姓名：_____　學號：_____

一、選擇題(共 45 points)

1. A sample of 200 students at a Big-Ten university was taken after the midterm to ask them whether they went bar hopping the weekend before the midterm or spent the weekend studying, and whether they did well or poorly on the midterm. The following table contains the result. Which of the following statements is true?

	Did Well in Midterm	Did Poorly in Midterm
Studying for Exam	80	20
Went Bar Hopping	30	70

   A. Of those who went bar hopping the weekend before the midterm in the sample, 30% of them did well on the midterm..
   B. Of those who did well on the midterm in the sample, 15% of them went bar hopping the weekend before the midterm.
   C. 50% of the students in the sample spent the weekend studying and did well on the midterm.
   D. If the sample is a good representation of the population, we can expect 20% of the students in the population to spend the weekend studying and do poorly on the midterm.
   E. None of above

2. Which of the following statements is true?
   A. A research analyst was directed to arrange raw data collected on the yield of wheat, ranging from 40 to 93 bushels per acre, in a frequency distribution. He should choose 30 as the class interval width.
   B. Ogives are plotted at the midpoints of the class groupings.
   C. The percentage polygon is formed by having the lower boundary of each class represent the data in that class and then connecting the sequence of lower boundaries at their respective class percentages.
   D. To determine the width of class interval, divide the number of class groups by the range of the data.
   E. None of above

3. Thirty-six of the staff of 80 teachers at a local intermediate school are certified in Cardio-Pulmonary Resuscitation (CPR). In 180 days of school, about how many days can we expect that the teacher on bus duty will likely be certified in CPR?
   A. In the range (75, 100] days　B. In the range (50, 75] days　C. In the range (25, 50] days
   D. In the range (0, 25] days　E. None of the above

4. A debate team of 4 members for a high school will be chosen randomly from a potential group of 15 students. Ten of the 15 students have no prior competition experience while the others have some degree of experience. What is the probability that at most 1 of the members chosen for the team have some prior competition experience?
   A. In the range (0.75, 1.00]　B. In the range (0.50, 0.75]　C. In the range (0.25, 0.50]
   D. In the range (0.00, 0.25]　E. None of the above

5. There are two houses with almost identical characteristics available for investment in two different neighborhoods with drastically different demographic composition. The anticipated gain in value when the houses are sold in 10 years has the following probability distribution. Which of the following statements is true?

Probability	Returns	
	Neighborhood A	Neighborhood B
.25	$22,500	$30,500
.40	$10,000	$25,000

.35	$40,500	$10,500

A. The expected value gain if you invest in both houses is $21,300.
B. The total standard deviation of value gain if you invest in both houses is $8,213.40.
C. If you can invest half of your money on the house in neighborhood A and the remaining on the house in neighborhood B, the portfolio risk of your investment is $16,925.
D. If you can invest 30% of your money on the house in neighborhood A and the remaining on the house in neighborhood B, what is the portfolio risk of your investment is $18,675.
E. None of above

6. You were told that the amount of time lapsed between consecutive trades on the New York Stock Exchange followed a normal distribution with a mean of 15 seconds and that the probability that the time lapsed between two consecutive trades to fall between 16 to 17 seconds was 13%. The probability that the time lapsed between two consecutive trades would fall below 13 seconds was 7%. Which of the following statements is true?
   A. The probability that the time lapsed between two consecutive trades will be longer than 17 seconds is 0.13.
   B. The probability that the time lapsed between two consecutive trades will be between 13 and 16 seconds is 0.453.
   C. The probability is 20% that the time lapsed will be shorter 11 seconds.
   D. The middle 86% of the time lapsed will fall between 13 seconds and 17 seconds.
   E. None of above

7. Patients arriving at an outpatient clinic follow an exponential distribution with mean 4 minutes. What is the probability that a randomly chosen arrival to be between 5 minutes and 15 minutes?
   A. In the range (0.75, 1.00] B. In the range (0.50, 0.75] C. In the range (0.25, 0.50]
   D. In the range (0.00, 0.25] E. None of the above

