Single-Factor Studies

- **Productivity Improvement Example.** An economist compiled data on productivity improvements last year for a sample of firms producing electronic computing equipment. The firms were classified according to the level of their average expenditures for research and development in the past three years (low, moderate, high). The results of the study is given in the table below (productivity improvement is measured on a scale from 0 to 100).

	1	2	3	4	5	6	7	8	9	10	11	12
Low	7.6	8.2	6.8	5.8	6.9	6.6	6.3	7.7	6.0			
Moderate	6.7	8.1	9.4	8.6	7.8	7.7	8.9	7.9	8.3	8.7	7.1	8.4
High	8.5	9.7	10.1	7.8	9.6	9.5						

1. What is the objective of the study?

2. Identify the following

a. Experimental units. ____________________

b. Response variable. ____________________

c. Factor. ____________________

d. Treatments. ____________________

- **Null and Alternative Hypotheses**.

H_0 : *The factor has NO effect.* vs. H_1 : *The factor has an effect.*

- **ANOVA Idea.** ANOVA is based on separating the total variation observed in the data into two parts: variation *among the group means* and variation *within groups.* If the variation among groups is large relative to the variation within groups, we have evidence against the null hypothesis.

- **Assumptions**.

1. The samples are independent SRS from each population.

2. The populations are assumed to be normal.

3. The standard deviations are equal. [Rule of thumb: *The largest standard deviation should not be more than twice the smallest standard deviation.*]

ANOVA Table

Source	DF	Sum of Squares	Mean Square	F
Treatment (Between)	$r-1$	$SSTR=\sum_{i=1}^{r} n_i(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2$	$MSTR=\frac{SSTR}{r-1}$	$F_{\text{obs}}=\frac{MSTR}{MSE}$
Error (Within)	n_T-r	$SSE=\sum_{i=1}^{r}(n_i-1)s_i^2$	$MSE=\frac{SSE}{n_T-r}$	
Total	n_T-1	$SSTO=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{\cdot\cdot})^2$		

Note:

1. $SSTO=SSTR+SSE$.

2. Under H_0, $F_{\text{obs}}\sim f_{(r-1,n_T-r)}$.

3. $E(MSTR)=\sigma^2+\frac{\sum n_i(\mu_i-\mu_\cdot)^2}{r-1}$.

4. $E(MSE)=\sigma^2$.

5. $p-$value=Pr$(F_{(df_1=r-1,df_2=n_T-r)}\geq f_{\text{obs}})$

[*Recall that we reject the null hypothesis if the $p-$value is less than the level of significance (α).*]

- Complete the ANOVA table below:

ANOVA Table

Source	DF	Sum of Squares	Mean Square	F	$p-$value
Treatment					
Error				xxxxx	xxxxx
Total			xxxxx	xxxxx	xxxxx

Sample R commands:

```
low=c(7.6,8.2,6.8,5.8,6.9,6.6,6.3,7.7,6.0)
moderate=c(6.7,8.1,9.4,8.6,7.8,7.7,8.9,7.9,8.3,8.7,7.1,8.4)
high=c(8.5,9.7,10.1,7.8,9.6,9.5)
length(low)        # 9
mean(low)          # 6.877778
sd(low)            # 0.8135997

prod.imp=c(low,moderate,high)
mean(prod.imp)     # 7.951852

p.value=1-pf(f.obs,r-1,nt-r)
```

- **Homework.** Due Monday, Febuary 2, 2015.

1. Rehabilitation Therapy. A rehabilitation center researcher was interested in examining the relationship between physical fitness prior to surgery of persons undergoing corrective knee surgery and time required in physical therapy until successful rehabilitation. Patient records in the rehabilitation center were examined, and 24 male subjects ranging in age from 18 to 30 years who had undergone similar corrective knee surgery during the past year were selected for the study. The number of days required for successful completion of physical therapy and the prior physical fitness status (below average, average, above average) for each patient follow.

	1	2	3	4	5	6	7	8	9	10
Below Average	29	42	38	40	43	40	30	42		
Average	30	35	39	28	31	31	29	35	29	33
Above Average	26	32	21	20	23	22				

a. Complete the ANOVA table below: [5]

ANOVA Table

Source	DF	Sum of Squares	Mean Square	F	$p-$value
Treatment					
Error				xxxxx	xxxxx
Total			xxxxx	xxxxx	xxxxx

b. Test whether or not the mean number of days required for successful rehabilitation is that same for the three fitness groups. Use $\alpha = 0.05$ level of significance.

2. Analysis of variance methods are often used in clinical trials where the goal is to assess the effectiveness of one or more treatments for a particular medical condition. One such study compared three treatments for dandruff and a placebo. The treatments were 1% pyrithione zinc shampoo (PyrI), the same shampoo but with instructions to shampoo two times (PyrII), 2% ketoconazole shampoo (Keto), and a placebo shampoo (Placebo). After six weeks of treatment, eight sections of the scalp were examined and given a score that measured the amount of scalp flaking on a 0 to 10 scale. The response variable was the sum of these eight scores. An analysis of the baseline flaking measurements indicated that randomization of patients to treatments was successful in that no differences were found between the groups. At baseline there were 112 subjects in each of the three treatment groups and 28 subjects in the Placebo group. During the clinical trial, 3 dropped out from the PyrII group and 6 from the Keto group. No patients dropped out of the other two groups. A summary of the data is given below.

Treatments	n	$\bar{y}$	s
Pyr I	112	17.39	1.142
Pyr II	109	17.20	1.352
Keto	106	16.03	0.931
Placebo	28	29.39	1.595

a. Complete the ANOVA table below: [5]

ANOVA Table

Source	DF	Sum of Squares	Mean Square	F	$p-$value
Treatment					
Error				xxxxx	xxxxx
Total			xxxxx	xxxxx	xxxxx

b. Using $\alpha = 0.01$, test the hypothesis that there is no difference in the effectiveness of the four treatments.