Review Materials for Exam 1 (10/28/05)

Exam coverage: Chapter 1.1-1.3, 3.1-3.4, 4.1-4.4

PLEASE BRING SOME FORM OF PHOTO IDENTIFICATION (e.g. Bruin Card, Drivers License, etc.) ATTENDANCE WILL BE TAKEN. PLEASE REMEMBER TO BRING WRITING INSTRUMENTS AND A CALCULATOR. WE WILL PROVIDE AN EXAM PACKET AND THE NORMAL TABLE FROM THE TEXT.

A ONE SIDED 8.5" x 11" piece of paper with formulas/notes/examples/whatever is allowed into the exam. Typed, printed (laser/inkjet), cut and paste, handwritten, pencil, pen, highlight, colored pens, pencils, crayon is OK. ONE SIDE ONLY PLEASE.

WHAT FOLLOWS ARE MANY EXAMS WORTH OF QUESTIONS FROM PAST CLASSES, YOUR EXAM <u>WILL NOT</u> BE THIS LONG

1.   The next four questions use information from this statement, but each question is separate (i.e. you can get the first one wrong and it won't affect the others): The Medical College Admissions Test (MCAT) is constructed to be normally distributed with a mean of 9 and a standard deviation of 2. Approximately 20,000 people take the test every year. SHOW YOUR WORK FOR FULL CREDIT.

(a) Harvard Medical School only considers applicants with a test score of 12.1 or greater. How many of the test takers qualify for Harvard?

(b)   The lowest 1% of all test takers can enroll in Dr. Nick Riviera's School of Medicine. At and below what MCAT score is the lowest 1% of all test takers?

(c) You decided to take the MCAT and got a 12.4. Your cousin, who went to USC, wouldn't tell you his score, but also took the MCAT and told you he is score was 2.2 Z scores below yours. What percentage of test takers have scores between yours and your cousin's?

(d) After thinking it over, you decide not to apply to Medical School, but apply to Law School instead. And to your surprise, the UCLA Law School is willing to consider applicants with a valid MCAT score – with the following condition: All applicants must add 23 to their MCAT Score first and then multiply that score by 5. So for example, you got a 12.4, your new score is 177.

If you apply UCLA rules to ALL the MCAT scores, what are the new median, standard deviation and 75th percentile?

2.   The next four questions refer to this statement, but each question is separate (i.e. you can get the first one wrong and that won't affect the others): Corporate securities (or publicly traded stocks) are an investment opportunity for individuals as well as institutions. The 10,000 stocks available for investment to U.S. residents are normally distributed with a mean one-year return of -1% (this means you lost 1% of the value of your investment) and a standard deviation of 12%. SHOW YOUR WORK FOR FULL CREDIT.

(a)   What percentage of stocks had one-year returns between -16% and +2%? (5 points)

(b)   A stock is at 25th percentile (i.e. 25% of the stocks have returns equal to or lower than this stock), what is its one-year return? (5 points)

(c)   Financial advisors search for stocks within ±1.5 Z scores of a return of 8%. Approximately how many stocks out of the 10,000 qualify? (5 points)

(d)   Your financial advisor is able to predict the future with some accuracy and says that next year, stocks will be normally distributed and the median return will be +2% and a stock at the 90th percentile will have a return of 11%. Given this information, is it possible to calculate the standard deviation for next year? Please answer YES or NO. If you answer YES, please calculate the standard deviation. If you answer NO, please explain why this is not possible to calculate. (5 points)

3. A professor constructed a sample survey to estimate the percentage of USC undergraduates living at home. Two assistants were stationed at the "Tommy" Trojan statue (it's on the main plaza) and were instructed to interview all students who passed by at specified times. Many students would not speak with the assistants, in fact, only 369 out of 1500 approached, did. As it turned out, 39% of 369 students interviewed said they live at home with their parents the others live elsewhere. Does the investigator's procedure give an accurate estimate of the percentage of USC students who live with their parents?

First, answer yes or no and then explain your reasons for your choice. This does not need to be a long answer.

4. Classify the following variables as categorical or quantitative by checking the correct box, if it is a quantitative variable, further classify the variable as either discrete or continuous:

	Variable	Categorical	Quantitative	Discrete	Continuous
A	Hair Color				
B	Frozen Food Brand				
C	Number of students in a classroom				
D	Your age				

5. To study the effects of exercise on the grades of college students, a researcher wishes to compare the grade point averages of students at randomly selected colleges across the United States. The researcher selects students at random and after interviewing them to find out who exercises and who does not, chose 644 students of each (exercisers and non-exercisers). The researcher made sure the two groups of 644 were similar in racial composition, gender, major, and every subject had accumulated at least 120 units towards graduation. There were a total of 1,288 students in the study from approximately 40 colleges, their overall GPA was 3.22. The average GPA for the students who exercised was 3.34 and the standard deviation was .36.

   a.   What is the "treatment"?

   b.   What is the response or outcome variable?

   c.   Is this an observational study or an experiment?

   d.   From this study, an example of a sample statistic is:

   e.   What is the population of interest?

   f.   What is the population parameter of interest in this study?

