

Student: <u>Hà Thanh Nga</u>	Instructor: Lewis Njuaem	Assignment: Homework Assignment # 3
Date: <u>19.9.2023</u>	Course: ADMN 2100 (60) - APP BUS STATISTICS	NEUV

1. What does the Y-intercept (b_0) represent?

- ☒ A. The estimated average Y when $X = 0$
- ☐ B. The change in estimated average Y per unit change in X
- ☐ C. The variation around the sample regression line
- ☐ D. The predicted value of Y

2. A candy bar manufacturer is interested in trying to estimate how sales are influenced by the price of their product. To do this, the company randomly chooses 6 small cities and offers the candy bar at different prices. Using candy bar sales as the dependent variable, the company will conduct a simple linear regression on the data below.

<u>City</u>	<u>Price (\$)</u>	<u>Sales</u>
A	1.30	100
B	1.60	90
C	1.80	90
D	2.00	40
E	2.40	38
F	2.90	32

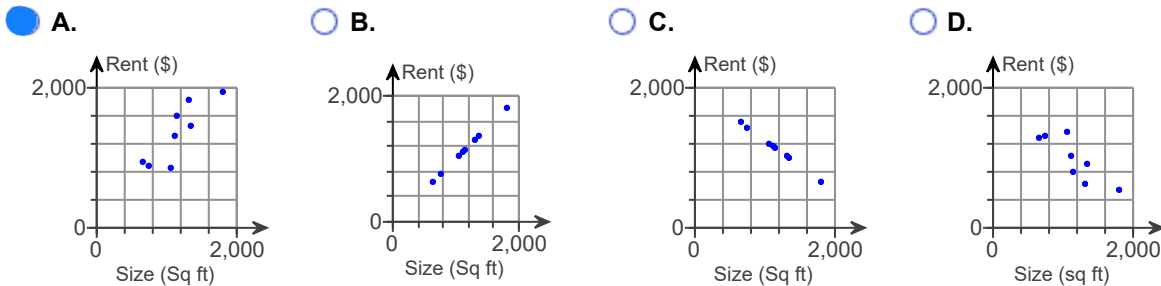
What is the estimated mean change in the sales of the candy bar if price goes up by \$1.00?

- ☐ A. - 3.810
- ☐ B. 0.784
- ☐ C. 161.386
- ☒ D. - 48.193

3. An agent for a residential real estate company in a suburb located outside a major city has the business objective of developing more accurate estimates of the monthly rental cost for apartments. Toward the goal, the agent would like to use the size of an apartment, as defined by square footage to predict the monthly rental cost. The agent selects a sample of 8 one-bedroom apartments and the data are shown. Complete parts (a) through (f).

Monthly Rent (\$)	900	1,450	850	1,600	1,950	950	1,825	1,300
Size (Square Feet)	750	1,350	1,050	1,150	1,800	650	1,300	1,100

a. Construct a scatter plot. Choose the correct graph below.



b. Use the least-squares method to determine the regression coefficients b_0 and b_1 .

$$b_0 = 200.9 \text{ (Round to one decimal place as needed.)}$$

$$b_1 = 1.0079 \text{ (Round to one decimal place as needed.)}$$

c. Interpret the meaning of b_0 and b_1 in this problem. Choose the correct answer below.

- ☒ **A.** For each increase of 1 square foot in space, the monthly rent is expected to increase by b_1 dollars. Since X cannot be zero, b_0 has no practical interpretation.
- ☐ **B.** For each increase of 1 square foot in space, the monthly rent is expected to increase by b_0 dollars. Since X cannot be zero, b_1 has no practical interpretation.
- ☐ **C.** For each increase of 1 square foot in space, the monthly rent is expected to increase by b_0 dollars. Apartments in this neighborhood cost at least b_1 dollars.
- ☐ **D.** For each increase of 1 square foot in space, the monthly rent is expected to increase by b_1 dollars. Apartments in this neighborhood cost at least b_0 dollars.

d. Predict the mean monthly rent for an apartment that has 1,000 square feet.

The predicted mean monthly rent for such an apartment is \$ 1204.64 .
(Round to the nearest cent as needed.)

e. Why would it not be appropriate to use the model to predict the monthly rent for apartments that have 500 square feet?

