MA 3710: Chapter 2 Handout
Md Mutasim Billah

Example 1: Histogram

Considering the following example, draw the histogram:

Suppose, 50 students run a mile, record times in minutes:

4:55.21 Jake, 5:28.20 Alyssa, 7:15.03 Hannah, …. , 9:31.07 Tyler.

Class	Frequency	
[4,5)	2	
[5,6)	5	
[6,7)	7	
[7,8)	11	
[8,9)	14	
[9,10)	11	
Total	50	

Example 2: Dot Plot

Considering the following example, draw the Dot Plot:

Suppose, we are given the breaking strength of carbon fibers in GPa (gigapascals) as follows: n=15 observations

{0.1, 2.0,2.3, 2.5, 3.2, 3.4, 3.4, 3.7, 3.9, 4.0, 4.1, 4.2, 4.2, 4.3, 4.4}

Example 3: Stem-and-Leaf Plot:

(a) Considering the above example, draw the Stem-and-Leaf Plot:

{0.1, 2.0,2.3, 2.5, 3.2, 3.4, 3.4, 3.7, 3.9, 4.0, 4.1, 4.2, 4.2, 4.3, 4.4}

(b) Suppose, we have collected a random sample of 15 observations, 7 were from Method A and 8 were from Method B. Considering the following data, draw a side-by-side Stem-and-Leaf Plot. Describe the plots as right-skewed, left-skewed, or symmetric. Also, compare the mean and median of the observations of method A & B.

Method A : n= 7	Method B : n= 8
{0.1, 2.0, 2.5, 3.2, 3.4, 3.4, 4.0	{2.3, 3.7, 3.9, 4.1, 4.2, 4.2, 4.3, 4.4}

Example 4: Box-Plot

Suppose, we have the following n = 7 observations:
{0, 11, 13, 25, 34, 40, 81}

For the above data set, find the values of the lower quartile, median, upper quartile, IQR, step, UIF, LIF, UOF, and LOF.
Use this information to draw a boxplot.

Chapter 3.1 (Probability): Handout
Md Mutasim Billah

Example 1: Let's we have a 6-sided die and we flip the die one time. Suppose,
E: An Even Number appears
F: A number less than or equal to 4 appears. Find P(E), and P(F).

Example 2: Suppose, we select a freshman at random at a big state University.
P(enrolled in Calculus) = P(C) = .3
P(enrolled in History) = P(H) = .5
P(enrolled in both classes) = P(H ∩ C) = .2

Draw a Venn diagram & find the following:

(a) $P(H \cup C)$:

(b) $P(\overline{H \cup C})$:

(c) $P(H \cap \bar{C})$:

(d) $P(\bar{H} \cup C)$:

(e) $P(C/H)$:

(f) Are C & H independent? Justify your answer.

(g) Are C & H mutually exclusive?

(h) H and ______ are mutually exclusive.

Example 3: Suppose, John is flipping a 2-sided coin. If he gets a tail in the first draw, he could stop the process. However, if he gets a head in the first draw, he is needed to flip the coin once more.

(a) Draw a tree diagram.

(b) Find P(Head appears/tail appears).

Example 4: 1% of the people on an island have Disease X. A test for the disease is 99% accurate.

(a) Given that someone on the island tests positive, what is the probability that they actually have the disease?

(b) Given that someone test positive three times in a row, what is the probability they actually have the disease?

Example 5: A company makes fluorescent light bulbs in three factories: Rockford, Sycamore, and Thompson. One bulb is selected at random.
Suppose, 70% are made in Rockford: P(R) = .70
20% are made in Sycamore: P(S) = .20
10% are made in Thompson: P(T) = .10

Let, W = A bulb is still working after 10,000 hours of use.
Find P(W/S), P(W/T), P(W), $P(S \cup \overline{W})$, P(W/T), and P(T/W).

Example 6: Suppose, A basket contains 10 balls: 7 red, 2 green, and 1 black. 3 balls will be drawn out one at a time at random. Considering with replacement and without replacement, find the following probabilities:

P(RRR), P(BGB), P(no black draws), and P(at least one green).

Example 7: The famous birthday problem:
Would you be surprised if two people in this room shared the same birthday?
Find P(at least one shared birthday).