6.   The next four questions use information from this statement, but each question is separate (i.e. you can get the first one wrong and it won't affect the others): A recent study showed that the gambling income of adults age 21 and over in the United States from all forms of legalized gambling (e.g. lottery, video poker, horse racing, casinos) is normally distributed with a mean of -250 dollars (a loss) and a standard deviation of $700.   SHOW YOUR WORK FOR FULL CREDIT.

(a) It is believed that the gamblers with the largest losses, that is, those with the lowest 10% of gambling income, should be considered gambling "addicts" and given some kind of treatment. How much money does a gambling adult need to lose to be considered an "addict"?

(b) What percentage of adults age 21 and over had gambling losses of at least $500 but not more than $1000?

(c)   What is the median gambling income? What is the variance for gambling income?

(d)   What percentage of gamblers actually break even or make money?

7. The Super Bowl is the number one party event of the year for Americans, exceeding even New Year's Eve celebrations. Suppose it is known that the typical party has 17 partygoers on average with a standard deviation of 3.3.On a typical Sunday afternoon, the average number of calories consumed in America is 600. Please assume that calories are normally distributed

The Hardvard School of Public Health decided to study the effects of attending Super Bowl Sunday parties on the caloric consumption of Americans. 850 Americans were selected by random-digit dialing and interviewed by telephone. 490 Americans reported that they had attended a Super Bowl party, 110 did not attend a party but watched the Super Bowl on television at home. The remainder did not attend a Super Bowl party or watch the game. The calories consumed by the partygoers had a mean 1,330 with a standard deviation of 600. The calories consumed by the non-party goers had a mean of 560 with a standard deviation of 100. Among the partygoers, 77% reported getting "drunk", only 5% of the non-party goers reported getting "drunk". The average party had 19 partygoers.

A. (2 points) The population of interest to the Hardvard School of Public Health is

       (a)   all Super Bowl Partygoers
       (b)   all Americans
       (c)   all Americans who watched the Super Bowl on television
       (d)   850 Americans
       (e)   490 Americans who reported that they had attended a Super Bowl Party
       (f)   110 Americans who had not reported attending a Super Bowl Party

B. (2 points) The parameter of greatest interest to the Hardvard School of Public Health is

       (a)   17 partygoers
       (b)   3.3 partygoers
       (c)   600 calories
       (d)   The percentage who report getting "drunk"
       (e)   The average number of calories consumed by Americans
       (f)   The average number of calories consumed by party goers on Super Bowl Sunday

C. (2 points) The sample of interest to the Hardvard School of Public Health is

       (a)   all Super Bowl Partygoers
       (b)   all Americans
       (c)   all Americans who watched the Super Bowl on television
       (d)   850 Americans
       (e)   490 Americans who reported that they had attended a Super Bowl Party
       (f)   110 Americans who had not reported attending a Super Bowl Party

D. (2 points) The statistic of greatest interest to the Hardvard School of Public Health is

       (a)   19 partygoers
       (b)   1330 calories
       (c)   560 calories
       (d)   850 Americans
       (e)   77% got drunk
       (f)   5% got drunk

E. (2 points) This is an example of a

       (a)   an Observational Study with historical controls
       (b)   an Observational Study utilizing multi-stage cluster sampling techniques
       (c)   an Observational Study that uses a random probability method for sample selection.
       (d)   a Randomized Experiment without Controls, but it is blind
       (e)   a Randomized Experiment without Controls, but it is double-blind
       (f)   a Randomized Controlled Experiment

8. Please indicate whether each statement is true or false **(one point each)**

	True	False	Statement
A			The total area of a histogram is always 100% when area is expressed as percentages
B			Larger samples are no better than smaller samples at preventing bias
C			A histogram is a graphical summary which represents percentages as areas
E			The area under the histogram between two values is equal to the percentage of cases in a class interval defined by those values
F			In a randomized controlled experiment utilizing a placebo, if the control group is comparable to the treatment group, then the difference in the responses of the two groups is likely to be a result of the treatment
G			Double blind experiments are better at preventing the placebo effect than single blind experiments
H			Confounding is not a source of bias
I			Random selection (or randomizing/randomization) is employed in sample designs because it is impartial however it does not minimize bias
J			If a large number of persons selected for a sample do not respond, problems of response bias are likely
K			A histogram with one peak is said to be unimodal, a histogram with two peaks is said to be bimodal

