IEOR151 Lab 3- ANOVA Review

- ANOVA: study the impact of certain factor
- Groups: different levels (treatments) of the factor
- H_0: all means are equal; H1: not all equal.
- Let m= number of groups, n_i=number of observations in the i-th group, $X_{i.}$=sample mean of the i-th group, n_T=total number of observations, $X_{..}$=sample mean of all observations.
- Between-group sum of squares
  - $SS_b = \sum_{i=1}^{m} n_i(X_{i.} - X_{..})^2, \frac{SS_b}{\sigma^2} \sim \chi^2_{m-1}$, if H_0 is true
  - Amount of variation that can be explained group difference

- Within-group sum of squares
  - $SS_W = \sum_{i=1}^{m} \sum_{j=1}^{n_i}(X_{ij} - X_{i.})^2, \frac{SS_W}{\sigma^2} \sim \chi^2_{n_T-m}$, SS_b and SS_W are independent
  - Amount of variation that cannot be explained by group difference
- Test statistics: $\text{F} - \text{ratio} = \frac{SS_b/(m-1)}{SS_W/(n_T-m)} \sim F_{m-1,n_T-m}$, if H_0 is true
- p-value= $P\{F_{m-1,n_T-m} \geq \text{observed F} - \text{ratio}\}$
- ANOVA table

source of variation	sum of squares	d.f.	mean square	F ratio	p-value
between groups	SS_b	$m-1$	$SS_b/(m-1)$	$\frac{SS_b/(m-1)}{SS_W/(n_T-m)}$	$P\{F_{m-1,n_T-m} \geq \text{F ratio}\}$
within groups	SS_W	n_T-m	$SS_W/(n_T-m)$		
total	$SS_b + SS_W$	n_T-1			

Example

A manufacturer of paper used for making grocery bags is interested in improving the tensile strength of the product. Product engineering thinks that tensile strength is a function of the hardwood concentration in the pulp and that the range of hardwood concentrations of practical interest is between 5 and 20%. A team of engineers responsible for the study decides to investigate four levels of hardwood concentration: 5, 10, 15, and 20%. They decide to make up six test specimens at each concentration level, using a pilot plant. All 24 specimens are tested on a laboratory tensile tester, in random order. The data from this experiment are shown below:

Hardwood Concentration (%)	Observations						Totals	Averages
	1	2	3	4	5	6		
5	7	8	15	11	9	10	60	10.00
10	12	17	13	18	19	15	94	15.67
15	14	18	19	17	16	18	102	17.00
20	19	25	22	23	18	20	127	21.17
							383	15.96

Use the ANOVA to test the hypothesis that different hardwood concentrations do not affect the mean tensile strength of the paper.