

**Statistics 104**  
**Practice Midterm Examination 2**  
**Autumn, 2004**

1. (20 points) An animal maintained in constant darkness for several months shows a rhythm of activity that is strongly periodic with a period near, but not exactly equal to 24 hours. The period is called the *circadian* period (from the Latin, *circa dia*). The circadian period varies from animal to animal.

Pittendrigh and Daan collected data to study whether the circadian period decreases with the age of the animal. Circadian periods of ten deermice at age 5 months,  $X_1, X_2, \dots, X_{10}$ , were measured to the nearest hundredth of an hour. Another set of measurements of ten deermice at age 14 months,  $Y_1, Y_2, \dots, Y_{10}$ , were also collected. Also computed are the 10 differences,  $D_1 = X_1 - Y_1, D_2 = X_2 - Y_2, \dots, D_{10} = X_{10} - Y_{10}$ . Summaries of the data are presented below.

Variable	N	MEAN	MEDIAN	TRMEAN	STDEV	SEMEAN
5 months	10	23.659	23.680	23.645	0.380	0.120
14 months	10	23.369	23.420	23.396	0.263	0.083
Diff	10	0.2900	0.3350	0.3188	0.2824	0.0893

For this problem, let  $\mu_x$  and  $\mu_y$  be the population means of the circadian period of deermice at age 5 months and 14 months respectively.

- (5 points) Analyze the data as though the two samples are independent. Conduct a hypothesis test at an  $\alpha = 0.05$  level to examine whether circadian period decreases with the age of the animal. State the relevant hypotheses and perform the appropriate test. Is it reasonable to believe that the circadian period decreases with the age of the animal?
- (5 points) Now analyze the data as though  $X$  and  $Y$  samples are paired. Construct a 95% confidence interval for the mean differences in circadian period. Assuming that this is the appropriate analysis, what conclusions can be made about whether circadian periods decrease with age?
- (5 points) Whether the  $X$  and  $Y$  samples should be considered as independent or paired depends on how the data were actually collected. In the description of the problem, it was deliberately made vague so that there is not enough information for you to decide which is the correct analysis. Describe below two different scenarios on how the data might have been collected to make the analysis in a) and b) appropriate for that particular scenario.
- (5 points) In reality, the data actually was paired. With the sample size of 10 used, what is the power to detect a differences in means of 0.35 if the true population standard deviation of the difference of the circadian differences was 0.3 with an alternative hypothesis of  $H_A: \mu_x - \mu_y > 0$ ?

2. (20 points) Consider the following two studies
- a) The Public Health Service studied the effects of smoking on health, in a large sample of representative households. For men and for women in each age group, those who had never smoked were on average somewhat healthier than the current smokers, but the current smokers were on average much healthier than the former smokers.
    - i) (5 points) What is the advantage of studying men and women and the different age groups separately?
    - ii) (5 points) The lesson seems to be that you shouldn't start smoking, but once you've started, don't stop. Comment briefly.
  - b) A group of college students believes that herb tea has remarkable restorative powers. To test this belief, they make weekly visits to a local nursing home, visiting with the residents and serving them herb tea. The nursing home staff reports that after several months many of the residents are more cheerful and healthy. A skeptical sociologist commends the students for their good deeds but scoffs at the idea that the herb tea helped the residents.
    - i) (5 points) Suggest possible reasons for the sociologists' lack of belief in the advantages of herb tea.
    - ii) (5 points) Suppose the students wish to do another study to try to convince the sociologist about the benefits of herb tea. What changes to the original study should be made to give more valid conclusions about the benefits of herb tea?
3. (15 points) One factory has four production lines to produce bicycles. Of the total production, line 1 produces 15%, line 2 produces 20%, line 3 produces 30% and line 4 produces 35%. The rate for *defective* products past the quality control for these four production lines are 5%, 4%, 3%, and 2% respectively.
- a) (5 points) What is the proportion  $p$  of defective bicycles in the factory's output?
  - b) (5 points) If a bicycle is found defective, what is the probability that it comes from production line 4?
  - c) (5 points) If an independent agency, like *Consumers' Report* buys 50 bicycles at random, what are the mean and standard deviation of the number of defective ones?

4. (15 points) During the summer of 1992, while Ross Perot was considering whether to run for the presidency, a number of polls were done to try and assess how popular Perot was to the voters. We will examine one poll taken in California on June 2. This poll asked 1112 voters in the Democratic primary and 612 voters in the Republican primary whether they would vote for Ross Perot if his name was on the ballot. Of the Democratic voters, 371 said they would vote for Perot, while 282 Republicans would vote for Perot.
- (5 points) Was there a difference in support for Perot between Democrats and Republicans in California on June 2nd? Construct a hypothesis test giving the test statistic, calculate the p-value (or bounds for it), and give a conclusion.
  - (5 points) Construct an approximate 95% confidence interval for  $p_R - p_D$ , the difference of the proportion of Republicans to the proportion of Democrats supporting Perot in California on June 2nd.
  - (5 points) Suppose that it was desired to have a 95% margin of error for the proportion of Republicans voting for Perot to be no more than 0.02. How many Republicans would need to be sampled to meet this criterion?