Chapter 11 Class Notes – Comparing Many ($I \geq 2$) Independent Sample Means

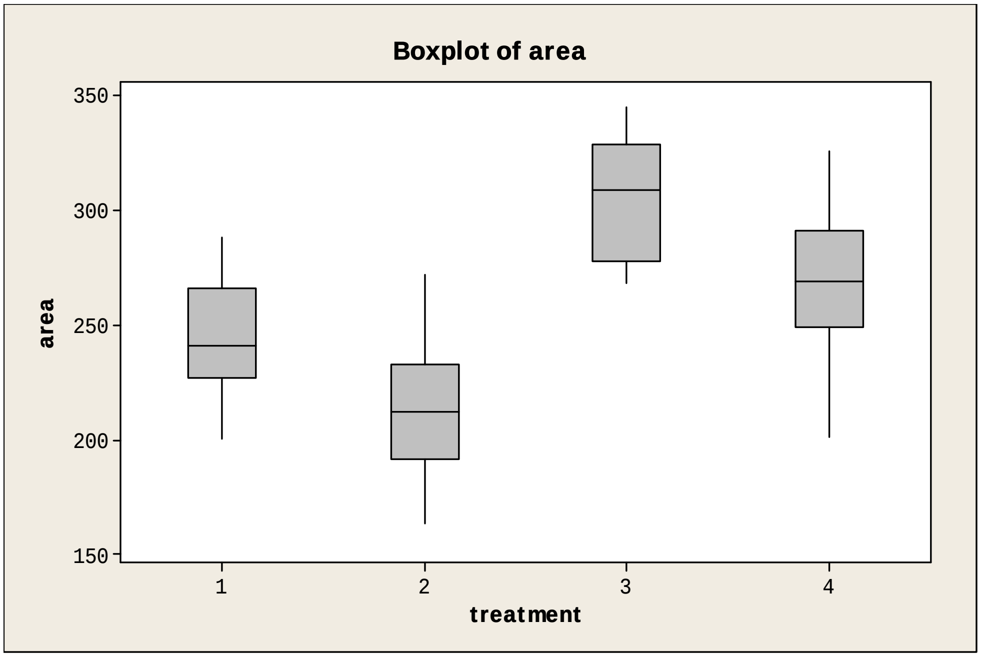
We'll skip §11.6 (One-way RBD), §11.7 (Two-way ANOVA) and §11.8 (Linear Combinations of Means), and cover the rest of this chapter.

In this chapter, we sample from $I \geq 2$ independent groups, and we wish to test whether the means ($\mu_1, \mu_2, \ldots \mu_I$) are equal. With regard to notation, let's consider an example: Soybeans, p.449 where there are k = 4 treatment groups:

- **Treatment 1: low light, control (n_1 = 13)**
- **Treatment 2: low light, stress (n_2 = 13)**
- **Treatment 3: moderate light, control (n_3 = 13)**
- **Treatment 4: moderate light, stress (n_4 = 13)**

This study is 'balanced' since $n_1 = n_2 = n_3 = n_4$, but this (balance) is certainly not a requirement. The response variable here is total leaf area (cm^2). In general, the data might (and do) look something like:

	Treatment			
	1	2	3	4
	y_{11} (264)	y_{21} (235)	y_{31} (314)	y_{41} (283)
	y_{12} (200)	y_{22} (188)	y_{32} (320)	y_{42} (312)
	…	…	…	…
	y_{1n_1} (230)	y_{2n_2} (202)	y_{3n_3} (273)	y_{4n_4} (257)
n	13	13	13	13
Mean	245.3	212.9	304.1	268.8
SD	27.0	29.7	26.9	35.2



Interestingly, the individual scores (in red) are not so important for the relevant test here – just the summary values (in green). But how do we analyze the data here?

