STAT 509 -- Analysis of Variance

• The Analysis of Variance (ANOVA) is most simply a method for comparing the means of several populations.

• It is commonly used to analyze experimental data arising from a Completely Randomized Design (CRD).

• The experimental units are the individuals on which the response variable is observed or measured.

• A specific experimental condition applied to the units is called a treatment.

• This experimental condition may be based on one or more factors, each of which has multiple levels. Each combination of factor levels is a different treatment.

Example: Plant growth study:
Experimental Units: A sample of plants
Response: Growth over one month
Factors: • Fertilizer Brand (levels: A, B, C)
• Environment (levels: Natural Sunlight, Artificial Lamp)
There are how many treatments?

(Could also have a quantitative factor…)

If 5 plants are assigned to each treatment (5 replicates per treatment), there are how many observations in all?

• A Completely Randomized Design is a design in which independent samples of experimental units are selected for each "treatment."
• That is, experimental units are assigned *at random* among the treatments.

Three Principles of the Design of Experiments

(1) Randomization: Assigning experimental units to treatments by random chance.
(2) Replication: Using multiple experimental units for each treatment to reduce sampling variation.
(3) Control: Reducing the effect of "lurking variables" on the response. Done by comparing numerous treatments, and sometimes by separating the units into "blocks" of similar units before the randomization.

Comparing Several Population Means

Suppose there are k treatments (usually $k \geq 3$), so that our data represent samples from k populations.

We want to test for any differences in mean response among the treatments.

Hypothesis Test:

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$
H_a: At least two of the treatment population means differ.

Q: Is the <u>variance within each group</u> small compared to the <u>variance between groups</u> (specifically, between group means)?

Pictures:

How do we measure the variance within each group and the variance between groups?

The Sum of Squares for Treatments (SST) measures variation <u>between</u> group means.

SST = $\sum_{i=1}^{k} n_i(\bar{Y}_i - \bar{Y})^2$

n_i **= number of observations in group *i***

$\bar{Y}_i$ **= sample mean response for group *i***

$\bar{Y}$ **= <u>overall</u> sample mean response**

SST measures how much each group sample mean varies from the overall sample mean.

The Sum of Squares for Error (SSE) measures <u>variation within groups</u>.

SSE = $\sum_{i=1}^{k}(n_i - 1)s_i^2$

s_i^2 **= sample variance for group *i***

SSE is a sum of the variances <u>of each group</u>, weighted by the sample sizes by each group.

To make these measures comparable, we divide by their degrees of freedom and obtain:

Mean Square for Treatments (MST) = $\frac{SST}{k - 1}$

Mean Square for Error (MSE) = $\frac{SSE}{n - k}$

The ratio $\frac{MST}{MSE}$ **is called the ANOVA F-statistic.**

If F = $\frac{MST}{MSE}$ **is much bigger than 1, then the variation between groups is much bigger than the variation within groups, and we would reject**
H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$ in favor of H_a.

Example: Newly hatched chicks were randomly placed into six groups, each group receiving a different feed supplement. Weights in grams after six weeks were measured.

<u>Response</u>: Weights in grams after six weeks
<u>6 treatments</u>: Six different feed supplements

Group Sample Means:

casein	horsebean	linseed	meatmeal	soybean	sunflower
323.58	160.20	218.75	276.91	246.43	328.92

Overall sample mean = 261.31.

Sample sizes for each group:
$n_1 = 12$, $n_2 = 10$, $n_3 = 12$, $n_4 = 11$, $n_5 = 14$, $n_6 = 12$
=> $n = 71$.

Sample variances for each group:

casein	horsebean	linseed	meatmeal	soybean	sunflower
4151.72	1491.96	2728.57	4212.09	2929.96	2384.99

- **We can use the formulas or software to obtain SST, SSE, MST, MSE, and F.**
- **This information is summarized in an <u>ANOVA table</u>:**

Source	df	SS	MS	F
Treatments	$k-1$	SST	MST	MST/MSE
Error	$n-k$	SSE	MSE	
Total	$n-1$	SS(Total)		

Note that df(Total) = df(Trt) + df(Error) and that SS(Total) = SST + SSE.

R code (using built-in `chickwts` data set):

```
> attach(chickwts)
> feed <- factor(feed)
> anova(lm(weight ~ feed))
```

For our example, the ANOVA table is:

In example, we can see F = 15.4 is "clearly" bigger than 1 … but how much bigger than 1 must it be for us to reject H_0?

<u>ANOVA F-test</u>:
If H_0 is true and all the population means are indeed equal, then this F-statistic has an F-distribution with numerator d.f. $k - 1$ and denominator d.f. $n - k$.

We would reject H_0 if our F is unusually large.
Picture:

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$
H_a: At least two of the treatment population means differ.

Rejection Region: F > F_α, where F_α based on (k – 1, n – k) d.f.

Assumptions for ANOVA F-test:

- **We have random samples from the *k* populations.**
- **All *k* populations are normal.**
- **All *k* population variances are equal.**

Example: Perform ANOVA F-test using α = .10.

If our F-test is significant, then which treatment means differ? We would then perform *multiple comparisons of means*. Tukey's multiple-comparisons procedure will simultaneously compare each pair of treatment means:

R code:

```
> TukeyHSD(aov(lm(weight ~ feed)),conf.level=0.95)
```