- ☐ **A.** The size of an apartment has no effect on the monthly rent, according to this model. There must be another factor that contributes to the rent price.
- ☐ **B.** An apartment with 500 square feet is outside the relevant range for the independent variable.
- ☐ **C.** The model predicts that the monthly rent for an apartment that has 500 square feet would be unrealistically low.
- ☒ **D.** The correlation between an apartment's size and its monthly rent is too weak to use this model for such a prediction.

f. Two people are considering signing a lease for an apartment in this neighborhood. They are trying to decide between two apartments, one with 1,000 square feet for a monthly rent of \$1,275 and the other with 1,200 square feet for a monthly rent of \$1,425. Based on (a) through (d), which apartment is a better deal?

Based on (a) through (d), the apartment with (1) _____ square feet is the better deal.

- (1) ☒ 1,200
☐ 1,000

4. The following Excel tables are obtained when "Score received on an exam (measured in percentage points)" (Y) is regressed on "percentage attendance" (X) for 22 students in a Statistics for Business and Economics course. Which of the following statements is true?

Regression Statistics	
Multiple R	0.142620229
R Square	0.02034053
Standard Error	20.25979924
Observations	22

	Coefficients	Standard Error	T Stat	P-value
Intercept	39.39027309	37.24347659	1.057642216	0.302826622
Attendance	0.340583573	0.52852452	0.644404489	0.526635689

- ☒ A. Approximately 2% of the total variability in score received can be explained by percentage attendance.
- ☐ B. Approximately 14.26% of the total variability in percentage attendance can be explained by score received.
- ☐ C. Approximately 2% of the total variability in percentage attendance can be explained by score received.
- ☐ D. Approximately 14.26% of the total variability in score received can be explained by percentage attendance.

5. If $SSR = 24$ and $SSE = 16$, determine SST , then compute the coefficient of determination, r^2 , and interpret its meaning.

$SST =$

$r^2 =$ (Type an integer or a decimal. Do not round.)

Interpret the meaning of r^2 . Choose the correct answer below.

- ☒ A. It means that $r^2 \cdot 100\%$ of the variation in the dependent variable can be explained by the variation in the independent variable.
- ☐ B. It means that $1 - r^2$ of the variation in the dependent variable cannot be explained by the variation in the independent variable.
- ☐ C. It means that r^2 of the variation in the independent variable can be explained by the variation in the dependent variable.
- ☐ D. It means that $(1 - r^2) \cdot 100\%$ of the variation in the independent variable cannot be explained by the variation in the dependent variable.

6. Annual revenues were used to predict the current value of a MLB baseball team, using the accompanying data for a sample of 28 teams. The data are displayed in the accompanying table, and the regression equation and other values are shown below. Use this information to complete parts (a) through (c) below.

$$\hat{Y}_i = -1045.6274 + 8.5930X_i, \quad \sum_{i=1}^n Y_i = 43,505, \quad \sum_{i=1}^n Y_i^2 = 81,761,625, \quad \sum_{i=1}^n X_i Y_i = 14,744,525$$

¹ Click the icon to view the data table.

- a. Determine the coefficient of determination, r^2 , and interpret its meaning.

$r^2 =$ (Round to four decimal places as needed.)

What is the meaning of the coefficient of determination?

- ☐ A. It measures the variability in the actual annual revenue from the predicted annual revenue.
- ☐ B. It measures the variability in the actual franchise value from the predicted waiting times.
- ☐ C. It is the proportion of the variation in annual revenue that is explained by the variability in franchise value.
- ☒ D. It is the proportion of the variation in franchise value that is explained by the variability in annual revenue.

- b. Determine the standard error of the estimate.

$S_{YX} =$ (Round to four decimal places as needed.)

- c. How useful do you think this regression model is for predicting the value of a major sports franchise?

