ONLINE TEST BANK

ELLEN BREAZEL
Clemson University

STATISTICS FOR BUSINESS: DECISION MAKING AND ANALYSIS

SECOND EDITION

Robert Stine
The Wharton School of the University of Pennsylvania

Dean Foster
The Wharton School of the University of Pennsylvania

Part 1: Variation   Test A

Place your answer in the space provided.

An online distributor of healthcare products maintains data on all transactions. The information that is recorded is shown in the sample transaction record below:

Date of purchase	Customer Name	Customer #	Zip Code	Product #	Amount of Purchase
01/03/2010	Robert Jones	12943762	84321	6387450	28.50

A new row is added to the data table for each transaction.

Section 2.2 – Categorical and Numerical Data
[Objective: Distinguish categorical from numerical variables.]
_____**1.** Which of the variables in the data table are numerical?

(a) Product Number **(b)** Customer Number **(c)** Amount of Purchase
(d) All of these variables are numerical.

Section 2.3 – Recoding and Aggregation
[Objective: Identify when recoding or aggregating data are useful]
_____**2.** At the end of each month, a new data table is generated that has a row for each day of the month, and columns labeled Date, Total Number of Transactions, and Total Amount Purchased. The production of this new data table is an example of:

(a) Re-coding the data **(b)** Aggregating the data **(c)** Use of a Likert Scale **(d)** Ordinal variables

Section 2.4 – Time Series
[Objective: Recognize time series data.]
_____**3.** Which of the following uses of the data would result in a time series?
(a) Summarizing the total number of transactions for each zip code.
(b) A summary showing the total amount purchased by all customers during a holiday sales event.
(c) A summary showing the total amount purchased on each Friday during the past year.
(d) A list of customer numbers for which no purchase was recorded during the past year.
(e) None of the above would represent a time series.

A manufacturer of toaster ovens keeps track of all ovens that are returned for warranty repair. Each oven is classified by the cause of the problem resulting in the return. Within each category, the total number of ovens returned is recorded, and the total number that the manufacturer was able to repair is recorded (ovens that cannot be repaired are replaced with a new unit). The results for a given month are provided:

Cause	Number Returned	Number Repaired
Faulty electrical component	200	195
Faulty mechanical component	230	220
Improper assembly	320	205
Cosmetic defect	350	310
Other	100	90

Section 3.1 – Looking at Data
[Objective: Create, describe, and interpret the distribution of a categorical variable and link this distribution to variation.]

_____**4.** What percentage of all ovens are returned due to faulty components? (Round your percentage to one decimal place.)

(a) 16.7% **(b)** 35.8% **(c)** 19.2% **(d)** 43.0% **(e)** 47.8%

Section 3.2 – Charts of Categorical Data
[Objective: Choose an appropriate plot that shows the distribution of a categorical variable]

_____**5.** A Pareto Chart with vertical bars is to be constructed for the Number Repaired. What is the appropriate leftmost category for the chart?

(a) Faulty electrical component **(b)** Faulty mechanical component **(c)** Improper assembly
(d) Cosmetic defect **(e)** Other

_____**6.** Which of the following would be the best choice for showing the *proportion* of ovens returned for each type of defect?

(a) Boxplot **(b)** Bar chart **(c)** Pie chart **(d)** IQR **(e)** Contingency table

The operations manager of a bank has been monitoring the time required to assist customers who use the bank's drive-up facility. Over a one-month period, 525 customers using the facility were selected at random and the time to assist the customer was recorded. The histogram for the service times was nearly bell-shaped and symmetric. A summary of the data is given below:

Summary	**Time (minutes)**
Mean	6.8
Standard deviation	.85
Lower quartile	6.2
Median	6.9
Upper quartile	7.5

Section 4.4 – Shape of Distribution
[Objective: Use the empirical rule to link the mean and standard deviation to the concentration of data in a bell-shaped histogram.

_____**7.** Based on this data, the Empirical Rule indicates that 95% of all customer assistance times will be in which of the following intervals?

