = 0.5) +
  ggtitle("Histogram of samp1 with binwidth = 1") +
  xlab("samp1") +
  ylab("Frequency")
```




```r
# Histogram for samp2
ggplot(dat2, aes(x = x)) +
```

```
  geom_histogram(binwidth = 1, fill = "red", alpha = 0.5) +
  ggtitle("Histogram of samp2 with binwidth = 1") +
  xlab("samp2") +
  ylab("Frequency")
```



Question 4.2

Treating `samp1` and `samp2` as samples from Normal distributions $Normal(\mu_1, \sigma_1^2)$ and $Normal(\mu_2, \sigma_2^2)$, please give maximum likelihood estimates of μ_1, σ_1^2, μ_2, and σ_2^2.

Your answer here:

```
# Maximum likelihood estimates for samp1
mu1 <- mean(samp1)
sigma1_squared <- var(samp1)

# Maximum likelihood estimates for samp2
mu2 <- mean(samp2)
sigma2_squared <- var(samp2)

# Output the results
mu1
```

```
## [1] 10.0043
```

```
sigma1_squared
```

```
## [1] 4.940776
```

```
mu2
```

```
## [1] 12.4697
```

```
sigma2_squared
```

```
## [1] 9.415423
```

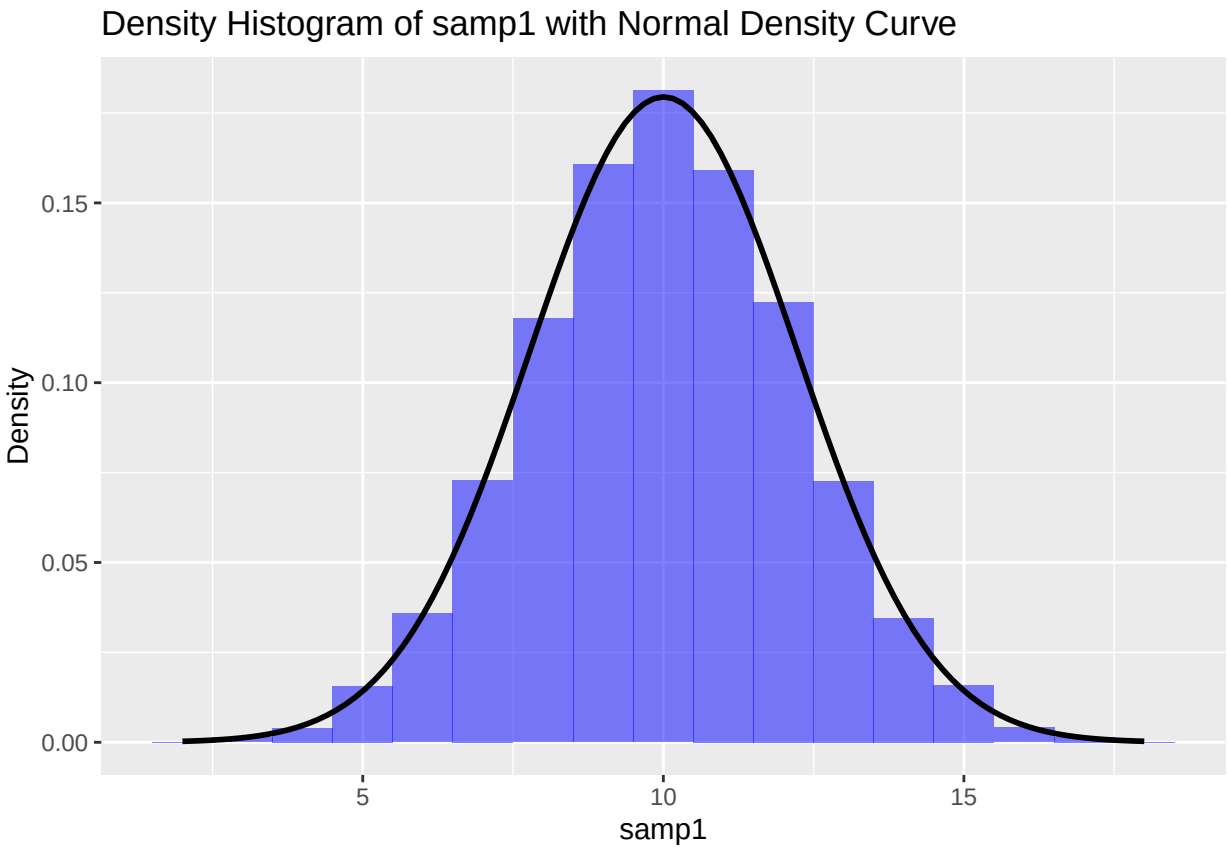
Question 4.3

The plotting methods from `continuous_probability_distributions_2_4_2.Rmd` and practice problem set 2 may be useful here.

For `samp1` please display the density histogram with density curve for $Normal(\mu_1, \sigma_1^2)$ superimposed.

Your code and plots here:

```
# Plot density histogram with density curve for samp1
ggplot(dat1, aes(x = x)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 1, fill = "blue", alpha = 0.5) +
  stat_function(fun = dnorm, args = list(mean = mu1, sd = sqrt(sigma1_squared)), color = "black", linew
  ggtitle("Density Histogram of samp1 with Normal Density Curve") +
  xlab("samp1") +
  ylab("Density")
```

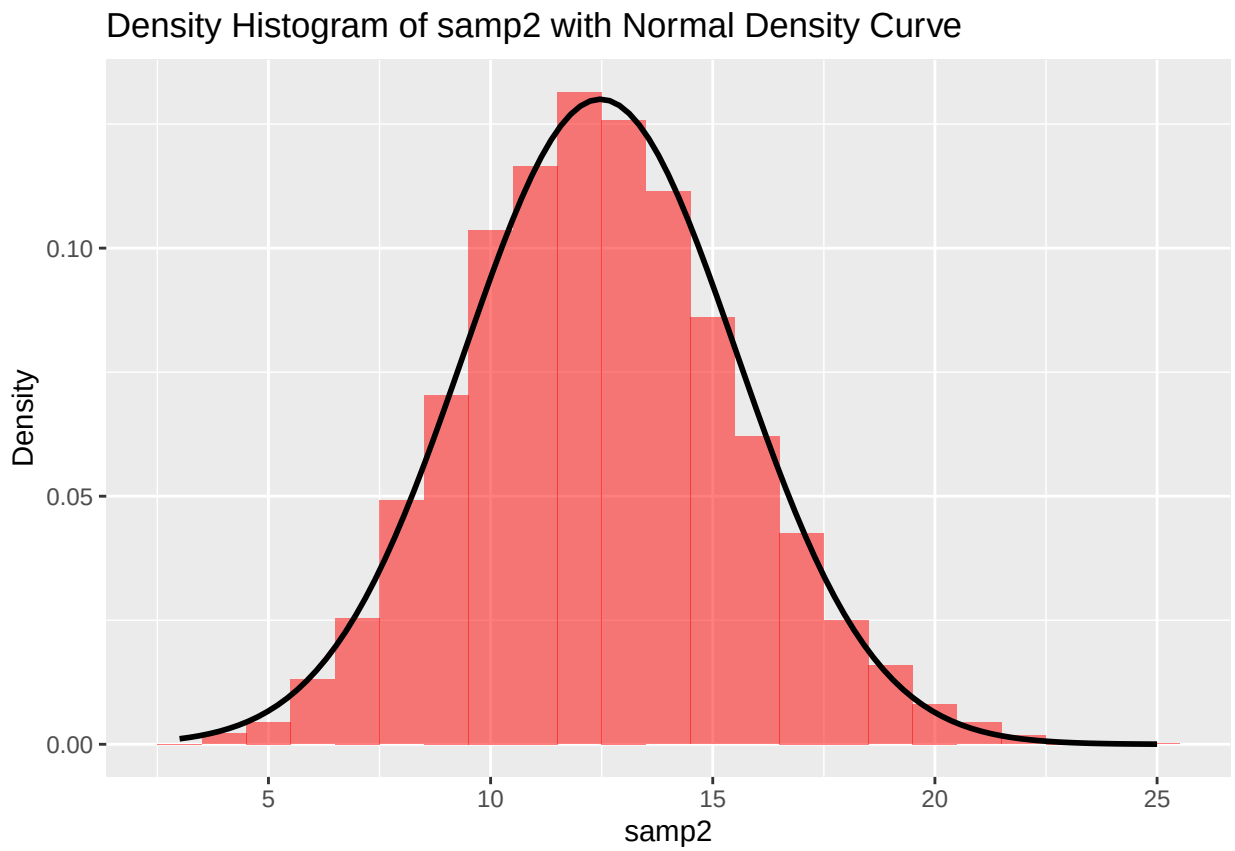


Question 4.4

For `samp2`: please display the density histogram with density curve for $Normal(\mu_2, \sigma_2^2)$ superimposed.

Your code and plots here:

```
# Plot density histogram with density curve for samp2
ggplot(dat2, aes(x = x)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 1, fill = "red", alpha = 0.5) +
  stat_function(fun = dnorm, args = list(mean = mu2, sd = sqrt(sigma2_squared)), color = "black", linew
  ggtitle("Density Histogram of samp2 with Normal Density Curve") +
  xlab("samp2") +
  ylab("Density")
```



Question 6

Let us consider $y_1, \ldots, y_n$ as samples from Normal distributions $Normal(\mu_1 = mx_1 + b, \sigma^2)$, ... , $Normal(\mu_n = mx_n + b, \sigma^2)$ where the input value x_i is a given constant, and the observations come in pairs (y_i, x_i). Based on the graph produced by the code chunk below answer these questions:

Question 6.1

Do you agree with the statement: "the line represents the fact that the peak of the normal distribution (the peak represents the average value) changes with each value of the input x_i and is given by the line $y_i = m \times x_i + b$." Why or why not? If not please write a revised statement.

Your answer here: Yes, I agree with the statement. The line $y_i = m \times x_i + b$ represents the expected value (or mean) of the normal distribution for each given input x_i. The peak of the normal distribution for each x_i corresponds to this mean value, indicating that the average value of the observations y_i changes linearly with x_i. Thus, the line accurately reflects how the mean of the normal distribution varies with x_i.

Question 6.2

Do you think that every observation would lie exactly on the line? Why or why not? Would it be more likely that an observation would lie close to the line or far away from the line for a given (fixed) value of the input x_i?

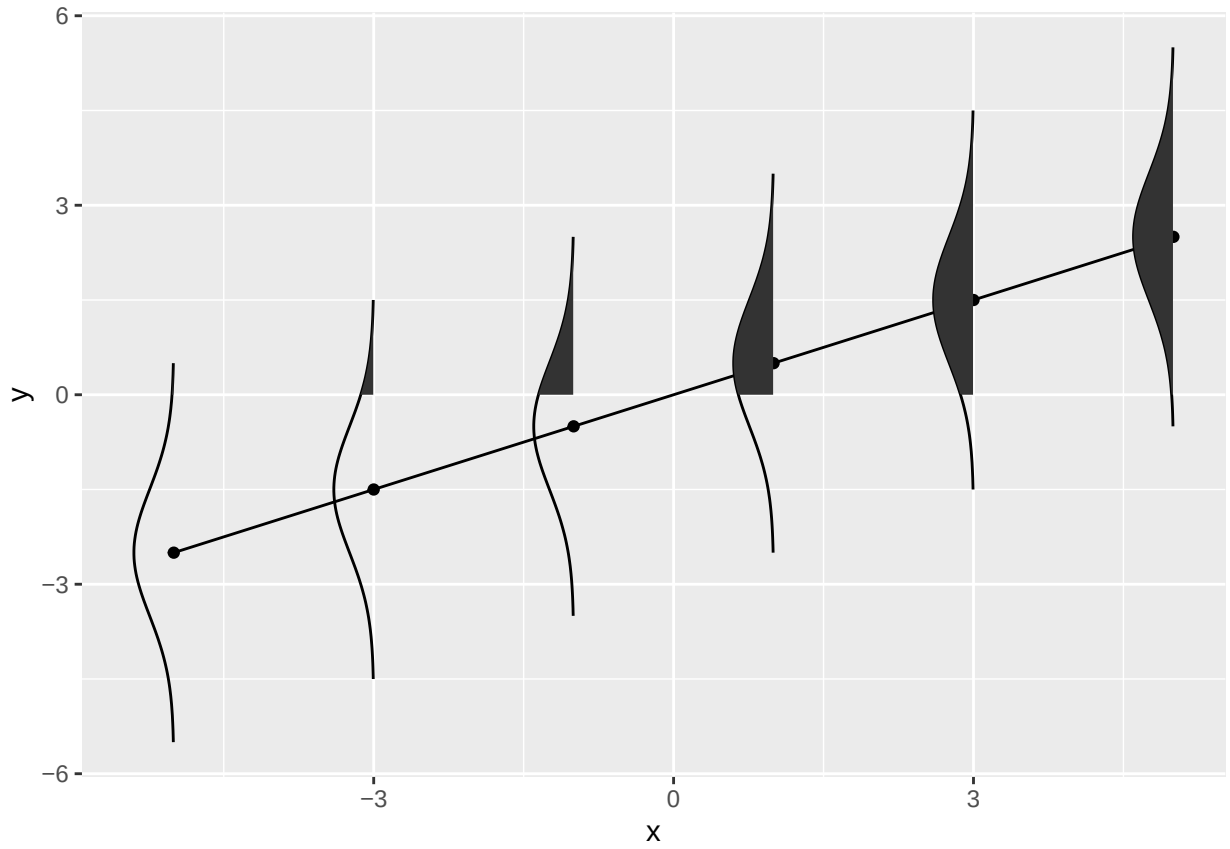
```r
x <- seq(1, 11, 2)
y <- x*0.5

