total grade: 4.9/8.0

you are welcome to present Ex. 2 in the class

Home Work 10

1) a) The least square estimator $\hat{\beta}$ can be written as,

[0.3/1.5]

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_n)(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}$$ ✓

the linear regression model is not well understood. Please look at Ch. 3.6

where $\bar{x}$ and $\bar{y}$ are the sample means of x and y respectively.

Now, let $S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x}_n)^2$, Then can rewrite

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_n) y_i}{S_{xx}}$$

This is simple algebraic manipulation.

Next,

$$E(\hat{\beta}) = E\left(\frac{\sum_{i=1}^{n} (x_i - \bar{x}_n) y_i}{S_{xx}}\right)$$ ✓

$$= \frac{1}{S_{xx}} \sum_{i=1}^{n} E\left((x_i - \bar{x}_n) y_i\right)$$ ✗

Now $E((x_i - \bar{x}_n) y_i) = Cov(x_i - \bar{x}_n, y_i) + E(\bar{x}_n) E(y_i)$

✗

Since $\bar{x}$ and y_i are independent. (assumption is that there is no correlation between x and y)

, $Cov(x_i - \bar{x}, y_i) = Cov(x_i, y_i)$

Atlas

CamScanner

No: Date: / /

Assuming that the errors are uncorrelated with x_i (which is a common assumption in linear regression), $Cov(x_i, y_i) = 0$

Therefore $E((x_i - \bar{x}) y_i) = E(\bar{x}) E(y_i) = \bar{x} E(y_i)$

Substituting this back into the expression for $E(\hat{\beta})$

$$E(\hat{\beta}) = \frac{1}{S_{xx}} \sum_{i=1}^{n} \bar{x} E(y_i)$$

Since $\bar{x}$ is a constant with respect to the summation, ~~[illegible]~~

$$E(\hat{\beta}) = \frac{\bar{x}}{S_{xx}} \sum_{i=1}^{n} E(y_i) = \frac{\bar{x} \cdot n E(y_i)}{S_{xx}}$$

Since $E(y_i)$ is the expected value of the error term. and assuming that the errors have an expected value of ~~zero~~ 0, Then

$$E(\hat{\beta}) = \frac{n \cdot \bar{x} \cdot 0}{S_{xx}} = 0$$

Therefore $E(\hat{\beta}) = 0$ and $\hat{\beta}$ is an unbiased estimator for β.

$$\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x}$$

Above proved, $E(\hat{\beta}) = 0$, and since $\bar{y}$ and $\bar{x}$ are constants, then,

Y_i are random variables while x_i are constants

$$E(\hat{\alpha}) = E(\bar{y} - \hat{\beta} . \bar{x})$$
$$= \bar{y} - \bar{x} \, E(\hat{\beta})$$

$$E(\hat{\beta}) = 0$$

$$E(\hat{\alpha}) = \bar{y} - \bar{x} \cdot 0$$
$$E(\hat{\alpha}) = \bar{y}$$

Therefore $E(\hat{\alpha}) = \bar{y}$ and $\hat{\alpha}$ is an unbiased estimator for α, Both $\hat{\alpha}$ and $\hat{\beta}$ are unbiased estimators for α and β.

b) $$\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{S_{xx}}$$

Let's consider the distribution of β.

Since ϵ_i are assumed to be normally distibuted $y_i = \alpha + \beta x_i + \epsilon_i$ is also normally distributed. The sum of normal distributed variables are also normaly distributed. Therefore $\sum_{i=1}^{n} (x_i - \bar{x}) y_i$ is is normaly distributed.

Now $\hat{\beta}$ is a linear combination of normal random variables divided by a constant (S_{xx}) and the result is also normally distributed. The variance of $\hat{\beta}$ is given by

Atlas

No: Date: / /

$$Var(\hat{\beta}) = \frac{\hat{\sigma}}{S_{xx}}$$ why?