(a) [5.1, 8.5] **(b)** [5.95, 7.65] **(c)** [4.3, 9.5] **(d)** [4.2, 9.4] **(e)** [5.2, 8.6]

Section 4.3 – Boxplot
[Objective: Calculate, interpret, and contrast the interquartile range (IQR) and the standard deviation (SD).]

_____**8.** Which of the following statements is accurate concerning the summarized service times:
(a) 50% of the service times recorded are greater than 6.8 minutes.
(b) The middle 50% of the service times are in the interval [5.95, 7.65].
(c) The smallest 25% of the service times recorded are less than 5.95 minutes.
(d) The IQR for the service times is 1.7 minutes.
(e) The largest 25% of the service times recorded are greater than 7.5 minutes.

[Objective: Interpret a boxplot and link it to the distribution.]

_____**9.** A boxplot is to be constructed for the service times that were recorded. Which of the following service times would be considered high outliers for the boxplot?

(a) 9.1 minutes **(b)** 9.2 minutes **(c)** 9.4 minutes **(d)** 9.6 minutes
(e) All of (a)-(d) would be outliers in the boxplot.
(f) None of (a)-(d) would be outliers in the boxplot.

A manufacturer of digital cameras uses 3 different assembly plants. The manufacturer also maintains a repair facility for cameras that are returned for warranty repair. At the repair facility, any camera that is returned as defective is classified by the plant at which it was produced, and whether the cause of the defect is due to improper assembly or a defective component. The contingency table below summarizes the information on the cameras sent to the repair facility over the previous month.

			Plant		
		1	**2**	**3**	**Total**
Cause of defect	**Improper assembly**	120	140	110	370
	Faulty component	130	160	70	360
	Total	250	300	180	730

Section 5.1 – Contingency Tables

[Objective: Connect marginal distributions of a contingency table to bar charts and distributions of a single categorical variable]

_____**10**. Which plant accounts for the highest percentage of all the improperly assembled cameras?
(a) Plant 1 **(b)** Plant 2 **(c)** Plant 3
(d) All plants account for an equal percentage of the improperly assembled cameras.
(e) The percentages cannot be determined from the given information.

[Objective: Choose between row and column percentages to illustrate the presence of association between categorical variables]

_____**11.** Which plant has the largest percentage of improperly assembled cameras among its defectives?
(a) Plant 1 **(b)** Plant 2 **(c)** Plant 3
(d) All plants have an equal percentage of improperly assembled cameras.
(e) The percentages cannot be determined from the given information.

Section 5.3 –Strength of Association

[Objective: Calculate and interpret measures of association for categorical variables.]

_____**12.** Using the data in the table, it is determined that χ^2=10.52 and Cramer's V=.12. Which of the following statements is an appropriate interpretation of these results?
(a) There is no association between the cause of the defect and the plant in which the camera was produced.
(b) There is a strong association between the cause of the defect and the plant in which the camera was produced.
(c) There is some association between the cause of the defect and the plant in which the camera was produced, but it is not very strong.
(d) The χ^2 value indicates that the expected number of defective cameras in each cell and the observed number of defective cameras in each cell are equal for every cell.
(e) None of these interpretations of the results are appropriate.

The production manager at a plant that produces light bulbs is attempting to determine if there is a relationship between the weekly production quota (Q) set each week by upper management (based on sales projections determined by expected demand), and the percentage (P) of defective bulbs that are produced each week (based on defective bulbs returned for replacement). Data gathered each week over the past 2 years resulted in the following summary, using Q as the explanatory variable, and P as the response variable:

	Q	**P**
Mean	1250	3.5%
Standard Deviation	60	0.8%

Correlation: r =.96

Section 6.3 – Measuring Association
[Objective: Calculate and interpret the amount of linear association using covariance and correlation.]

