Topic 4: Introduction to ANOVA

Montgomery: chapter 3

Prof. Lingling An
University of Arizona

Motivation

- We discussed method for comparing two conditions or treatments.
- But how about compare more than two conditions or levels of a factor?

What if there are more than two levels of a single factor?

- There are lots of practical situations where there are either more than two levels of interest, or there are several factors of simultaneous interest
- The *t*-test does not directly apply
- The **analysis of variance** (ANOVA) is the appropriate analysis “engine” for these types of experiments

ANOVA - Analysis of Variance

- Extends independent-samples t test
- Compares the means of groups of independent observations
  - Don't be fooled by the name
  - ANOVA does not compare variances
  - The name "analysis of variance" stems from a **partitioning** of the total variability in the response variable into components
  - Can compare more than two groups
- The ANOVA was developed by Fisher in the early 1920s, and initially applied to agricultural experiments. Now it's used widely.

ANOVA – Null and Alternative hypotheses

Say the sample contains ***a*** independent groups

- ANOVA tests the null hypothesis

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_a$$

  – That is, "the group means are all equal"

- The alternative hypothesis is

$$H_1: \mu_i \neq \mu_j \text{ for some } i, j$$

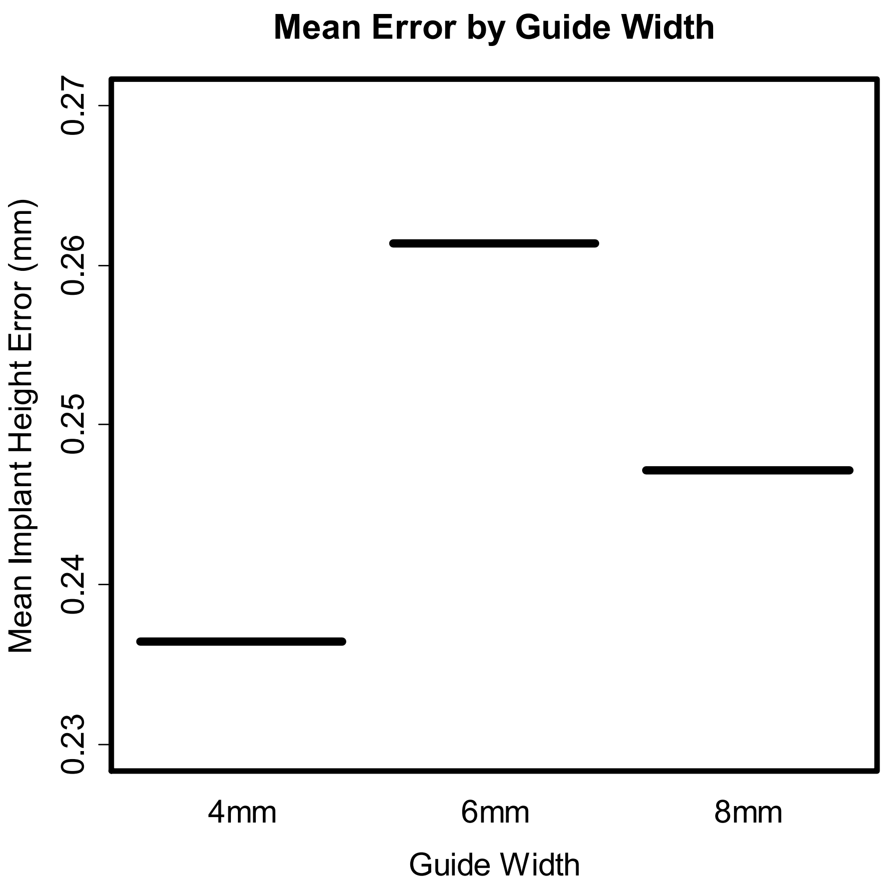
  – or, "the group means are *not* all equal"

Example: Accuracy of Implant Placement

Implants were placed in a manikin using placement guides of various widths.

15 implants were placed using each guide.

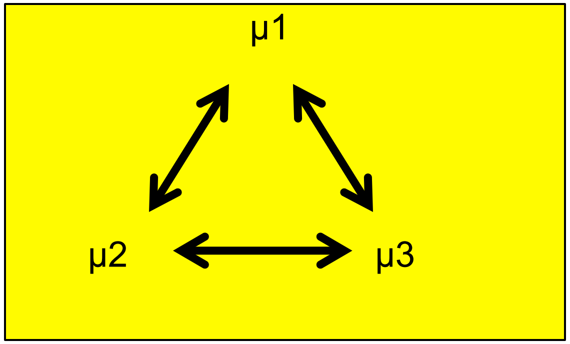
Error (discrepancies with a reference implant) was measured for each implant.



- Does **changing** the guide change the mean height error?
- Is there an **optimum** level for guide?
- We would like to have an objective way to answer these questions
- The *t*-test really doesn't apply here – more than two factor levels
  - Pairwise comparisons will inflate type I error

Why pairwise comparisons inflates type I error?

- Each time a hypothesis test is performed at significance level *α*, there is probability *α* of rejecting in error.
- Performing multiple tests increases the chances of rejecting in error *at least once*.
- For example:
  - if you did 3 independent hypothesis tests at the $\alpha = 0.05$
  - If, in truth, H_0 were true for all three.
  - The probability that *at least one* test rejects H_0 is 14.3%
    - P(at least one rejection) = 1-P(no rejections) = $1-.95^3 = 0.143$