9.   Here are three histograms, assume they have been correctly drawn:

0 50	0 50	0 50
(A)	(B)	(C)
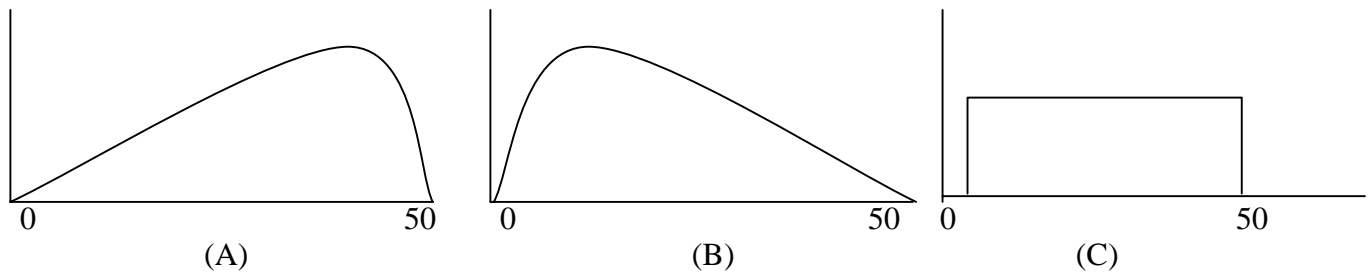
Match **each** histogram above to the best choice listed below: **(2 points each, 6 points total)**

> **i.**   The average is smaller than the median.
> **ii.**   The average is equal to the median
> **iii.**   The average is larger than the median
> **iv.**   Cannot determine the average for this graphic
> **v.**   Cannot determine the median for this graphic

10. In an observational study (choose one) **(2 points)**
   a. Investigators do not assign subjects to treatment or to control groups
   b. There isn't a control group
   c. Investigators can establish association but not causation
   d. Confounding factors cannot be controlled
   e. All of the above are true
   f. Only A and C are true
   g. Only D is false

11. The next 2 questions refer to this statement, but each question is separate (i.e. you can get the first one wrong and its result will not affect the others): You are on the verge of investing some of your hard-earned money in the stock market and you are examining two funds, let's call them A and B. Your investment adviser, I'll call him The Oracle, gives you some information on their performance (as measured by percentage returns over many, many days). Fund A has mean return of -3% with a standard deviation of 15%. It had a minimum of –78% and a maximum of 72%. Fund B has a mean return of 5% with a standard deviation of 2%. It had a minimum of –5% and a maximum of 15%. Assume both funds are normally distributed. SHOW YOUR WORK FOR FULL CREDIT.

A. You need to invest in a fund that spends as much time as possible giving returns in excess of 8%. Which fund is more likely to do this **(7 points)**

B. Oracle says you know what? You need to take inflation into account in all of your calculations. So subtract 4% (professor: just subtract 4, don't worry about the percentage sign) from all of the returns and then multiply by 2. So for example, on a given day, Fund A returned –11%, so following Oracle's instructions subtracting 4 yields –15% and multiplying by 2 yields –30%. If you do this, what are the new mean, median, and standard deviations for funds A and B? **(6 points)**

12.   Some computer output from a database of 907 movies produced between 1996-2000. The variable totalgrossreceipts is the total amount of money earned (in millions) domestically.

```
                    totalgrossreceipts
-------------------------------------------------------------
       Percentiles       Smallest
  1%      .437118         .235049
  5%     1.184727         .242093
 10%     2.019237         .258212      Obs                    907
 25%     5.672903         .297527      Sum of Wgt.            907

 50%      17.0678                      Mean              33.85867
                          Largest      Std. Dev.         48.02333
 75%     42.42211         293.5017
 90%      90.4436         306.1693     Variance          2306.241
 95%     125.6034         431.0883     Skewness          4.075231
 99%     215.3973         600.787      Kurtosis          32.94395
```

A.   This variable most probably came from a(n): (2 points)
        (a)   Observational Study with no treatments or factors
        (b)   Study confounded by voluntary response bias
        (d)   Experiment which utilizes simple random sampling
        (e)   Observational Study which utilizes multistage cluster sampling
   (f)   None of the above

B. (1 point) Are there outliers present in this variable? (circle one) YES   NO
C. (3 points) Justify your answer in the space below (an answer utilizing numbers is required for full credit)

D. Please calculate the range (not interquartile range) and list the values of the quartiles (i.e. Q1, Q2, Q3) for this variable in the space below. (4 points)