8. Which of the following statements is true?
   A. A statistic is usually unobservable while a parameter is usually observable.
   B. Using the number of registered voters who turned out to vote for the primary in Iowa to predict the number of registered voters who will turn out to vote in Vermont's primary is an example of descriptive statistics.
   C. Compiling the number of registered voters who turned out to vote for the primary in Iowa is an example of descriptive statistics.
   D. A professor computed the sample average exam score of 20 students and used it to estimate the average exam score of the 1,500 students taking the exam was an example of descriptive statistics.
   E. None of above

9. The chancellor of a major university was concerned about alcohol abuse on her campus and wanted to find out the proportion of students at her university who visited campus bars on the weekend before the final exam week. Her assistant took a random sample of 250 students. What type of variable is the portion of students in the sample who visited campus bars on the weekend before the final exam week?
   A. a categorical random variable B. a discrete random variable
   C. a parameter D. a continuous random variable
   E. None of the above

10. Which of the following statements is true?
   A. You have collected information on the market share of 5 different search engines used by U.S. Internet users in May 2007. The best for presenting the information is A histogram.
   B. A survey of 150 executives were asked what they think is the most common mistake candidates make during job interviews. Six different mistakes were given. The best for presenting the information is a bar chart.

C. Data on the number of credit hours of 20,000 students at a public university enrolled in a Spring semester were collected. The best for presenting the information is a Pareto chart.
D. Data on the number of part-time hours students at a public university worked in a week were collected. The best chart for presenting the information is a side-by-side bar chart.
E. None of above

11. Health care issues are receiving much attention in both academic and political arenas. A sociologist recently conducted a survey of citizens over 60 years of age whose net worth is too high to qualify for Medicaid and have no private health insurance. The ages of 25 uninsured senior citizens were as follows. Which of the following statements is true?

60	61	62	63	64	65	66	68	68	69	70	73	73
74	75	76	76	81	81	82	86	87	89	90	92	–

A. The coefficient of variation of the ages of the uninsured senior citizens is 16.13%.
B. The interquartile range of the ages of the uninsured senior citizens is 12 years.
C. 25% of the senior citizens sampled are older than 71.5 years of age.
D. The distribution of the sample appears to be slightly positive or right-skewed.
E. None of above

12. The data in the following table represent the total rate of return of four companies from 2006 to 2009. Which of the following statements is true?

Year	Company A	Company B	Company C	Company D
2009	25.30	26.40	45.40	29.40
2008	-15.01	-22.10	-21.58	-20.90
2007	-5.44	-11.90	-1.03	-10.97
2006	-6.20	-9.10	-3.02	-10.89

A. The geometric mean rate of return for Company C is 2.28%.
B. The geometric mean rate of return for Company B is -3.74%.
C. The geometric mean rate of return for Company A is 1.24%.
D. The geometric mean rate of return for Company D is -4.17%.
E. None of above

13. The probability that house sales will increase in the next 6 months is estimated to be 0.25. The probability that the interest rates on housing loans will go up in the same period is estimated to be 0.74. The probability that house sales or interest rates will go up during the next 6 months is estimated to be 0.89. Which of the following statements is true?
A. The probability that both house sales and interest rates will increase during the next 6 months is 0.185.
B. The probability that house sales will increase but interest rates will not during the next 6 months is 0.15.
C. The probability that neither house sales nor interest rates will increase during the next 6 months is 0.195.
D. The events of increase in house sales and increase in interest rates in the next 6 months are statistically independent.
E. None of above

14. A company has 2 machines that produce widgets. An older machine produces 23% defective widgets, while the new machine produces only 8% defective widgets. In addition, the new machine produces 3 times as many widgets as the older machine does. Given a randomly chosen widget was tested and found to be defective, what is the probability it was produced by the new machine?
A. In the range (0.00, 0.25]　B. In the range (0.25, 0.50]　C. In the range (0.50, 0.75]
D. In the range (0.75, 1.00]　E. None of the above

15. According to the record of the registrar's office at a state university, 35% of the students are

freshman, 25% are sophomore, 16% are junior and the rest are senior. Among the freshmen, sophomores, juniors and seniors, the portion of students who live in the dormitory are, respectively, 80%, 60%, 30% and 20%. Which of the following statements is true?