We might be tempted to perform the $_4C_2 = 6$ separate two-sample t-tests using the methods from Chapter 7. But there are important problems with this approach:

- **It is cumbersome and timely**
- **Each test at the α level produces what size overall α level? As shown on p.416, the overall α level would be 20% – very high indeed**
- **Since *we assume equal variances here*, there is information in each treatment group about σ^2, so the data must be analyzed simultaneously.**

And the following F test does just that.

To illustrate the spirit of this F test, here is a smaller example:

Weight Gain of Lambs on p.420

	Weight Gains		
	Diet 1	**Diet 2**	**Diet 3**
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	
Sample size, n_i	3	5	4
Sum, $\sum_{j=1}^{n_i} y_{ij}$	33	75	48
Mean, $\bar{y}_i$	11	15	12
Within SS, $\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$	38	98	74
Sample variance, s_i^2	19	24.5	24.667

For the study sample size $n_{\bullet} = n_1 + n_2 + \cdots + n_I$, the grand mean is $\bar{\bar{y}} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i} y_{ij}}{n_{\bullet}}$, and this is *not the average of the treatment means*. Here, $n_{\bullet}$ = 12 and the grand mean is $\bar{\bar{y}}$ = (33 + 75 + 48) / 12 = 13.

To form the requisite F test statistic, we need two components:

- **The Within groups SS (Sum of Squares):**

With degrees of freedom $df(within) = n_{\bullet} - I$,
$SS(within) = \sum_{i=1}^{I}(n_i - 1)s_i^2$ and the Mean Square is then
$MS(within) = \frac{SS(within)}{df(within)} = \frac{\sum_{i=1}^{I}(n_i-1)s_i^2}{n_{\bullet}-I}$. Then σ^2 is estimated by
$s_{POOLED}^2 = MS(within)$ and σ by $s_{POOLED} = \sqrt{MS(within)}$

- **The Between groups SS (Sum of Squares):**

With degrees of freedom $df(between) = I - 1$,
$SS(between) = \sum_{i=1}^{I} n_i(\bar{y}_i - \bar{\bar{y}})^2$, so the Mean Square is then
$MS(between) = \frac{SS(between)}{df(between)} = \frac{\sum_{i=1}^{I} n_i(\bar{y}_i-\bar{\bar{y}})^2}{I - 1}$

The relevant F test statistic is obtained by dividing $MS(between)$ by $MS(within)$: this test statistic has the F distribution with numerator df = (I - 1) and denominator df = ($n_{\bullet}$ - I). ANOVA results are usually presented in a table such as the following ANOVA Table:

Source	df	SS	MS	F
Between Groups	I - 1	$SS(between)$	$MS(between)$	$\frac{MS(between)}{MS(within)}$
Within Groups	$n_{\bullet}$ - I	$SS(within)$	$MS(within)$	
Total	$n_{\bullet} - 1$	$SS(total)$		

In this Table, note that the degrees of freedom sum down (Total df = Between df + Within df) – the same goes for SS (sum of squares).

For the above Lambs example, we calculate by hand:

- **Within**: $df = 12 - 3 = 9$, $SS(within) = 2 \times 19 + 4 \times 24.5 + 3 \times 24.67 = 38 + 98 + 74 = 210$, $MS(within) = \frac{210}{9} = 23.33$. Here σ^2 is estimated by $MS(within) = 23.33$, and σ is estimated by $s_{POOLED} = \sqrt{23.33} = 4.8305$.
- **Between**: $df = 3 - 1 = 2$, $SS(between) = 3 \times (11 - 13)^2 + 5 \times (15 - 13)^2 + 4 \times (12 - 13)^2 = 36$, $MS(between) = \frac{36}{2} = 18$

Here, the TS is $F_{2,9} = \frac{18.0}{23.33} = 0.7714$, and the ANOVA Table is:

Source	df	SS	MS	F
Between Groups	2	36	18	0.7714
Within Groups	9	210	23.33	
Total	11	246		

The One-Way ANOVA Model (§11.3)