_____**13.** Which of the following is an appropriate interpretation of the correlation between Q and P?

(a) A given value for a production quota will allow us to accurately predict the percentage of defective bulbs 96% of the time.
(b) An increase in the production quota will be accompanied by an increase in the percentage of defective bulbs 96% of the time.
(c) The data suggest a strong linear association between production quota and the percentage of defective bulbs.
(d) Using P as the explanatory variable and Q as the response variable would result in a correlation of r = -.96.
(e) Larger production quotas cause higher percentages of defectives bulbs for a given week.

Section 6.4 – Summarizing Association with a Line
[Objective: Find the line that summarizes the linear association that is measured by the correlation and use it to predict one numerical variable from another.]

_____**14.** For a given week prior to a holiday, upper management is considering requesting a production quota of 1430 bulbs (note that this is 3 standard deviations above the mean production quota). Based on the summary data above, what would be the predicted percentage of defective bulbs at this production level?

(a) 2.880% **(b)** 3.680% **(c)** 6.380% **(d)** 5.900% **(e)** 5.804%

Section 6.5 – Spurious Correlation
[Objective: Distinguish association from causation.]

_____**15.** The production manager says the following to you: “Based on the results, it appears that an increase in the weekly production quota causes an increase in the percentage of defective bulbs produced. Do you agree with that?” Which of the following would be appropriate answers to her question using a proper interpretation of the data and the concept of association between variables?

(a) Yes, I agree because the correlation between Q and P is close to 1.
(b) Yes, I agree because the correlation between Q and P is positive.
(c) No, I do not agree because association between Q and P does not imply causation.
(d) No, I do not agree because there could be other factors that affect the percentage of defective bulbs that are produced.
(e) Both (a) and (b) are appropriate responses.
(f) Both (c) and (d) are appropriate responses.

The following data were collected from a survey of 10 randomly selected college students.

Student ID	**Facebook User**	**# hours of study per week**
244701130	yes	8
302896051	no	5
734077249	yes	11
891072704	yes	5
730265917	yes	9
894866913	no	6
644678646	no	1
369417477	yes	1
388511718	yes	2
554470987	no	1

[Section 4.1 - Summaries of Numerical Variables]

______**16.** Find the mean number of hours per week this sample of students studies.
(a) 5 **(b)** 1 **(c)** 4.9 **(d)** 5.4

______**17.** Find the median number of hours per week this sample of students studies.
(a) 5 **(b)** 1 **(c)** 4.9 **(d)** 5.4

______**18.** Find the standard deviation of the number of hours per week this sample of students studies.
(a) 3.45 **(b)** 3.63 **(c)** 11.90 **(d)** 13.18

______**19.** True or False: The 85th percentile of a dataset is the data value such that 85% of the dataset is greater than that value and 15% of the dataset is less than that value.
(a) True **(b)** False

______**20.** True or False: The 50th percentile of a data set has to be a value in the data set.
(a) True **(b)** False

Answer Key

1. C
2. B
3. C
4. B
5. D
6. C
7. A
8. E
9. D
10. B
11. C
12. B
13. C
14. E
15. F
16. C
17. A
18. B
19. B
20. B

Part 1: Variation Test B

Place your answer in the space provided.

An auto parts supply store maintains a record of each transaction. The information below shows two sample transactions and the data that is recorded:

Date of Purchase	**Customer Name**	**Item #**	**Quantity**	**Unit Price ($)**	**Total Cost ($)**
01/05/2010	Tom Smith	13587	1	87.50	87.50
01/05/2010	EZ Auto Repair	21543	2	47.90	95.80

A new row is added to the data table for each transaction.

Section 2.2 – Categorical and Numerical Data
[Objective: Distinguish categorical from numerical variables.]
_____**1.** The 'Item #' is an example of which type of variable?