Let's consider the expression for $\hat{\alpha}$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x}$$

$\bar{y}$ and $\hat{\beta}$ are linear combinations of normally distributed variables. Therefore, $\hat{\alpha}$ is also normally distributed. The variance of $\hat{\alpha}$ is given by,

$$Var(\hat{\alpha}) = \sigma^2 \left(\frac{1}{n} + \frac{(\bar{x})^2}{S_{xx}} \right)$$

$$Var(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}$$

$$Var(\hat{\alpha}) = \left(\frac{1}{n} + \frac{(\bar{x})^2}{S_{xx}} \right)$$

So, both $\hat{\alpha}$ and $\hat{\beta}$ are normally distributed and their variances are as specified.

CamScanner

No: Date: / /

c) $\hat{\beta} = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$ and $\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x}$

Now, let's use the hint and express the covariane of $\hat{\alpha}$ and $\hat{\beta}$ in terms of convariances of individuals observations,

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = \text{Cov}\left(\bar{y} - \hat{\beta} \cdot \bar{x}, \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{S_{xx}}\right)$$

Using the Covariance property,

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = \sum_{i=1}^{n} \text{Cov}\left(\bar{y} - \hat{\beta} \cdot \bar{x}, (x_i - \bar{x}) y_i\right) \cdot \frac{1}{S_{xx}}$$

~~Let's simplify~~

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = \sum_{i=1}^{n} \left[\text{Cov}(\bar{y}, (x_i - \bar{x}) y_i) - \text{Cov}(\hat{\beta} \cdot \bar{x}, (x_i - \bar{x}) y_i)\right]$$

$\text{Cov}(\bar{y}, (x_i - \bar{x}) y_i)$ Since $\bar{y}$ is a constant with respect to y_i, the covariance term is ~~$(x_i - \bar{x})$~~. $(x_i - \bar{x}) \cdot \text{Cov}(\bar{y}, y_i)$ assuming independence,

this is not constant

$\text{Cov}(\bar{y}, y_i) = 0$

$\text{Cov}(\hat{\beta} \cdot \bar{x}, (x_i - \bar{x}) y_i)$, this involves product of $\hat{\beta}$ and $\bar{x}$ with y_i and $(x_i - \bar{x})$. This covariance term can be simplified by treating $\hat{\beta}$ and $\bar{x}$ as constants.

CamScanner

No :

Then the only non zero term involves $(x_i - \bar{x}) \cdot \frac{\sigma^2}{S_{xx}}$ which leads to desired result,

$$cov(\hat{\alpha}, \hat{\beta}) = \frac{-\sigma^2 \bar{x}}{S_{xx}}$$ ×

03)

a) PMF is $f_\pi(x) = \pi^x (1-\pi)^{1-x}$ for Bernoulli

[1.1/1.5]

Now let $x_1, x_2, \ldots, x_n$ be a random sample from this distribution.

The log function for the entire sample is

$$L(\pi : x_1, x_2, \ldots x_n) = \sum_{i=1}^{n} \log f_\pi(x_i)$$

$$L(\pi : x_1, x_2, \ldots x_n) = \sum_{i=1}^{n} \left[x_i \log(\pi) + (1-x_i) \log(1-\pi) \right]$$ ✓

The ~~ob~~ Fisher Information $I(\pi)$ is given by the second derivative of log function.

$$\frac{d}{d\pi} l(\pi) = \sum_{i=1}^{n} \left[\frac{x_i}{\pi} - \frac{1-x_i}{1-\pi} \right]$$ ✓

$$\frac{d^2}{d\pi} l(\pi) = \frac{d}{d\pi} \left[\sum_{i=1} \left(\frac{x_i}{\pi} - \frac{1-x_i}{1-\pi} \right) \right]$$

✓

$$\frac{d^2}{d\pi} l(\pi) = \sum_{i=1}^{n} \left[-\frac{x_i}{\pi^2} + \frac{(1-x_i)}{(1-\pi)^2} \right]$$

CamScanner

No :

b) Since $x_1, x_2, \ldots x_n$ are independent iid the expected Fisher information for the entire sample is the sum of the individual expected Fisher information, $I(\pi) = -E\left[\frac{d^2}{d\pi^2} L(\pi)\right]$

$$I \text{ expected } (\pi) = \cancel{n} \cdot E\left[\frac{x_i}{\pi^2} + \frac{(1-x_i)}{(1-\pi)^2}\right]$$

$$E(x_i) = \pi$$

$$I \exp (\pi) = n \cdot \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)$$ since x_i follows Bernoulli distn with parameter π.