13. A pair of Geographers are studying the distribution of age for each of the 50 states in the U.S. They used a box plot to summarize the medians for each state by geographic region. Here is a box plot of the results:
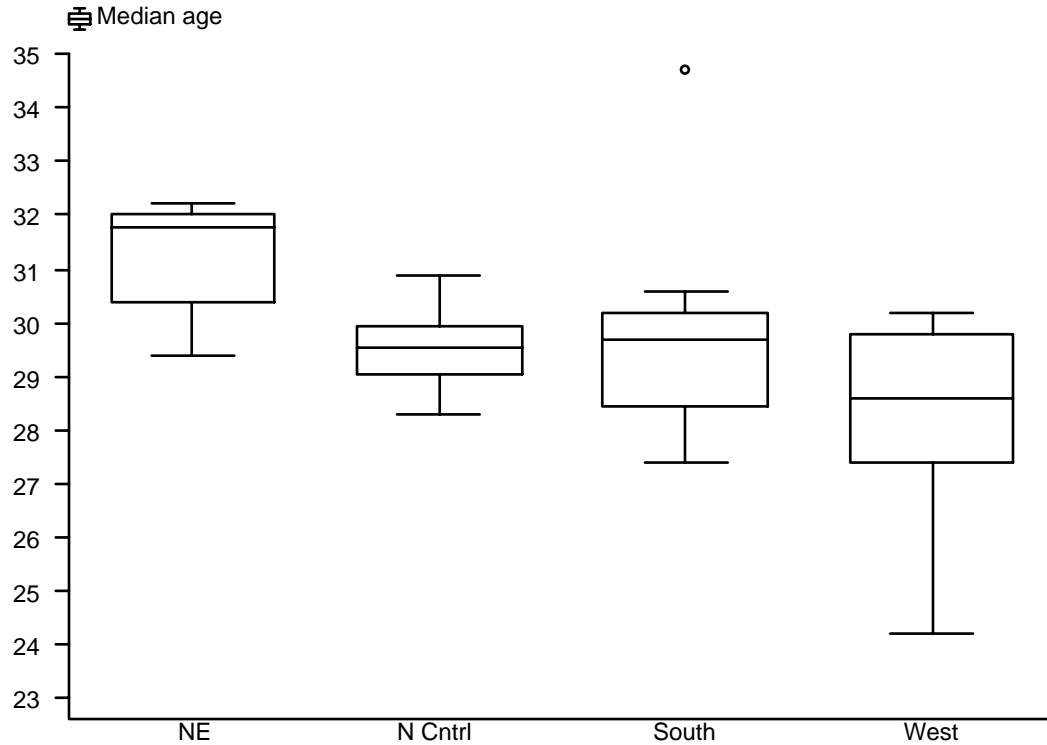
Using the box plot, please answer the following questions:

a)   Is there enough information present to estimate the inter-quartile range (IQR) for the North Central (N Cntrl) region? (circle one)

        YES   NO

        If you answered "YES" please give an estimate of that value in the space below, if you answered "NO" please explain why it is not possible to estimate the IQR using a box plot. (3 points total)

b)   Which region appears to be the most symmetrical in age? (circle one) (2 points)

            NE   N Cntrl   South   West   Not enough information

c)   Which region probably has the most skewness in age? (circle one) (2 points)

            NE   N Cntrl   South   West   Not enough information

d)   Which region has the most "negative skewness" in age? (circle one) (2 points)

            NE   N Cntrl   South   West   Not enough information

e)   Is there enough information in the box plot to accurately estimate the mean for the West region? (circle one)

