36-201 INTRODUCTION TO STATISTICAL REASONING - FINAL

May 7, 2004

You must show your work and explain your steps in order to get full credit.

Carry calculations to completion, and use enough decimal places to get an accurate answer. You should also comment on the numerical results whenever it seems appropriate.

You may use one (two sided) sheet of notes (8.5 by 11 inches) and a calculator. You may not share a calculator, pencil, paper or anything else during the exam.

Your Student ID: | | | | | | | | | |

Your First Name: _____ **Your Last Name:** _____

Your Section: _____

Your Signature: _____

Grader use:

Problem	Total	Correct
1-25	100	
26	10	
27	10	
28	14	
29	18	
30	17	
31	29	

Problem	Total	Correct
32	18	
33	20	
34	25	
35	23	
36	25	
37	28	
38	38	
Total	375	

***** Do not turn this page until instructed to do so.*****

Your Student ID: | | | | | | | | | |

PLEASE CHECK THAT YOU HAVE PAGES 2 THROUGH 27 + A TABLE

PART I: Multiple Choice Questions (4 points each)

Questions 1 and 2 are related to the following study:

A store asked 250 of its customers whether they were satisfied with the service or not. The response was also classified according to the gender of the customer. We want to study whether there is a relationship between satisfaction and gender.

1. A meaningful display of the data from this study will be

   (a) side–by–side box plots
   (b) a pie-chart
   (c) a histogram
   (d) a scatterplot
   (e) a two-way table

2. The statistical procedure that should be used to analyze this data is:

   (a) a 2-sample t-test
   (b) a paired t-test
   (c) ANOVA
   (d) a chi-squared test
   (e) simple linear regression

3. Which of the following statistical measures is **not** affected by outliers in the data?

   (a) the standard deviation.
   (b) the mean.
   (c) the correlation coefficient.
   (d) the median.
   (e) all of the above.

Your Student ID: | | | | | | | | | |

Questions 4 and 5 are related to the following information:

A study was conducted in order to compare the effectiveness of four leading brands of pain relievers. The researchers tested the four brands of medicine on four groups of randomly selected patients (each of the groups received one of the brands), and recorded the relief time (in minutes) for each patient.

4. A meaningful display of the data from this study will be

   (a) side–by–side box plots
   (b) a pie-chart
   (c) a histogram
   (d) a scatterplot
   (e) a two-way table

5. The statistical procedure that should be used to analyze this data is:

   (a) a 2-sample t-test
   (b) a paired t-test
   (c) ANOVA
   (d) a chi-squared test
   (e) simple linear regression

6. A magazine reported that 4000 adult Americans mailed in a questionnaire answering the following question: "Do you think your coworkers should bathe more often?" Of those responding, 55% said yes. This survey:

   (a) gives pretty accurate and reliable results because it is based on a large simple random sample.
   (b) probably understates the percentage of people who think that their coworkers should bathe more often.
   (c) probably overstates the percentage of people who think that their coworkers should bathe more often.
   (d) uses voluntary response as its sampling method
   (e) both (c) and (d)

Your Student ID: | | | | | | | | | |

Questions 7 and 8 are related to the following study:
The director of admissions in a small college administered a newly designed entrance test to 20 students selected at random from the new freshman class. The purpose of this study was to determine whether student's grade point average (GPA) at the end of the the freshman year can be predicted from the entrance test score.