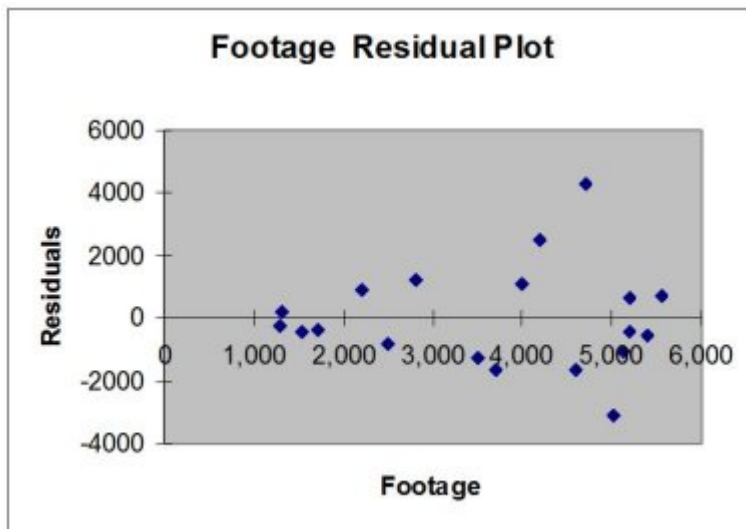
Since the value of r^2 is (1) _____ the regression model (2) _____ for predicting the value of a major sports franchise.

1: Data Table

Team	Revenue	Value
Boston	434	2700
Chicago White Sox	269	1350
Cleveland	271	920
Detroit	275	1200
Houston	299	1450
Kansas City	246	950
Los Angeles Angels	350	1750
Minnesota	249	1025
New York Yankees	526	3700
Oakland	216	880
Seattle	289	1400
Tampa Bay	205	825
Texas	298	1550
Toronto	278	1300
Arizona	253	1350
Atlanta	275	1500
Chicago Cubs	434	2650
Cincinnati	229	915
Colorado	248	1000
Los Angeles Dodgers	462	2750
Miami	206	940
Milwaukee	239	925
New York Mets	332	2000
Philadelphia	325	1650
Pittsburgh	265	1250
St. Louis	310	1800
San Diego	259	1125
San Francisco	428	2650

- (1) ☐ close to 0, (2) ☒ should be very useful
☒ close to 1, ☐ is not very useful
☐ equal to 0,
☐ equal to 1,
-

7. Based on the residual plot below, you can conclude that there might be a violation of which of the following assumptions?



- ☐ A. Independence of errors
- ☐ B. Homoscedasticity
- ☐ C. Normality of errors
- ☒ D. Linearity of the relationship

8. What do the residuals represent?

- ☐ A. The predicted value of Y for the average X value
- ☐ B. The difference between the actual Y values and the mean of Y
- ☐ C. The square root of the slope
- ☒ D. The difference between the actual Y values and the predicted Y values

9. A box office analyst seeks to predict opening weekend box office gross for movies. Toward this goal, the analyst plans to use YouTube trailer views as a predictor. For each of 65 movies, the YouTube trailer view count, the number of YouTube trailer views from the release of the trailer through the Saturday before a movie opens, and the opening weekend box office gross (in \$millions) are provided. The simple linear regression model for predicting movie weekend box office gross based on YouTube trailer views is $\hat{Y}_i = -1.009 + 1.408X_i$. Perform a residual analysis for these data. Based on these results, evaluate whether the assumptions of regression have been seriously violated.

² Click the icon to view the movie data.

Which of the assumptions of regression, if any, have been seriously violated? Select all that apply.

- ☐ A. The assumption of linearity has been violated because the data are clearly curvilinear.
- ☐ B. The assumption of equal variance has been violated because the variability of the residuals is much greater when the views are larger.
- ☐ C. The assumption of independence has been violated because errors from later time periods appear to be related to those from earlier time periods.
- ☒ D. The assumption of normality has been violated because the normal probability plot does not appear to be a straight line.
- ☐ E. The assumptions of linearity, independence, normality, and equal variance do not appear to have been seriously violated.