(a) Numerical variable **(b)** Ordinal variable **(c)** Categorical variable **(d)** Interval variable
(e) Both (a) and (d)

Section 2.3 – Recoding and Aggregation
[Objective: Identify when recoding or aggregating data are useful]
_____**2.** The manager of the store decides to create a new column from the column labeled 'Customer Name'. The column will be labeled 'Customer Type' and will have 3 possible entries: 'I' if the customer is an individual consumer; 'C' if the customer is a commercial repair shop; and 'G' if the customer is a government agency. The production of this new column from the existing data is an example of:

(a) Re-coding the data **(b)** Aggregating the data **(c)** Use of a Likert Scale
(d) Use of a categorical variable **(e)** The new column involves both (a) and (d)

_____**3.** At the end of the month, the manager produces a table with columns labeled 'Date' and 'Total Sales Revenue,' and a row for each day of the month. A graph is produced showing the daily sales revenues in chronological order. Which of the following techniques and concepts have not been used in this process?

(a) Aggregating the data **(b)** Frequency of measurement **(c)** Time series **(d)** Timeplot
(e) All of the techniques and concepts in (a)-(d) have been used.
(f) Only the techniques and concepts in (c) and (d) have been used.

A retailer of outdoor equipment and sportswear provides only these methods of purchasing items: Internet, phone, mail-order, and in-store. A survey of customers that have made purchases in the last month asked what methods of purchase the customers used. The percentage of customers who indicated each method of purchase are given below:

Method	**Percentage**
Internet	24%
Phone	32%
Mail-order	18%
In-store	38%

Section 3.2 – Charts of Categorical Data
[Objective - Choose an appropriate plot that shows the distribution of a categorical variable]

_____**4.** The manager would like to summarize the data using a pie chart. The assistant manager has given him several reasons why a pie chart should not be used:

1. The categories are not mutually exclusive.
2. A customer could have used more than one method of purchase during the month.
3. There is no "Other" category included.
4. The sum of the percentages exceeds 100%.

Which of these reasons are valid reasons for why a pie chart is not appropriate for summarizing this data?

(a) Only 1 and 2 **(b)** Only 3 **(c)** Only 1, 2, and 4 **(d)** Only 1, 2, and 3
(e) All of the reasons provided are appropriate reasons why a pie chart should not be used.

_____**5.** The manager also creates a table summarizing the number of purchases by each method over the past month:

Method	**Number of Purchases**
Internet	450
Phone	510
Mail-order	320
In-store	730

Which of the following could be used to display the proportions of purchases by each method?

(a) Pie chart **(b)** Relative frequency table **(c)** Contingency table
(d) Boxplot of purchases **(e)** Both (a) and (b) could be used.

Section 4.1 – Summaries of Numerical Variables
[Objective - Calculate, interpret, and contrast the interquartile range (IQR) and the standard deviation (SD).]

_____**6.** The variation in the number of purchases by each method could be seen by summarizing the data above using which of the following:

(a) Relative frequency table **(b)** Contingency table **(c)** Boxplot of purchases **(d)** IQR
(e) Standard deviation of purchases

The consumer research department for a chain of retail grocery stores has collected data from a large sample of consumers in the areas where their stores are located. One piece of information that the consumers were asked to provide is how much money they spend each month on products that are advertised as being "health foods." One of the researchers provided the Five Number Summary for the amounts spent on these products: ($70, $270, $350, $390, $430).

Section 3.1 – Looking at Data
[Objective - Create, describe, and interpret the distribution of a categorical variable and link this distribution to variation.]

_____**7.** Which of the following measures cannot be determined from the Five Number Summary provided?

(a) The range **(b)** The IQR **(c)** The middle 50% of the amounts **(d)** The mean amount.
(e) All of these measures can be determined.

Section 4.3 – Boxplot
[Objective: Interpret a boxplot and link it to the distribution.
_____**8.** Which of the following would be true of the boxplot that is created using the Five Number Summary provided?

(a) The boxplot would indicate a low outlier.
(b) The boxplot would indicate a high outlier.
(c) The boxplot would indicate both a low and high outlier.
(d) The median value would be located in the center of the box.
(e) None of (a)-(d) are true.
(f) All of (a)-(d) are true.