7. At the end of the year when all the data is available, a meaningful graphical display of this data will be

   (a) side–by–side box plots
   (b) a pie-chart
   (c) a histogram
   (d) a scatterplot
   (e) a two-way table

8. To supplement the graphical display, a meaningful numerical measure for this data is:

   (a) the median
   (b) the correlation coefficient
   (c) standard deviation
   (d) conditional percentage
   (e) the Chi-squared statistic

9. According to the central limit theorem, the sampling distribution of $\bar{X}$:

   (a) is normal regardless of the shape of the population distribution, as long as the sample size, n, is large enough.
   (b) is normal regardless of the sample size, if it is known that the population is normal.
   (c) has the same mean as the population mean.
   (d) has a S.D. that is smaller than the population S.D.
   (e) all of the above

10. The main advantage of experiments over observational studies is that

    (a) a well-designed experiment can give good evidence that the treatment actually causes the response
    (b) an experiment can compare two or more groups
    (c) an experiment is always cheaper
    (d) an experiment is always shorter
    (e) we can study the relationship between two or more explanatory variables

Your Student ID: | | | | | | | | | |

11. A publishing company wanted to test whether the typing speed of secretaries differs when using word processor A or word processor B. A random sample of 25 secretaries was selected and the typing speeds (in words per minute) were recorded for each secretary when using word processor A and then when using word processor B. (Which word processor is used first is determined for each secretary by a coin flip). The statistical procedure that should be used to analyze these data is:

    (a) a 2-sample t-test
    (b) a paired t-test
    (c) ANOVA
    (d) a chi-squared test
    (e) simple linear regression

12. A publishing company wanted to test whether the typing speed of secretaries differs when using word processor A or word processor B. The typing speeds (in words per minute) are recorded for a random sample of 25 secretaries using word processor A, and for another (different) random sample of 25 secretaries using word processor B. The statistical procedure that should be used to analyze these data is:

    (a) a 2-sample t-test
    (b) a paired t-test
    (c) ANOVA
    (d) a chi-squared test
    (e) simple linear regression

13. The most important reason for the use of random allocation of subjects to the different treatments is:

    (a) to ensure that each subject in the trial receives the best possible treatment
    (b) to guarantee that approximately same number of subjects are assigned to each treatment group
    (c) to protect the experimenters from legal action in the event that the experiment goes drastically awry.
    (d) to ensure that the different treatment groups are as similar as possible in every way $except$ for the treatment received.
    (e) to guarantee that the results of the experiment can never be duplicated.

Your Student ID: | | | | | | | | | |

14. When conducting a survey, it is important to use *random* sampling in order:

    (a) to get a sample that represents the population well.
    (b) to reduce bias resulting from poorly worded questions.
    (c) to reduce bias resulting from poorly ordered questions.
    (d) to reduce bias resulting from sensitive questions.
    (e) None of the above.

15. High blood pressure is unhealthy. Here are the results of one of the studies that link high blood pressure to death from cardiovascular disease. The researchers classified a group of white males aged 35 to 64 as Low blood pressure or High, then followed the subjects for five years. The following two-way table gives the results of the study

	Blood pressure		
Cardiovascular death?	Low	High	Total
Yes	21	55	76
No	2655	3283	5938
Total	2676	3338	6014

    In this example it would be appropriate to calculate:

    (a) conditional row percentages
    (b) conditional column percentages
    (c) the correlation coefficient r
    (d) the 5 number summary of both variables
    (e) None of the above

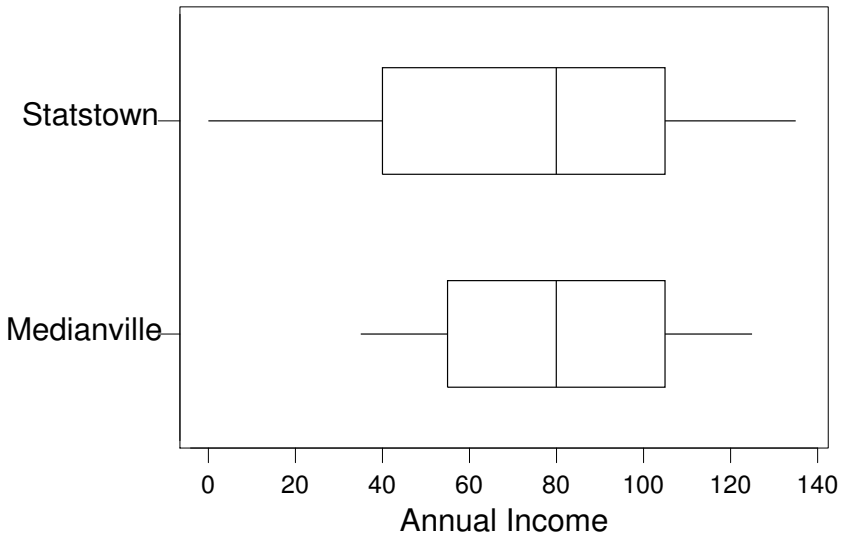
16. Let X be a Binomial random variable with $n = 600$ and $p = 0.4$. According to the *Normal Approximation to the Binomial*, X is approximately