2: Movie Data

Movie	Opening Weekend Gross (\$mil)	YouTube Trailer Views (millio
The Mummy	32.246	57.897
It Comes At Night	6.001	10.785
Megan Leavey	3.768	10.099
Captain Underpants: The First Epic Movie	23.852	8.725
Wonder Woman	103.251	84.205
Pirates of the Caribbean: Dead Men Tell No Tales	62.983	34.99
Baywatch	18.504	21.764
Everything, Everything	11.727	5.55
Diary of a Wimpy Kid: The Long Haul	7.126	3.836
Alien: Covenant	36.161	45.615
Snatched	19.542	7.791
King Arthur: Legend of the Sword	15.371	28.187
Lowriders	2.404	4.496
Guardians of the Galaxy Vol. 2	146.51	57.324
How to Be a Latin Lover	12.252	7.394
The Circle	9.034	11.145
Sleight	1.702	11.175
Born in China	4.79	0.508
Free Fire	0.994	1.061
Unforgettable	4.785	5.387
The Promise	4.096	6.354
Phoenix Forgotten	1.816	7.714
The Fate of the Furious	98.787	30.87
The Case for Christ	3.968	0.28
Smurfs: The Lost Village	13.21	8.124
The Boss Baby	50.199	52.292
The Zookeeper's Wife	3.289	3.886
Ghost in the Shell	18.676	31.055
CHIPS	7.723	7.081
Life	12.502	13.55

Power Rangers	40.3	59.296
Beauty and the Beast	174.751	80.077
The Belko Experiment	4.137	5.546
Kong: Skull Island	61.025	35.309
The Shack	16.172	2.532
Logan	88.412	44.196
Before I Fall	4.69	4.989
Get Out	33.377	6.63
Rock Dog	3.705	0.942
Collide	1.513	2.258
The Great Wall	18.47	11.327
Fist Fight	12.202	8.966
A Cure for Wellness	4.357	15.177
John Wick: Chapter 2	30.436	13.714
The LEGO Batman Movie	53.003	31.231
Fifty Shades Darker	46.607	52.612
Rings	13.003	16.235
The Space Between Us	3.776	6.884
A Dog's Purpose	18.223	11.698
Gold	3.471	2.827
Resident Evil: The Final Chapter	13.602	23.075
Split	40.011	12.606
20th Century Women	1.385	0.826
xXx: Return of Xander Cage	20.13	27.536
The Founder	3.404	7.273
The Resurrection of Gavin Stone	1.207	3.323
Monster Trucks	10.951	4.267
Sleepless	8.344	3.79
Patriots Day	11.614	7.597
The Bye Bye Man	13.501	12.912
Live By Night	5.106	7.067
Silence	1.985	5.02
Hidden Figures	22.8	7.739
Underworld: Blood Wars	13.689	16.795
A Monster Calls	2.08	7.643

10. What does it mean if b_1 is computed to be + 2.0 in a multiple regression problem involving two independent variables?

- ☐ A. It means that the relationship between X_1 and Y is significant.
- ☐ B. It means that the estimated mean of Y is 2 when X_1 equals zero.
- ☐ C. It means that the estimated mean of Y increases by 2 units for each increase of 1 unit of X_1 , without regard to X_2 .
- ☒ D. It means that the estimated mean of Y increases by 2 units for each increase of 1 unit of X_1 , holding X_2 constant.

11. A professor of industrial relations believes that an individual's wage rate at a factory (Y) depends on his performance rating (X_1) and the number of economics courses the employee successfully completed in college (X_2). The professor randomly selects 6 workers and collects the following information. An employee who took 12 economics courses scores 10 on the performance rating. What is her estimated expected wage rate?

Employee	Y (\$)	X_1	X_2
1	10	3	0
2	12	1	5
3	15	8	1
4	17	5	8
5	20	7	12
6	25	10	9

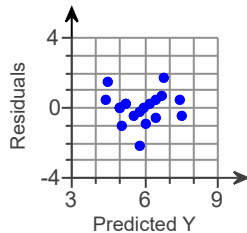
-
- ☐ A. 12.20
- ☒ B. 24.87
- ☐ C. 25.70
- ☐ D. 10.90
-

12. The production of wine is a multi-billion-dollar worldwide industry. In an attempt to develop a model of wine quality as judged by wine experts, data was collected from red wine variants from a particular type of foreign wine. The multiple linear regression model $\hat{Y}_i = -0.48 + 0.719X_{1i} - 15.044X_{2i}$ was developed from a sample of 20 wines, where Y_i is the wine quality, measured on a scale from 0 (very bad) to 10 (excellent) based on the alcohol content (%), X_{1i} , and the amount of chlorides, X_{2i} . Use the regression model and the accompanying data to complete parts (a) through (e) below.