Therefore

$$I \text{ ex } (\pi) = \cancel{n} \cdot \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right) ✓$$

c) $$f_\pi(x_i) = \pi^{x_i}(1-\pi)^{1-x_i}$$

$$L(\pi : x_1, x_2 \ldots x_n) = \sum_{i=1}^{n} \log f_\pi(x_i)$$

$$\frac{\partial}{\partial \pi} L(\pi : x_1, x_2, \ldots x_n) = \sum_{i=1}^{n}\left(\frac{x_i}{\pi} - \frac{1-x_i}{1-\pi}\right)$$

set it to zero and find the estimator

$$I(\pi) = -E\left(\frac{\partial^2}{\partial \pi^2} L(\pi; x_i)\right]$$

$$= -E\left[-\frac{x_i}{\pi^2} - \frac{1-x_i}{(1-\pi)^2}\right]$$

$$= E\left[\frac{x_i}{\pi^2} + \frac{1-x_i}{(1-\pi)^2}\right]$$

CamScanner

No : ________ Date: __/__/____

$$I(\pi) = \frac{1}{\pi} + \cancel{\frac{1-\pi}{(1-\pi)^2}} \frac{(1-\pi)}{(1-\pi)^2}$$

$$Var(\hat{\pi}_n) \geq \frac{1}{n I(\pi)}$$

$$Var(\hat{\pi}_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$$

$$E(x_i) = \pi,$$

Since x_i's are assume to be iid,

$$Var(\hat{\pi}_n) = \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i)$$

$$Var(\hat{\pi}_n) = \frac{1}{n^2}\sum_{i=1}^{n} \pi(1-\pi)$$

$$Var(\hat{\pi}_n) = \frac{1}{n^2} n\pi(1-\pi)$$

$$Var(\hat{\pi}_n) = \frac{\pi(1-\pi)}{n} ✓$$

Compare to the Cramer Rao bound

$$\frac{1}{n I(\pi)} = \frac{1}{n\left(\frac{1}{\pi} + \frac{(1-\pi)}{(1-\pi)^2}\right)}$$

$$\frac{1}{n I(\pi)} = \frac{1-\pi^2}{n}$$

Cramer – Rao bound states that

No:

$$Var(\hat{\pi}_n) \geq \frac{1}{nI(\pi)}$$

$$\frac{\pi(1-\pi)}{n} \geq \frac{(1-\pi)^2}{n}$$

$$\pi - \pi^2 \geq 1 - 2\pi + \pi^2$$

$$2\pi^2 - 3\pi + 1 \leq 0$$

CamScanner

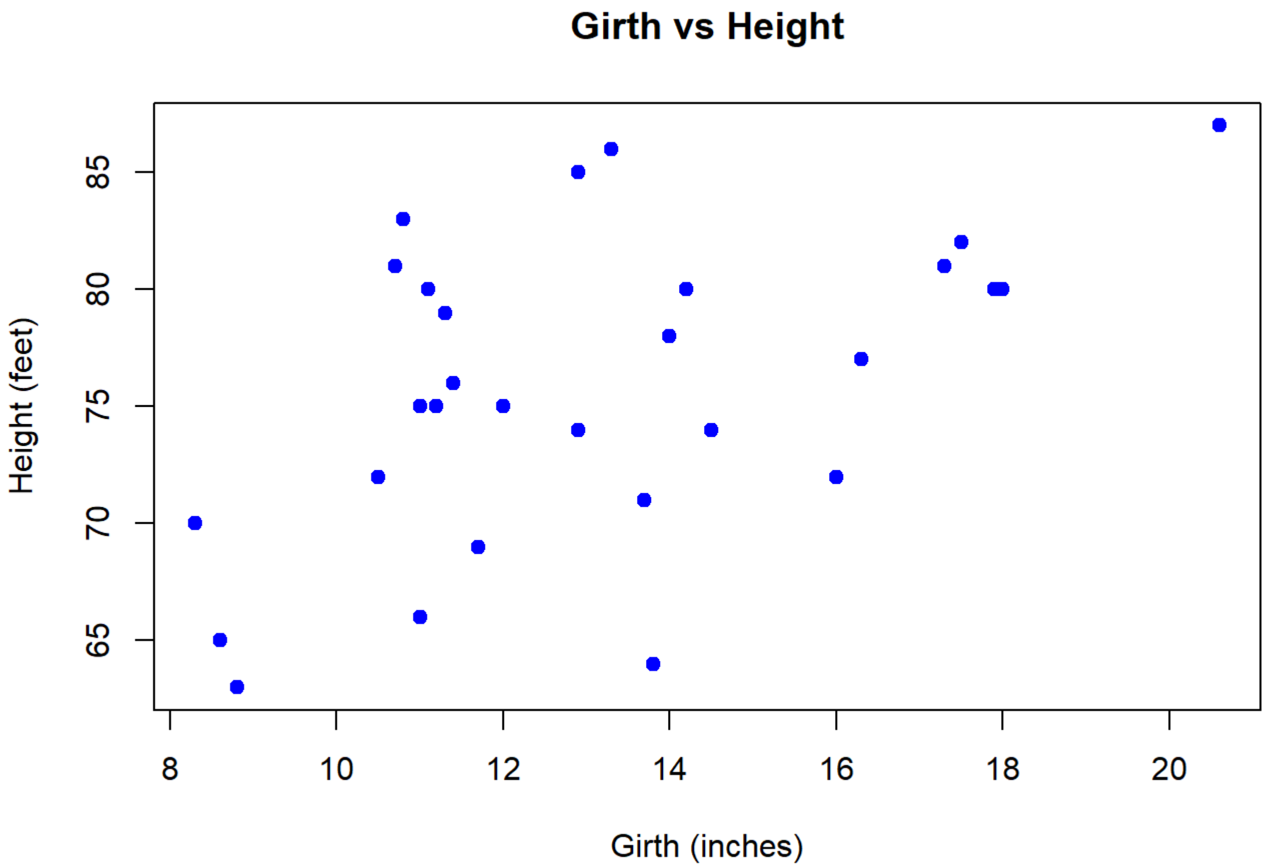
R Notebook

[1.3/1.5] you are welcome to present Ex. 2 in the class

2 - (a)

```
data(trees)

#Scatterplot
plot(trees$Girth, trees$Height, main="Girth vs Height",
     xlab="Girth (inches)", ylab="Height (feet)", pch=19, col="blue")
```

does not show all the data use instead plot(trees)