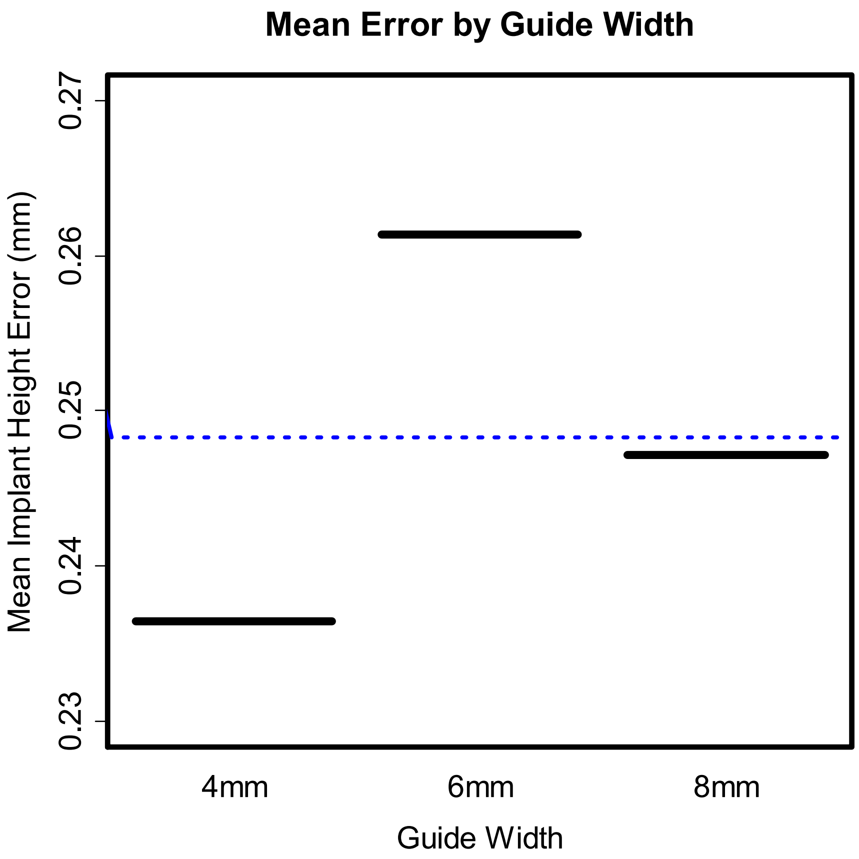


Example: Accuracy of Implant Placement -2

The overall mean of the entire sample was 0.248 mm.

This is called the "grand" mean, and is often denoted by $\bar{\bar{X}}$ or $\bar{X}..$

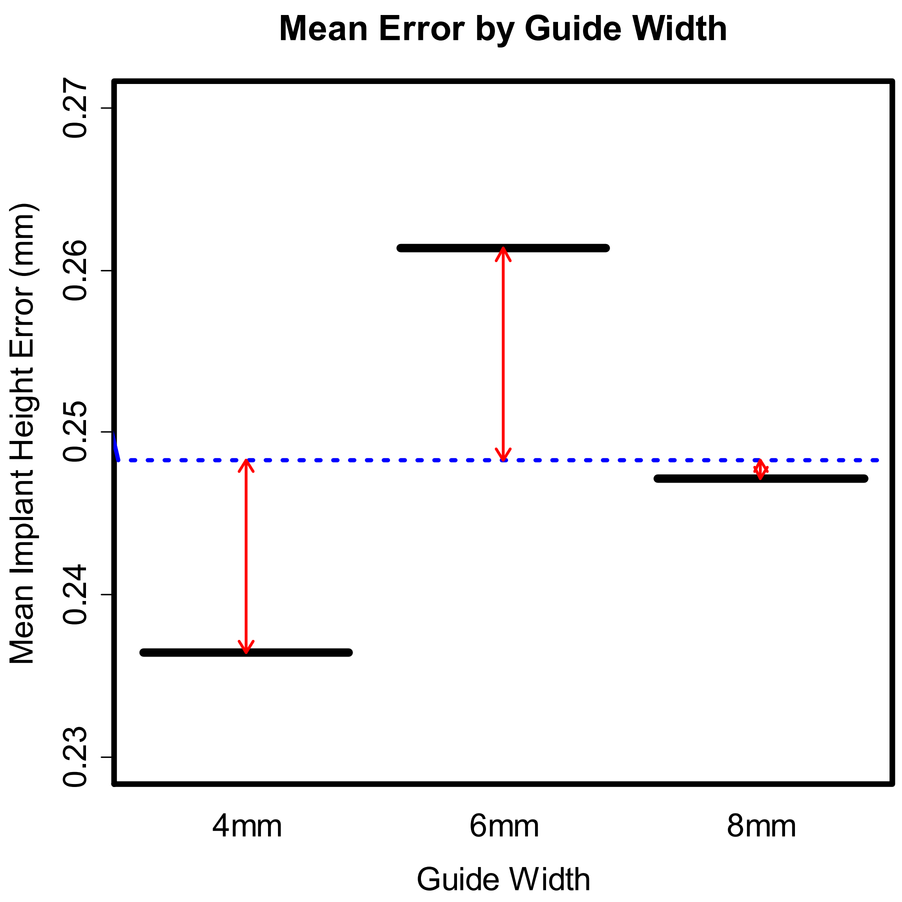
If H_0 were true then we'd expect the group means to be close to the grand mean.



Example: Accuracy of Implant Placement -3

The ANOVA test is based on the combined distances from $\bar{\bar{X}}$.

If the combined distances are large, that indicates we should reject H_0.



The ANOVA Statistic

To combine the differences from the grand mean we

- Square the differences
- Multiply by the numbers of observations in the groups
- Sum over the groups

$$\mathrm{SS_B} = 15\left(\overline{X}_{4mm} - \overline{\overline{X}}\right)^2 + 15\left(\overline{X}_{6mm} - \overline{\overline{X}}\right)^2 + 15\left(\overline{X}_{8mm} - \overline{\overline{X}}\right)^2$$

where the $\overline{X}_*$ are the group means.

"SS_B" = ***S***um of ***S***quares ***B***etween groups

Note: This looks a bit like a variance.

How big is big?

- For the Implant Accuracy Data, $SS_B = 0.0047$
- Is that big enough to reject H_0?
- As with the *t* test, we compare the statistic to the variability of the individual observations.
- In ANOVA the variability is estimated by the Mean Square Error, or MSE

MSE: Mean Square Error

The Mean Square Error is a measure of the variability after the group effects have been taken into account.

$$MSE = \frac{1}{N-a}\sum_j \sum_i \left(x_{ij} - \overline{X}_j\right)^2$$

where x_{ij} is the i^{th} observation in the j^{th} group.