            YES   NO

            If you answered "YES" please give an estimate of that value in the space below, if you answered "NO" please explain why not. (3 points total)
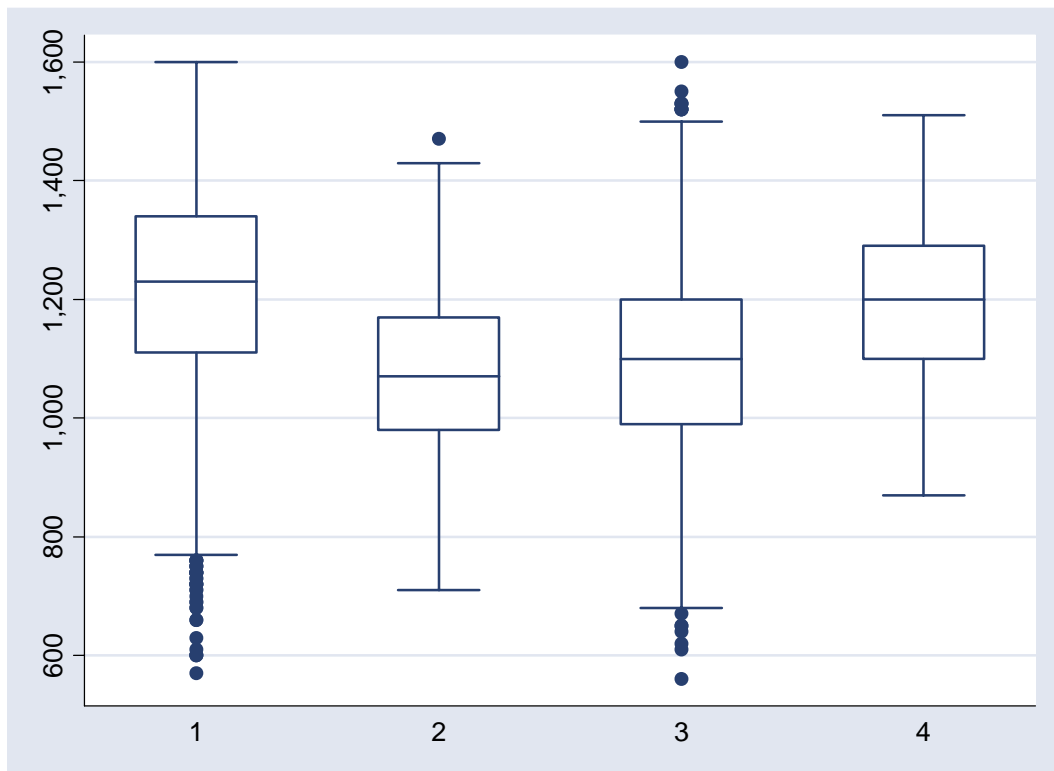
14.   The horizontal axis should be labeled "GROUP" and the vertical axis should be labeled "POINTS". The dark dots should actually look like open circles or asterisks. Using the box plot shown above, please answer the following questions:

a) Is there enough information present to estimate the inter-quartile range (IQR) for group 3? (circle one)

          YES   NO

        If you answered "YES" please give an approximate estimate of that value in the space below, if you answered "NO" please explain why it is not possible to estimate the IQR in this situation. (3 points total)

b) Which group appears to be the most symmetrical of the four groups ? (circle one) (2 points)

                    1   2   3   4   Not enough information

c) Which group appears to be the most left skewed? (circle one) (2 points)

                    1   2   3   4   Not enough information

15.   In a hypothetical experiment, a new drug was compared with "standard therapy" treatment for patients suffering from inoperable cancer. These types of patients volunteered for the experiment and were randomized into treatment and control groups. The difference in survival time (in months) was selected as the response variable. Which of the following best describes the primary reason to randomize patients into treatment or control groups? (Choose the one best answer) (2pts)

(a)   to prevent the bias introduced when the patients know what type of treatment they are receiving
(b)   to prevent the placebo effect from confounding the results of the experiment
(c)   to create "double blinding" when neither the investigators nor the patients know what type of treatment the patients are receiving
(d)   to create two groups that are similar at the start of the experiment on both known and unknown
              factors associated with survival time.
(e)   to eliminate the selection bias resulting from the fact that all of the patients had inoperable cancer
(f)   all of the above

16. Suppose an examination is very easy and all but a few students in a class received very high scores (total possible points was 150). Which statement below most correctly describes the relationship between the mean and median? (2 pts)

(a)   The mean is lower than the median for this examination
(b)   The mean and the median are approximately equal
(c)   The median will be exactly one-half the value of the mean
(d)   The median is lower than mean for this examination
(e)   There is not enough information to describe the relationship between the mean and median
(f)   None of the above

17. Suppose an examination is very difficult and all but a few students in a class received very low scores (total possible points was 150). If these scores were graphed, the distribution of these scores would be described as: (2 pts)

(a)   Normal
(b)   Symmetric or symmetrical
(c)   Left skewed
(d)   Right skewed
(e)   Uniform
(f)   None of the above

18. A consulting is firm is hired to count the number of passengers on buses in the city of Los Angeles. The first bus measured had 25 passengers, the second bus had 35 and the third had 55 passengers. The difference in the number of passengers between the first and second bus is equivalent to half the difference between the second and third bus. This is an example of a: (2 pts)

(a)   Quantitative variable with discrete values
(b)   Quantitative variable with continuous values
(c)   Quantitative variable with descriptive values
(d)   Qualitative variable with descriptive values
(e)   Qualitative variable with ordered values
(f)   None of the above

19. There are approximately 10000 publicly traded mutual funds. Morningstar.com gives a 0 to 5 star rating for each of the funds, but a new rival investment guide has come out with a 0 to 3 "thumbs up" rating for each of the funds. Apparently 1500 of the 10000 funds received a "3", 1500 of the 10000 received a "2", 3000 of the 10000 funds received a "1" and the rest received a "0" thumbs up. I have some software that produces a relative frequency table, but I have messed it up during the writing of this exam:

```
thumbs up |     Freq.      Percent        Cum.
----------+---------------------------------
       0 |                             40.00
       1 |     3,000       30.00        70.00
       2 |     1,500       15.00        85.00
       3 |     1,500       15.00       100.00
----------+---------------------------------
   Total |    10,000      100.00
```