The one-way ANOVA model can be written in two equivalent ways:

(1) **Model function one: $y_{ij} = \mu_i + \varepsilon_{ij}$, for $i = 1, 2 \ldots I$ (groups) and $j = 1, 2, \ldots, n_i$ (subjects within groups)**

(2) **Model function two: $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, for $i = 1, 2 \ldots I$ (groups) and $j = 1, 2, \ldots, n_i$ (subjects within groups)**

The connection between these expressions is to let $\tau_i = \mu_i - \mu$.
In the first form, we can specify the null and alternative hypotheses:

$H_0: \mu_1 = \mu_2 = \cdots = \mu_I$
H_A: **at least one of the μ_i differs from the others**

In the second instance, and in light of the fact that $\tau_i = \mu_i - \mu$, these hypotheses are equivalently written:

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_I = 0$$
$$H_A: \text{at least one of the } \tau_i \text{ differs from zero}$$

For the Lambs, the estimates are: $\hat{\mu} = 13, \hat{\mu}_1 = 11, \hat{\mu}_2 = 15, \hat{\mu}_3 = 12, \hat{\tau}_1 = 11 - 13 = -2, \hat{\tau}_2 = 15 - 13 = 2, \hat{\tau}_3 = 12 - 13 = -1.$

It is important to reiterate that the *best* estimate of σ^2 is $s_P^2 = MS(within)$, and to point out that we can also write:

$$s_P^2 = MS(within) = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots (n_I - 1)s_I^2}{(n_1 - 1) + (n_2 - 1) + \cdots (n_I - 1)}$$

The denominator in this expression is ($n_\bullet$- I), but written in this form, it is then easy to see that s_P^2 is a <u>weighted average of the treatment sample variances</u> ($s_1^2, s_2^2, \ldots s_I^2$).

Equally important: when H_0 is true, $MS(between)$ also estimates σ^2, so then the F test statistic should then be around 1.

<u>§11.4. The Global F Test</u> – As mentioned above, the above null hypothesis (in either form) is tested using the test statistic

$$F_s = \frac{MS(between)}{MS(within)}$$

When H_0 is true, this TS has the $F_{df(between),df(within)}$ distribution; F tables are on pp. 628-637. Notice that for this distribution, there is both a numerator and denominator degrees of freedom. As for χ^2 tests, the rejection region for this test is <u>*only in the right tail*</u>.

For the above Soybean example, it turns out that $F_s(3,48) = 21.46$; from p.630, p-value < 0.0001 since 21.46 > 9.13. ***We conclude that not all the Soybean treatment total leaf area means are the same.***

For the above Lambs example, $F_s(2,9) = 0.77$, and from p.629, p-value > 0.20 since 0.77 < 1.93. ***Based on these data, we cannot reject the claim that the average weight gains are the same for the three lamb diets.***

Note that when k = 2,

- **the independent samples non-directional t-test assuming $\sigma_1^2 = \sigma_2^2$ (i.e., the non-directional pooled t-test), and**
- **the ANOVA global F-test**

are equivalent! That is, they give the same p-value and decision.

To illustrate, see the handout and the Diet (Bean versus Oat) example. For the pooled t-test, we get $t_s = 2.76, p = 0.012$; for the ANOVA F-test, we get $F_s = 7.64, p = 0.012$. Thus we note the relationship:

$$F_s = t_s^2$$

This is connection is only true when for the pooled t-test, the hypotheses are $H_0: \mu_1 = \mu_2$ and $H_A: \mu_1 \neq \mu_2$; the connection does not apply when the alternative is directional or if the variances are not assumed to be equal.

This connection is proven algebraically on p.2 of the same handout.

