Question 1 *(20 points)*

The monthly closing values of the Dow Jones Industrial Average (DJIA) for the period beginning in January 1950 are given in the CSV file *Dow*. According to Wikipedia, the Dow Jones Industrial Average, also referred to as the Industrial Average, the Dow Jones, the Dow 30, or simply the Dow, is one of several stock market indices created by Charles Dow. The average is named after Dow and one of his business associates, statistician Edward Jones. It is an index that shows how 30 large, publicly owned companies based in the United States have traded during a standard trading session in the stock market. It is the second oldest U.S. market index after the Dow Jones Transportation Average, which Dow also created.

The Industrial portion of the name is largely historical, as many of the modern 30 components have little or nothing to do with traditional heavy industry. The average is price-weighted, and to compensate for the effects of stock splits and other adjustments, it is currently a scaled average. The value of the Dow is not the actual average of the prices of its component stocks, but rather the sum of the component prices divided by a divisor, which changes whenever one of the component stocks has a stock split or stock dividend, so as to generate a consistent value for the index.

Along with the NASDAQ Composite, the S&P 500 Index, and the Russell 2000 Index, the Dow is among the most closely watched benchmark indices for tracking stock market activity. Although Dow compiled the index to gauge the performance of the industrial sector within the U.S. economy, the index's performance continues to be influenced not only by corporate and economic reports, but also by domestic and foreign political events such as war and terrorism, as well as by natural disasters that could potentially lead to economic harm.

a) *4 points* Compute the average Dow return over the period given on the dataset.

b) *4 points* Plot a histogram and intuitively comment on whether or not the returns are normally distributed.

c) *4 points* Do the returns adhere to the Empirical Rule? Verify this using the Empirical Rule like the code for In-class Exercise.

d) *4 points* Plot a boxplot of returns and the five-number summary. Interpret the first, second, and third quartiles.

e) *4 points* Identify the mild and extreme outlier returns using the IQR. Include whether each monthly return is a mild, an extreme outlier, or not an outlier. Sort the results displaying the outliers first.

Question 2 *(20 points)*

Melvyl Stores, a division of National Clothing, is a chain of women's apparel stores operating throughout the country. The chain recently ran a promotion in which discount coupons were sent to customers of other National Clothing stores. Data collected for a sample of 100 in-store credit card transactions at Melvyl Stores during one day while the promotion was running are contained in the CSV file *Melvyl*. The Proprietary Card method of payment refers to charges made using a National Clothing charge card. Customers who made a purchase using a discount coupon are referred to as promotional customers and customers who made a purchase but did not use a discount coupon are referred to as regular customers. Because the promotional coupons were not sent to regular Melvyl Stores customers, management considers the sales made to people presenting the promotional coupons as sales it would not otherwise make. Of course, Melvyl also hopes that the promotional customers will continue to shop at its stores. Most of the variables in the data set are self-explanatory, but two of the variables require some clarification.

Items: The total number of items purchased
Net Sales: The total amount ($) charged to the credit card

Melvyl's management would like to use this sample data to learn about its customer base and to evaluate the promotion involving discount coupons.

Use the techniques of descriptive statistics and EDA to help management develop a customer profile and to evaluate the promotional campaign. Write your interpretations and narratives. Don't just display the numerical values. Consider including the following in your analysis.

a) *8 points* Frequency distributions for each of the key variables: number of items purchased, net sales, method of payment, gender, marital status, and age.

b) *4 points* A crosstabulation of type of customer (regular or promotional) versus net sales. Comment on any similarities of differences observed.

c) *6 points* Descriptive statistics on net sales and descriptive statistics on net sales by various classifications of customers.

d) *2 points* Descriptive statistics regarding the relationship between age and net sales.

