Stat 218 - Day 29
More on ANOVA: Checking Technical Conditions

Recall that analysis of variance is a technique or comparing means among several groups. We test H_0: $\mu_1 = \mu_2 = \ldots = \mu_I$ by constructing an ANOVA table and determine the test statistic as $F_s = \frac{MS(between)}{MS(within)} = \frac{SS(between)/(I-1)}{SS(within)/(n^*-I)}$, where SS(within) = $\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\bullet})^2 = \sum_{i=1}^{I}(n_i - 1)s_i^2$

and SS(between) = $\sum_{i=1}^{I} n_i(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2$.

Example: Hot dogs
A study compared sodium contents and calories of hot dogs classified as beef, poultry, and meat. Summary statistics for the "calories" variable are:

	Beef	Poultry	Meat	*Overall*
Sample size	20	17	17	*54*
Sample mean	156.85	118.76	158.71	*145.44*
Sample standard deviation	22.64	22.55	25.24	*29.38*

(a) Use these summary statistics to construct the ANOVA table.

(b) Test whether the sample data provide evidence of a difference in mean calorie amounts among these three types of hot dogs. State the hypotheses, and report the test statistic and *P*-value. Also summarize your conclusion.

(c) Use Minitab (`HotDogs.mtw`) to perform an ANOVA on the sodium contents. Record the ANOVA table, state the hypotheses, and report the test statistic and *P*-value. Also summarize your conclusion.

Technical conditions:
The technical conditions required for the validity of the ANOVA procedure and F-test are:

- that the data can be regarded as independent random samples from the populations
- that the underlying populations follow normal distributions
- that the standard deviations of those populations are the same for all groups

To check the normality condition, examine histograms and normal probability plots of the sample data. To check the equal standard deviations condition, determine whether the ratio of the largest to the smallest sample standard deviation is less than 2.

Example: Crash test dummies
The Minitab worksheet `crash.mtw` contains data on automobile crash test results. Response variables are measurements of injury extent on head (`c5`), chest (`c6`), left leg (`c7`), and right leg (`c8`). Explanatory variables include whether the dummy was on the driver or passenger side (`c9`), protective devices in the car (`c10`), number of doors on the car (`c11`), year of make (`c12`), and size of car (`c14`). We will first investigate whether the head injury measurements appear to differ based on number of doors.

(a) Construct visual displays of the head injury measurements by the number of doors. Describe the distributions, paying particular attention to the question of whether the head injury measurements appear to differ significantly among the three groups.

(b) Does it appear that the technical conditions of the ANOVA procedure and F-test are satisfied with these data? Explain.

(c) Transform the head injury measurements by using logarithms. Re-examine the distributions of these measurements by the number of doors. Do the technical conditions appear to be satisfied now? Explain.

(d) Construct the ANOVA table and interpret the results. Do the data provide strong evidence that head injury measurements differ based on the number of doors on the vehicle? Explain.

The following (incomplete) ANOVA table pertains to the log of head injury measurements, with the vehicle's *year of make* as the explanatory variable:

One-way ANOVA: log(HeadIC) versus Year

Source	DF	SS	MS	F	P
Year		0.2152	0.0538		0.291
Error	333				
Total	337	14.5724			

(e) Fill in the four missing entries of the ANOVA table.

(f) How many groups were there for year of make? Explain how you know.

(g) What conclusion would you draw from this analysis? Explain.