Workers at a plant assemble electric power drills. Each worker is responsible for the entire assembly of a drill, and they must record the amount of time required for the assembly of each drill. The assembly times for all of the drills assembled in a given month are summarized below:

Summary	**Time (hours)**
Mean	10.20
Standard deviation	0.60
Median	10.40
IQR	0.92

Section 4.4 – Shape of a distribution
[Objective: Use the empirical rule to link the mean and standard deviation to the concentration of data in a bell-shaped histogram.]
_____**9.** A histogram of the times appeared bell-shaped and symmetric. The plant manager is interested in identifying any worker that has assembled a drill in an "exceptionally short amount of time." The manager defines this amount of time as being any time that is among the fastest 2.5% of all assembly times for all the workers. Using the Empirical Rule, what assembly times would be among the fastest 2.5% of all times?

(a) Only times of 9.48 hours or less **(b)** Any time between 9.0 and 10.2 hours
(c) Only times of 8.56 hours or less **(d)** Only times of 9.0 hours or less.
(e) Only times of 9.6 hours or less.

Section 4.2 – Histograms and distribution of numerical data
[Objective: Prepare, describe, and interpret a histogram that summarizes the distribution of a numerical variable.]
_____**10.** A bank's loan department is looking at the amounts of the loans made to individuals and small businesses over the last year. One loan officer makes the observation that in the summary of the data, the mean amount of all loans is much larger than the median amount for the loans. Which of the following would explain the fact that the mean amount is much larger than the median amount?

(a) The histogram of the loan amounts was heavily skewed to the right.
(b) The histogram of the loan amounts was heavily skewed to the left.
(c) The standard deviation of the loan amounts was very small.
(d) The IQR of the loan amounts was very large.
(e) The histogram of the loan amounts was nearly bell-shaped and symmetric.

The human resource director of a large corporation is using a contingency table to summarize the results of a random sample of 2500 employees. Each employee was classified by position as Clerical, Technical, and

Management. Then it was determined if the employee had filed more than $5,000 in health care claims in the previous year. The results are provided below:

		Position			
		Clerical	**Technical**	**Management**	**Total**
Over $5k in claims	**No**	240	860	55	1155
	Yes	465	720	160	1345
	Total	705	1580	215	2500

Section 5.1 – Contingency Tables
[Objective: Connect marginal distributions of a contingency table to bar charts and distributions of a single categorical variable]

_____**11.** Among all the employees in the sample, which position is least likely to file more than $5,000 in health care claims?

(a) Clerical **(b)** Technical **(c)** Management
(d) All positions are equally likely to file more than $5000 in health care claims.

[Objective: Link conditional distributions in a table to stacked bar charts and mosaic plots]

______**12.** Among all the employees who filed more than $5,000 in health care claims, what percentage were filed by employees in Management positions? (Round the percentage to one decimal place.)

(a) 6.4% **(b)** 74.4% **(c)** 11.9% **(d)** 8.6% **(e)** 16.0%

Section 5.3 – Strength of Association
[Objective: Calculate and interpret measures of association for categorical variables.]

_____**13.** If we assume there is no association between the amount of health care claims filed by an employee and the employee's position, how many employees would we expect to find in a sample of 2500 employees that are in clerical positions and filed more than $5,000 in health care claims? (Round your answer to one decimal place.)

(a) 250.2 **(b)** 131.1 **(c)** 352.5 **(d)** 379.3 **(e)** 416.7

A drug manufacturer has sales representatives working out of offices all over the country. The corporate vice president for sales has collected data from a large sample of sales representatives to see if there appears to be an association between the number of years of experience in drug sales and the monthly sales amount generated by the representative. The VP uses the most recent monthly sales amount for the analysis. The summary of the data is provided below:

	Years Experience	**Monthly Sales**
Mean	8.6	$42,500
Standard Deviation	.8	$3,200

Correlation: r =.78

Section 6.4 – Summarizing Association with a line
[Objective: Find the line that summarizes the linear association that is measured by the correlation and use it to predict one numerical variable from another.]