Question 3 *(20 points)*

The CSV file *SuperMarketTransactions* contains over 14,000 transactions made by supermarket customers over a period of approximately two years. (The data are not real, but real supermarket chains have huge data sets similar to this one.) Column A contains the date of the purchase, column B is a unique identifier for

each customer, columns C–G contain information about the customer, columns H–J contain the location of the store, columns K–M contain information about the product purchased, and the last two columns indicate the number of items purchased and the amount paid.

a) *2 points* Using these data, create side-by-side box plots for revenues broken down by state or province. Are these distributions essentially symmetric or skewed?

b) *2 points* Note that these box plots include revenues from countries besides the United States. Do whatever it takes to create side-by-side box plots of revenue for only states within the United States.

c) *2 points* Summarize the total revenue by state.

d) *2 points* Show total revenue by product category.

e) *2 points* Show the number of transactions per state.

f) *2 points* Proportion of shoppers who have more than one child.

g) *2 points* Total revenue for January and February 2016. Show your summary in one view, which means you should not show January 2016 revenue and February 2016 review in separate result sets.

h) *2 points* Create a crosstab table showing the relationship between gender and product family.

i) *2 points* Proportion of shoppers who are single and own a home.

j) *2 points* Show the distribution (frequency and percentage) of gender.

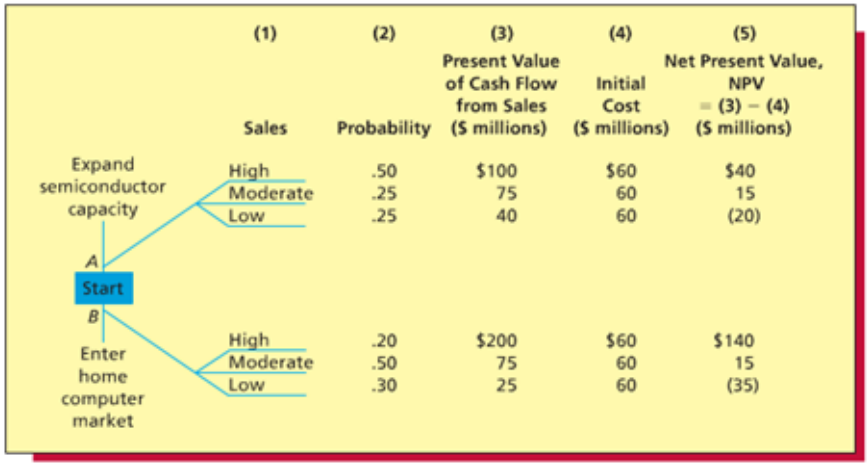
Question 4 *(20 points)*

The term **churn** is very important to managers in the cellular phone business. Churning occurs when a customer stops using one company's service and switches to another company's service. Obviously, managers try to keep churning to a minimum, not only by offering the best possible service, but by trying to identify conditions that lead to churning and taking steps to stop churning before it occurs. For example, if a company learns that customers tend to churn at the end of their two-year contract, they could offer customers an incentive to stay a month or two before the end of their two-year contract. The file *CellphoneMarket* contains data on over 2000 customers of a particular cellular phone company. Each row contains the activity of a particular customer for a given time period, and the last column indicates whether the customer churned during this time period. Use the various concepts in the EDA to learn (1) how these variables are distributed, (2) how the variables in columns B–R are related to each other, and (3) how the variables in columns B–R are related to the Churn variable in column S. Write a short explanation of your findings, including any recommendations you would make to the company to reduce churn.

Question 5 *(14 points)*

In the book *Foundations of Financial Management* (7th ed.), Stanley B. Block and Geoffrey A. Hirt discuss a semiconductor firm that is considering two choices: (1) expanding the production of semiconductors for sale to end users or (2) entering the highly competitive home computer market. The cost of both projects is \$60 million, but the net present value of the cash flows from sales and the risks are different.

Figure 1 below gives a tree diagram of the project choices. The tree diagram gives a probability distribution of expected sales for each project. It also gives the present value of cash flows from sales and the net present value (NPV = present value of cash flow from sales minus initial cost) corresponding to each sales alternative. Note that figures in parentheses denote losses.



Source: S. B. Block and G. A. Hirt, *Foundations of Financial Management*, 7th ed., p. 387. Copyright © 1994. Reprinted by permission of McGraw-Hill Companies, Inc.

Figure 1: Investment Options.

a) *2 points* Find the expected net present value of expanding semiconductor business project.

b) *2 points* Determine variance and standard deviation of the net present value

c) *2 points* Find the expected net present value of entering home computer market.

d) *2 points* Determine variance and standard deviation of the net present value

e) *2 points* Which project has the higher expected net present value?

f) *2 points* Which project carries the least risk? Explain.

g) *2 points* Calculate the relative variation for each project choice. Compare them to see which project carries the least risk. Is your response consistent with the part f) above?