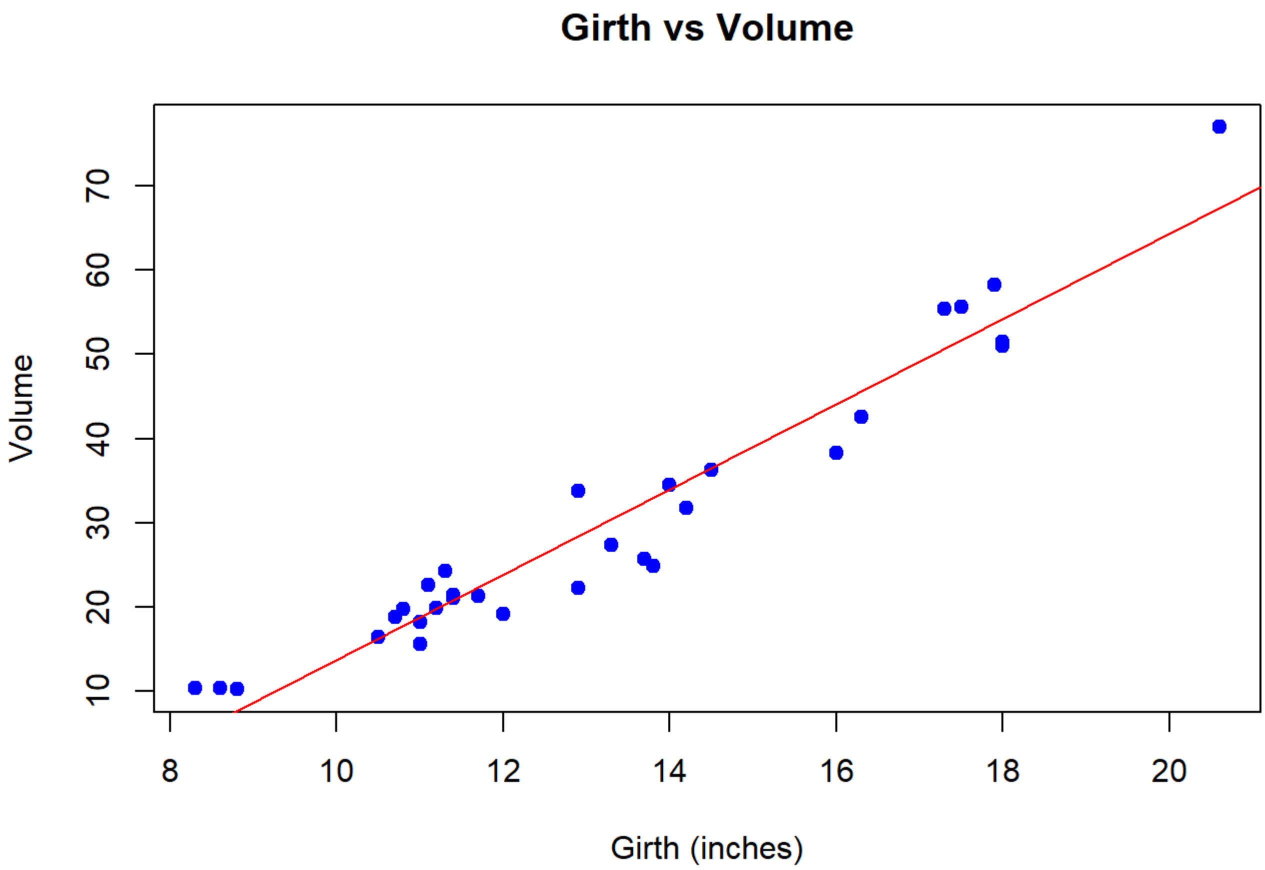


2 - (b)

```
x <- trees$Girth
y <- trees$Volume

#Linear regression model
fit <- lm(y ~ x)

#Summary of the regression model
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.065 -3.107  0.152  3.495  9.587
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -36.9435     3.3651  -10.98 7.62e-12 ***
## x             5.0659     0.2474   20.48  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.252 on 29 degrees of freedom
## Multiple R-squared:  0.9353, Adjusted R-squared:  0.9331
## F-statistic: 419.4 on 1 and 29 DF,  p-value: < 2.2e-16
```

```
plot(x, y, main="Girth vs Volume",
     xlab="Girth (inches)", ylab="Volume", pch=19, col="blue")

#Regression line to the scatterplot
abline(fit, col="red")
```


Girth vs Volume
Volume
10
20
30
40
50
60
70
8
10
12
14
16
18
20
Girth (inches)

2 - (c)