Example 8: Suppose, we are rolling two fair die. Let, one die is red and the other is blue. There are 36 equally likely outcomes in the sample space:

(red = 1, blue = 1)
(red = 2, blue = 1)
(red = 3, blue = 1)
…
(red = 5, blue = 6)
(red = 6, blue = 6)

The table below has 36 cells corresponding to these 36 outcomes. The sum corresponding to each outcome appears in the appropriate cell.

		red die					
		1	**2**	**3**	**4**	**5**	**6**
	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
blue die	**3**	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Find each of the following probabilities:

P(sum=7)

P(sum=2)

$P(blue = 6 \cap sum \geq 11)$

$P(sum = even \cup red = 1)$

$P(red = 4/sum = 10)$

Identify whether the following event is “independent”, “mutually exclusive”, or “neither”. Justify your answer.

Red=5 and sum=9

Chapter 3.2 (Random Variables and Distributions): Handout
Md Mutasim Billah

Example 1: Suppose, in a basket we have 10 balls. 7 of them are red, 1 is black, and the remaining are green. We are going to draw two balls from the basket at random WR.

Let, Y = The number of red ball draws $=\{0,1,2\}$

(a) Define probability mass function (p.m.f.).

(b) Find $P(Y=y) = P(y)$.

(c) Now, let’s consider to draw repeatedly WR until we have found the first black ball. Suppose, Y_B = The number of draws required to observe the first black ball = {1,2,3, 4,……}. Discuss this scenario as a geometric series.

(d) Define population mean (expected value of a discrete random variable. Find $E(Y)$ and $E(Y_B)$.

Example 2: Four buses carrying 40, 33, 25, and 50 students. (148 total)
*Select one student at random
Suppose, Y_1 = # of students on bus he/she rode

*Select one bus driver at random
Y_2 = # of students on bus he/she drove

Compare $E(Y_1)$ and $E(Y_2)$. Which one is greater?

Example 3: Y = outcome of a loaded die = {1, 2, 3, 4, 5, 6}

Y	P(Y=y)			
1	.1			
2	.1			
3	.2			
4	.2			
5	.1			
6	.3			

Find μ_y and σ_y.

What if we have n=1000 die rolls?

Example 4: Suppose, in a basket we have 10 balls. 7 of them are red, 1 is black, and the remaining are green. We are going to draw two balls from the basket at random WR.

Let, Y = The number of red ball draws $=\{0,1,2\}$. Find σ_y.

Example 5: Let's play Roulette! A roulette wheel is divided into 38 slices.
1-36 (18 reds, 18 blacks), 0, 00 (2 greens). Let, Y= Profit from \$1 bet on red. In the long run, do you expect to win money, lose money, or break even? Why?

Chapter 3.3 (Discrete Random Variables): Handout
Md Mutasim Billah

Binomial Distribution

Example 1: Are the following random variables binomial?

a) The number of trials in 6-coin flips
A = {0,1,2,3,4,5,6}

b) The number of flips needed to have a total of 3 tails
B = {3,4, 5,….}

c) The number of red draws if we draw 4 balls out of 7 red, 2 green, and 1 black ball at random WOR
C = {1,2,3,4}

d) Same as above, but WR
D = {0,1,2,3,4}

Example 2:

A company makes fluorescent light bulbs. Suppose, W = a bulb is working after 10,000 hours of use. P(W) = .86 (assuming bulbs are independent)

In a batch of n = 5 {0,1,2,3,4,5} bulbs, what is the probability exactly 4 will last 10,000 hours?

Example 3: A class has 9 women and 6 men. If the class meets 10 times with perfect attendance at each meeting, and a student is selected at random at each meeting, what is the probability that a woman is selected 7 times (out of the 10 times)?

Example 4: Geometric Distribution

A new radio communications system attempts to send a signal to a receiver in the basement of the EERC once a minute. The probability of success on each attempt (independently) is .25.

Suppose, Y = # of attempts needed to observe first success.

(a) What kind of discrete random variable is Y?

(b) What is the probability that the first successful free throw is on the fourth attempt?

(c) What is the probability that at least two attempts will be needed to observe the first success?

Example 5: Poison Distribution

Suppose, Y = # of lightning strikes to ME-EM per year. Y~Poisson(λ= 0.125= 1/8)

(a) Find μ= E[Y] and interpret the result.

(b) P(no strikes for the next year):

(c) P(at least one strike within the next year):

Chapter 3.4 (Continuous Random Variables): Handout
Md Mutasim Billah

Example 1: Uniform Distribution

Suppose, Y = weight (grams) of a randomly selected car part.
$Y \sim UNIF(2.97, 3.01)$

(a) Find $P(2.99<Y<3.00)$

(b) Find μ_y & σ_y

Example 2: Exponential Distribution

Suppose, Y_{Poi} = # of lightning strikes to ME-EM per year. $Y_{Poi} \sim$ Poisson(λ= 0.125= 1/8).
Let, Y_{Exp} = # of years until the next strike (i.e. waiting time until next strike)

(a) Find $E(Y_{Exp})$ & construct the probability density function of Y_{Exp}. Check whether it is a true p.d.f.