_____**14.** The manufacturer is considering hiring a sales representative who has 6.6 years of experience in drug sales. Based on the summary data, what monthly sales volume would be predicted for this sales representative?

(a) $48,740 **(b)** $36,260 **(c)** $34,500 **(d)** $37508
(e) No prediction can be made unless $r = 1$ or $r = -1$.

_____**15.** Which of the following statements can you determine to be an appropriate interpretation of the summary of the data provided?

(a) Among all sales representatives in the sample, those with more than 8.6 years of experience had monthly sales greater than $42,500.
(b) 95% of all sales representatives in the sample had monthly sales between $36,100 and $48,900.
(c) Among the sales representatives in the sample with less than 8.6 years of experience, 78% of them had monthly sales less than $42,500.
(d) There is no evidence of a linear association between the years of experience and the monthly sales amounts for the sales representatives in this sample.
(e) None of these statements can be determined to be appropriate.

Suzanne is training for a marathon. Training consists of runs at various distances. Throughout her training, Suzanne has recorded the time and distance for each of her runs. Her collected data can be found below.

Distance (miles)	**Time (minutes)**
3.2	27
5.6	58
7.5	65
13.2	136
18.3	200

Section 4.1 – Summaries of Numerical Variables
[Objectives: Calculate, interpret, and contrast the mean and the median; Calculate, interpret, and contrast the interquartile range (IQR) and the standard deviation (SD).]

_____**16.** Find the mean and standard deviation of the distance that Suzanne ran.
(a) Mean = 7.5; standard deviation = 6.124
(b) Mean = 9.56; standard deviation = 6.124
(c) Mean = 9.56; standard deviation = 5.477
(d) Mean = 7.5; standard deviation = 6.543

______**17.** Find the mean and standard deviation of the time that Suzanne ran.
(a) Mean = 65; standard deviation = 78.675
(b) Mean = 65; standard deviation = 70.369
(c) Mean = 97.2; standard deviation = 62.57
(d) Mean = 97.2; standard deviation = 69.96

Section 6.1 – Scatterplots
[Objective: Recognize and describe the strength and direction of association between two numerical variables from a scatterplot, as well as tell whether there is any association between the variables.]

______**18.** In Suzanne's training, which variable is the explanatory variable and which is the response variable?
(a) Distance – explanatory; time – response
(b) Distance – response; time – explanatory

Section 6.3 – Measuring Association
[Objective: Calculate and interpret the amount of linear association using covariance and correlation.]

______**19.** The covariance between distance and time for Suzanne's data is 426.935. Using your calculations from questions 15 and question 16, what is the correlation between distance and time?
(a) 0.876
(b) 0.459
(c) 0.996

(d) 0.927

[Objective: Recognize that correlation, unlike measures of association for categorical variables, only measures linear association.]

_____**20.** Which of the following is the best interpretation of the value found in question 19?

(a) There is a strong association between distance (in miles) and time (in minutes).
(b) There is not a strong linear association between distance (in miles) and time (in minutes).
(c) There is a strong linear association between distance (in miles) and time (in minutes).
(d) There is a moderate linear association between distance (in miles) an time (in minutes).

Answer Key

1. C
2. E
3. A
4. C
5. E
6. A
7. D
8. A
9. D
10. A
11. C
12. C
13. D
14. B
15. A
16. B
17. D
18. A
19. C
20. C

Chapter 2: Data Quiz A

Objectives:

- Organize data into a table with multiple variables (columns) and cases (rows).
- Distinguish categorical from numerical variables. Be aware that some categorical variables (ordinal) define an ordering of the cases.
- Recognize time series data.
- Identify when recoding or aggregating data are useful.

The camping permit at a state park asks that the person who is registering supply the following: Date, Name, Group Size, and the Zip Code for the person filling out the permit. The park management maintains a data table that records this information for each person who registers.