A. If a randomly selected student lives in the dormitory, the probability that the student is not a freshman is 0.532.
B. If a randomly selected student does not live in the dormitory, the probability that the student is a junior or a senior is 0.359.
C. The probability that a randomly selected student is a junior or senior who lives in a dormitory is 0.112.
D. The probability that a randomly selected student is a freshman who lives in a dormitory is 0.207.
E. None of above

二、計算題：(共 55 points)

1. (12 points) A computer intrusion detection system (IDS) is designed to provide an alarm whenever an intrusion (e.g. unauthorized access) is being attempted into a computer system. Consider a system with three independently operating intrusion detection systems (a triple IDS), A, B, and C. A sample of 300 message scans were selected from this system and the results are listed and summarized in the spreadsheet "**Intruder**" of Stat_2016_1_Mid.xlsx. The values for five columns (variables) are defined as follows:

Column	Column Name	Description
A	Sample #	Sample No.
B	Yes or No	Scan indicating an intruder or not, Y: Yes; N: No
C	System A	An alarm is sounded by system A or not, 1: Yes; 0: No
D	System B	An alarm is sounded by system B or not, 1: Yes; 0: No
E	System C	An alarm is sounded by system C or not, 1: Yes; 0: No

(1) (4 points) Please construct three joint probability tables with two variables, **"Yes or No"** vs. "**System A**", **"Yes or No"** vs. "**System B**", and **"System A"** vs. "**System B**", respectively. Can you conclude that "**System A**" and **"System B"** are independent?

(2) (2 points) If there is an intruder, what is the probability that at least two of the three systems sound an alarm?

(3) (2 points) If there is no intruder, what is the probability that at least one of the three systems sound an alarm?

(4) (2 points) If there is an intruder and **System A** sounds an alarm, what is the probability that **System B** also sounds an alarm?

(5) (2 points) If there is no intruder and **System B** sounds an alarm, what is the probability that **System C** also sounds an alarm?

2. (17 points) Diamonds are categorized according to the four C's: Carats, clarity, color, and cut. Each diamond stone that is sold on the open market is provided a certificate by an independent diamond assessor that lists these characteristics. Data for 308 diamonds were extracted from Singapore's Business Times are located in the spreadsheet "**Diamonds**" of Stat_2016_1_Mid.xlsx. The values for three columns (variables) are defined as follows:

Column	Column Name	Description
A	CARAT	The measure of Carat
B	COLOR	Color is classified as D, E, F, G, H, or I
C	CLARITY	Clarity is classified as IF, VVS1, VVS2, VS1, or VS2
D	CERT	Three independent certification groups: GIA, HRD or IGI
E	PRICE	Price of the diamond

(1) (3 points) Please construct three side-by-side bar charts of the "**CLARITY**" for the three independent certification groups: GIA, HRD or IGI. Can you conclude that "**CLARITY**" and "**CERT**" are independent?

(2) (1 point) Please find the 90^{th} percentile for the "**PRICE**".

(3) (1 point) Please find the relative standing of the "**PRICE**" = $5961.

(4) (3 point) Construct a scatter plot: "**CARAT**" and "**PRICE**". Compute and interpret the coefficient of correlation between "**CARAT**" and "**PRICE**". What type of relationship do you detect from this scatter plot?

(5) (3 points) Please construct three box plots of the "**PRICE**" for the three independent certification groups: GIA, HRD or IGI.

(6) (3 point) Use these three box plots in (5) to determine which diamonds have the unusually higher or lower "**PRICE**" (outliers), respectively.

