Review for Final
The final is cumulative, but will mostly have questions mostly from chapters 5, 6, and 7. You may have two pages (front and back) of notes, and a calculator. The z-table and t-table will be provided. If English is not your native language you may use a translation dictionary, but nothing with texting/internet capability.

The test will contain:

1) topics from chapters 1-4 (see the midterm review for details)

2) Confidence interval questions. Make sure you know how to compute and interpret a confidence interval for each type of situation:
   a) paired data (section 5.1)
   b) one sample mean (large sample), quantitative variable, using the normal curve (section 4.2)
   c) one sample mean (small sample), quantitative variable, $\sigma$ is unknown, using t-distribution (section 5.3)
   d) difference of two sample means, quantitative variable, using t-distribution (section 5.4)
   e) one sample, categorical variable, $p$ is unknown (section 6.1)
   f) difference of two sample proportions, categorical variable (section 6.2)

3) Hypothesis testing. Do all the steps: (write $H_0$, $H_a$, gather data, do computations, get $p$-value, decision, conclusion). Make sure you know how to do a hypothesis test for each of the following situations:
   a) paired data (section 5.1)
   b) one sample mean (large sample), quantitative variable, using the normal curve (section 4.3)
   c) one sample mean (small sample), quantitative variable, $\sigma$ is unknown, using t-distribution (section 5.3)
   d) two sample means, quantitative variable, using t-distribution (section 5.4)
   e) one sample, categorical variable, $p$ is unknown (section 6.1)
   f) two sample proportions, categorical variable (section 6.2)

4) Two quantitative variables (chapter 7)
   a) describe a scatterplot
   b) match correlation to a scatterplot
   c) calculate a least-squares regression line
   d) use a regression line to predict the value of a response variable
   e) use a residual plot to determine if a line is a good fit
   f) calculate residuals
   g) determine if extrapolation is valid in a given situation
   h) use $R^2$ to explain the amount of variation in the response variable can be explained by the explanatory variable.
Sample Questions

1. (multiple choice) A level 90% confidence interval is
   a. any interval with margin of error ± 0.90.
   b. an interval computed from sample data by a method that has probability 0.90 of producing
      an interval containing the true value of the parameter of interest.
   c. an interval with margin of error ± 0.90, which is also correct 90% of the time.
   d. an interval computed from sample data by a method guaranteeing that the probability the
      interval computed contains the parameter of interest is 0.90.

2. A medical researcher treats 100 subjects with high cholesterol with a new drug. The average
   decrease in cholesterol level is \( \bar{x} = 80 \) after two months of taking the drug. Assume that the
   decrease in cholesterol after two months of taking the drug follows a normal distribution, with
   unknown mean and standard deviation \( \sigma = 20 \).
   a) What is the 90% confidence interval?
   b) Multiple choice: Which of the following would produce a confidence interval with a smaller
      margin of error than the 90% confidence interval you computed above?
      a. Give the drug to only 25 subjects rather than 100, since 25 people are easier to
         manage and control.
      b. Give the drug to 500 subjects rather than 100.
      c. Compute a 99% confidence interval rather than a 90% confidence interval. The
         increase in confidence indicates that we have a better interval.
      d. None of the above.
   c) Suppose the true population mean decrease in cholesterol level was \( \mu = 76 \). Make a sketch of the
      sampling distribution in this situation.
   d) Based on your answer from part c, what is the probability of getting a sample mean \( \bar{x} \) that is
      greater than or equal to 80?

3. Can aspirin help prevent heart attacks? A large medical experiment involving 22,000
   men was conducted. 11,000 randomly selected men took an aspirin every other day,
   while the other 11,000 took a placebo every other day. After several years, it was found
   that the aspirin group had significantly fewer heart attacks than the placebo group.
   a) Let \( p_1 \) = the percentage of men in the aspirin group who had a heart attack, and
      let \( p_2 \) = the percentage of men in the placebo group who had a heart attack. Write
      the null and alternate hypothesis for this experiment
   b) Explain what the phrase “significantly fewer heart attacks” means.

4. (multiple choice) A university administrator obtains a sample of the academic records of past
   and present scholarship athletes at the university. The administrator reports that no significant
   difference was found in the mean GPA (grade point average) for male and female scholarship
   athletes \( (P = 0.287) \). This means
   a. the GPAs for male and female scholarship athletes are identical, except for 28.7% of the
      athletes.
   b. the maximum difference in GPAs between male and female scholarship athletes is 0.287.
   c. the chance of obtaining a difference in GPAs between male and female scholarship athletes
      as large as that observed in the sample if there is no difference in mean GPAs is 0.287.
   d. the chance that a pair of randomly chosen male and female scholarship athletes would have
      a significant difference in GPAs is 0.287.
5. In 1998, the Nabisco Company advertised that every 18-ounce bag of Chips Ahoy cookies contained at least 1000 chocolate chips. Students at the Air Force Academy purchased some randomly selected bags of cookies and counted the number of chocolate chips. Their data is given below:

1219 1214 1087 1200 1419 1121 1325 1345
1244 1258 1356 1132 1270 1295 1135

Find the 95% confidence interval for the average number of chocolate chips per bag.