x <- x - mean(x)
y <- y - mean(y)

df <- data.frame(x, y)

# For every row in `df`, compute a rotated normal density centered at `y` and shifted by `x`
curves <- lapply(seq_len(NROW(df)), function(i) {
  mu <- df$y[i]
  range <- mu + c(-3, 3)
  seq <- seq(range[1], range[2], length.out = 100)
  data.frame(
    x = -1 * dnorm(seq, mean = mu) + df$x[i],
    y = seq,
    grp = i
  )
})
# Combine above densities in one data.frame
curves <- do.call(rbind, curves)

ggplot(df, aes(x, y)) +
  geom_point() +
  geom_line() +
  # The path draws the curve
  geom_path(data = curves, aes(group = grp)) +
  # The polygon does the shading. We can use `oob_squish()` to set a range.
  geom_polygon(data = curves, aes(y = scales::oob_squish(y, c(0, Inf)),group = grp))
```



Your answer here: No, not every observation would lie exactly on the line. This is because the observations y_i are samples from normal distributions with means $y_i = m \times x_i + b$ and a common standard deviation σ. The normal distribution introduces variability around the mean, meaning that while most observations will be close to the line, due to the nature of the normal distribution, they will not be exactly on it. It is more likely for an observation to lie close to the line rather than far away, as the probability density is highest near the mean and decreases as you move further away from it. The spread of the observations around the line is determined by the standard deviation σ; smaller σ means observations are closer to the line, while larger σ means observations are more spread out.

Question 7

Let us consider $y_1, \ldots, y_n$ as samples from Normal distributions $Normal(\mu_1 = mx_1 + b, \sigma^2)$, $\ldots$, $Normal(\mu_n = mx_n + b, \sigma^2)$. Provide the work to show that the likelihood function $L(y_1, ..., y_n)$ is given by

$$(2\pi\sigma^2)^{-n/2} e^{\sum_{i=1}^{n} -\frac{(y_i-\mu_i)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-n/2} e^{\sum_{i=1}^{n} -\frac{(y_i-(mx_i+b))^2}{2\sigma^2}}$$

Your answer here: To show that the likelihood function $L(y_1, \ldots, y_n)$ for samples from Normal distributions $textNormal(\mu_i = mx_i + b, \sigma^2)$ is given by

$$L(y_1, \ldots, y_n) = (2\pi\sigma^2)^{-n/2} e^{\sum_{i=1}^{n} -\frac{(y_i-\mu_i)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-n/2} e^{\sum_{i=1}^{n} -\frac{(y_i-(mx_i+b))^2}{2\sigma^2}}$$

The probability density function for a normal distribution $y_i \sim \text{Normal}(\mu_i, \sigma^2)$ is given by:

$$f(y_i|\mu_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu_i)^2}{2\sigma^2}}$$

Given $n$independent samples $y_1, y_2, \ldots, y_n$from normal distributions with means $\mu_i = mx_i + b$ and common variance σ^2, the joint likelihood function $L(y_1, \ldots, y_n | m, b, \sigma^2)$ is the product of the individual probability density functions:

$$L(y_1, \ldots, y_n | m, b, \sigma^2) = \prod_{i=1}^{n} f(y_i | \mu_i, \sigma^2)$$