§11.5. Applicability of Methods – Assumptions and Requirements

1. The one-way ANOVA methods we have considered are valid for CRD's but not RBD's (see §11.6 for RBD's);
2. Require random samples from I populations or one random sample with independent assignment to I 'treatment' groups;
3. The I samples have been independently chosen;
4. The response variable for the I groups has Normal distributions;
5. The population variances are equal: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_I^2$

To reiterate, the global F-test is the much better way to test the null hypothesis ($H_0: \mu_1 = \mu_2 = \cdots = \mu_I$) as compared with separate t-tests. For the Soybean illustration, the 6 individual pooled t-test do not control the 'overall α level' and each t-test has denominator df = 24. On the other hand, the global F-test has denominator df = 48, and so this F test has higher power.

Another illustration: the Altman & Bland example (see handout p.3). The hypotheses are $H_0: \mu_{cont} = \mu_{cr} = \mu_{uc}$ versus H_A: at least one μ_i differs from the others. Here is the ANOVA table:

Source	df	SS	MS	F	p-value
Between Groups	2	5,174,310	2,587,155	7.34	0.002
Within Groups	39	13,743,776	352,405		
Total	41	18,918,086			

Using our text (p.629), we could say that 0.001 < p-value < 0.01. We conclude that not all the means are equal, ***but which means differ***?

In Section 11.9, we consider MCP's (multiple comparison procedures). The list of MCP's is long – we will use Minitab to perform the Tukey HSD ("Honestly Significant Different") MCP and the Fisher LSD ("Least Significant Difference") MCP – and will perform by hand the Bonferroni procedure.

From the handout (p.5), we see that Tukey's method yields:

```
CR          UC          CONT
----------------
            --------------------
```

The above easy manner to summarize our findings is called the 'underline method'; it conveys that here we retain $\mu_{cr} = \mu_{uc}$ and $\mu_{uc} = \mu_{cont}$ but we reject $\mu_{cr} = \mu_{cont}$, and conclude $\mu_{cr} \neq \mu_{cont}$.

For these data (handout p.6), Fisher's LSD method yields:

```
CR          UC          CONT
----------------        ---------
```

(Both CR and UC means differ from control.) Notice that the Tukey's HSD method and the Fisher's LSD method do not yield the same result; this is not unusual. Some statisticians feel that *Tukey's method is too conservative, and Fisher's is too liberal*, and so might use the Bonferroni method discussed below (but see handout p.8).

To illustrate the Bonferroni, here is an illustration from p.473 exercise 11.9.1 (Eggplants) and the response variable is eggplant dry weight (g) without roots. There are I = 5 soil treatments and we have k = ${}_5C_2$ =

10 pairs of means. We want to control the experiment-wise type I error rate at $\alpha_{ew} = 0.05$, so the Bonferroni method sets each comparison-wise error rate at $\alpha_{cw} = \frac{\alpha_{ew}}{k} = \frac{0.05}{10} = 0.005$. Thus, each of the comparison-wise CI's has level 99.5%, and the relevant t′ values come from Table 11 on p.638; in this table, df = df(within).

This study is balanced with each $n_i = 9$; we are given $s_P^2 = MS(within) = 0.2246$, and conclude that df(within) = 40. The means are: 4.37 (a), 4.76 (b), 3.70 (c), 5.41 (d), and 5.39 (e).

From Table 11, $t_{40,0.025/10} = 2.971$. (Without the Bonferroni adjustment, we would have wrongly used t = 2.021 from Table 4.)

Ten Bonferroni CI's need to be obtained, and the 95% Bonferroni CI for $\mu_e - \mu_a$ is

$$(5.38 - 4.37) \pm 2.971 \times \sqrt{0.2246} \times \sqrt{\frac{1}{9}+\frac{1}{9}}$$

This is 1.01 ± 0.6637 or (0.35 ,1.67); since the CI does not contain zero, we conclude that $\mu_e \neq \mu_a$. That leaves nine Bonferroni CI's to go, but the good news is that – since this study is balanced and all the pairs of treatments have the same sample sizes – we only need the one ME = 0.6637.

Next, try the Bonferroni method on the Altman & Bland data and also on the homework exercises.