Question 6 *(15 points)*

An office machine costs $7,500 to replace unless a mysterious, hard-to-find problem can be found and fixed. Repair calls from any service technician cost $500 each, and you're willing to spend up to $2,000 to get this machine fixed. You estimate that a repair technician has a 27% chance of fixing it.

a) *5 points* Create a probability model for the number of visits needed to fix the machine or exhaust your budget of $2,000.

b) *5 points* Find the expected number of service technicians that will be called in.

c) *5 points* Find the expected amount spent on this machine. (You must spend $7,500 for a new one if you do not find a vendor that repairs yours.)

Question 7 *(3 points)*

A manager at 24/7 Fitness Center is strategic about contacting open house attendees. With her strategy, she believes that 40% of the attendees she contacts will purchase a club membership. Suppose she contacts 20 open house attendees.

a) *1 point* What is the probability that exactly 10 of the attendees will purchase a club membership?

b) *1 point* What is the probability that no more than 10 of the attendees will purchase a club membership?

c) *1 point* What is the probability that at least 15 of the attendees will purchase a club membership?

Question 8 *(4 points)*

Airline travelers should be ready to be more flexible as airlines once again cancel thousands of flights this summer. The Coalition for Airline Passengers Rights, Health, and Safety averages 400 calls a day to help stranded travelers deal with airlines. Suppose the hotline is staffed for 16 hours a day.

a) *1 point* Calculate the average number of calls in a one-hour interval, 30-minute interval, and 15-minute interval.

b) *1 point* What is the probability of exactly six calls in a 15-minute interval?

c) *1 point* What is the probability of no calls in a 15-minute interval?

d) *1 point* What is the probability of at least two calls in a 15-minute interval?

Question 9 *(10 points)*

Sampling is a very common practice in quality control. Sampling plans are created to control the quality of manufactured items that are ready to be shipped. To illustrate the use of a sampling plan, suppose that a chip manufacturing company produces and ships electronic computer chips in lots, each lot consisting of 1000 chips. This company's sampling plan specifies that quality control personnel should randomly sample 50 chips from each lot and accept the lot for shipping if the number of defective chips is four or fewer. The lot will be rejected if the number of defective chips is five or more.

Find the probability of accepting a lot as a function of the actual fraction of defective chips. In particular, let the actual fraction of defective chips in a given lot equal any of 0.02, 0.04, 0.06, 0.08, 0.10, 0.12, 0.14, 0.16, 0.18. Then compute the lot acceptance probability for each of these lot defective fractions.

Create a bar plot showing how the acceptance probability varies with the fraction defective. Comment on what you observe.

A revised sampling plan call for accepting a given lot if the number of defective chips found in the random sample of 50 chips is five or fewer.Repeat the above exercise and summarize any notable differences between the two bar plots.

Question 10 *(16 points)*

Madison, an appliance store in the United States, sells window air conditioner units. Based on past sales experience, Madison stocks five window air conditioner units for the coming week. Assume that no orders for additional air conditioners will be made until next week. The weekly consumer demand for this the air conditioner unit has the probability distribution given below.

Number demanded	**Probability**
0	0.05
1	0.05
2	0.08
3	0.16
4	0.30
5	0.16
6	0.10
7	0.05
8	0.05
9 or more	0.00

Table 1: Weekly air conditioner demand.

a) *5 points* Let X be the number of window air conditioner units left at the end of the week (if any), and let Y be the number of special stockout orders required (if any), assuming that a special stockout order is required each time there is a demand and no unit is available in stock. Find the probability distributions of X and Y.

b) *4 points* Find the expected value of X and the expected value of Y.

c) *5 points* Assume that this appliance store makes a \$60 profit on each air conditioner sold from the weekly available stock, but the store loses \$20 for each unit sold on a special stockout order basis. Let Z be the profit that Madison earns in the coming week from the sale of window air conditioners. Find the probability distribution of Z.

d) *2 points* Find the expected value of Z.