(7) (3 points) Compare the three side-by-side bar charts in (1) and the three box plots of the "**PRICE**" in (5). What similarities and differences are there in the distributions of the "**CLARITY**" and "**PRICE**" for the three independent certification groups: GIA, HRD or IGI?

3. (11 points) Chemical and manufacturing plants sometimes discharge toxic-waste materials such as DDT into nearby rivers and streams. These toxic can adversely affect the plants and animals inhabiting the river and the riverbank. The U.S. Army Corps of Engineers conducted a study of fish in the Tennessee River (in Alabama) and its three tributary creeks: Flint Creek, Limestone Creek, and Spring Creek. A total of 144 fish were captured and are listed in the spreadsheet "**DDT**" of Stat_2016_1_Mid.xlsx and the columns of this spreadsheet are defined as follows:

Column	Column Name	Description
A	RIVER	River/Creek where each fish was captured
B	SPECIES	Species (channel catfish, largemouth bass, or smallmouth buffalo fish)
C	LENGTH	The length of the fish captured (in centimeters)
D	WEIGHT	The weight of the fish captured (in grams)
E	DDT	DDT concentration (parts per million)

(1) (3 points) Please construct a histogram for the "**WEIGHT**" with **8** classes. Interpret the shape of this histogram.

(2) (2 point) Find the mean, median, and modal class for the "**WEIGHT**". Interpret these values.

(3) (2 points) Please compute the coefficient of skewness and the coefficient of kurtosis for the "**WEIGHT**". Interpret these values.

(4) (2 point) Which rule, Empirical or Chebyshev, is better to use to describe these data ("**WEIGHT**")? Please explain the reason and find the 2*s* upper and lower bound.

(5) (1 points) Based on the conclusion of (4), find the probability that the "**WEIGHT**" will fall between 1000 and 1400 grams.

(6) (1 points) Based on the conclusion of (4), find the probability that the "**WEIGHT**" will be less than 500 grams or larger than 1750 grams.

4. (8 points) As part of a project targeted at improving the services of a local bakery, a management consultant monitored customer arrivals and purchases for several Saturdays and Sundays. Using the arrival data, she estimated the number of customer arrivals during weekends follows a Poisson Process and the average time between two customer arrivals is 2 minutes. She also found that there is a 20% of the chance that each customer leaves without purchasing anything from the bakery. The bakery opens from 9:00am to 6:00pm every Saturday and from 10:00am to 4:00pm every Sunday
   (1) (2 points) What is the probability that more than 45 customers per hour enter the bakery during weekend?
   (2) (2 points) What is the probability that the time between two customer arrivals to the bakery is more than 1.5 minutes during weekend?
   (3) (2 points) What is the probability that more than 20 customers make purchases if 30 customers enter the bakery in one hour?
   (4) (2 points) Please compute the average number of arrivals to the bakery during one weekend (Saturday and Sunday). Use this average number to compute the average number of customers who make purchases in the bakery during one weekend (Saturday and Sunday).

5. (7 points) You are trying to set up a portfolio that consists of a corporate bond fund and a common stock fund. The information about the annual return (per $1000) of each of these investments under different economic conditions is listed in the spreadsheet "**Portfolio**" of Stat_2016_1_Mid.xlsx along with the probability that each of these economic conditions will occur.
   (1) (2 points) Suppose that you wanted to create a portfolio that consists of corporate bonds and common stocks, please find the portfolio expected return and portfolio risk when the proportion invested in corporate bond is 0.1.
   (2) (2 points) Suppose that you wanted to create a portfolio that consists of corporate bonds and common stocks, please find the portfolio expected return and portfolio risk when the proportion invested in corporate bond is 0.5.
   (3) (2 points) Suppose that you wanted to create a portfolio that consists of corporate bonds and common stocks, please find the portfolio expected return and portfolio risk when the proportion invested in corporate bond is 0.9.
   (4) (1 points) On the basis of the results of (1) to (3), which portfolio would you recommend? Explain.