Math 158 – Linear Models Spring 2009
Jo Hardin
One Way ANOVA

**FORMAL MODEL & NOTATION:**

$$\left.\begin{array}{l} Y_{11}, Y_{12}, \ldots Y_{1n_1} \text{ are a SRS of size } n_1 \text{ from } N(\mu_1, \sigma^2) \text{ distribution} \\ Y_{21}, Y_{22}, \ldots Y_{2n_2} \text{ are a SRS of size } n_2 \text{ from } N(\mu_2, \sigma^2) \text{ distribution} \\ \cdots \\ Y_{r1}, Y_{r2}, \ldots Y_{rn_r} \text{ are a SRS of size } n_r \text{ from } N(\mu_r, \sigma^2) \text{ distribution} \end{array}\right\} = \left\{\begin{array}{l} Y_{ij} \sim N(\mu_i, \sigma^2) \\ Y_{ij} = \mu_i + \epsilon_{ij} \\ \epsilon_{ij} \sim N(0, \sigma^2) \end{array}\right.$$

unknown parameters:
$\mu_1, \mu_2, \ldots, \mu_r, \sigma^2$

useful notation:
$\overline{Y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$
$\overline{Y}_{\cdot\cdot} = \frac{1}{n_T}\sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij}$
$\mu_{\cdot} = \sum_{i=1}^{r} \frac{n_i}{n_T}\mu_i$

**ESTIMATION**

The value of $\mu_i$ that minimizes: $\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\mu_i)^2$ is the least squares estimate of $\mu_i$. That is, $\hat{\mu}_i = \overline{Y}_{i\cdot}$.

Residual sum of squares (within sum of squares) = SSE = $\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i\cdot})^2$

**THEORY**

| Variable | | Distribution | Assumptions |
|---|---|---|---|
| $\overline{Y}_{i\cdot}$ | $\sim$ | $N(\mu_i, \sigma^2/n_i)$ | CLT, normal pop, $= \sigma_i^2$ |
| SSE/$\sigma^2$ | $\sim$ | $\chi^2_{df=n-r}$ | normal pop, $= \sigma_i^2$ |
| SSTR/$\sigma^2$ | $\sim$ | $\chi^2_{df=r-1}$ | normal pop, $= \sigma_i^2$, under $H_o$ |
| SSTO/$\sigma^2$ | $\sim$ | $\chi^2_{df=n-1}$ | normal pop, $= \sigma_i^2$, under $H_o$ |

$$E(MSE) = \sigma^2$$
$$E(MSTR) = \sigma^2 + \sum_{i=1}^{r}\frac{n_i(\mu_i - \mu_\cdot)^2}{(r-1)}$$

F-distribution:
If, in fact, $H_0$ is true (that is, $\mu_1 = \mu_2 = \ldots = \mu_r$), then we know our sums of squares are independent $\chi^2$ random variables. Therefore:

$$\text{F-stat} = \frac{MSTR}{MSE} \sim F_{(r-1, n_T - r)}$$

Note: if $H_0$ is not true, then MSTR will be too big and the distribution of F-stat will shift to the right of $F_{(r-1,n_T-r)}$.