Section 2.1 – Data Tables
[Objective: Organize data into a table with multiple variables (columns) and cases (rows).]

1. On a particular day, 27 persons filled out a permit application. How many rows will be in the data table for that day?
(a) 27
(b) 4
(c) 108
(d) 31

2. On a particular day, 27 persons filled out a permit application. How many columns will be in the data table for that day?
(a) 27
(b) 4
(c) 108
(d) 31

Section 2.2 – Categorical and Numerical Data
[Objective: Distinguish categorical from numerical variables. Be aware that some categorical variables (ordinal) define an ordering of the cases.]

3. Identify each piece of information by the type of variable it represents:

DATE

NAME

GROUP SIZE

ZIP CODE

Section 2.3 – Recoding and Aggregation
[Objective: Identify when recoding or aggregating data are useful.]

4. For a mid-summer report, the park manager decides to use the Zip Code to generate a column for the data table that is labeled “INSTATE” with categories “Yes” and “No.” This column will identify the person registering as being from within the state or from a different state. This procedure is an example of ______________ data.
(a) aggregating
(b) recoding
(c) time series

(d) observing

5. At the end of the summer, the park manager creates a new data table using the information from each day's permit applications. The new data table consists of Date and the Total Number of Campers on that date. This is an example of ____________ the data, and generating results in data that are referred to as a _______________.

(a) recoding; cross section
(b) aggregating; cross section
(c) recoding; time series
(d) aggregating; time series

Section 2.4 – Time Series
[Objective: Recognize time series data.]

6. The manager wishes to use the data table from question 4 to produce a graph showing the Total Number of Campers for each day of the summer. What type of graph is most appropriate for this data?

(a) Bar graph
(b) Time Series
(c) Time plot
(d) Histogram

Section 2.2 – Categorical and Numerical Data
[Objective: Distinguish categorical from numerical variables. Be aware that some categorical variables (ordinal) define an ordering of the cases.]

7. Each camper at the park is asked to fill out a survey which reads as follows: "We are interested in knowing your return status. Are you planning to return to this park for camping next summer? Circle **the number** corresponding to your response." Campers are also asked to supply their Zip Code. The camper will circle one of the **numbers** below, depending on their status.

NO	UNLIKELY	UNSURE	LIKELY	YES
1	2	3	4	5

The summary of the data from the responses consists of the "Zip Code," and the "Return Status" of the camper. What type of variable is "Return Status"?

(a) Ordinal
(b) Numerical
(c) Categorical
(d) Likert

Publishers track sales data from Amazon.com. Typical tracking variables include book purchased, date of purchase, form of purchase (hardback, paperback, ebook, used), rating of purchase, and any comments

Section 2.4 – Time Series
[Objective: Recognize time series data.]

8. From the information provided, give an example of two variables that would result in time series data.

9. From the information provided, give an example of two variables that would results in cross-sectional data.

Section 2.3 – Recoding and Aggregation
[Objective: Recognize time series data.]

10. An author wants to look at the Amazon.com data pertaining to her book. She creates a table that includes form of purchase, the frequency of each form, and the total amount of purchase for each form. What is this an example of?

(a) Aggregating the data
(b) Recoding the data
(c) Observing the data
(d) Graphing the data

Answers:

1. A
2. B
3. Date – Ordinal; Name – Categorical; Group Size – Numerical; Zip Code – Categorical
4. B
5. D
6. C
7. A
8. Date of Purchase and an additional Variable (Answers will vary)
9. Answers will vary.
10. A

Chapter 2: Data — Quiz B

Objectives:

- Organize data into a table with multiple variables (columns) and cases (rows).
- Distinguish categorical from numerical variables. Be aware that some categorical variables (ordinal) define an ordering of the cases.
- Recognize time series data.
- Identify when recoding or aggregating data are useful.

A medical center obtains the following information for each patient that visits the office: Date, Name, Gender, an Identification Number, the Procedure Performed, and the Total Charge for the visit. A data table is used to organize the information collected each day.

