

## Statistics 106: Analysis of Variance

### Sample First Examination

#### Fall Quarter, 2002

An agronomist planted four varieties of wheat, each variety in three plots. The following yields (in pounds per plot) were obtained. The following sums of squares were obtained for the response variable  $Y_{ij}$ , denoting the yield of variety  $i$  in plot  $j$ :

$$\sum_{i,j} (Y_{ij} - \bar{Y}_{i.})^2 = 34.00$$

$$\sum_{i,j} (Y_{ij} - \bar{Y}_{..})^2 = 510.25$$

**Table 1: Yields of Four Varieties of Wheat**

| Variety | Yields |    |    |
|---------|--------|----|----|
| A       | 63     | 61 | 59 |
| B       | 56     | 52 | 51 |
| C       | 45     | 45 | 42 |
| D       | 58     | 55 | 58 |

The following questions were of interest:

1. If the idea is to determine whether the mean yields for the four varieties differed, write out the linear models for the two hypotheses of interest. Give the assumptions for the two models and indicate how you might test whether the assumptions are valid.
2. Summarize the results of the experiment in an ANOVA table and explain the meaning of each component. What are the error sums of squares under the two models identified in 1?
3. Test the hypothesis that the yields of the four varieties do not differ. Use  $\alpha = .05$  and explain what this value means.
4. Use Tukey's method to calculate how large the pairwise comparisons of the sample means would have to be in order to be statistically significant. Which differences are significant if you test the differences at a joint significance level of  $\alpha = .05$ ?
5. Construct a 99% interval for the difference between the average yields of varieties A and D and the average of varieties B and C.
6. Which variety is best? Give a 99% confidence interval for the mean of the best variety.
7. What would the overall confidence level be for all statements jointly in 4-6? Be sure to include the effect of focusing on the best yield.