姓名：_____　　學號：_____

1. (35 points) Pelican Store, 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 store who own National Clothing charge cards by email. Data collected for a sample of 100 in-store credit card transactions at Pelican Store during one day while the promotion was running are contained in the spreadsheet "Pelican" of Stat_2019_1_Mid.xlsx. 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 customer. Because the promotional coupons were not sent to regular Pelican Stores customers, management considers the sales made to people presenting the promotional coupons as sales it would not otherwise make. Of course, Pelican also hopes that the promotional customers will continue to shop at its stores. The columns of this spreadsheet are defined as follows:

Column	Column Name	Description
A	Customer	Customer Sample Sequence Number
B	Type_of_Customer	0 for Promotional; 1 for Regular
C	Item	The total number of items purchased
D	Net_Sales	The total amount ($) charged to the credit card
E	Method_of_Payment	1 for Proprietary Card; 2 for MasterCard; 3 for Visa
F	Age	Age of the sampled customer

(1) (4 points) Construct histogram for the "**Net_Sales**" with 8 classes.
(2) (3 points) Find the mean, median, and modal class for the "**Net_Sales**". Interpret these values.
(3) (2 points) Please find the 95th percentile for the "**Net_Sales**". Please find the relative standing of the "**Net_Sales**" = $95.2.
(4) (2 points) Please find the coefficients of skewness and kurtosis for the "**Net_Sales**". Explain these coefficients and compare them with the histogram in **(1)**.
(5) (4 points) Which rule, Empirical or Chebyshev, is better to use to describe these data ("**Net_Sales**")? Please explain the reason and find a reasonable % of upper and lower bound.
(6) (2 points) Please construct a stem-and-leaf plot for the "**Net_Sales**" using the number of tens as the stem and the number of ones as the leaf. Please describe the distribution of the data set (central tendency, variation, and skewness) and explain similarities and differences after comparing the stem-and-leaf plot with the histogram in **(1)** and the coefficients of skewness and kurtosis in **(4)**.
(7) (3 points) Construct a scatter-plot for "**Net_Sales**" and "**Age**", with "**Age**" on the horizontal axis and "**Net_Sales**" on the vertical axis. What type of relationship do you detect?
(8) (3 points) Compute and interpret the coefficient of correlation between "**Net_Sales**" and "**Age**". Does the coefficient confirm the relationship you detect in question **(7)**?
(9) (4 points) Please classify the sample list into two groups based on "Type_of_Customer". Construct two box plots for "**Net_Sales**" for these two groups. What similarities and differences are there in these two box plots? Use the two box plots to determine which customers have the unusually higher or lower "**Net_Sales**" (outliers).
(10) (6 points) Please classify the sample list into three groups based on "Method_of_Payment". Construct three box plots for "**Net_Sales**" for these three groups. What similarities and differences are there in these three box plots? Use the three box plots to determine which customers have the unusually higher or lower "**Net_Sales**" (outliers).
(11) (2 points) Based on the results of **(9)** and **(10)**, can you conclude that there is a big effect on using the promotional coupons? Explain.

2. (15 points) The number of cases (a case has 24 bottles or cans) of different types of soft drinks sold in the United Sates in 2014 is summarized in the spreadsheet "Soft_drinks" of Stat_2019_1_Mid.xlsx. The columns of this spreadsheet are defined as follows:

Column	Column Name	Description
A	Brand	The brand name of the soft drink
B	Company	The company name holding the brand
C	Cases	The number of cases (in millions) sold

   (1) (3 point) Construct a chart to display the amounts of cases sold by each brand. Describe your impression of the chart. Do you think that you should present every brand in your chart?
   (2) (3 point) Prepare a chart that should emphasize the shares of the market held by the different brands. What is your interpretation regarding to this chart?
   (3) (4 point) Prepare a chart that compares the share of the market held by each of the three companies. Is this chart different from the chart prepared in (2)?
   (4) (5 point) Prepare a chart that contrasts of the amounts of cases of diet and regular soft drinks sold by these three companies. Did you see a relationship between company and diet/regular?