    (a) Normal with mean $\mu = 600$ and standard deviation $\sigma = 0.4$.
    (b) Normal with mean $\mu = 0.4$ and standard deviation $\sigma = .02$.
    (c) Normal with mean $\mu = 12$ and standard deviation $\sigma = 240$.
    (d) Normal with mean $\mu = 240$ and standard deviation $\sigma = 12$.
    (e) None of the above, since the rules of thumb (which guarantee that the normal approximation "works"), are not satisfied in this case.

Your Student ID: | | | | | | | | | |

Questions 17 through 20 are related to the following information:

The boxplots below display annual incomes (in thousands) for households in two cities.



Choose from the following options in answering questions 16-19 below.

A=impossible to tell
B=Medianville
C=Statstown
D=equal

17. Which city has more households? _____

18. Which city has greater variability in income? _____

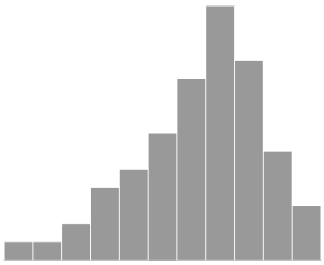
19. Which city has a greater percentage of households with annual income above $80,000? _____

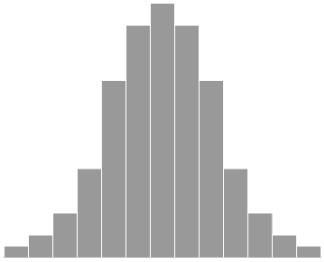
20. Which city has a greater percentage of households with annual income between $50,000 and $80,000? _____

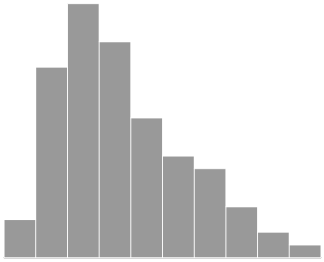
Your Student ID: | | | | | | | | | |

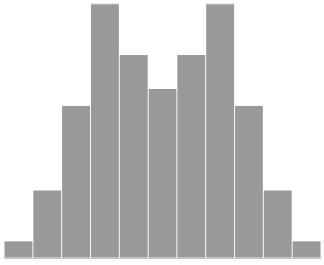
21. For each of the four distributions below, decide whether:
    A. the mean is approximately equal to the median, or
    B. the mean is greater than the median, or
    C. the mean is less than the median

    (Clearly mark A,B, or C next to each graph)









Your Student ID: | | | | | | | | | |

22. If you were answering this problem with a random guess, the probability of getting the correct answer would be:

    (a) 1/2
    (b) 1/3
    (c) 1/4
    (d) 1/5
    (e) None of the above

23. If A and B are two *disjoint* events such that $P(A) = 1/3$ and $P(B) = 1/2$, then the probability of both events A *and* B occurring is equal to:

    (a) 1/6
    (b) 5/6
    (c) 2/3
    (d) 0
    (e) cannot determine without additional information

24. If A, and B are two *independent* events such that $P(A) = .3$ and $P(B) = .4$, then the probability of neither event A nor event B occurring $[P(A^c \text{ and } B^c)]$ is equal to:
    (**Hint:** Use the space below to create a probability table using the above information)

