NAME: _____

Math 124 FALL 2004: Section 11 MWF 2-3
Final Exam Date: Dec 13, 2004

Instructions: Answer questions 1-8. Show all work where reasonable to do so. You have 2 hours 30 minutes. To allow others to fully concentrate at the end please do not leave in the last 10 minutes. You should submit your pages of notes with your test paper.

Question 1. (*25 points*)

Define each of the following and explain why or how it is used (in the context of an experiment)

(a.) *Randomization*

(b.) *Placebo*

(c.) *Matched pairs*

(d.) *Single-blind* and *Double-blind*

(e.) *Replication*

Question 2. (*25 points*)

Few people want to eat discolored french fries. Potatoes are kept refrigerated before being cut to make fries to prevent spoiling. However, immediate processing of cold potatoes causes discoloration due to a complex chemical reaction. The potatoes must therefore be brought to room temperature before processing. Potatoes may be either fresh picked, stored for a month at room temperature or stored for a month in a refrigerator. They will either be sliced and cooked immediately or after waiting for an hour.

(a.) Identify the factors, their levels, the treatments and a response variable for this experiment.

(b.) Describe and outline the design of this experiment (you may assume that you have 24 large potatoes).

Question 3. (*25 points*)

Suppose that you have the following data

x	12.8	12.9	12.9	13.6	14.5	14.6	15.1	17.5	19.5	20.8
y	5.5	6.2	6.3	7.0	7.8	8.3	7.1	10.0	10.8	11.0

where each x and y are a pair of measurements taken on the same individual. Note that $\sum_{i=1}^{n} y_i = 80$, $\sum_{i=1}^{n} y_i^2 = 675.16$ and $\sum_{i=1}^{n} x_i y_i = 1282.74$.

(a) Compute $\sum_{i=1}^{n} x_i$, $\sum_{i=1}^{n} x_i^2$, $\bar{x}$ and $\bar{y}$.

(b) Compute the standard deviations s_x and s_y.

(c) Compute the correlation between x and y.

(d) Interpret your correlation. What does it say about the relationship between x and y.

Question 4. (*25 points*)

A machine produces nails in a factory. Acceptable nails have a mean length 1.25in. A quality inspector working for the factory takes a random sample of 50 nails from all the nails produced each hour and decides whether or not the machine is producing acceptable nails. If the machine is

not producing acceptable nails it needs to be adjusted. For one particular hour the mean length of the 50 nails was 1.27in with standard deviation 0.05 inches.

(a) What is $\text{SE}_{\bar{x}}$?

(b) State appropriate H_0 and H_1 (null and alternative hypotheses) to test whether the machine is producing acceptable nails.

(c) Compute the test statistic and its P-value.

(d) Would you recommend that the machine be adjusted? Explain your answer.

Question 5. (*25 points*)

A biologist is doing research on how a particular growth hormone affects weight gain in mice. She takes 39 mice each 7 days old and randomly chooses 17 of these mice to receive the growth hormone and the remaining 22 mice to not receive any growth hormone (these are the control group mice). Both groups of mice are fed the same diet. After 14 days she weighs each of the mice and records their weights. The group of mice that received the growth hormone had mean weight gain of 26.3g and standard deviation 3.3g. The control group had mean weight gain of 24.4g and standard deviation 2.8g.

For this question let μ_1 be the mean weight gain of any mouse who receives the growth hormone and let μ_2 be the mean weight gain of any mouse who receives the same treatment as the control group.

(a) What is the Standard Error of the difference between the two sample means $\bar{x}_1$ and $\bar{x}_2$?

(b) Do mice who receive the growth hormone have higher weight gain than the control mice? State appropriate H_0 and H_1 (null and alternative hypotheses) for answering this question

(c) Compute the test statistic and its P-value. State in words what you conclude about the effects of the growth hormone based on this study.

Question 6. (*25 points*)

A clinical trial was conducted to examine the effectiveness of daily doses of aspirin in the treatment of strokes. Patients were randomized into treatment and control groups. Neither the physicians nor the patients knew whether they were receiving the aspirin or a placebo tablet. After six months of treatment, the attending physicians evaluated each patient's progress as favorable or unfavorable. Of the 78 patients in the aspirin group, 63 had favorable outcomes. The control group of 77 patients had 43 patients with favorable outcomes. Let p_1 be the proportion of patients in the treatment group (aspirin) who had favorable outcomes after six months. Let p_2 be the proportion of patients in the control group (placebo) who had favorable outcomes after six months.

(a) What z^* should we use for a 98% confidence interval?

(b) Compute $\text{SE}\,(\hat{p}_1 - \hat{p}_2)$ and give the 98% confidence interval for the difference $p_1 - p_2$.

(c) Would you conclude that using aspirin helps in the treatment of strokes? Explain your answer. Note that you may do this either by using your confidence interval from part (b) or by carrying out the appropriate hypothesis test.

Question 7. (*25 points*)

A clinical research study collected data on emphysema patients. Specifically, they measured the number of years smoked (x) and the lung capacity of the patient (evaluated on a scale of 0 to 100). In all, 10 patients were studied and their measurements were as follows:

x	25	36	22	15	48	39	42	31	28	33
y	45	40	50	70	25	30	30	45	70	65

Based on this data the following quantities were calculated: $\bar{x} = 31.9$, $\bar{y} = 47$, $s_x = 9.87$, $s_y = 16.70$ and $r = -0.77$.

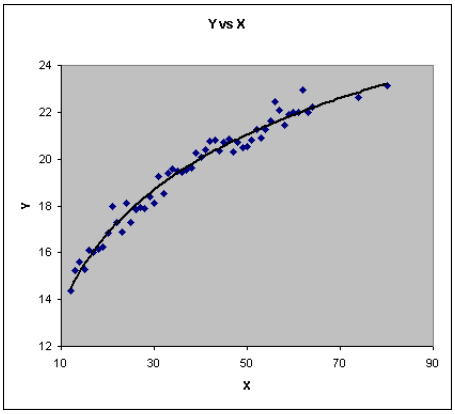
(a) Compute the intercept and slope for the regression line $\hat{y} = \beta_0 + \beta_1 x$.

(b) Suppose, using a computer, you find that $\text{SE}_{b_1} = 0.3789$. Test whether or not the number of years smoked can be used to predict lung capacity with this linear regression model.

Question 8. (*25 points*)

A scientist working for a corporation has collected some data and wants your statistical advice. To protect the company's corporate secrets she does not reveal to you the exact variables studied. Instead, she refers to the explanatory variable as X and the response variable as Y. As part of your report to her on the dataset, you need to answer the following questions:

(a) Suppose you create the following scatterplot for the data:



What transformation would you recommend for this data to make it more linear?

(b) Suppose you fit a linear regression to the non-transformed data. Sketch, roughly the pattern you would expect to see on a residuals plot.

(c) Using the transformed dataset you fit a regression line. Would it be sensible to use the fitted regression line to predict the value of Y at $X = 50$? How about at $X = 100$. Make sure to clearly explain your answer.