MSE: Mean Square Error - 2

The Mean Square Error is a measure of the variability after the group effects have been taken into account.

$$MSE = \frac{1}{N-a} \sum_j \sum_i \left(x_{ij} - \overline{X}_j \right)^2$$

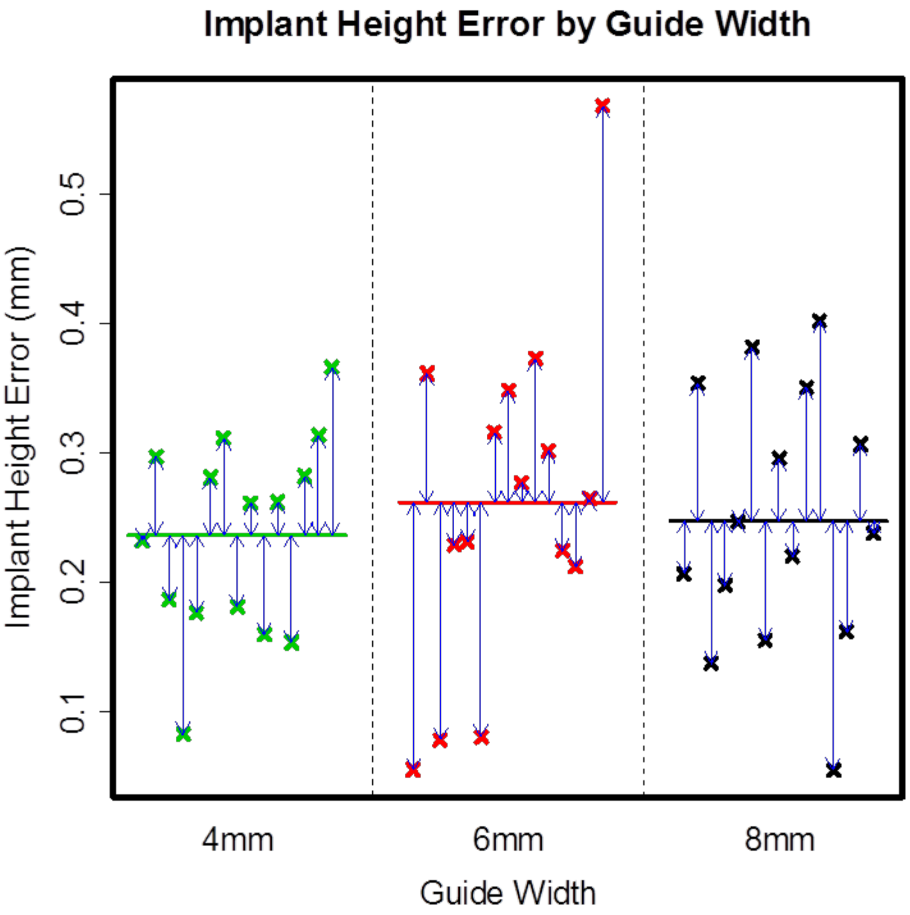
where x_{ij} is the i^{th} observation in the j^{th} group.

MSE: Mean Square Error - 3

The Mean Square Error is a measure of the variability after the group effects have been taken into account.

$$MSE = \frac{1}{N-a} \sum_j \sum_i \left(x_{ij} - \overline{X}_j \right)^2$$

Note that the variation of the means seems quite small compared to the variance of observations within groups



Notes on *MSE*

- If there are only two groups, the *MSE* is equal to the pooled estimate of variance used in the equal-variance *t* test.

- ANOVA assumes that all the group variances are equal.

- Other options should be considered if group variances differ by a factor of 2 or more.

ANOVA F Test

- The ANOVA F test is based on the F statistic

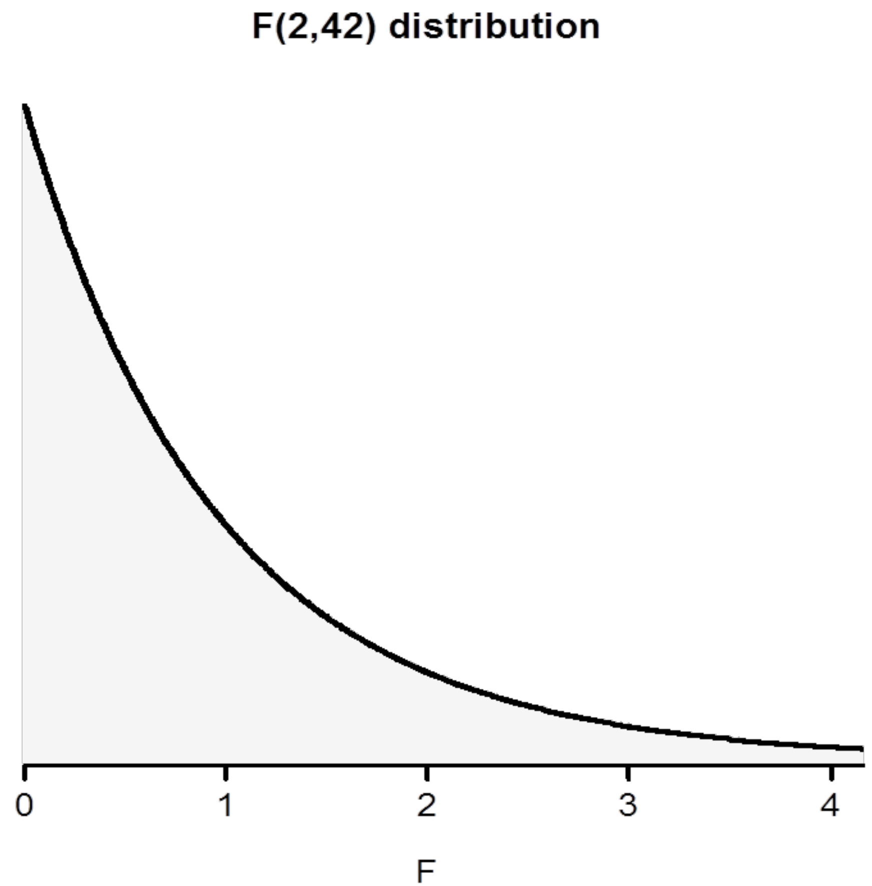
$$F = \frac{SS_B / (a-1)}{MSE}$$

where a is the number of groups.