(b) Find
P(no lightning strikes for the next year) =

(c) Find
P(waiting time until next strike is one year or more) =

(d) Find
P(we will observe at least one lightning strike within the next 66 months) =

Chapter 3.5 (The Normal Distribution): Handout
Md Mutasim Billah

Example 1:
If $Z \sim N(\mu=0, \sigma=1)$, then find the following probabilities:

(a) $P(Z < -0.77)$

(b) $P(Z > -0.77)$

(c) $P(-1.05 < Z < 2.41)$

(d) $P(-1 < Z < 1)$; $P(-2 < Z < 2)$; $P(-3 < Z < 3)$

Example 2:
Suppose, $Z \sim N(\mu=0, \sigma=1)$, then find the value of Z if:

(a) $P(Z < z) = 0.0708$

(b) $P(Z > z) = 0.6480$

(c) $P(-z < Z < z) = 0.95$

(d) $P(-1.21 < Z < z) = 0.7330$

Example 3:
Suppose, Z ~ N(μ=0, σ=1). Answer the following questions:

(a) What is the 90^{th} & 10^{th} percentile of Z?

(b) What is the 75^{th} & 25^{th} percentile of Z?

Example 4:
Suppose, Y = Mass (in grams) of a randomly selected newborn baby, where,
Y ~ N(μ= 3400, σ= 5000).

(a) What is the probability that a randomly selected baby mass is greater than 4170 grams?

(b) What is the probability that a randomly selected baby mass is less than 3800 grams?

Example 5: Fill in the blanks! ☺

87.7% of all babies are lighter than Mary. Mary's mass is the __________________________ and the mass equals __________ grams.

Example 6:
Suppose, Y = Gallons of milk placed in a plastic jug, where, $Y \sim N(\mu=1.02, \sigma=.01)$.

(a) Find
P(a randomly selected jug will be under filled) =

(b) 80% of all milk jugs contain a volume of milk in an interval (A,B) centered around $\mu = 1.02$. What is (A,B)?

Chapter 3.6 (Random Behavior of Means): Handout
Md Mutasim Billah

Example 1:

Suppose, Y = Gallons of milk placed in a plastic jug, where, $Y \sim N(\mu=1.02, \sigma=.01)$.

(a) If we randomly select n=2 jugs and compute the mean $\bar{Y}$, then find $P(\bar{Y} < 1)$.

(b) What happens to $P(\bar{Y} < 1)$ as sample size increases?

Example 2:
Suppose, Y = weight (grams) of a randomly selected car part, where, Y ~ UNIF(2.97, 3.01).

(a) If you select n=30 car parts at random & compute $\bar{Y}$, then find P(2.987 < $\bar{Y}$ < 2.991)

(b) If you select n=50 car parts at random & compute $\bar{Y}$, then find P(2.987 < $\bar{Y}$ < 2.991)

Chapter 3.7 (Random Behavior of means when the variance is unknown)

Example 3: Suppose, we have collected a random sample of n=5 observations as follows: {3,7,10,11,14}. Find the sample variance and sample standard deviation.

Y	Y- $\bar{Y}$	(Y- $\bar{Y}$)^2	Y^2
3			
7			
10			
11			
14			
Total			

Example 4: Without calculation, answer the following questions:

(a) {10,10,10,12,12,12}; Find $\bar{Y}$ & s.

(b) {0,0,0,0,100,100,100,100}; Find $\bar{Y}$ & s.

(c) Given, {3,7,10,11,14}; $\bar{Y}$ = 9, s^2 = 17.5, s = 4.18. Then, for {13,17,20,21,24}, find $\bar{Y}$, s^2, & s.

Rule:

(d) Given, {3,7,10,11,14}; $\bar{Y}$ = 9, s^2 = 17.5, s = 4.18. Then, for {30,70,100,110,140}, find $\bar{Y}$, s^2, & s.

Rule:

Chapter 4.1 (Estimation): Handout
Md Mutasim Billah

Example 1: Let us consider the following situation: $\bar{y}$ is normal (population is normal) or approximately normal (sample is large, $n \geq 30$). Therefore, $\bar{y} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$. For now, assume that σ is known. Construct 95% confidence interval for μ.