    (a).12
    (b).28
    (c).42
    (d).18
    (e).58

25. Let A and B be two *independent* events. If $P(A) = .5$, what can you say about $P(A \mid B)$?

    (a) cannot find it since $P(B)$ is not known
    (b) cannot find it since $P(A \ \ and \ \ B)$ is not known
    (c) cannot find it since both $P(B)$ and $P(A \ \ and \ \ B)$ are not known
    (d) it is equal to .5
    (e) it is equal to .25

Your Student ID: | | | | | | | | | |

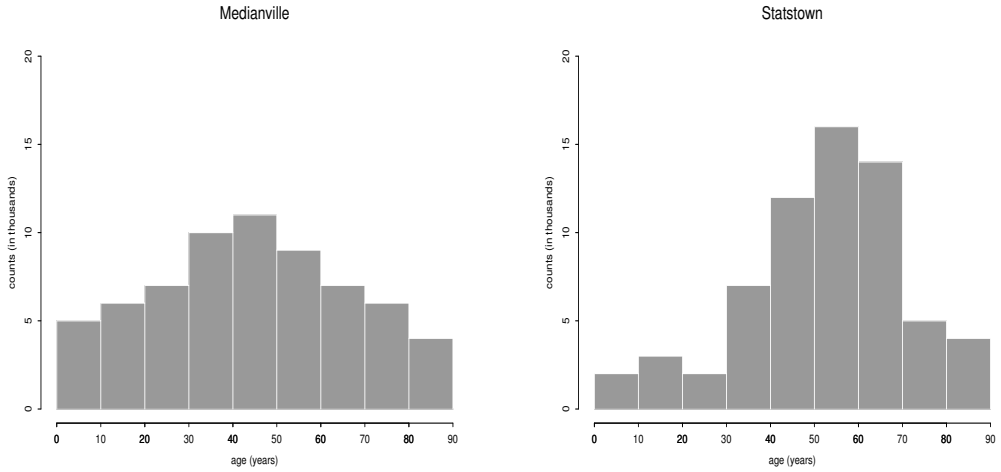
PART II

26. (10 points) A survey taken in a large statistics class contained the question "What's the fastest you have driven a car (in mph)?" The five-number summary for the 87 males surveyed is:

min=60 Q_1=95 Median=110 Q_3=120 Max=150

Should the largest observation in this data set be classified as an outlier? **Show work!**

27. (10 points) The histograms below show the distribution of ages of the people living in two cities - Statstown and Medianville:



Both cities have a "Summer Fair" and need to choose a sample of 9 committee members to organize it. Instead of randomly choosing 9 people from the entire population, both cities decide (independently) to randomly choose one person from each of the 9 age groups that are represented in the histograms. In which of the two cities will the resulting sample be more representative of the entire city (in terms of age). **Explain!**

Your Student ID: | | | | | | | | | |

28. The *College Student Journal* (December 1992) investigated differences in traditional and non-traditional students, where non-traditional students are generally defined as those 25 years old or older.
Based on the study results, we can assume that the GPA of non-traditional college students in the U.S. has mean $\mu = 3.5$ and standard deviation of $\sigma = .5$.

    (a) (6 points) According to the Central Limit Theorem, what is the sampling distribution of $\bar{\mathbf{X}}_{\mathbf{400}}$, the average GPA of 400 randomly chosen non-traditional students?

    (b) (8 points) How likely is $\bar{\mathbf{X}}_{\mathbf{400}}$, the average GPA of a random sample of 400 non-traditional college students to be above 3.57?

Your Student ID: | | | | | | | | | |

29. A dashboard warning light is supposed to flash red if a car's oil pressure is low $(\mathbf{L})$. In a certain model, the probability of the light flashing $(\mathbf{F})$ when it should (i.e., when the pressure is really low) is 0.95; 2% of the the time, though, it flashes for no apparent reason (i.e. when the oil pressure is normal). Assume that in this model there is 10% chance that the oil pressure is really low.