Section 2.1 – Data Tables
[Objective: Organize data into a table with multiple variables (columns) and cases (rows).

1. For a given day, 20 patients visited the office. How many columns would the data table have for that particular day?
(a) 20 **(b)** 6 **(c)** 120 **(d)** 26

Section 2.2 – Categorical and Numerical Data
[Objective: Distinguish categorical from numerical variables. Be aware that some categorical variables (ordinal) define an ordering of the cases.

2. The Identification Number for each patient represents which type of variable:
(a) Numerical **(b)** Categorical **(c)** Time Series **(d)** Observation

Section 2.3 – Recoding and Aggregation
[Objective: Identify when recoding or aggregating data are useful.]

3. At the end of each month, a new column is created for the data table using the information contained in "Procedure Performed." The new column is labeled "Procedure Type" with categories "Surgical" and "Non-Surgical." This modification is an example of:
(a) Aggregation **(b)** An ordinal variable **(c)** Cases **(d)** Re-coding **(e)** A Likert Scale

4. At the end of each quarter, a data table is created with a column for "Gender" with categories "Male" and "Female," a column for the Total Number of each gender that visited the center during that quarter, and a column with the Total Charges for each gender that quarter. This procedure is an example of:
(a) Aggregation **(b)** Re-coding **(c)** A Likert Scale **(d)** Time Series
(e) Change in frequency

Section 2.4 – Time Series
[Objective: Recognize time series data.]

5. Another data table is created at the end of the quarter with a column for "Date" and a column for the "Total Charges" for that date. This summary provides:
(a) An aggregation **(b)** A time series **(c)** A Likert scale **(d)** An ordinal variable
(e) A timeplot

Section 2.2 – Categorical and Numerical Data
[Objective: Distinguish categorical from numerical variables. Be aware that some categorical variables (ordinal) define an ordering of the cases.]

6. Before leaving the center, patients are asked to respond to a survey concerning the amount of time they had to wait before seeing a doctor. The survey reads as follows: "Please circle the appropriate **number** below concerning the length of time you waited to see a doctor. Was the length of time you had to wait:"

Brief	About what you expected	Long	Inconveniently Long
1	2	3	4

Based on the results, another column is added to the data table with the heading "Length of Wait" and the number circled by the patient is recorded. Which type of variable is "Length of Wait"?

(a) Numerical **(b)** Ordinal **(c)** Nominal **(d)** Re-coded **(e)** An aggregation

A credit card company creates a table of its 50,000 customers. The table records the Account Number, Payment Due, Total Expenditures, Due Date, Paid (Yes or No), Amount Paid and Credit Limit.

Section 2.1 – Data Tables
[Objective: Organize data into a table with multiple variables (columns) and cases (rows).]

7. How many cases will this table have?

(a) 7 **(b)** 8 **(c)** 50,000 **(d)** cannot determine from information given

Section 2.2 – Categorical and Numerical Data
[Objective: Distinguish categorical from numerical variables. Be aware that some categorical variables (ordinal) define an ordering of the cases.]

8. In the table, what type of variable is Payment Due?

(a) Numerical **(b)** Categorical **(c)** Time Series **(d)** Observation

Section 2.3 – Recoding and Aggregation
[Objective: Identify when recoding or aggregating data are useful.

9. Which of the following is a good example of aggregation?

(a) Creating a new column titled Overdue that records a "YES" if a customer is past due on their payment and a "NO" if the customer paid the amount due by the due date.

(b) Create an additional table of Credit Limit, the frequency in each credit limit bracket, and the total expenditures in each credit limit bracket.

Section 2.4 – Time Series
[Objective: Recognize time series data.]

10. A financial advisor wants to look at a client's net returns over the past 10 years for each of the client's mutual funds. Determine (a) if the situation describes time series or cross-sectional data (b) Give a name to each variable in the data table and determine if the variable is categorical or numerical.