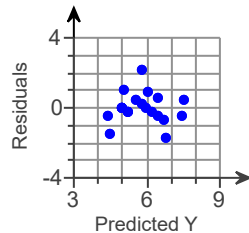
³ Click the icon to view the table of data.

a. Plot the residuals versus \hat{Y}_i . Choose the correct graph below.

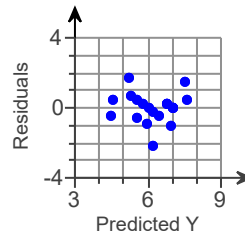
☐ A.



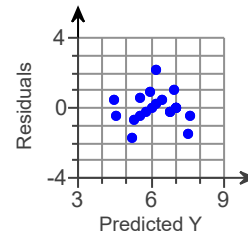
☒ B.



☐ C.

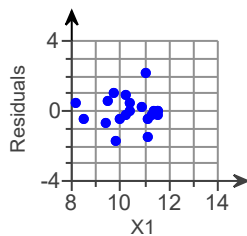


☐ D.

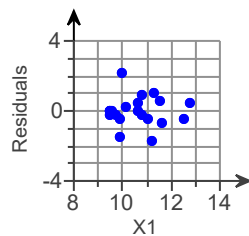


b. Plot the residuals versus X_{1i} . Choose the correct graph below.

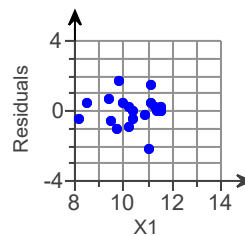
☐ A.



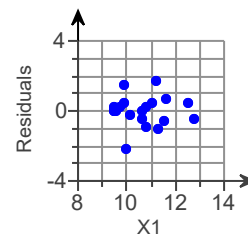
☒ B.



☐ C.

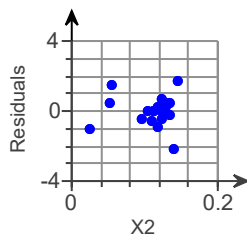


☐ D.

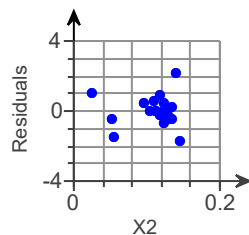


c. Plot the residuals versus X_{2i} . Choose the correct graph below.

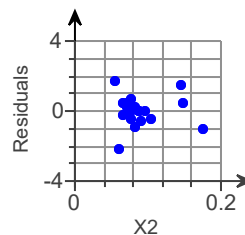
☐ A.



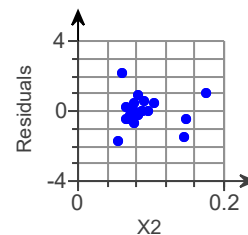
☐ B.



☐ C.



☒ D.



d. In the residuals plots created in parts (a) through (c), is there any evidence of a violation of the regression assumptions?

- ☐ A. Yes, because there appears to be a pattern in at least one of the residual plots.
- ☐ B. No, because there appears to be a pattern in at least one of the residual plots.
- ☐ C. Yes, because the residuals appear randomly distributed in all the plots.
- ☒ D. No, because the residuals appear randomly distributed in all the plots.

e. Should you compute the Durbin-Watson statistic for these data? Explain

- ☐ A. Yes, because the data were collected in a specific time-order, so the Durbin-Watson test statistic should be computed.
- ☐ B. No, because the data were collected in a specific time-order, so the Durbin-Watson test

statistic should not be computed.

- ☐ C. Yes, because there is no indication that the data were collected in a specific time order.
- ☒ D. No, because there is no indication that the data were collected in a specific time order.

3: Wine Quality, Alcohol Content, and Chloride Data

Wine Quality	Alcohol Content	Amount of Chlorides
8.0	12.8	0.078
6.0	10.8	0.07
5.0	9.5	0.073
5.0	9.7	0.082
5.0	9.6	0.097
7.0	11.5	0.091
5.0	11.2	0.055
6.0	11.3	0.174
5.0	9.5	0.089
7.0	12.5	0.072
6.0	11	0.067
6.0	11.6	0.077
4.0	9.9	0.147
6.0	10.6	0.105
8.0	10	0.061
3.0	9.9	0.145
6.0	10.1	0.067
6.0	10.6	0.077
7.0	10.8	0.081
5.0	9.5	0.089