Substituting the probability density function:

$$L(y_1, \ldots, y_n | m, b, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \mu_i)^2}{2\sigma^2}}$$

The product of the exponentials can be combined into a single exponential term, and the product of the constants can be simplified as follows:

$$L(y_1, \ldots, y_n | m, b, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^{n} e^{-\frac{(y_i - \mu_i)^2}{2\sigma^2}}$$

$$L(y_1, \ldots, y_n | m, b, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{\sum_{i=1}^{n} -\frac{(y_i - \mu_i)^2}{2\sigma^2}}$$

Since $\mu_i = mx_i + b$, we can substitute this into the expression:

$$L(y_1, \ldots, y_n | m, b, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{\sum_{i=1}^{n} -\frac{(y_i - (mx_i + b))^2}{2\sigma^2}}$$

Simplifying the constant term:

$$\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n = (2\pi\sigma^2)^{-n/2}$$

Therfore, the likelihood function is:

$$L(y_1, \ldots, y_n | m, b, \sigma^2) = (2\pi\sigma^2)^{-n/2} e^{\sum_{i=1}^{n} -\frac{(y_i - (mx_i + b))^2}{2\sigma^2}}$$

Question 8

In the previous problem the mean is always changing for each observation y_i let us assume that the mean value $\mu_i = mx_i + b$ where x_i is an input variable. Substitute the formula $\mu_i = mx_i + b$ into the likelihood equation in 3.1.2 and calculate the partial derivatives $\frac{dL}{dm}$ and $\frac{dL}{db}$. Solve for m and b respectively.

Your answer here: To show that the likelihood function $L(y_1, \ldots, y_n)$ for samples from Normal distributions Normal$(\mu_i = mx_i + b, \sigma^2)$ is given by

$$L(y_1, \ldots, y_n) = (2\pi\sigma^2)^{-n/2} e^{\sum_{i=1}^{n} -\frac{(y_i - (mx_i + b))^2}{2\sigma^2}}$$

we need to start with the probability density function of the normal distribution and then extend it to the likelihood function for multiple independent normal samples. The probability density function for a normal distribution $y_i \sim$ Normal(μ_i, σ^2) is given by:

$$f(y_i | \mu_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \mu_i)^2}{2\sigma^2}}$$

Given n independent samples $y_1, y_2, \ldots, y_n$ from normal distributions with means $\mu_i = mx_i + b$ and common variance σ^2, the joint likelihood function $L(y_1, \ldots, y_n | m, b, \sigma^2)$ is the product of the individual probability density functions:

$$L(y_1, \ldots, y_n | m, b, \sigma^2) = \prod_{i=1}^{n} f(y_i | \mu_i, \sigma^2)$$

Substituting the probability density function:

$$L(y_1, \ldots, y_n | m, b, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \mu_i)^2}{2\sigma^2}}$$

The product of the exponentials can be combined into a single exponential term, and the product of the constants can be simplified as follows:

$$L(y_1, \ldots, y_n | m, b, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^{n} e^{-\frac{(y_i - \mu_i)^2}{2\sigma^2}}$$

$$L(y_1, \ldots, y_n | m, b, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{\sum_{i=1}^{n} -\frac{(y_i - \mu_i)^2}{2\sigma^2}}$$