Example 2: We are interested in μ = population mean weight among all MTU students.

$n = 30$, $\bar{y} = 150$ pounds, $\sigma = 25$ (known). Construct 95% confidence interval for μ.

Example 3: True/False Game:

(a) We sampled 95% of all MTU students.

(b) We know that $\mu = 150$

(c) 95% of all MTU students have a weight between 141.05 & 159.95.

(d) We know that $141.05 < \mu < 158.95$

(e) We are 95% confident that $141.05 < \mu < 158.95$

Example 4: What about other confidence levels for μ? (like 90% and 99%)

Example 05: We are interested in μ = population mean weight among all MTU students.

n = 30, $\bar{y}$ = 150 pounds, σ = 25 (known). 95% confidence interval for μ was 150 ± 8.95.

(a) If we want the error margin to equal a specific small number B, what sample size is necessary?

(b) If we want the error margin to qual to 3 pounds with 95% confidence, what should be the sample size of MTU students?

Example 06: Suppose, μ = Population mean cholesterol level among 70-years-old American men. n = 19 (as small n, we must assume that population is normal), $\bar{y}$ = 44.2, σ is unknown, s = 3.8. Find a 90% confidence interval for μ.

Example 07: Suppose, $n = 16$, σ is unknown, 80% confidence interval for μ was (48,54). What was $\bar{y}$ and s?

Example 08: Suppose, σ is unknown, $n = 75$, construct a 95% confidence interval for μ.

CH 4.2 (Hypothesis Testing) Handout
Md Mutasim Billah

Example 1: 3 balls will be drawn at random *with replacement* from basket A or Basket B; you don't know which one.

Suppose,

H_0: The balls are from basket A vs H_a: The balls are from basket B

Basket A	Basket B
9 red balls 1 green ball	1 red ball 4 green balls

Let us reject the null hypothesis (H_0) if we observe 3 green balls in a row. Find the probability of type I error (α) and probability of type II error (β).

Example 2:
Purchasing shipments of iron ore pellets:
μ= mean weight among all shipments being purchased (unknown)
σ = 10 pounds (known)
Supplier claims that "the average shipment weighs 80 pounds". You suspect you may be getting shortchanged.

H_0: $\mu = 80$ vs

Three options for H_a:

$\mu < 80$	Left-tailed or lower-tailed	CC: Decrease (↓)
$\mu > 80$	Right-tailed or upper-tailed	CC: Increase (↑)
$\mu \neq 80$	Two-tailed	CC: Change/Different

Choose the appropriate alternative hypotheses (H_a) and perform a hypothesis test.

One Sample Z-test for μ:

1) **Assumption:** Population Standard Deviation σ is known and sample mean $\bar{y}$ is normally distributed or approximately normal.
2) **Hypothesis:** Choose H_0, H_a, and α = Level of significance

H_0: $\mu = \mu_o$ vs

Three options for H_a:

$\mu < \mu_o$	Left-tailed or lower-tailed	CC: Decrease (↓)
$\mu > \mu_o$	Right-tailed or upper-tailed	CC: Increase (↑)
$\mu \neq \mu_o$	Two-tailed	CC: Change/Different

3) **Test Statistic:** $\boldsymbol{Z} = \frac{(\bar{y} - \mu o)}{\sigma/\sqrt{\boldsymbol{n}}}$
4) **Critical/Rejection Region:** (α = area, H_a = Location)

5) **Conduct Experiment & find Value of the test statistic.**

6) **Conclusion:** If test statistic falls in the rejection region, then reject H_0, otherwise, fail to reject H_0.

   Reject H_0: "We have significant evidence at level α that H_0 is false"
   Fail to reject H_0: "We do not have significant evidence at level α that H_0 is false; i.e. data is not significant".

   Sometimes, we also compute a confidence interval for μ after a rejection of H_0 to report the location of the plausible values for μ.

Chapter 6: Handout

Topic: Correlation Matrix

Example: Once upon a time is Illinois!

ID #				
Pop Quizzes				
Exam 1				
Exam 2				
	ID #	**Pop Quizzes**	**Exam 1**	**Exam 2**

Example: MA 3710 Data (Fall 2008)

HW Total (1st 6)			
Exam 1			
Exam 2			
	HW Total (1st 6)	**Exam 1**	**Exam 2**

Topic: Computing Correlation Coefficient (r)

Example: An Engineering student has designed a new electric clothes dryer. Maybe there is something special about this dryer, maybe it is energy efficient. However, the student is having a problem with the temperature setting (maybe there is a knob to select the temperature from the setting menu) and the actual temperature inside the dryer.