Using the information above, please assume that the ratings do not change and these are all of the funds in existence. Please answer the following questions:

a.   What is the mean number of "thumbs up" and what is the median number of "thumbs up" for this population of funds?

b.   You select 3 funds at random with replacement from the 10000.   What is the probability (or chance) that at least 2 of the 3 funds you select will have at least a 2 "thumbs up" rating (i.e. score 2 or more "thumbs up")?

c.   You select 3 funds at random again, what is the chance that all 3 will receive "0" thumbs up?

20. You work for a credit card issuer and it is your job to issue cards to new customers. Since you also go to school, you decide to randomly issue cards to college students. Suppose it is known that 30% of all college students will eventually fail to pay their credit card debt within the first year of possessing a credit card.

A. You issue credit cards to 3 students selected at random, what is the chance that at least one of them will fail to pay their credit card debt within the first year? Assume independence.

        (a)   10% or .10
        (b)   30% or .30
        (c)   34% or .34
        (d)   66% or .66
        (e)   70% or .70
        (f)   90% or .90
        (g)   less than 10% (less than .10)
        (h)   greater than 90% (greater than .90)

B. Suppose once a student fails to pay on a credit card, the chance that the student will fail to pay on the next credit card that is issued rises to 60%. If a student has not failed to pay on a credit card, the chance still remains 30% for the next card that is issued. A randomly chosen student has been issued two cards. What is the chance that the student will fail to pay on at least one of the two cards? (Hint: a tree might be helpful here):

        (a)   60% or .60
        (b)   30% or .30
        (c)   36% or .36
        (d)   49% or .49
        (e)   51% or. 51
        (f)   64% or .64
        (g)   90% or .90
        (h)   greater than 90% (greater than .90)

C. You issue credit cards to 5 students selected at random, what is the chance that all 5 will be able to pay their credit card within the first year? Please show your work for full credit.

D.   You issue credit cards to 2 students selected at random, please fill in the missing information in this table:

# of student failures	probability
0	
1	
2	
Total	

21. The computer output for analysis variable OHNO is being shown on a screen in a big lecture hall.

```
. summarize ohno, detail
                          ohno
-------------------------------------------------------------
      Percentiles      Smallest
  1%    -10.65347      -15.3681
  5%    -9.295065      -15.03784
 10%    -8.56562       -14.96494      Obs              1000000
 25%    -7.346793      -14.91875      Sum of Wgt.      1000000

 50%    -5.999892                     Mean             -5.999883
                       Largest        Std. Dev.         2.00111
 75%    -4.650739      2.982908
 90%    -3.436283      3.037329       Variance         4.004441
 95%    -2.707139      3.061257       Skewness          .000419
 99%    -1.351807      3.272228       Kurtosis         3.000024
```

Please answer the following questions based on the Stata results for variable OHNO. You may round the numbers given above to two decimal places. For example, -8.56562 can be rounded to –8.57

The professor says "The Stata results suggest that variable OHNO is normally distributed"

B)   Using the Stata results above, determine how the professor came to this conclusion. Identify at least 2 conditions.

C)   Does this variable have outliers (circle one):   YES   NO

       Using information from the Stata results, show how you arrived at your choice.

D)   Please calculate the interquartile-range for variable OHNO

22. Here is the distribution of poor adults who live in a neighborhood in Los Angeles:

```
          poor |      Freq.     Percent        Cum.
    -----------+---------------------------------
    white male|       500       10.00        10.00
  white female|       750       15.00        25.00
non-white male|     1,500       30.00        55.00
non-white female|   2,250       45.00       100.00
    -----------+---------------------------------
         Total |     5,000      100.00
```

UCLA is planning a research study on the neighborhood and will need 3 residents to serve as community advisors. UCLA decides to select 3 at random. Assume independence for all parts of this problem please.

   a.   What is the probability that they will select at least 2 male residents? (3 points)

   b.   What is the probability that none of the three selected residents are white? (3 points)

   c.   What is the probability that the first white male selected is the third randomly selected resident? (3 points)

   d.   Suppose UCLA decides it really needs 7 residents, not 3. Six residents are selected at random, none of them are white females. Is the probability of selecting a white female as the 7th resident (3 points)
          A.   One-Sixth the probability of selecting a non-white male?
          B.   One-Third the probability of selecting a non-white female?
          C.   Equal to the probability of selecting a non-white female?
          D.   Equal to the probability of selecting a resident from the other 3 categories?
          E.   Two times the probability of selecting a non-white male?
          F.   Six times the probability of selecting a white male?

