Assignment 10 - ANOVA

your name goes here

Due: Wednesday, April 25, 2018

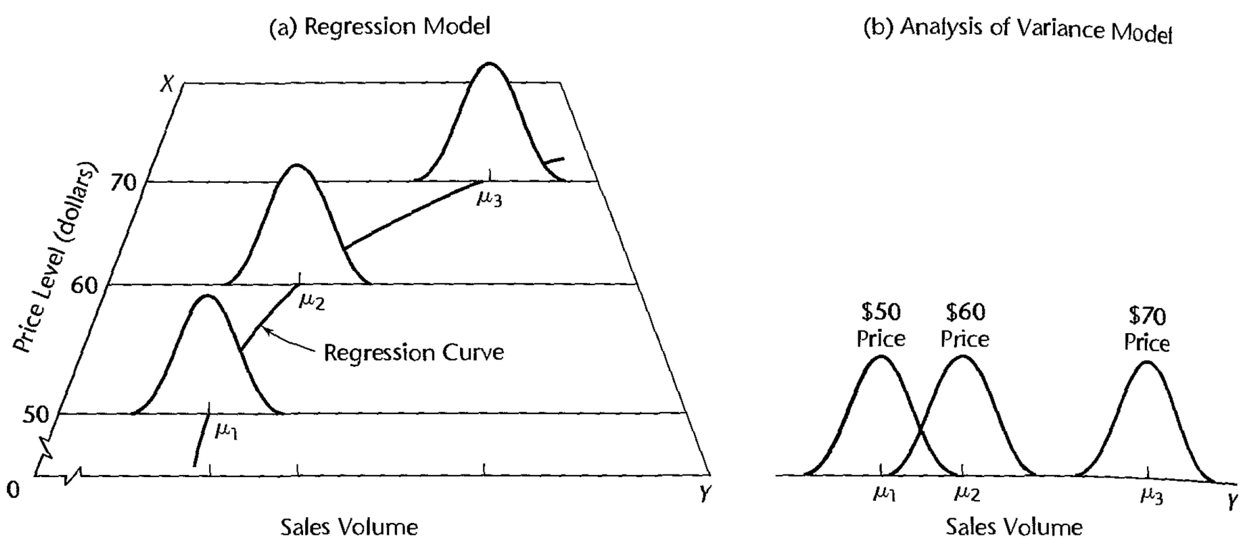
Summary

The last homework assignment of the semester introduces you to the ideas of ANalysis Of VAriance (ANOVA). We will only scratch the surface with one- and two-way models, but know that you have the tools to learn a lot more about more sophisticated nested models, cross-over designs, split-plot designs, and random effects. The key is that with each different model, the variability associated with the observations (which translates to the variability associated to the coefficients / inference) needs to be calculated differently.

Assignment

1. Refer to Figure 16.1a. Could you determine the mean sales level when the price level is \$68 if you knew the true regression function? Could you make this determination from Figure 16.1b if you only knew the values of the parameters μ_1, μ_2, and μ_3 of ANOVA model (16.2)? What distinction between regression models and ANOVA models is demonstrated by your answers?

FIGURE 16.1 **Relation between Regression and Analysis of Variance Models.**



2. A market researcher, having collected data on breakfast cereal expenditures by families with 1, 2, 3, 4, and 5 children living at home, plans to use an ordinary regression model to estimate the mean expenditures at each of these five family size levels. However, the researcher is undecided between

fitting a linear or a quadratic regression model, and the data do not give clear evidence in favor of one model or the other. A colleague suggests: "For your purposes you might simply use an ANOVA model." Is this a useful suggestion? Explain.

3. Consider the following linear combinations of interest in a single-factor study involving four factor levels:

   (i)
$$\mu_1 + 3\mu_2 - 4\mu_3$$

   (ii)
$$0.3\mu_1 + 0.5\mu_2 + 0.1\mu_3 + 0.1\mu_4$$

   (iii)
$$\frac{\mu_1 + \mu_2 + \mu_3}{3} - \mu_4$$

   (a) Which of the linear combinations are contrasts? State the coefficients for each of the contrasts.

   (b) Give an unbiased estimator for each of the linear combinations. Also give the estimated variance of each estimator assuming that $n_i \equiv n$.

4. Consider a single-factor study with $r = 5$ treatments and sample sizes $n_i \equiv 5$.

   (a) Find the T, S, and B multipliers if $g = 2, 5$, and 10 pairwise comparisons are to be made with a 95 percent family confidence coefficient. What generalization is suggested by your results? (n.b. use the functions: `qt`, `qf`, and `qtukey(1-alpha, r, n-r)`)

   (b) What would be the T, S, and B multipliers for sample sizes $n_i \equiv 20$? Does the generalization obtained in part (a) still hold?

5. Questionnaire color: In an experiment to investigate the effect of color of paper (blue, green, orange) on response rates for questionnaires distributed by the "windshield method" in supermarket parking lots, 15 representative supermarket parking lots were chosen in a metropolitan area and each color was assigned at random to five of the lots. The response rates (in percent) follow. Assume that ANOVA model (16.2) is appropriate.

	j				
i	1	2	3	4	5
Blue	28	26	31	27	35
Green	34	29	25	31	29
Orange	31	25	27	29	28

```
rsp.rate <- c(28, 26, 31, 27, 35, 34, 29, 25, 31, 29, 31, 25, 27, 29, 28)
ppr.col <- c(rep("blue", 5), rep("green", 5), rep("orange", 5))
paper = data.frame(rsp.rate, ppr.col)
```

   (a) Make a dot plot of the data. Do the factor level means appear to differ? Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

   `library(ggplot2)`

   `ggplot(paper, aes(x = yourxvariable, y = youryvariable)) + geom_somegraphyouwant`

   (b) Obtain the fitted values.

(c) Obtain the residuals.

(d) Obtain the analysis of variance table.

(e) Conduct a test to determine whether or not the mean response rates for the three colors differ. Use level of significance $\alpha = .10$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

(f) Conduct one F-test to see if blue and green (together) are different from orange (include hypotheses, test stat, p-value, conclusion, context).

(g) When informed of the findings, an executive said: "See? I was right all along. We might as well print the questionnaires on plain white paper, which is cheaper." Does this conclusion follow from the findings of the study? Discuss.

6. Refer to Questionnaire color problem above. Regression model (16.75 - note a totally different set of notation!) is to be employed for testing the equality of the factor level means. Feel free to use pencil and paper here.

   (a) Set up the Y, X, and β matrices (see eq (16.73)).

   (b) Obtain $X\beta$. Develop equivalent expressions of the elements of this vector in terms of the cell means, μ_i (see eq (16.74)).

   (c) Obtain the fitted regression function (use the previous problem) (see eq (16.75)). What is estimated by the intercept term?

   (d) Conduct the test for equality of factor level means; use $\alpha = .10$. State the alternatives, decision rule, and conclusion.