X: Temperature setting (F)

Y: Actual interior temperature during cycle (F)

The student has carefully measured several readings for X along with the values of Y. The dataset is below:

X	140	145	150	155	160
Y	170	174	172	178	176

a) Draw a scatterplot with the above dataset and interpret it.

b) Compute SS_{XX}, SS_{YY}, SS_{XY}, and r. Interpret the correlation coefficient. Does the interpretation of r match with the interpretation of the scatterplot you have drawn in part (a)?

X	**Y**			
140	170			
145	174			
150	172			
155	178			
160	176			
$\sum$**X** = 750	$\sum$**Y** = 870			

Chapter 4.4 (Inference for a single proportion): Handout
Md Mutasim Billah

Example 01: Confidence Interval for proportion

Filling one-gallon milk jugs. Given that: n = 100, Y = 80

(a) What proportion of jugs does our filling process successfully place at least one gallon of milk into?

(b) Construct a 95% confidence interval for p.

(c) If the sample had been 800 out of 1000, the 95% CI would have been _____________

*** If we want the error margin to equal some small value B, what sample size is necessary?

$$B = Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \rightarrow \boxed{n = \hat{p}\hat{q}\left(\frac{Z_{\alpha/2}}{B}\right)^2}$$

If you have no prior estimate of p, use $\hat{p} = .50$

(d) If we wanted a 99% CI for p to have an error margin of .01, what should n be?

(e) If we had no prior estimate of p and still we are interested to construct a 99% CI for p to have an error margin of .01, what should n be?

Example 02: Hypothesis Testing for proportion

A gambler has invented a new strategy for a game where players are supposed to win 49% of the time. He hopes his strategy will win more often. To test it, he plays the game 80 times and wins 44 of them.

(a) Choose the appropriate null and alternative hypotheses and perform a hypothesis test at α =.05.

(b) If the gambler had won 55 out of 100 games, the p-value would have been ____________

*** P-values for fixed $\hat{p} = 0.55$ with different sample size:

$$\hat{p} = .55$$

Y	n	Z	p-value
44	80	1.07	.1423
55	100	1.20	.1151
275	500	2.68	.0037
550	1000	3.80	$\approx \frac{1}{14000}$

Chapter 4.8 (Inference for variances): Handout
Md Mutasim Billah

Example 03: Hypothesis Testing for variance

Manufacturing sticks of dynamite – waiting time (seconds) for detonation is designed to be very close to 25, and obviously the variance should be very small. The manufacturer wants $\sigma^2 = 0.25$ sec ($\sigma = 0.50$) and is afraid that it might be too big.

Further Information: n= 20, $\overline{Y} = 25$, $s^2 = 0.2916$

Choose the appropriate null and alternative hypotheses and perform a hypothesis test at α =.01.

Handout: Chapter 6
Md Mutasim Billah

Topic: Correlation Matrix

Example 1: Once upon a time is Illinois!

ID #				
Pop Quizzes				
Exam 1				
Exam 2				
	ID #	**Pop Quizzes**	**Exam 1**	**Exam 2**

Example 2: MA 3710 Data (Fall 2008)

HW Total (1st 6)			
Exam 1			
Exam 2			
	HW Total (1st 6)	**Exam 1**	**Exam 2**

Topic: Computing Correlation Coefficient (r)

Example 3: An Engineering student has designed a new electric clothes dryer. Maybe there is something special about this dryer, maybe it is energy efficient. However, the student is having a problem with the temperature setting (maybe there is a knob to select the temperature from the setting menu) and the actual temperature inside the dryer.

X: Temperature setting (F)

Y: Actual interior temperature during cycle (F)

The student has carefully measured several readings for X along with the values of Y. The dataset is below:

X	140	145	150	155	160
Y	170	174	172	178	176

a) Draw a scatterplot with the above dataset and interpret it.

b) Compute SS_{XX}, SS_{YY}, SS_{XY}, and r. Interpret the correlation coefficient. Does the interpretation of r match with the interpretation of the scatterplot you have drawn in part (a)?

X	**Y**			
140	170			
145	174			
150	172			
155	178			
160	176			
$\sum$**X** = 750	$\sum$**Y** = 870			

c) Fit a simple linear regression model. What are the interpretation of the slope and the intercept?

d) If we set the temperature setting as X = 152 °F, on average, what can we expect the actual temperature (Y) to be?

e) Compute $\sum$ e and check whether the sums is equal to zero or not. Also compute $\sum e^2$.