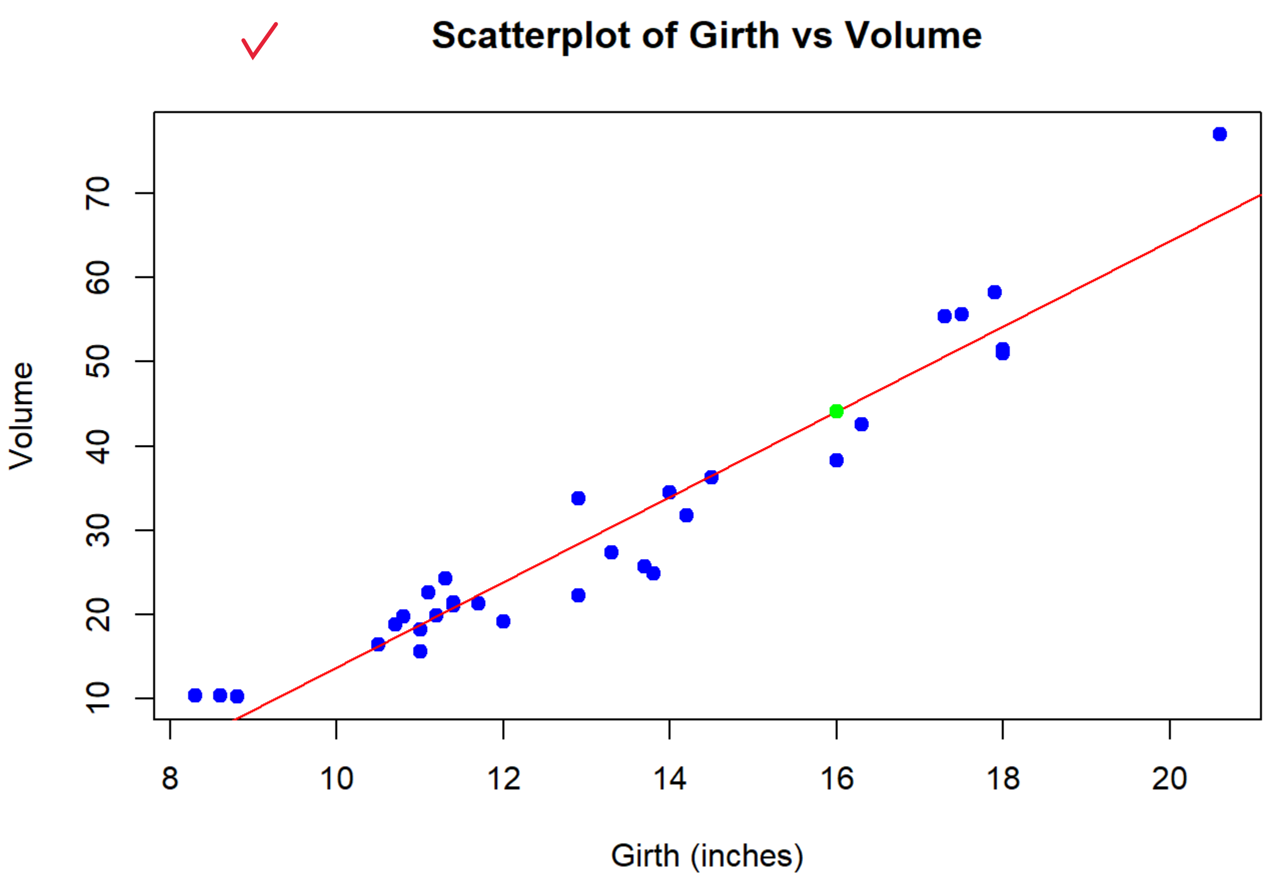
```
new_girth <- 16

#Making a prediction
prediction <- predict(fit, newdata = data.frame(x = new_girth), interval = "prediction")
cat("Predicted Volume (using predict()):", prediction[1], "\n")
```

```
## Predicted Volume (using predict()): 44.11024
```

```
#Computation using coefficients
coefficients <- coef(fit)
direct_computation <- coefficients[1] + coefficients[2] * new_girth
cat("Predicted Volume (direct computation):", direct_computation, "\n")
```

```
## Predicted Volume (direct computation): 44.11024
```

```
plot(x, y, main="Scatterplot of Girth vs Volume",
     xlab="Girth (inches)", ylab="Volume", pch=19, col="blue")
abline(fit, col="red")

#Adding the new point for the tree with girth size 16 inches
points(new_girth, prediction[1], col="green", pch=19)
```



