**Correlation and Regression Practice**

Solutions are at the end of this document.

**Part 1. Correlation**

Interpret the correlation coefficients below. The correlation between ….

1. sales rep experience and sales rep job performance = .39
2. age and hours of sporting events watched on television = -.07
3. number of days before Christmas and shopping mall traffic in Nov and Dec = -.42
4. foot traffic outside a donut shop and donut shop sales = .53
5. age and attendance at live music events = -.49
6. number of page views on an online retail site and size of promotional discount = .22

**Part 2. Simple regression**

1. The owner of a fitness center wants to know if younger or older members use the fitness center facilities more often.

• dependent variable = frequency of fitness center visits (visits per month)

• independent variable = client age (in years)

Here’s the regression model:

Visits per month = 12.19 - .21 \* age

R = .39

1. To what extent is satisfaction with an online clothing retailer associated with future purchases from the retailer?

In April, an online women’s clothing retailer surveyed customers about their satisfaction with their purchases from the company, using a 7-point satisfaction scale (1=completely dissatisfied, 7=completely satisfied). Purchases were then tracked for the following 12 months.

• dependent variable = dollar amount of purchases for 12 months beginning in May

• independent variable = satisfaction measured in April

Here’s the regression model:

Purchases (in dollars) = 131.56 + 61.91 \* satisfaction rating

R = . 44

**Part 3. Permissible variables in multiple regression**

1. The sales manager at a company that sells cement pipe to commercial construction companies wants to know what variables are associated with salesperson success. The manager has access to the following variables about each sales representative:
2. employee ID number
3. how long the sales rep has been employed by the company
4. how many clients each sales rep serves
5. the sales region for each rep (regions are numbered 1 through 14 based on geographic location)
6. most recent performance rating (100-point rating scale)
7. age
8. gender
9. size of the accounts the sales rep handles (small, medium, or large)

Indicate which variables are appropriate to enter into a multiple regression model (based on their level of measurement).

**Part 4. Multiple regression interpretation**

1. The regression model below was developed to identify the characteristics of successful retail locations for a chain of sandwich shops. Data came from internal company records.

dependent variable: sales per square foot

independent variables:

• average foot traffic per day in thousands of passersby (000)

• distance to nearest competitor shop (in city block units, where a city block is 320 feet long)

• employment rate for the city in which the shop is located (percentage of the labor force in the city

that has a job)

regression model:

sales per square foot = 58.06 + 12.77 (foot traffic) + 10.95 (competitor distance) + .19 (% employed)

R = . 51

1. The regression model below examines the relationship between the following variables and the amount of video streaming service use.

dependent variable: number of hours per week a person uses a video streaming service

independent variables:

• hours of paid employment per week

• live music event attendance (1 = has attended within the last month, 0= has not attended; binary)

• movie theater attendance (1 = has attended within the last month, 0= has not attended; binary)

regression model:

video streaming hours = 4.28 - .11 (employment hours) + .32 (if attended music event) + 1.31 (if

went to movie theater)

R = . 37

**SOLUTIONS**

**Part 1. Correlation**

All correlation coefficients in these examples (except #2) show low to moderate degrees of association, which is typical in analysis of marketing variables.

1. sales rep experience and sales rep job performance = .39

sales reps with more experience had higher job performance; those with less experience had lower job performance

1. age and hours of sporting events watched on television = -.07

this correlation is so low that we can say there is essentially now relationship between age and watching sporting events; older and younger people watch sporting events on television in the same amounts

1. number of days before Christmas and shopping mall traffic in Nov and Dec = -.42

the farther Christmas is away (more days before Christmas), the lower the shopping mall traffic

1. foot traffic outside a donut shop and donut shop sales = .53

donut shops with more foot traffic outside have higher sales

1. age and attendance at live music events = -.49

younger people attend more live music events; older people attend fewer

1. number of page views on an online retail site and size of promotional discount = .22

pages with higher promotional discounts had more page views

**Part 2. Simple regression**

1. Each one year increase in a gym member’s age is associated with .21 fewer visits to the gym per month.

A member who is 40 visits the gym 2.1 fewer times per month, on average, than a member who is 30.

15.2% (or .392) of the variation in gym visits is explained by members’ age.

The intercept of 12.19 is not directly interpretable because an age of 0 is not within the relevant range.

1. Each one point increase in the satisfaction rating a customer gives the retailer is associated with $61.91 of additional purchases from the retailer in the following 12-month period.

19.4% (or .442) of the variation in customer purchases explained by customers’ satisfaction level.

The intercept of 131.56 is not directly interpretable because the satisfaction rating scale doesn’t include a value of 0.

1. Permissible variables
2. employee ID number: not permissible; these numbers are arbitrary and don’t represent a quantity
3. how long the sales rep has been employed by the company: permissible (ratio level of measurement)
4. how many clients each sales rep serves: permissible (ratio level of measurement)
5. the sales region for each rep (regions are numbered 1 through 14 based on geographic location): not permissible; these numbers are arbitrary and don’t represent a quantity
6. most recent performance rating (100-point rating scale): permissible (interval level of measurement)
7. age: permissible (ratio level of measurement)
8. gender: permissible (binary variable in this situation)
9. size of the accounts the sales rep handles (small, medium, or large): not permissible (ordinal level of measurement)
10. Multiple regression model:

sales per square foot = 58.06 + 12.77 (foot traffic) + 10.95 (competitor distance) + .19 (% employed)

Each 1,000 increase in foot traffic is associated with a $12.77 increase in sales per square foot, on average.

Each increase of one city block in distance to the nearest competitor sandwich shop is associated with an increase of $10.95 in sales per square foot.

Each percentage increase in employment is associated with an increase of $0.19 in sales per square foot.

The intercept of 58.06 is not directly interpretable because there were no shops with 0 on the foot traffic, competitor distance, and employment variables.

This model explains 26.0% (.512) of the variation in sandwich store sales.

1. Multiple regression model:

video streaming hours = 4.28 - .11 (employment hours) + .32 (if attended music event) + 1.31 (if

went to movie theater)

Each hour of paid employment per week is associated with .11 fewer hours using a video streaming service, on average (for example, a person working 40 hours per week would use a streaming service 4.4 fewer hours per week than some who is not working).

People who attended a live music event within the last month used a video streaming service for .32 hours more per week, on average, than people who did not attend a live music event.

People who went to a movie theater within the last month used a video streaming service for 1.31 hours more per week, on average, than people who did not go to a movie theater.

The intercept of 4.28 indicates that people who were not employed, did not attend a music event, and did not go to a movie theater, used a video streaming service, on average, 4.28 hours per week.

This model explains 13.7% (.372) of the variation in video streaming service use.