

**MGMT 230: Introductory Statistics**

**Take Home Final Exam**

**Due *in my office* Tuesday, May 5, 12pm (late exams get zero credit).**

**Name (15 points):**

For every problem, clearly state if any further assumptions need to be made to solve the problem, and if so, what assumption. To show your work, whenever you use a formula to answer a question, write down the formula first, then write down the formula with the correct numbers plugged in.

For every hypothesis test, conduct every step of a hypothesis test, and also report and interpret the p-value for every test. Use any significance level you think is appropriate.

1. Suppose a computer processor manufacturer is interested in estimating the proportion of processors that are defective. A sample of 1500 processors is taken and 8 are found to be defective.
  - (a) (5 points) If the true population proportion of defective processors is 0.4% what is the probability of finding at least 8 defective processors in the sample of 1500?
  - (b) (10 points) If the true population proportion of defective processors is 0.4%, describe the sampling distribution of the sample estimate. Describe the shape of the distribution, the center of the distribution, and the standard deviation of the distribution. Also describe what the sampling distribution means.
  - (c) (15 points) Test the hypothesis that the proportion of defective processors is less than 1%.
2. Suppose the city of La Crosse is interested in determining whether the price of homes have declined since 2005. A sample of 50 home sales from 2005 had a mean sale price of \$160,000 and a standard deviation of \$25,000. A sample of 40 home sales from 2009 had a mean sale price of \$155,000 and a standard deviation of \$20,000. Assume that the population variance of home sales has been the same in 2005 and 2009.
  - (a) (10 points) Construct and interpret a 90% confidence interval for the difference in mean sale prices between 2007 and 2005.
  - (b) (15 points) Test the hypothesis that the average sale price was lower in 2007 than in 2005.
3. (5 points) List the assumptions behind regression analysis. Which assumptions are absolutely necessary? What assumptions are allowed to be violated such that the estimates of your regression coefficients are still consistent?
4. (5 points) What happens to the standard deviation of the sampling distribution of any statistic when you increase the sample size?

5. (25 points) In April 1979 the Albuquerque Police Department began a special enforcement program aimed at countering DWI (driving while intoxicated) accidents. The program was composed of a squad of police officers, Breath Alcohol Testing devices, and random checkpoints to test drivers for driving while intoxicated. The police department collected data on the number of injuries and accidents that occurred during nighttime hours for 29 weeks prior to instituting the program, and 23 weeks during the DWI prevention program. Surprisingly, they found there were more accidents and injuries during the program. The police department attributed this to the fact there is usually more traffic during that part of the year. This hypothesis was confirmed by finding that the total gasoline sold in Albuquerque during the weeks of the DWI prevention program was greater than the 29 weeks prior to the program.

The police department is therefore interested in determining whether injuries as a ratio of gasoline consumption (a proxy measure for the percentage of injuries over the total amount of traffic) decreased as a result of the DWI prevention program. The Excel sheet on the class website shows the actual number of nighttime injuries for each week, and the total amount gasoline consumption (in millions of gallons) for each week in Albuquerque.

The data for on injuries and accidents is given in the Excel file, `albuquerque.xls`, posted on the class website. Test the hypothesis that the DWI program was effective at decreasing the number of nighttime injuries (as measured per unit of gasoline consumption).