- Under H_0 the F statistic has an "F" distribution, with a-1 and N-a degrees of freedom (N is the total number of observations)
  - In this case N=45

Implant Data: F test p-value

To get a p-value we compare our *F* statistic to an F(2, 42) distribution.

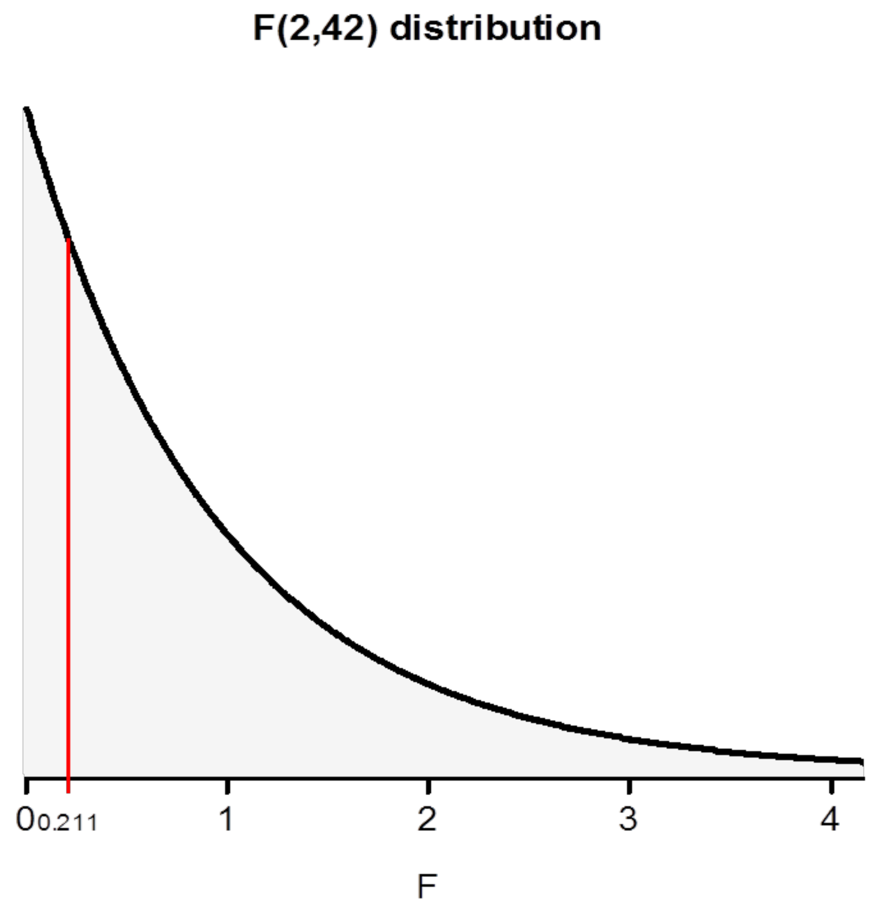


Implant Data: F test p-value - 2

To get a p-value we compare our *F* statistic to an F(2, 42) distribution.

In our example

$$F = \frac{.0047/2}{.0467/42} = .211$$



Implant Data: F test p-value - 3

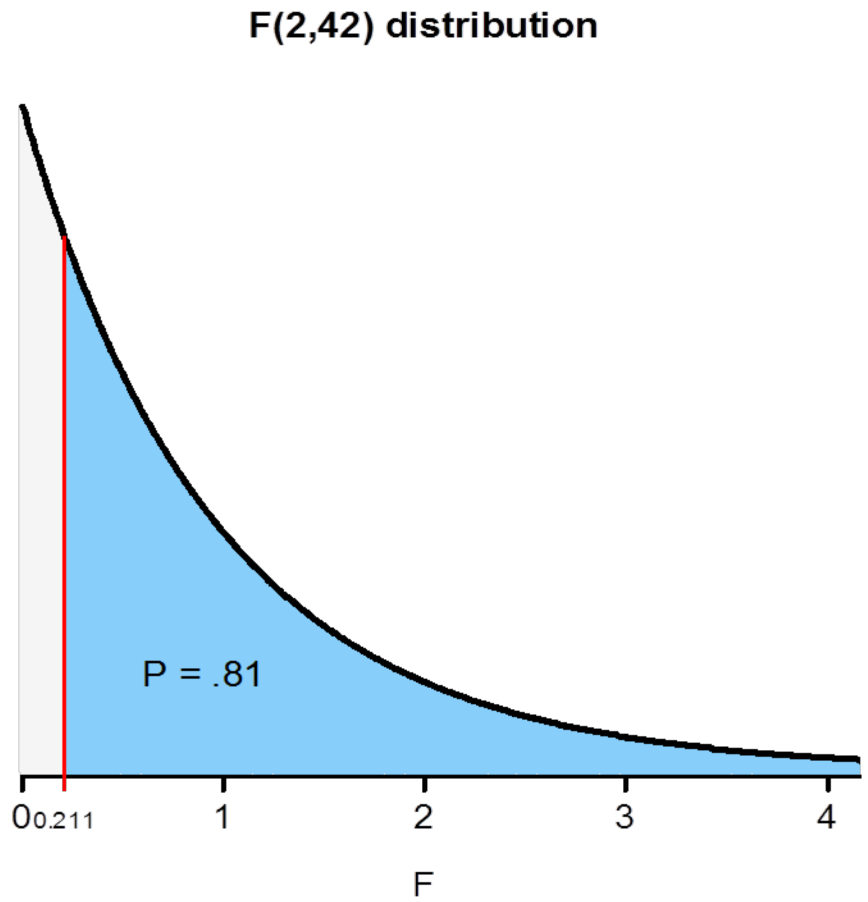
To get a p-value we compare our *F* statistic to an F(2, 42) distribution.

