

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:

<u>Date of purchase</u>	<u>Customer Name</u>	<u>Customer #</u>	<u>Zip Code</u>	<u>Product #</u>	<u>Amount of Purchase</u>
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			Total
		1	2	3	
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 $\chi^2=10.52$ 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.