23. October is traditionally the beginning of flu season in the United States and this year there is a shortage of flu vaccine. Suppose it is known that 67% of all American adults are concerned about the shortage. In response to public concerns, the medical community has assigned grades to Americans based on their risk of death from flu: "A" for the elderly and babies; "B" for persons with chronic illness; "C" for pregnant women; "D" for those with low risk and "F" for those with no risk. The average duration of illness for all groups combined is 4 days with a standard deviation of 3 days. Only persons with grade "A" (highest risk) are asked to get shots, all others have been asked to wait.

A large pharmaceutical firm has developed an experimental vaccine designed to prevent babies from becoming ill from the flu and needs to test its effectiveness. To do so, a simple random sample of 1,200 grade "A" Americans was selected. After interviewing the 1,200, it was discovered that 700 were elderly. The babies were randomized into two groups: 200 controls and 300 treatment. The controls were given a shot that was nothing more than vitamins, the treatment group was given the experimental vaccine that looked exactly like the vitamins. The nurses who administered the shots could not distinguish the vitamins from the vaccine. After 30 days, 20% of the controls had become ill and 15% of the treatment group had become ill. The illness had an average duration of 6 days for the controls and duration was not normally distributed. The controls had an average age of 14.2 months and the treatment group had an average age of 14.8 months.

A. (2 points) The population of greatest interest to the pharmaceutical firm is:

B. (3 points) The parameter of greatest interest to the pharmaceutical firm is:

C. (2 points) The sample of interest to the pharmaceutical firm is:

D. (3 points) The statistic of greatest interest to the pharmaceutical firm is:

E. (2 points) This is an example of a
     (g)   an Observational Study with historical controls
     (h)   an Observational Study utilizing multi-stage cluster sampling techniques
     (i)   an Observational Study that uses a random probability method for sample selection.
     (j)   a Randomized Experiment without Controls, but it is blind
     (k)   a Randomized Experiment without Controls, but it is double-blind
     (l)   a Randomized Controlled Experiment

24.   Please use the graphic to answer the questions below: (all are 2 points each)
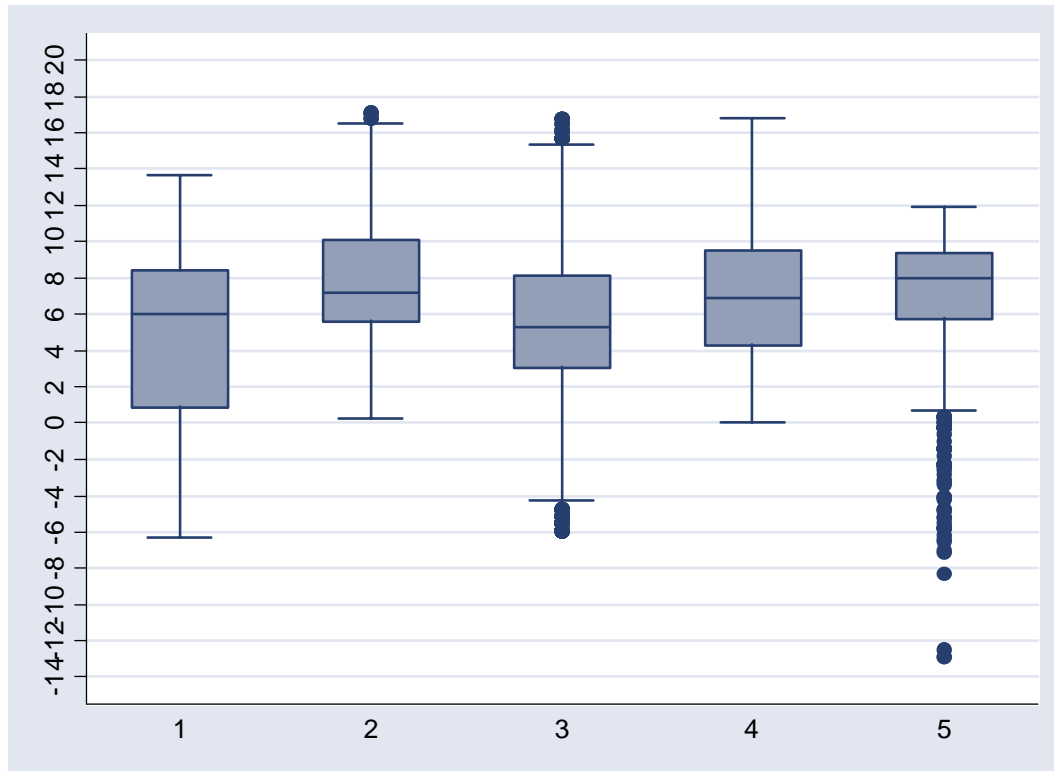
a.   Which box plot has the largest (or greatest) range? (circle one)