    (a) (6 points) Make a tree diagram for this problem. Be clear in your notations!

    (b) (6 points) What is the probability that the warning light is flashing?

    (c) (6 points) Given that the warning light is flashing, What is the probability that a driver needs to be concerned (i.e. the oil pressure is really low)?

Your Student ID: | | | | | | | | | |

30. Your favorite pizza place "Slice - 201" delivers only one kind of pizza which is sold for \$10, and costs the pizza place \$4 to make. "Slice - 201" has the following policy regarding delivery:
• If delivery is made more than half an hour (but less than an hour) after the order has been placed you get the pizza for half price (i.e. for \$5)
• If delivery is made more than an hour after the order has been placed you get the pizza for free.
Experience has shown that only in 10% of the time delivery takes between half an hour to an hour, and in only 5% of the time delivery takes more than an hour.
Let the random variable X be the net gain (profit) of "Slice - 201" for one delivered pizza.

(a) (6 points) Find the probability distribution of X (i.e. a table of the possible values of X and the corresponding probabilities)

(b) (5 points) Find the mean gain per pizza for "Slice - 201"? (i.e. find μ_X).

(c) (6 points) If "Slice - 201" wanted to increase the mean gain per pizza to \$5.45, how much should they charge for the pizza (instead of \$10)?

Your Student ID: | | | | | | | | | |

31. It is known that roughly 8% (0.08) of males have some sort of color vision deficiency (also known as color-blindness). A random sample of 40 males was chosen for a study about color-blindness. Let $\mathbf{X}$ be the number of males (out of the 40) who are color-blind.

    (a) (5 points) Explain why X is a Binomial random variable and give the values of n and p.

    (b) (6 points) What is the probability that exactly 2 of the 40 males in the study are color-blind.

    (c) (6 points) What is the probability that at least one of the 40 males in the study is color-blind?

    (d) (4 points) What are the mean and standard deviation of X? (i.e., find μ_X and σ_X)

    (e) (8 points) The study has two stages. First, each of the 40 males is taking a test to check whether he is color-blind or not (cost: \$50 per test). Then, all those who were found to be color-blind are given another, more detailed, test in order to detect the specific nature/type of their color-blindness (cost: \$250 per test). Find the mean and standard deviation of the **total cost** of this study.

Your Student ID: | | | | | | | | | |

32. The faculty senate at a large university wanted to estimate p, the proportion of the students who thought a foreign language should be required for everyone. The Dean of the Engineering School offered to cooperate in conducting the survey, and a random sample of 450 students was selected from all the students enrolled in engineering classes. 171 out of these 450 students favored the foreign language requirement.

    (a) (5 points) What is the point estimate for p in this case? (i.e., if you had to estimate p based on this study by a $single$ number, what would it be?)

    (b) (8 points) Find a 95% confidence interval for p, and interpret it in context.

    (c) (5 points) What features of this survey should make you nervous about the use of this confidence interval to estimate the true proportion of the university students who favor the foreign language requirement?

Your Student ID: | | | | | | | | | |

33. Math SAT (SATM) scores of seniors from a large metropolitan area follow a normal distribution with $unknown$ mean μ and standard deviation 100. A random sample of 400 seniors yields an average SATM score of 480.

(a) (5 points) find a 95% confidence interval for μ.

(b) (5 points) Which of the following will provide a shorter (narrower) confidence interval than the one in (a)? (Circle **all** the correct choices):

(i) Use a sample size of 100.
(ii) Use a sample size of 1000.
(iii) Use a different sample of size 400.
(iv) Use a 90% confidence interval
(v) Use a 99% confidence interval

(c) (5 points) How large a sample of seniors is needed in order to estimate μ with a 95% confidence interval of $length$ 15?

(d) (5 points) We want to test:

$$H_0 : \mu = 500$$
$$H_a : \mu \neq 500$$

for $\alpha = 5\%$.

