

STATISTICS 210 – Fall 2008

Midterm Exam: Wednesday November 5, 2008

- The exam is closed book, closed notes. You may use a calculator.
- Tables are provided separately.
- Problem values are written to the left of each problem. The total number of points is equal to 110.
- Two important reminders:
  - show your work so that you can receive partial credit;
  - budget your time so that you don't miss problems you know how to do.
- Please do not write your solutions on the exam paper.
- Good luck!

1. Seeing violence on television is believed to be associated with violent behavior. In one study (published in the *Journal of the American Academy of Child and Adolescent Psychiatry*) children in grades 3-8 in eleven public schools in Ohio were asked how much television they watch per day and how often they demonstrate violent behaviors. The data are self-reported. The response is a score from 0 to 15 with higher numbers meaning greater levels of violent behavior. The children were put into five groups based on daily television viewing. Summaries of the reported amount of violent behavior by children in the five groups are provided below.

Daily Hours of Television	sample size	sample mean	sample s.d.
< 1	277	2.94	2.73
1-2	526	2.59	2.44
3-4	666	2.87	2.36
5-6	310	3.10	2.45
> 6	488	4.03	2.81

- (a) Identify the population parameters to which the analysis of variance results in this problem would apply.
- (b) An incomplete analysis of variance table is provided below. Complete the table and carry out a test of the hypothesis that all group means are the same. Be sure to state your conclusion.

Source	d.f.	SS
Groups		
Error		
Total	2266	15188

- (c) Examine the sample mean and s.d. for each group. What can you infer about the distribution of response scores within each group? Does your discovery about the distribution within each group indicate a problem for the analysis of variance. Explain.
- (d) It is natural to wonder if there is a linear relationship between the dose (amount of TV watched) and the response (amount of violent behavior).
- i. We can assess whether there is a linear trend with the contrast defined by  $c = (-2, -1, 0, 1, 2)$ . Test whether the data support a linear trend or not. Carefully state your result.
  - ii. There is one strange feature of this linear contrast. The weight on the middle group is **zero**. This means the sample contrast is the same regardless of the value of this group mean. Does the test of the contrast depend on the middle group mean in any way (e.g., what would happen if the number changed from 2.87 to 12.87)? Explain/interpret what you discover.

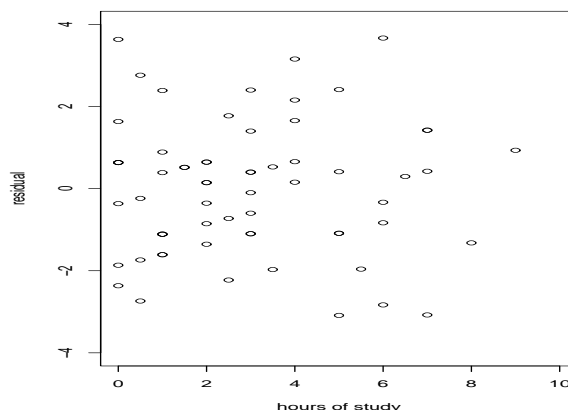
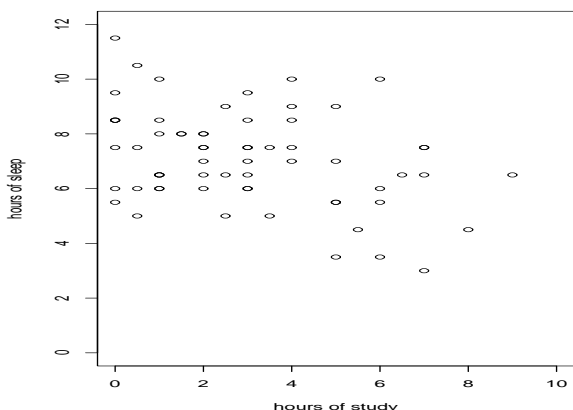
2. A survey in an undergraduate statistics class asked each student to record the number of hours they studied during the previous 24 hours and the number of hours they slept during the previous 24 hours. There are 64 students who supplied responses. A scatterplot of the data are shown at the left below. The table below presents summary statistics for each variable and regression results. A scatterplot of regression residuals versus hours of study is shown at the right below. The standard error of the regression is  $s_e = 1.6363$ .

#### Summary statistics

Variable	mean	s.d.
study	2.99	2.34
sleep	7.10	1.73

#### Regression results

Variable	Param.Est.	Std.Err.
intercept	7.8659	0.3333
sleep	-0.2555	0.0879



- (a) Interpret the slope of the regression for the University's student newspaper (i.e., keep it simple!).
- (b) Give a 95% confidence interval for the slope.
- (c) What conclusion would you reach about a test of the hypothesis of zero slope at the  $\alpha = .05$  level? Explain how you know.
- (d) How many hours of sleep would you predict for a student that studied for 3 hours in the last 24 hours period? Also give an **approximate** 95% prediction interval for this prediction. (No formula is required – but you should justify your approximation.)
3. A 1994 paper in the journal *Pediatrics* compared intelligence quotient (IQ) test scores of children born to mothers who smoked during pregnancy (10+ cigarettes per day) to test scores of children born to mothers who did not smoke during pregnancy. IQ test score results at 4 years of age are summarized in the table below.

Group	n	mean	s.d.
Smokers	47	103.1	12.9
Nonsmokers	66	113.3	14.0

- (a) Give a 95% confidence interval for the difference in 4-year old mean IQ scores of smoking and nonsmoking mothers and provide a one sentence summary of the result.
- (b) Does **this study** prove a causal link between smoking during pregnancy and children's intelligence? Explain.
- (c) A question about replication of studies:

The study authors also compared 5-year old mean IQ scores. They find the means are 113.8 (nonsmoker) and 106.9 (smokers), the pooled s.d. is 13.5, the resulting  $t$  statistic is 2.68 and the hypothesis of equal means is rejected with  $p$ -value less than .01. A physician decides to repeat the experiment within her practice. The results of her smaller study (15 nonsmokers and 10 smokers) produce  $t = 1.25$  which is not significant, even at the .10 level. She decides that her study does not confirm the *Pediatrics* study. Is her conclusion correct? Briefly discuss. (Hint: It may be useful to recall from HW 1 that the  $t$  statistic can be written as the effect size (difference in means divided by s.d.) times a factor related to sample size.)