                1   2   3   4   5   can't be accurately determined

b.   Which box plot has the largest (or highest) median? (circle one)

                1   2   3   4   5   can't be accurately determined

c.   Which box plot has the largest (or greatest) interquartile range? (circle one)

                1   2   3   4   5   can't be accurately determined

d.   Which box plot has the largest (or highest) mean? (circle one)

                1   2   3   4   5   can't be accurately determined

e.   Which box plot has the largest (or highest) Q3 or 3rd Quartile? (circle one)

                1   2   3   4   5   can't be accurately determined

f.   Which box plot appears to have the strongest left skewness? (circle one)

                1   2   3   4   5   can't be determined

g.   Which box plot appears to have the strongest right skewness? (circle one)

                1   2   3   4   5   can't be determined

25.   There are 20,000 restaurants in the County of Los Angeles, 50% of them received a letter grade of "A" during inspections, 40% received either a B or a C grade and 10% failed their inspections. Restaurant grades are not normally distributed. My financial adviser, the Oracle, has hired you as a temporary personal assistant. Your job is to schedule his next 100 dinners (Oracle never eats at home). Unfortunately, you didn't know about the rating system and you never eat out because you don't have the money. So you listened to your best friend and picked 100 restaurants at random from an internet database of the 20,000 restaurants in Los Angeles. The Oracle will give you +3 points if you choose "A" restaurants, +1.25 points if you choose "B" or "C" restaurants, and -20 points if you choose a restaurant with a failing grade. Treat your restaurant selections as if they were a simple random sample of restaurants.

A. Construct a probability distribution for this problem

B. What is the mean of this probability distribution?

C. What is the standard deviation of this probability distribution?

26. You know that every UCLA student will definitely get a job after graduation. The only uncertainty is the salary. Suppose this is what you know about the job prospects of UCLA students after graduation:

There is a 35% chance that the salary will be $20,000 per year; a 45% chance that it will be $90,000 per year; and a 20% chance that it will be $40,000 per year.   Suppose you draw a random sample of 121 UCLA students.

    a.   What is the probability distribution for the variable salary?

    b.   Find the mean of the probability distribution.

    c.   What is the Standard Deviation of probability distribution?

27. It's a family tradition: your professor goes to Las Vegas every year for Thanksgiving. A new casino has opened and they are playing a modified roulette game that has 40 possible numbers that can be spun on a wheel. To play you bet $4 and you get to choose 4 numbers. If the wheel lands on any number that you chose, you win $10. If the wheel does not land on a number you choose but on one of 8 "special numbers" you don't win or lose anything. If the wheel lands on any of the remaining numbers (not the ones you choose or the special numbers), you lose your bet of $4. Suppose the typical person plays 25 times and it could be treated as if it were a simple random sample.

a.   This game of modified roulette can be represented by a probability distribution, please construction one in the space below.

b.   Find the mean of this game.

c.   Find the standard deviation of this same game.

28. Some friends take you to a casino and you are confronted with two games.

GAME A works like this: you can bet $8 on a number, and if your number comes up, you win $11, if not, you lose your $8. Your number comes up 35% of the time.

GAME B works like this: you can bet $3 on a number, and if your number comes up, you win $2, if not, you lose your $3. Your number comes up 55% of the time.

A.   Please construct probability distributions for each game in the space below. Please label them clearly so we know which represents GAME A and which represents GAME B.

B.   Please calculate the mean and standard deviation for each game. Again please label them clearly so we know which is which.

29. A student issued the following command for analysis variable API03

```
. summarize api03, detail
                        API Score
-------------------------------------------------------------
      Percentiles      Smallest
  1%         616          522
  5%         670          580
 10%         704          590        Obs                   684
 25%         750          598        Sum of Wgt.           684

 50%         807                     Mean             800.9444
                      Largest        Std. Dev.        74.33433
 75%         854          964
 90%         894          974        Variance         5525.593
 95%         915          980        Skewness        -.2930783
 99%         952          989        Kurtosis         2.939588
```

Please answer the following three questions based on the Stata results for variable API03.

A.   Using the Stata results above, please list the values of a five number summary for variable API03 ordered from the lowest to the highest (5 POINTS)

B.   Which statement best describes this distribution based on the Stata results given above: (3 POINTS)

        a. API score is right skewed
        b. API score is left skewed
        c. API score is not skewed
        d. API score is negatively skewed right
        e. API score is positively skewed left
        f. At least two of the above are true

C.   Does the variable API03 have outliers (circle one):   YES   NO   (ONE POINT)

Using information from the Stata results above, show how you arrived at your conclusion. (3 POINTS)