3. (10 points) Fresh Food Market retail store offer its customers store credit cards that can earn them bonus gifts when customers make purchases at the store. Currently 60 customers are shopping in the store and half of these customers already have a store credit card. Fresh Food Market also estimates that customers arrive at the rate of 60 customers per hour following a Poisson process during the peak period of 5:00 pm to 8:00 pm. Fresh Food Market also found that about 80% of the customers return to Fresh Food Market to shop more than twice. Fresh Food Market opens 3 checkout counters during the peak period and each customer checks out items at one counter. The checkout time for each customer follows a normal distribution with mean = 2 minutes and standard deviation = 1.2 minutes.
   (1) (2 points) If employees offer store credit cards to 6 of the customers currently in the store, what is the probability that all of those selected already have a card? If a sample of 10 customers is randomly selected for applying store credit card, what is the probability that at least four of those selected already have a card?
   (2) (2 points) If it takes a counter 5 minutes to complete the checkout of items for one customer, how many customers on average do you expect to arrive by that time? What is the probability that none will arrive by that time?
   (3) (2 points) What is the expected inter-arrival time between two customers? What is the probability that the inter-arrival time between two customers is more than 2 but less than 5 minutes?
   (4) (2 points) In a sample of 60 customers, what is the probability that more than 20 but less than 40 customers return to shop at Fresh Food Market for more than twice? What is the probability that more than 50 customers return to shop at Fresh Food Market for more than twice?
   (5) (2 points) What is the probability that it takes 2 to 4 minutes to checkout items for one customer? What is the probability that it takes more than 5 minutes to checkout items for one customer?

4. (12 points) A manufacturing firm receives shipments of parts from three different suppliers, V1, V2, and V3. Currently, 60% of the parts purchased by the company are from supplier V1, 25% are from supplier V2, and the remaining 15% are from supplier V3. The quality of the purchased parts varies with the source of supply. Historical data suggest that the quality ratings of the three suppliers are shown in the following table.

Supplier	Supplier V1	Supplier V1	Supplier V3
% of Good Parts	98	96	95

(1) (4 points) What is the probability that one randomly selected part is good? What is the probability that one randomly selected part is defect?

(2) (4 points) If one randomly selected part is good, what is the probability that it was from Supplier V1? If one randomly selected part is good, what is the probability that it was from Supplier V2? If one randomly selected part is good, what is the probability that it was from Supplier V3?

(3) (4 points) If one randomly selected part is defect, what is the probability that it was from Supplier V1? If one randomly selected part is defect, what is the probability that it was from Supplier V2? If one randomly selected part is defect, what is the probability that it was from Supplier V3?

5. (16 points) A day trader is offered the opportunity to buy shares in two companies, IBM and Microsoft. Assume that a share of IBM and Microsoft stock costs \$100 today. The random variable X denotes the change in the value of stock in IBM while the random variable Y represents the change in Microsoft. The joint probability distribution of X and Y is given in the following table.

$Y \setminus X$	$x = -5$	$x = 0$	$x = 5$
$y = 4$	0.00	0.11	0.07
$y = 0$	0.03	0.62	0.02
$y = -4$	0.06	0.07	0.02

(1) (3 points) If the day trader wants to invest all on IBM, what are the expected value and the standard deviation of this investment's value change?

(2) (3 points) If the day trader wants to invest all on Microsoft, what are the expected value and the standard deviation of this investment's value change?

(3) (4 points) If the day trader wants to invest equally on these two stocks, what are the expected value and the standard deviation of this investment's value change?

(4) (2 points) Based on the above results, what recommendation would you give to the day trader?

(5) (4 points) The day trader learns from an expert and will invest 30% on IBM and 70% on Microsoft. Based on this new information, will you change your recommendation to the day trader about the percentages invested on these two stocks?

6. (12 points) Cooper Realty is a small real estate company located in Albany, New York, specializing primarily in residential listings. They recently became interested in determining the likelihood of one of their listings being sold within a certain number of days. A list of 400 homes the company sold in previous years is listed in the spreadsheet "Cooper" of Stat_2019_1_Mid.xlsx. The columns of this spreadsheet are defined in the following table. Please answer the following questions based on this list.

Column	Column Name	Description
A	House_No	House number assigned by the FIFS order
B	Days	Days listed until sold (1: Under 30 days; 2: 31-90 days; 3:over 90 days)
C	Price	Initial Asking Price (1: Under $150K; 2: $150K-$199,999; 3: $200K-$250K; 4:over $250K)

(1) (2 points) What is the probability that one randomly selected house is listed for more than 90 days before being sold? What is the probability that one randomly selected house is sold under 30 days?

(2) (2 points) What is the probability that one randomly selected house initially asked for $200K-$250K? What is the probability that one randomly selected house initially asked for under $150K?

(3) (1 point) What is the probability that one randomly selected house either is sold under 30 days or initially asked for under $150K or both?

(4) (1 point) What is the probability that one randomly selected house is listed for more than 90 days before being sold and asked for $200K-$250K?

(5) (2 points) What is the probability that the house selected initially asked for $150K-$199,999, given that the house is listed for 31-90 days before being sold?

(6) (2 points) What is the probability that the house selected is listed for 31-90 days before being sold, given that the house initially asked for $150K-$199,999?

(7) (2 points) Are days and price independent?