6. In April, 2004, 432 out of a sample of 800 likely voters in Washington State planned to vote for Senator Patty Murray this November.
   a) Find the 95% confidence interval of the percentage of Washington State voters who plan to vote for Patty Murray.

7. An association of Christmas tree growers in Indiana sponsored a sample survey of Indiana households to help improve the marketing of Christmas trees. Two random samples were taken: one from urban areas, and one from rural areas. In the urban areas, 89 out of a sample of 261 households had a natural Christmas tree. In the rural areas, 64 out of a sample of 160 had a natural Christmas tree. Is this evidence that a higher percentage of households in rural areas than urban areas have natural Christmas trees?
   a) What would be the appropriate null and alternate hypotheses?
   b) Compute the z-statistic.
   c) What is the P-value?
   d) Let $\alpha = 0.05$, should the null hypotheses be accepted or rejected?
   e) What is the conclusion?

8. Do middle-aged male executives have a different average blood pressure than the general population of middle-aged males? The National Center for Health Statistics reports that the mean systolic blood pressure for males 35 to 44 years old is 128. The medical director of a company looked at the records of 72 company executives in this age group and found the mean systolic blood pressure of this sample is 126.07, and the standard deviation of this sample was found to be 15.2. Is this evidence that middle-aged male executives have a different blood pressure than 128?
   a) What would be the appropriate null and alternate hypotheses?
   b) Carry out a hypothesis test (Let $\alpha = 0.05$). What is your conclusion?
Answers to Sample Questions

1) B

2a) 80 ± 3.29.
b) B
c) a normal curve, with a mean = 76 and standard deviation = 2
d) Use table A, z = 2.00, \ P(\text{sample mean} \geq 80) = 0.0228

3a) \ H_0: p_1 = p_2 \text{ and } \ H_a: p_1 < p_2
b) The difference in the percentages too large to be explained only by random sampling; there is enough evidence that aspirin really does prevent heart attacks.

4) C

5) mean = 1238.188
sd = 94.282
n = 16
df = 15
t* = 2.131 (based on 95% confidence and df = 15)
\ SE = \frac{s}{\sqrt{n}} = \frac{94.282}{\sqrt{16}} = 23.5705
\bar{x} ± t_{\alpha/2}(SE) = 1238.188 ± (2.131)(23.5705) = 1238 ± 50 (rounded off)

6a) \hat{p}' = 0.54, \ n = 800, z* = 1.96 (based on 95% confidence)
\ SE = \sqrt{\frac{p'q'}{n}} = \sqrt{\frac{(0.54)(0.46)}{800}} = 0.017621
\hat{p} ± z_{\alpha/2}SE = 0.54 ± (1.96)(0.017621) = 54% ± 3.5% (rounded off)

7) Let \ p_1 = \text{the proportion of urban households that have natural Christmas trees}
Let \ p_2 = \text{the proportion of rural households that have natural Christmas trees}
\ p'_1 = 0.340996, \ p'_2 = 0.4
\ n_1 = 261, \ n_2 = 160
\ p'_{\text{pool}} = \frac{89 + 64}{261 + 160} = 0.36342

a) \ H_0: p_1 = p_2, \ H_a: p_1 < p_2
b) \ SE = \sqrt{\frac{(0.36342)(0.63658)}{261} + \frac{(0.36342)(0.63658)}{160}} = 0.048294
\ z = \frac{0.340996 - 0.4}{0.048294} = -1.22
c) From looking on the table, the P-value is 0.1112
d) do not reject \ H_0
e) There is not enough evidence that rural households have a higher percentage of natural Christmas trees.

8 a) $H_0: \mu = 128$, $H_a: \mu \neq 128$

b) $SE = \frac{s}{\sqrt{n}} = \frac{15.2}{\sqrt{72}} \approx 1.79134$

t = \frac{\bar{x} - \mu_0}{SE} = \frac{126.07 - 128}{1.79134} = -1.077$

df = 71, but on the table, use df = 60 (round down). P-value > 0.20 (2-tail)

Do not reject $H_0$

There is not enough evidence that a middle-aged male executive has a blood pressure that is different than other middle-aged men.