	Y	$\widehat{Y} = 126 + 0.32X$	Errors/Residuals $Y-\widehat{Y} = e$	e^2
140	170			
145	174			
150	172			
155	178			
160	176			

f) Perform the hypothesis testing for β_1.

MA 3710 – Md Mutasim Billah
Regression Analysis: Preparing for Project 02

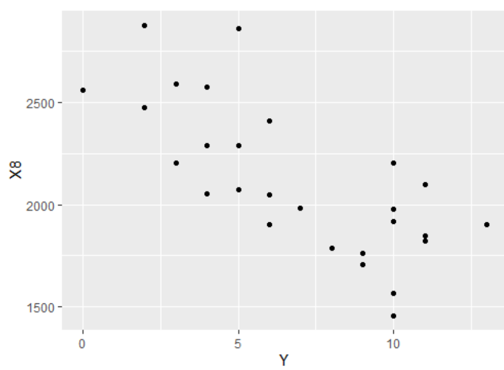
Simple Linear Regression Analysis:

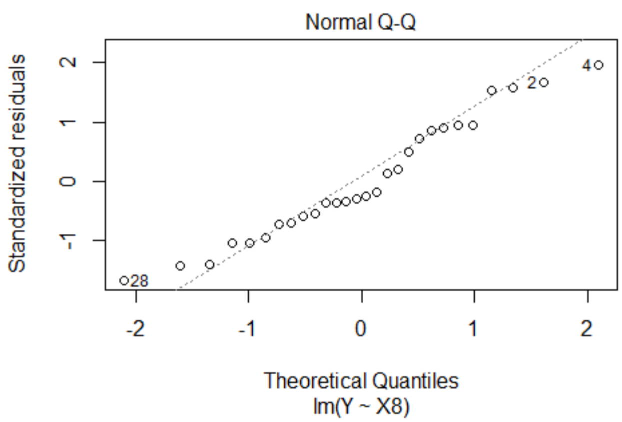
Dataset b1 gives data concerning the performance of the 28 National Football League teams in 1976. It is suspected that the number of yards gained rushing by opponents (X8) has an effect on the number of games won by a team (Y).

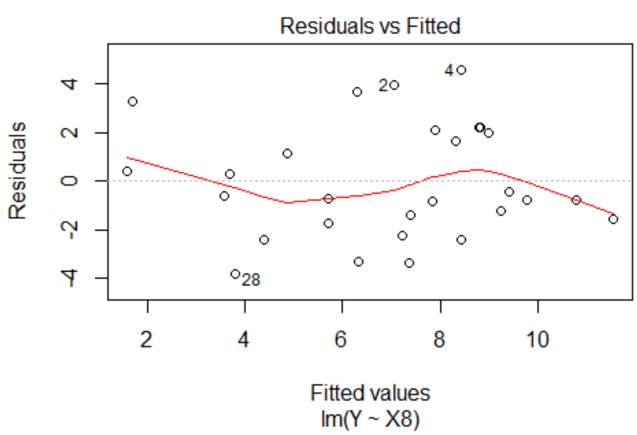
(a) Fit a simple linear regression model relating games won y to yards gained rushing by opponents X8. Interpret the intercept and slope. Do you think that the number of yards gained rushing by opponents has any influence on the number of games won by a team (i.e. x4 is a significant variable to y)? Hints: Answer this question in terms of the p-value of T statistic.

```
Call:
lm(formula = Y ~ X8)

Residuals:
   Min     1Q Median     3Q    Max 
-3.804 -1.591 -0.647  2.032  4.580 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 21.788251   2.696233   8.081 1.46e-08 ***
X8          -0.007025   0.001260  -5.577 7.38e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.393 on 26 degrees of freedom
Multiple R-squared:  0.5447,	Adjusted R-squared:  0.5272 
F-statistic:  31.1 on 1 and 26 DF,  p-value: 7.381e-06
```

(a) Interpret the following figures:







(b) Construct the analysis-of-variance table and test for significance of regression. In general, do you think that the number of yards gained rushing by opponents has any influence on the number of games won by a team (i.e. x4 is a significant variable to y)?
Hints: Answer this question in terms of the p-value of F statistic.

```
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
X8         1 178.09 178.092  31.103 7.381e-06 ***
Residuals 26 148.87   5.726
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(c) What percent of the total variability in Y is explained by this model?