Since $\mu_i = mx_i + b$, we can substitute this into the expression:

$$L(y_1, \ldots, y_n | m, b, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{\sum_{i=1}^{n} -\frac{(y_i - (mx_i + b))^2}{2\sigma^2}}$$

Simplifying the constant term:

$$\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n = (2\pi\sigma^2)^{-n/2}$$

Hence, the likelihood function is:

$$L(y_1, \ldots, y_n | m, b, \sigma^2) = (2\pi\sigma^2)^{-n/2} e^{\sum_{i=1}^{n} -\frac{(y_i - (mx_i + b))^2}{2\sigma^2}}$$

To find the maximum likelihood estimates for m and b, we need to take the partial derivatives of the log-likelihood function with respect to m and b and set them to zero.

$$\ell = \log(L) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - (mx_i + b))^2$$

$$\frac{\partial \ell}{\partial m} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{\partial}{\partial m}(y_i - (mx_i + b))^2$$

Using the chain rule:

$$\frac{\partial}{\partial m}(y_i - (mx_i + b))^2 = 2(y_i - (mx_i + b))(-x_i)$$

Substituting back:

$$\frac{\partial \ell}{\partial m} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(y_i - (mx_i + b))(-x_i) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - (mx_i + b))x_i$$

$$\frac{\partial \ell}{\partial b} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{\partial}{\partial b}(y_i - (mx_i + b))^2$$

Using the chain rule:

$$\frac{\partial}{\partial b}(y_i - (mx_i + b))^2 = 2(y_i - (mx_i + b))(-1)$$

Substituting back:

$$\frac{\partial \ell}{\partial b} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(y_i - (mx_i + b))(-1) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - (mx_i + b))$$

Set the partial derivatives to zero to find the maximum likelihood estimates for m and b:

$$\frac{\partial \ell}{\partial m} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - (mx_i + b))x_i = 0$$

$$\frac{\partial \ell}{\partial b} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - (mx_i + b)) = 0$$

Solving for m and b:

$$\sum_{i=1}^{n}(y_i - (mx_i + b))x_i = 0$$

$$\sum_{i=1}^{n}(y_i - (mx_i + b)) = 0$$

$$\sum_{i=1}^{n} y_i - m\sum_{i=1}^{n} x_i - nb = 0$$

Solving these two linear equations:

$$b = \frac{\sum_{i=1}^{n} y_i - m\sum_{i=1}^{n} x_i}{n}$$

$$m = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i - nb}$$

Question 9

From the previous problem we know the maximum likelihood estimate for the slope and intercept are given by:

$$\frac{\frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum x_i^2 - \bar{x}^2}$$

and the intercept is

$$\bar{y} - m\bar{x}$$

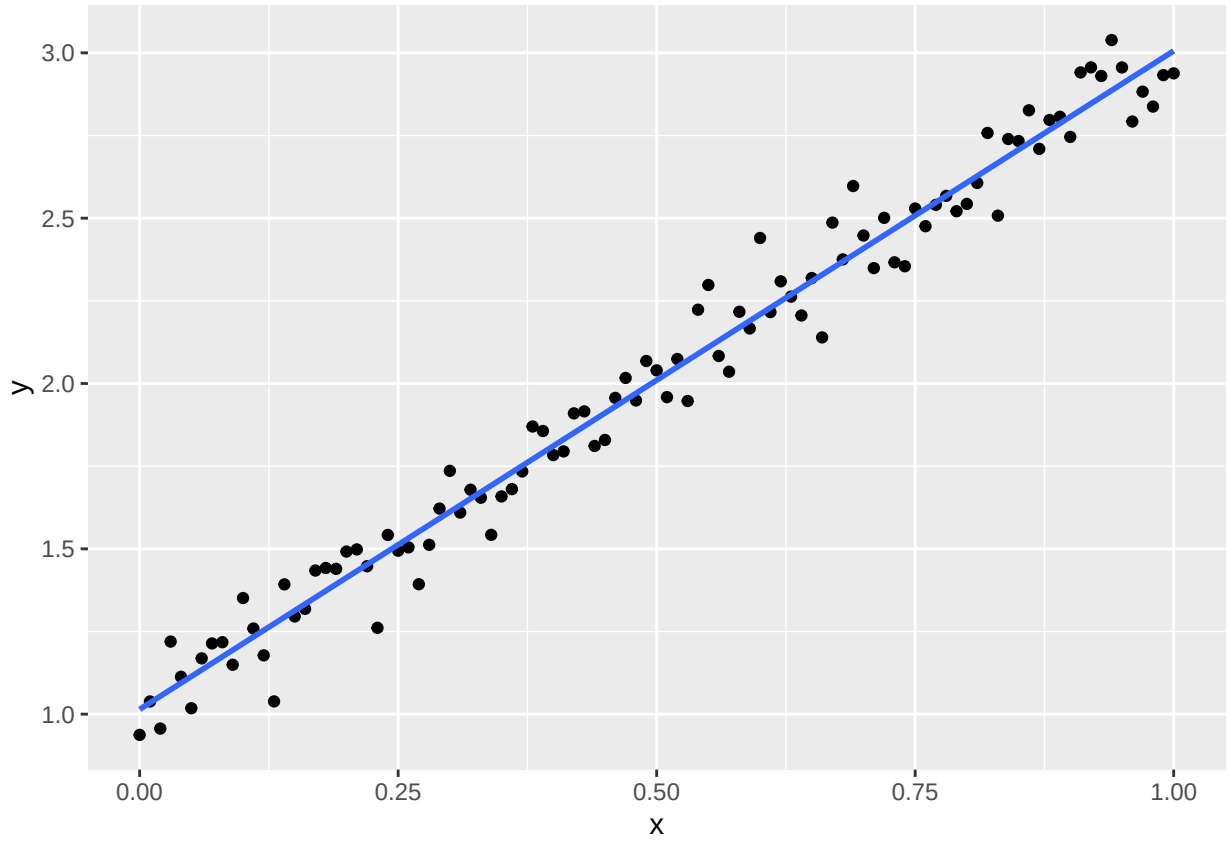
Uncomment b and fill in the formula for the estimate of b. Explain why the points do not lie exactly on the (regression) line, the estimates are not exactly the same as the expected value slope of 2 and intercept of 1. Note that you should get the same estimates for the slope and intercept as provided by the built in lm() command.

Your answer and code here:

```r
set.seed(1)
x = seq(0,1,by=.01)
y = 2*x+1+rnorm(length(x),0,.1)

ggplot(data=data.frame(x,y),aes(x=x,y=y)) + geom_point() + geom_smooth(method=lm, se = FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```r
m = (mean(y*x)-mean(y)*mean(x))/(mean(x^2)-mean(x)^2)
print(m)
```

```
## [1] 1.991375
```

```r
# Uncomment the two lines below and fill in the correct equation.
b = mean(y) - m * mean(x)
print(b)
```

```
## [1] 1.014479
```

```r
#Compare your answer with the built in R command
linear = lm(y~x)
linear$coefficients
```

```
## (Intercept)           x
##    1.014479    1.991375
```

Question 10

Data and Background

This data set lists the individual observations for 934 children in 205 families in 1886 recorded by Galton. One question by Galton is concerning the relation between heights of parents and their offspring. It is from this investigation that Galton coined the phrase: "regression towards the mean."

Remark: According to Dominique Aubert-Marson (in the article Sir Francis Galton: the founder of eugenics), "Not only was Sir Francis Galton a famous geographer and statistician, he also invented "eugenics" in 1883. Eugenics, defined as the science of improving racial stock, was developed from a new heredity theory, conceived by Galton himself, and from the evolution theory of Charles Darwin, transposed to human society by Herbert Spencer. Galton's eugenics was a program to artificially produce a better human race through regulating marriage and thus procreation."

Question 10.1

In the code chunk below create a linear model with the input variable being 'midparentHeight' and 'childHeight' as the output variable. According to this model how does a one unit increase of the variable 'midparentHeight' affect 'childHeight'? What is the slope estimate and what would a slope of zero imply about the relationship between the two variables? Print the linear regression coefficients in the code chunk below as well.

Your answer and code here:

```
attach(GaltonFfamilies)
head(GaltonFamilies)
```

```
##   family father mother midparentHeight children childNum gender childHeight
## 1    001   78.5   67.0           75.43        4        1   male        73.2
## 2    001   78.5   67.0           75.43        4        2 female        69.2
## 3    001   78.5   67.0           75.43        4        3 female        69.0
## 4    001   78.5   67.0           75.43        4        4 female        69.0
## 5    002   75.5   66.5           73.66        4        1   male        73.5
## 6    002   75.5   66.5           73.66        4        2   male        72.5
```

```
# ?lm  provides help page for lm function
model = lm(childHeight ~ midparentHeight, data = GaltonFamilies)
print(model)
```

```
##
## Call:
## lm(formula = childHeight ~ midparentHeight, data = GaltonFamilies)
##
## Coefficients:
##     (Intercept)  midparentHeight
##         22.6362           0.6374
```

```
summary(model)$coefficients
```

```
##                   Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)     22.6362405  4.2651074  5.307308 1.390930e-07
## midparentHeight  0.6373609  0.0616076 10.345491 8.053865e-24
```

According to the linear regression model, the slope estimate of 0.6374 indicates a positive relationship between `midparentHeight` and `childHeight`. So, Yes! A one-unit increase in `midparentHeight` is associated with an increase of approximately 0.6374 units in `childHeight`.

Question 10.2

In the code chunk below we calculate a confidence interval for the slope of the regression equation shown in the previous question. What does the confidence interval imply about the slope and relationship between the variables? Is it possible that there is a negative or no relationship between the variable 'midparentHeight' and 'childHeight' based on the interval?

Your answer and code here:

```
confint(model)
```

```
##                     2.5 %     97.5 %
## (Intercept)     14.2659135 31.0065676
## midparentHeight  0.5164552  0.7582666
```

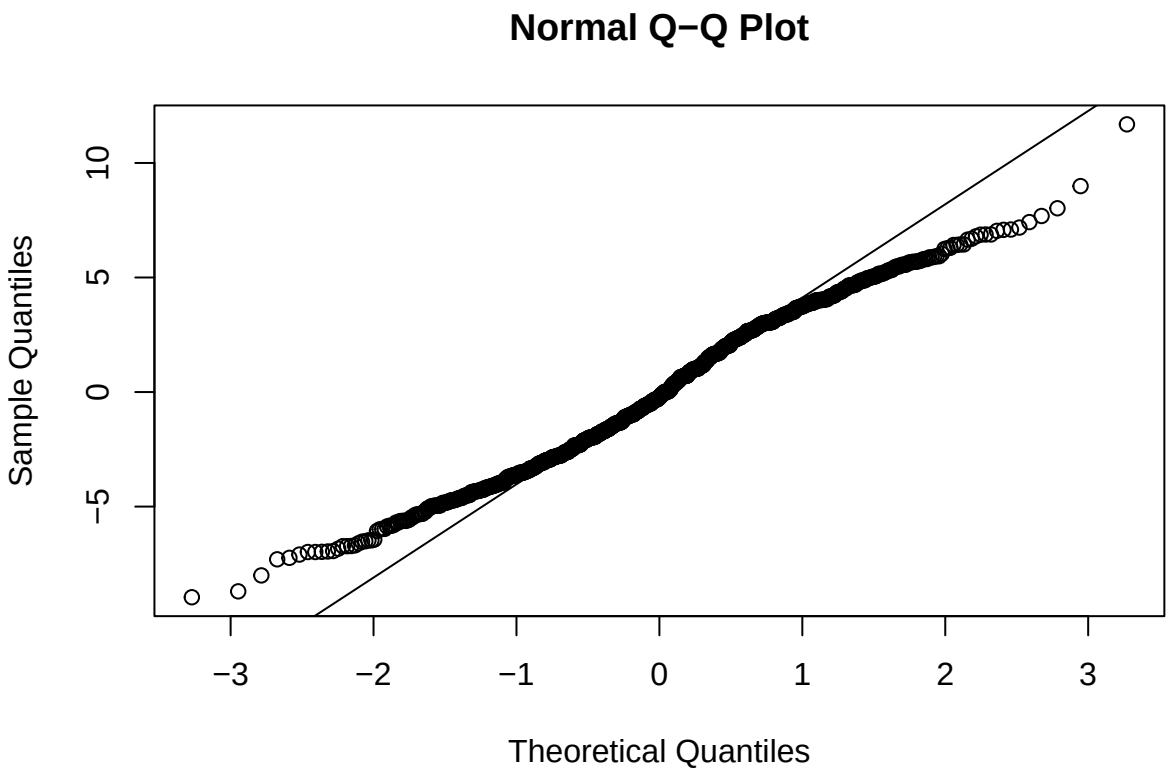
The confidence interval for the slope provides strong evidence that there is a positive linear relationship between `midparentHeight` and `childHeight`. The estimates suggest that for every one-unit increase in `midparentHeight`, `childHeight` increases by an amount between 0.516 and 0.758 units. This positive relationship is statistically significant, and the data do not support the possibility of a negative or non-existent relationship between the variables.

Question 10.3

One of the main assumptions of a linear regression model is that the error terms are normally distributed. The errors for a specified linear model and set of data is called the residuals. Assess the normality of the residuals by making a qqplot and performing the shapiro test on the residuals. Comment on if you believe the residuals are normal or not.

Your answer and code here:

```
#Make qqplot with qq line below
qqnorm(model$residuals)
qqline(model$residuals)
```




```
#Perform shapiro test belwo
#model$residuals
shapiro.test(model$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  model$residuals
## W = 0.98714, p-value = 2.636e-07
```

Based on the results of the Shapiro-Wilk test, which has a very low p-value, we reject the null hypothesis that the residuals are normally distributed. This indicates that there is significant evidence to suggest that the residuals are not normally distributed.