In our example

$$F = \frac{.0047/2}{.0467/42} = .211$$

The p-value is

$$P\left(F(2,42) \;\; > \;.211\right) = 0.81$$



ANOVA Table

Results are often displayed using an ANOVA Table

Source of Variation	Sum of Squares	df	Mean Square	F	P-value
Between Groups	.005	2	.002	.211	.811
Within Groups	.466	42	.011		
Total	.470	44			

Pop Quiz!: Where are the following quantities presented in this table?

Sum of Squares Between (SS_B)

Mean Square Error (MSE)

F Statistic

p value

ANOVA Table - 2

Results are often displayed using an ANOVA Table

Source of Variation	Sum of Squares	df	Mean Square	F	P-value
Between Groups	.005	2	.002	.211	.811
Within Groups	.466	42	.011		
Total	.470	44			

Sum of Squares Between (SS_B)

Mean Square Error (*MSE*)

F Statistic

p value

Statistic model: ANOVA

- The name "analysis of variance" stems from a **partitioning** of the total variability in the response variable into components that are consistent with a **model** for the experiment
- The basic single-factor ANOVA model is

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases}$$

Balanced design

$\mu =$ an overall mean, $\tau_i = ith$ treatment effect,

$\varepsilon_{ij} =$ experimental error, $NID(0, \sigma^2)$

τ_i is constant and $\sum \tau_i = 0$ => Fixed effect model

Models for the Data

There are two ways to write a model for the data:

$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ is called the effects model

Let $\mu_i = \mu + \tau_i$, then

$y_{ij} = \mu_i + \varepsilon_{ij}$ is called the means model

Regression models can also be employed

Notations for ANOVA

- **Total variability** is measured by the total sum of squares:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$$

- The basic ANOVA partitioning is:

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}[(\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})]^2$$

$$= n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$$

$$SS_T = SS_{Treatments} + SS_E$$

$$\|$$

$$SS_B$$

Notations for ANOVA - 2

$$SS_T = SS_{Treatments} + SS_E$$

$$\text{or} \quad SS_T = SS_B + SS_E$$

- A large value of $SS_{Treatments}$ reflects large differences in treatment means
- A small value of $SS_{Treatments}$ likely indicates no differences in treatment means
- Formal statistical hypotheses are:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_a$$

$$H_1 : \text{at least one "=" does not hold}$$

for means model

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_a = 0$$

$$H_1 : \text{at least one is not } 0$$

for effect model

Notations for ANOVA -3

- While sums of squares cannot be directly compared to test the hypothesis of equal means, **mean squares** can be compared.
- A mean square is a sum of squares divided by its degrees of freedom:

$$df_{Total} = df_{Treatments} + df_{Error}$$

$$an - 1 = a - 1 + a(n-1)$$

$$MS_{Treatments} = \frac{SS_{Treatments}}{a-1}, MS_E = \frac{SS_E}{a(n-1)}$$

- If the treatment means are equal, $\mathrm{MS}_{\mathrm{Teatments}}$ =0.

The Analysis of Variance is Summarized in a Table

■ **TABLE 3.3**

The Analysis of Variance Table for the Single-Factor, Fixed Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Between treatments	$SS_{\text{Treatments}} = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2$	$a - 1$	$MS_{\text{Treatments}}$	$F_0 = \dfrac{MS_{\text{Treatments}}}{MS_E}$
Error (within treatments)	$SS_E = SS_T - SS_{\text{Treatments}}$	$N - a$	MS_E	
Total	$SS_{\text{T}} = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$	$N-1$		

- Computing…see text, p74
- The **reference distribution** for F_0 is the $F_{a-1,\, a(n-1)}$ distribution
- **Reject** the null hypothesis (equal treatment means) if

$$F_0 > F_{\alpha, a-1, a(n-1)}$$

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} \tag{3.8}$$

$$SS_{\text{Treatments}} = \frac{1}{n} \sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N} \tag{3.9}$$

$$SS_E = SS_T - SS_{\text{Treatments}} \tag{3.10}$$

Some words for coding the observations

- If every observation *subtracts* the same constant, then sums of squares do not change, so we can get the same conclusion.

- If we *multiply* each observation by the same constant, then the sums of squares change. But the F ratio is equal to the F ratio for the original data. It implies that we can still get the same conclusion.

Parameter estimation

- Estimates for parameters:

$$\hat{\mu} = \overline{y}_{..}$$

$$\hat{\tau}_i = (\overline{y}_{i.} - \overline{y}_{..})$$

$$\hat{\epsilon}_{ij} = y_{ij} - \overline{y}_{i.} \qquad (\text{ residual })$$

So $y_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\epsilon}_{ij}$.

- So estimator of mean of the ith treatment is:

$$\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \bar{y}_{i.}$$

- And the 100(1-α)% confidence interval for the ith treatment mean:

$$\bar{y}_{i.} - t_{\alpha/2,N-a}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_{i.} + t_{\alpha/2,N-a}\sqrt{\frac{MS_E}{n}}$$

- C.I. For the difference of two treatment means:

$$\bar{y}_{i.} - \bar{y}_{j.} - t_{\alpha/2,N-a}\sqrt{\frac{2MS_E}{n}} \le \mu_i - \mu_j \le \bar{y}_{i.} - \bar{y}_{j.} + t_{\alpha/2,N-a}\sqrt{\frac{2MS_E}{n}}$$

Model for unbalanced experiment

- More general model for unbalanced experiment :

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \; i = 1,2,..,a$$
$$j = 1,2,\ldots,n_i$$

- Notation
  - $y_{i.} = \sum_{j=1}^{n_i} y_{ij} \rightarrow \overline{y}_{i.} = y_{i.}/n_i$ (treatment sample mean, or row mean)
  - $y_{..} = \sum\sum y_{ij} \rightarrow \overline{y}_{..} = y_{..}/N$ (grand sample mean)
- Decomposition of y_{ij}: $\quad y_{ij} = \overline{y}_{..} + (\overline{y}_{i.} - \overline{y}_{..}) + (y_{ij} - \overline{y}_{i.})$

- Estimates for parameters:

$$\hat{\mu} = \overline{y}_{..}$$

$$\hat{\tau}_i = (\overline{y}_{i.} - \overline{y}_{..})$$

$$\hat{\epsilon}_{ij} = y_{ij} - \overline{y}_{i.} \qquad \text{(residual)}$$

So $y_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\epsilon}_{ij}$.