Using the confidence interval; that you found in part (a): (circle one)

(i) We reject H_0 since 500 falls inside the confidence interval.
(ii) We do not reject H_0 since 500 falls inside the confidence interval.
(iii) We reject H_0 since 500 falls outside the confidence interval.
(iv) We do not reject H_0 since 500 falls outside the confidence interval.

Your Student ID: | | | | | | | | | |

34. Social psychologists at the University of California at Berkeley wanted to study the effect that staring at drivers would have on driver behavior (**Source:** Ellsworth, Carlsmith and Henson, (1972). "The stare as a stimulus to flight in human subjects: A series of field experiments", *Journal of Personality and Social Psychology*, Vol. 21, pp.302-311).
In a randomized experiment, the researchers either stared or did not stare at the drivers of automobiles stopped at the campus stop sign. The researchers then timed how long it took each driver to proceed from the stop sign to a mark on the other side of the intersection. (crossing times are measured in seconds).

The researchers would like to use the data in order to determine whether there are differences in crossing times between the two groups.

(a) (5 points) The following two outputs are available. Circle the one which represents the appropriate way to analyze the results of this study as it was conducted. Explain briefly!

Two-Sample T-Test and CI: Stare (NO/YES)

```
Two-sample T for Stare

             N      Mean     StDev   SE Mean
No          20      6.62      1.13      0.25
Yes         20     5.595     0.653      0.15

Difference = mu (No ) - mu (Yes)
Estimate for difference:  1.025
95% CI for difference: (0.430, 1.620)
T-Test of difference = 0 (vs      ): T-Value = 3.516 P-Value = 0.001
```

Paired T-Test and CI

```
Paired T for No - Yes

                  N      Mean     StDev   SE Mean
No Stare         20     6.620     1.127     0.252
Stare            20     5.595     0.653     0.146
Difference       20     1.025     1.041     0.233

95% CI for mean difference: (0.538, 1.512)
T-Test of mean difference = 0 (vsł       ): T-Value = 4.40 P-Value = 0.000
```

Your Student ID: | | | | | | | | | |

(b) (4 points) State the appropriate hypotheses H_0 and H_a. Define clearly any parameter(s) that you are using.

(c) (8 points) Give the test statistic and p-value for this test, and state your conclusion in context.

(d) (4 points) Explain how you could have used the confidence interval that is in the output to reach the same conclusion you did in part (c).

(e) (4 points) Interpret the confidence interval that is the output, and explain what it tells you about the nature of the effect that staring has on the driver's behavior.

Your Student ID: | | | | | | | | | |

35. A study was conducted in order to determine whether males and females **differ** in the amount of time they spend watching TV between 8:00 PM and 11:00 PM. For convenience, the researchers decided to randomly choose and get data from 40 married couples. Each of the 80 subjects that participated in the study reported the total number of minutes he/she watched TV between 8:00 PM and 11:00 PM during a given week.

    (a) (5 points) The following two outputs are available. Circle the one which represents the appropriate way to analyze the results of this study as it was conducted. Explain briefly!

Two-Sample T-Test and CI: Males, Females

```
Two-sample T for Males vs Females

          N      Mean     StDev   SE Mean
Males    40       343       222         35
Females  40       355       210         33

Difference = mu Males - mu Females
Estimate for difference:  -12.0
95% CI for difference: (-108.2, 84.3)
T-Test of difference = 0 (vs not =): T-Value =       P-Value = 0.805
```

Paired T-Test and CI: Males, Females

```
Paired T for Males - Females

                  N      Mean     StDev   SE Mean
Males            40     343.3     222.2      35.1
Females          40     355.2     209.9      33.2
Difference       40    -11.98     47.50      7.51

95% CI for mean difference: (-27.17, 3.22)
T-Test of mean difference = 0 (vs not = 0): T-Value =        P-Value = 0.119
```

Your Student ID: | | | | | | | | | |

(b) (4 points) State the appropriate hypotheses H_0 and H_a. Define clearly any parameter(s) that you are using.

(c) (9 points) Find the test statistic, give the p-value, and state your conclusion in context.