As we can see we are getting the same answer in both the cases. This consistency is expected and indicates that both methods are correctly predicting the volume of a tree with a girth size of 16 inches based on the linear regression model. It’s a good sign that the model and prediction process are working as expected.

7 - (a) [0.7/1]

```
set.seed(123)

#Parameters for simulation
p_true <- 0.05
n <- 60
num_simulations <- 10000

#Function to calculate the confidence interval
calculate_ci <- function(x) {
  p_hat <- sum(x) / length(x)
  se <- sqrt(p_hat * (1 - p_hat) / length(x))
  margin_of_error <- 1.96 * se
  lower_bound <- p_hat - margin_of_error
  upper_bound <- p_hat + margin_of_error
  return(c(lower_bound, upper_bound))
}

#Performing simulations
coverage_count <- 0

for (i in 1:num_simulations) {
  sample_data <- rbinom(n, 1, p_true)
  ci <- calculate_ci(sample_data)
  if (p_true >= ci[1] & p_true <= ci[2]) {
    coverage_count <- coverage_count + 1
  }
}

#The true coverage probability
coverage_probability <- coverage_count / num_simulations
cat("True Coverage Probability:", coverage_probability, "\n")
```

```
## True Coverage Probability: 0.8062
```

7 - (b)

```
set.seed(123)

# Parameters
p_true <- 0.05
n <- 60
alpha <- 0.05
num_simulations <- 10000

# Function to calculate the Wilson confidence interval
calculate_wilson_ci <- function(x, n, alpha) {
  p_hat <- sum(x) / length(x)
  z_alpha_half <- qnorm(1 - alpha / 2)
  sqrt_term <- sqrt(p_hat * (1 - p_hat) / n)
```

some terms missig here

```
  lower_bound <- (p_hat + (z_alpha_half^2) / (2 * n) - z_alpha_half * sqrt_term) /
                 (1 + (z_alpha_half^2) / n)

  upper_bound <- (p_hat + (z_alpha_half^2) / (2 * n) + z_alpha_half * sqrt_term) /
                 (1 + (z_alpha_half^2) / n)

  return(c(lower_bound, upper_bound))
}

#Performing simulations
coverage_count <- 0

for (i in 1:num_simulations) {
  sample_data <- rbinom(n, 1, p_true)
  wilson_ci <- calculate_wilson_ci(sample_data, n, alpha)
  if (p_true >= wilson_ci[1] & p_true <= wilson_ci[2]) {
    coverage_count <- coverage_count + 1
  }
}

#The coverage probability for Wilson CI
coverage_probability <- coverage_count / num_simulations
cat("Wilson Confidence Interval Coverage Probability:", coverage_probability, "\n")
```

```
## Wilson Confidence Interval Coverage Probability: 0.8765
```

Q.4] a]

[0.8/1.5]

The likelihood function for given density is

~~$L(\lambda)$~~

$$L(\lambda) = \prod_{i=1}^{n} \left(\frac{\lambda}{2\sqrt{x_i}} e^{-\lambda \sqrt{x_i}} \right)$$

taking log-likelihood

$$\ell(\lambda) = \sum_{i=1}^{n} \left(\log\left(\frac{\lambda}{2\sqrt{x_i}}\right) - \lambda \sqrt{x_i} \right)$$

differentiate w/ respect to λ & set ~~to~~ equal to zero

$$\frac{dl}{d\lambda} = \sum_{i=1}^{n} \left(\frac{1}{\lambda} - \sqrt{x_i} \right) = 0$$