- *For stat students*
  - It can be verified that

$$\sum_{i=1}^{a} n_i \hat{\tau}_i = 0; \qquad \sum_{j=1}^{n_i} \hat{\epsilon}_{ij} = 0 \text{ for all } i.$$

Another example: Tensile Strength

- Investigate the tensile strength of a new synthetic fiber. The factor is the weight percent of cotton used in the blend of the materials for the fiber and it has five levels.

percent	tensile strength						
of cotton	1	2	3	4	5	total	average
15	7	7	11	15	9	49	9.8
20	12	17	12	18	18	77	15.4
25	14	18	18	19	19	88	17.6
30	19	25	22	19	23	108	21.6
35	7	10	11	15	11	54	10.8

SAS code

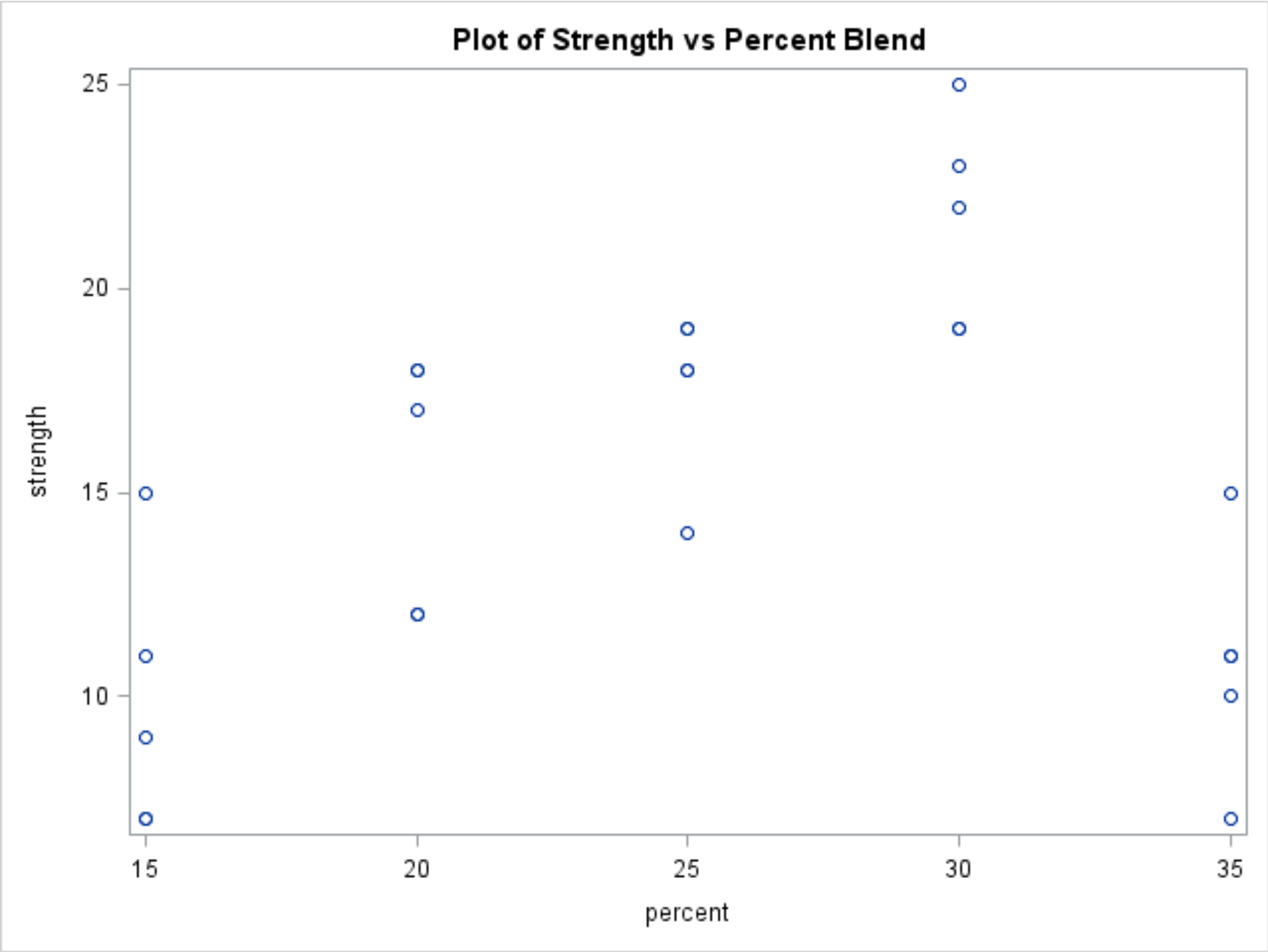
```
options ls=75 ps=60 nocenter;

data one;
 infile 'D:\TEACHING\T_stat571B\lab\sas_data
\tensile.dat';
 input percent strength time;
 run;

title1 'Tensile Strength  example';
proc print data=one;
run;
```