(d) (5 points) As stated, the purpose of this study was to check whether males and females differ in the amount of time they spend watching TV between 8:00 PM and 11:00 PM. Do you think that the way this study was designed by the researchers leads to reliable results? If yes, explain why. If no, explain why and suggest a different way to design this study which will lead to more reliable results.

Your Student ID: | | | | | | | | | |

36. A large company buys thousands of light bulbs every year. The company is currently considering four brands of light bulbs to choose from. Before the company decides which light bulbs to buy, it wants to investigate if the mean life of the four types of light bulbs is the same or not. The company's research department randomly selected 7 bulbs of each type and tested them. The data is the number of hours (in hundreds) that each of the bulbs in each brand survived before burning out. The following *minitab* output is available.

One – way ANOVA

```
SOURCE     DF        SS        MS        F        p
FACTOR      3     97.54     32.51     5.44    0.005
ERROR      24    143.43      5.98
TOTAL      27    240.96

                                        INDIVIDUAL 95 PCT CI'S FOR MEAN
                                        BASED ON POOLED STDEV
  LEVEL      N      MEAN     STDEV  -----+---------+---------+---------+-
Brand I      7    23.143     2.268            (-------*------)
Brand II     7    20.714     2.289  (-------*------)
Brand III    7    24.571     2.299                  (------*-------)
Brand IV     7    25.714     2.870                      (-------*------)
                                        -----+---------+---------+---------+-
POOLED STDEV =    2.445              20.0      22.5      25.0      27.5
```

(a) (5 points) What are the null and alternative hypotheses of interest in this case? Define clearly the parameters that you are using.

(b) (4 points) Is it reasonable to assume that the assumption of equal standard deviations is satisfied in this case? Explain!

Your Student ID: | | | | | | | | | |

(c) (8 points) Test your hypotheses for $\alpha = .05$ (give the test statistic and P-value), and state your conclusions in context.

(d) (8 points) Give an interpretation of the value of the P-value in the context of the problem. (**Note:** I am not asking you to repeat what you said in part (c) about whether you can or cannot reject H_0, but rather explain what the value of the P-value tells you about the data you observed).

Your Student ID: | | | | | | | | | |

37. A study was done in order to determine whether the type of crime a person is convicted of is related to whether that person is a drinker or not. Data was collected on the number of people convicted of the five crimes listed, classified according to whether they were drinkers or not.

The following (edited) Minitab output summarizes the results:

```
Expected counts are printed below observed counts |
                                                  |
                          non-                    |
                Drinker   Drinker   Total         |
                                                  |(c) - show work!
Arson              50        43       93         |
                49.04     43.96                   |
                                                  |
Rape               88        62      150         |
                79.09     70.91                   |
                                                  |
Violence          155       110      265         |
                139.72    125.28                  |
                                                  |
Stealing          379       300      679         |
                          320.99                  |
                                                  |(d) - show work!
Fraud              63       144      207         |
                109.14     97.86                  |
                                                  |
Total             735       659     1394         |
                                                  |
                                                  |
ChiSq =  0.019 +  0.021 +                         |
         1.004 +  1.120 +                         |
         1.670 +  1.863 +                         |
               +  1.373 +                         |
        19.508 + 21.758 =                         |
                                                  |
df = 4  , P-value = 0.000                         |
```

(a) (4 points) Is this study an experiment? Why or why not?

(b) (4 points) State the appropriate null hypothesis and alternative hypothesis of this study.

Your Student ID: | | | | | | | | | |

(c) (4 points) What is the *expected* number of people who were drinkers convicted of stealing? (Show work on previous page, and complete the output).

(d) (4 points) Find the Chi-Squared test statistic. (Show work on previous page, and complete the output).

