Fruit Yield Example (From *Ott and Longnecker, 2010*)

An experiment was conducted to determine the effects of four different pesticides (Factor B) on the yield of fruit from three different varieties (Factor A) of a citrus tree. Eight trees from each variety were randomly selected from an orchard. The four pesticides were then randomly assigned to two trees of each variety and applications were made according to recommended levels. Yields of fruit (in bushels per tree) were obtained after the test period.

	Pesticide, B			
Variety, A	1	2	3	4
1	49	50	43	53
	39	55	38	48
2	55	67	53	85
	41	58	42	73
3	66	85	69	85
	68	92	62	99

Use the following commands to read the data into R and define the factors and response; the `sep='\t'` option indicates that columns are separated by tabs.

```
data = read.table('citrus.dat',header=T,sep='\t')
variety = as.factor(data[,1])
pesticide = as.factor(data[,2])
yield = data[,3]
```

Next, use `fit=lm(yield~variety+pesticide+variety*pesticide)` to fit the two-way ANOVA model with both factors and their interaction: $Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$

1. Briefly explain what each of the terms Y_{ijk}, $\mu_{..}$, α_i, β_j, and $(\alpha\beta)_{ij}$ represents.

2. In terms of ε_{ijk}, state the assumptions for the two-way ANOVA model.

3. How are the residuals defined for this model? Use `e=residuals(fit)` to save them. Also, check the assumptions in 2 with a QQ plot and a plot of residuals versus fits. Are they reasonably satisfied?

4. Use `interaction.plot(pesticide,variety,yield)`to plot the yield means for each combination of variety and pesticide. This plot will put pesticide on the horizontal axis and use separate lines for the varieties. Comment on the evidence for

   (a) interaction between variety and pesticide
   (b) a variety effect
   (c) a pesticide effect

5. Use `anova(fit)` to obtain the ANOVA table and relevant F-statistics. Is there significant evidence for interaction between variety and pesticide? Use $\alpha = .05$.

6. Recall the estimate for the effect of the ith variety is $(\overline{Y}_{i..} - \overline{Y}_{...})$, which averages over all pesticides. Comment on the appropriateness of this, given the results of the interaction test.

7. Conduct the test for a variety effect. State the null and alternative hypotheses in terms of the model parameters, and also state these words. What is your conclusion with $\alpha = .05$?

8. Obtain 95.0% simultaneous confidence intervals for the mean yield of each variety with a Bonferroni adjustment for multiplicity. *Hint:* `tapply(yield,variety,mean)`.

9. Use Tukey's procedure to conduct pairwise comparisons of mean yields for the varieties. Recall Tukey's critical value is
$$T = \frac{q_{a,ab(n-1)}(\alpha)}{\sqrt{2}}$$
where a is the number of levels of Factor A (variety), and $ab(n-1)$ is the degrees of freedom for error.

10. If the interaction term is not significant, researchers sometimes remove it and re-fit the model.

   (a) If the interaction term is removed, what happens to the sum of squares for the factors (main effects)? What happens to the sum of squares for error? *Hint: recall the relationship*
   $$SSTO = SSA + SSB + SSAB + SSE$$

   (b) If the interaction term is removed, what happens to the degrees of freedom for the main effects and for the error? How does this change the critical values for the tests?

   (c) Given the changes observed above, does removing the interaction term necessarily make the main effect tests more powerful? Explain.

   (d) Use `anova(lm(yield~variety+pesticide))` to fit the model without the interaction term, and compare them with the results in 5. Comment on this, given the evidence for interaction in this situation.