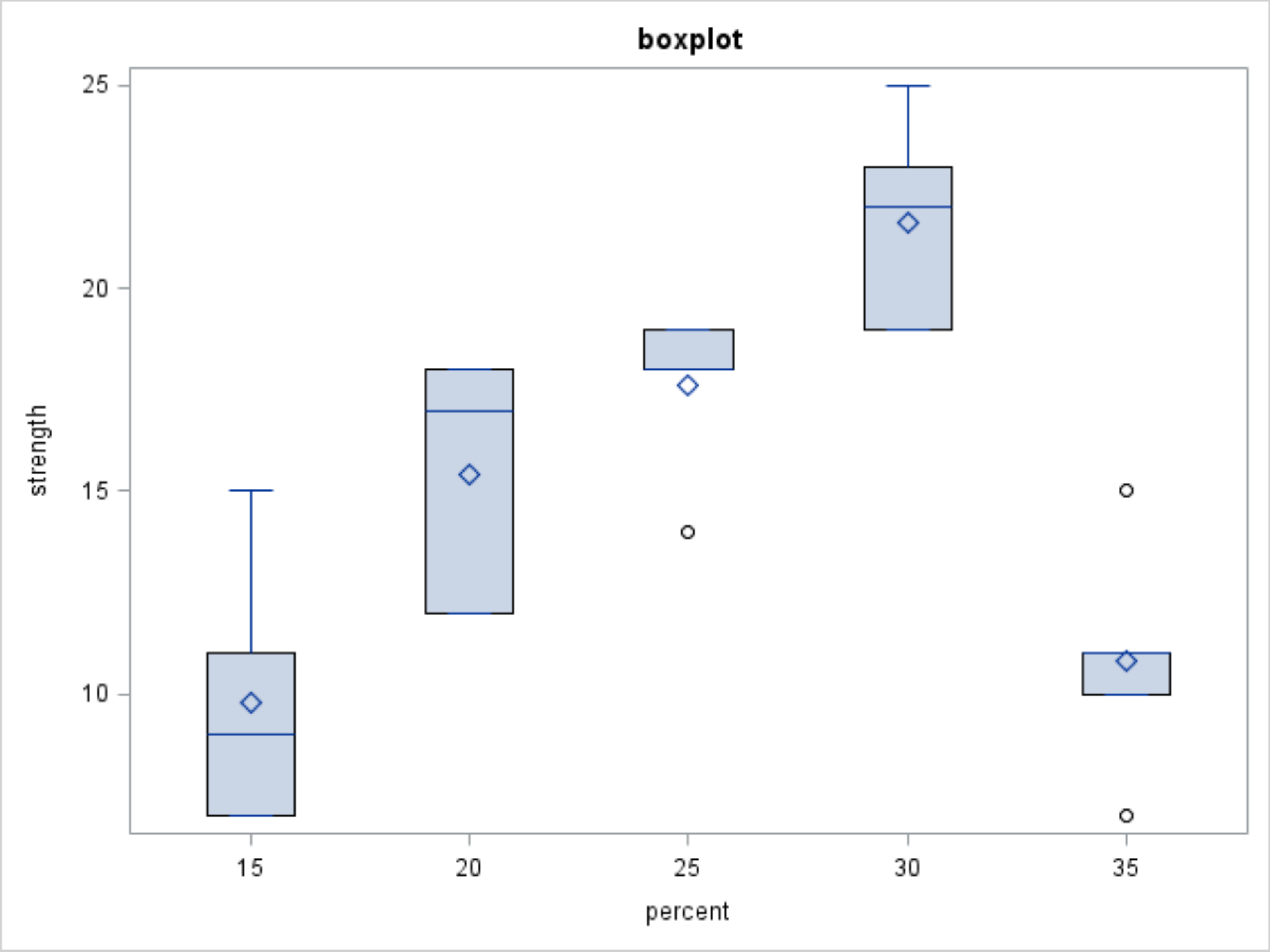
Tensile Strength example

Obs	percent	strength	time
1	15	7	15
2	15	7	19
3	15	15	25
4	15	11	12
5	15	9	6
6	20	12	8
7	20	17	14
8	20	12	1
9	20	18	11
10	20	18	3
11	25	14	18
12	25	18	13
13	25	18	20
14	25	19	7
15	25	19	9
16	30	19	22
17	30	25	5
18	30	22	2
19	30	19	24
20	30	23	10
21	35	7	17
22	35	10	21
23	35	11	4
24	35	15	16
25	35	11	23

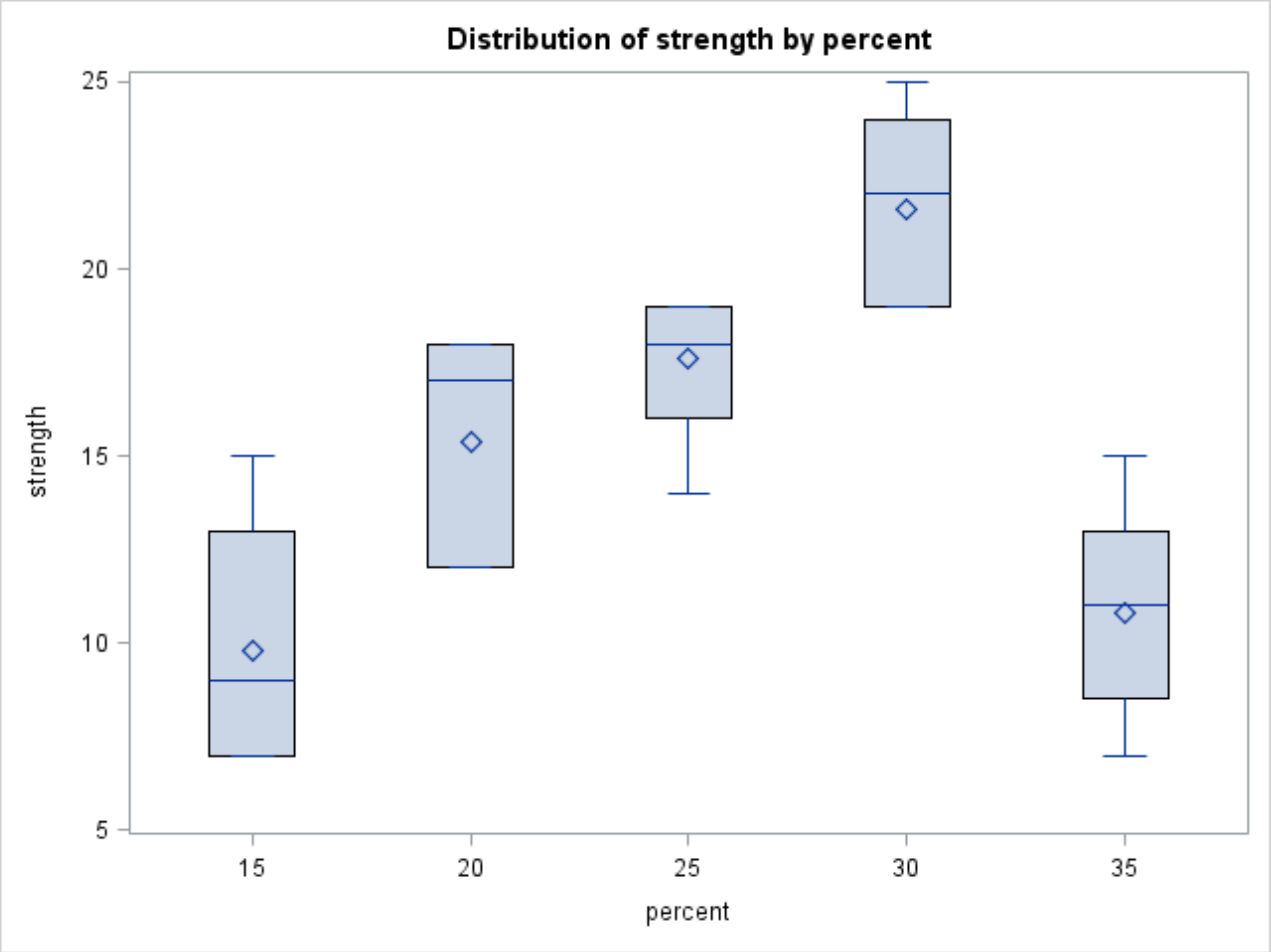
```
symbol1 v=circle;
title1 'Plot of Strength vs Percent Blend';
proc sgplot data=one;
scatter x=percent y=strength;
run;

title1 'boxplot';
proc sgplot data=one;
vbox strength/category=percent;
run;

proc boxplot;
 plot strength*percent/boxstyle=skeletal pctldef=4;
 run;
```