(e) (4 points) State your conclusion regarding the significance of the relationship between the type of crime a person is convicted of and whether that person is a drinker or not. (use $\alpha = 5\%$)

(f) (4 points) If we repeat the analysis using the data for the first four types of crime only (i.e., without "Fraud"), we get the following results (only the relevant part of the output is given):

```
ChiSq =  0.133 +  0.174 +
         0.112 +  0.146 +
         0.165 +  0.215 +
         0.076 +  0.099 = 1.120

df = 3, P-value = .7722
```

What would be your conclusion regarding the significance of the relationship between type of crime and drinking in this case?

(g) (4 points) It is obvious that whether or not you include "Fraud" in the study makes a big difference – you get different conclusions! Can you reason why "Fraud" is such a "key player" in the relationship between type of crime and drinking? (use common sense)

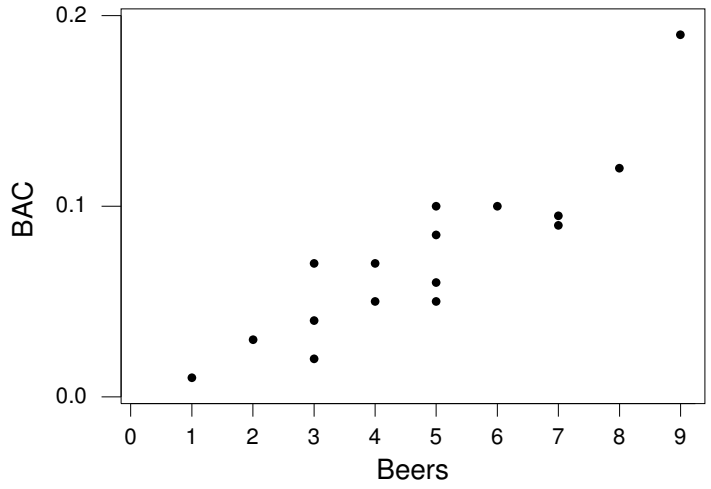
Your Student ID: | | | | | | | | | |

38. How well does the number of beers a student drinks predict his or her Blood Alcohol Content (BAC). Sixteen students at Ohio State University drank a randomly assigned number of cans of beers. Thirty minutes later, a police officer measured their BAC.

Below are two MINITAB outputs:
I. Scatterplot of the data.
II. Model fit of the simple regression of BAC vs. BEERS.

The questions follow the outputs on the next page.

I.



II.

Regression Analysis: BAC versus Beers

The regression equation is
BAC = - 0.0127 + 0.0180 Beers

Predictor	Coef	SE Coef	T	P
Constant	-0.01270	0.01264	-1.00	0.332
Beers	0.017964	0.002402	7.48	0.000

S = 0.02044 R-Sq = 80.0% R-Sq(adj) = 78.6%

Your Student ID: | | | | | | | | | |

(a) (4 points) Describe the relationship between BAC and $\textsc{Beers}$ as it appears in the scatterplot. (Be sure to mention the direction in context, and the form of the relationship.)

(b) (4 points) Find the value of the correlation coefficient r, and interpret it in context.

(c) (3 points) If beer consumption were measured in fluid onces (1 can = 12 FL. OZ.), what would be the value of the correlation coefficient r?

(d) (4 points) Write down the equation for the least squares regression line for these data. What is $\hat{\beta}_{\mathbf{0}}$, the estimate of the y-intercept, and what is $\hat{\beta}_{\mathbf{1}}$, the estimate of the slope?.

(e) (4 points) Give an interpretation of $\hat{\beta}_{\mathbf{1}}$ in the context of the problem.

(f) (4 points) Use the regression equation, to predict the "normal" BAC level (i.e., the BAC level when no beer is consumed). Comment on the result you got.

(g) (7 points) Test the significance of the linear relationship between BAC and $\textsc{Beers}$. Be sure to state the appropriate H_0 and H_a, give the test statistic, P-value, and to state your conclusion in context.

(h) (4 points) Find a 95% confidence interval for the slope β_1.

(i) (4 points) What is value of R^2 in this case? What does the value of R^2 tell us about the fit of the linear model to the data?