$$\frac{1}{\lambda} = \frac{1}{n} \sum_{i=1}^{n} \sqrt{x_i}$$

$$\lambda n = \frac{n}{\sum_{i=1}^{n} \cancel{\Sigma} \sqrt{x_i}}$$

For Fisher information $I_n(\lambda)$:-

$$I_n(\lambda) = -E\left[\frac{d^2 l}{d\lambda^2}\right]$$

finding second derivative of $l(\lambda)$:

$$\frac{d^2 l}{d\lambda^2} = -\sum_{i=1}^{n} \frac{1}{\lambda^2}$$

$$I_n(\lambda) = E\left[\sum_{i=1}^{n} \frac{1}{\lambda^2}\right]$$

$$= \frac{n}{\lambda^2}$$

$$\therefore I_n(\lambda) = \frac{n}{\lambda^2} \checkmark$$

b) Use the fisher information to approximate $Var(\hat{\lambda})$ as $n \to \infty$:

$$Var(\hat{\lambda}_n) \approx [I_n(\lambda)]^{-1}$$

$$= \frac{1}{n/\lambda^2}$$

$$= \frac{\lambda^2}{n} \checkmark$$

As $n \to \infty$, the variance goes to 'zero'.

Date
Page

5 a] Fisher information for σ

[0.7/1.5]

likelihood function :-
pdf of normal distribution

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

log-likelihood function :-

$$l(x_i/\sigma) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{x_i^2}{2\sigma^2}$$

$$\frac{\delta l(x_i/\sigma)}{\delta\sigma} = -\frac{1}{\sigma} + \frac{x_i^2}{\sigma^3}$$ ✓

Fisher information ($I\sigma$):

$$I\sigma = E\left[\left(\frac{\delta l(x_i/\sigma)}{\delta\sigma}\right)^2\right]$$

Taking ~~expectation~~ expectation. However xi follow standard normal distribution, $E[x_i^2] = 1$ & $E[x_i^4] = 3$

sigma missing here

$$\therefore I(\sigma) = E\left[\left(\frac{\delta l(x_i/\sigma)}{\delta\sigma}\right)^2\right] = \frac{\cancel{1}}{\sigma^2}$$

b] ML-Estimator $\hat{\sigma}_n$:

$$l(x_1, x_2, \ldots x_n/\sigma) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2$$

differentiating w respect to σ

$$\frac{\delta l(x_1, x_2, \ldots, x_n/\sigma)}{\delta\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n} x_i^2$$

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} x_i^2 = 0$$

✓

$$\hat{\sigma}_n = \sqrt{(1/n) \sum_{i=1}^{n} x_i^2}$$ ... by Centeral li~~m~~it theorem as $n \to \infty$ $\hat{\sigma}_n$ approches a normal dist.

c] Now consider $\theta = \sigma^2$ as unknow parameter

Using Invariance property of Fisher information i.e $I_e(\theta) = \frac{I(\sigma)}{4\cancel{\sigma}}$ for $\theta = \sigma^2$

$$I_e(\theta) = \frac{1}{\cancel{4\theta}}$$ ✗

d] MLE for $\theta(\hat{\theta}_n)$: find directly?

Applying invariance property $\hat{\theta}_n = \hat{\sigma}_n^2$ ✓

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

By law of large number $\hat{\theta}_n$ converges in probability to $E[x_i^2]$ which is θ.

Using properties of mean & variance, let's check unbiasedness

$$E[\hat{\theta}_n] = \frac{1}{n} \sum_{i=1}^{n} E[x_i^2] = \theta$$

$\hat{\theta}_n$ is unbiased, while $\hat{\sigma}_n$ is biased. This bias arises because taking square root of unbiased applying Jensen's inequality to $E[\sqrt{\hat{\theta}_n}]$

we get,

$$E[\hat{\sigma}_n] \leq \sqrt{E[\hat{\theta}_n]},$$

which indicating bias.