Scatter plot




boxplot
strength
25
20
15
10
15
20
25
30
35
percent

Side-by-Side boxPlot

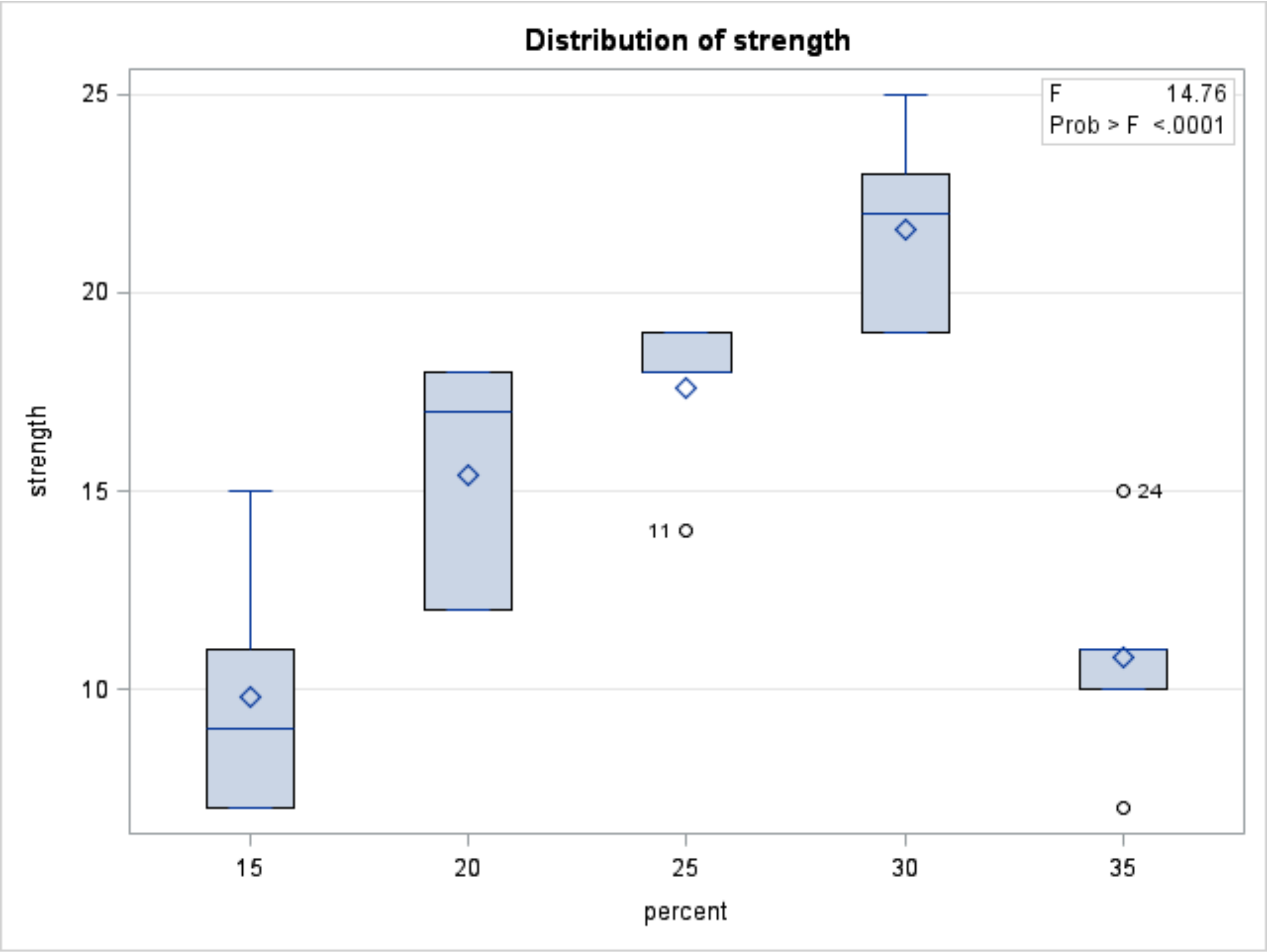


```
proc glm;
 class percent;
 model strength=percent;
 output out=oneres p=pred r=res;
run;
```

The GLM Procedure
Dependent Variable: strength

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	475.760000	118.9400000	14.76	<.0001
Error	20	161.200000	8.0600000		
Corrected Total	24	636.960000			

R-Square	Coeff Var	Root MSE	strength Mean
0.746923	18.87642	2.839014	15.04000



Distribution of strength
F 14.76
Prob > F <.0001
25
20
15
10
strength
11
24
15
20
25
30
35
percent

Last slide

- Read Sections: 3.1-3.3