مقایسه های ابتدایی

Design of Engineering Experiments Part 2 – Basic Statistical Concepts

- Simple **comparative** experiments
  - The hypothesis testing framework
  - The two-sample *t*-test
  - Checking assumptions, validity
- Comparing more that two factor levels…**the analysis of variance**
  - ANOVA decomposition of total variability
  - Statistical testing & analysis
  - Checking assumptions, model validity
  - Post-ANOVA testing of means
- **Sample size** determination

Portland Cement Formulation (page 23)

Table 2-1 Tension Bond Strength Data for the Portland Cement Formulation Experiment

j	Modified Mortar y_{1j}	Unmodified Mortar y_{2j}
1	16.85	17.50
2	16.40	17.63
3	17.21	18.25
4	16.35	18.00
5	16.52	17.86
6	17.04	17.75
7	16.96	18.22
8	17.15	17.90
9	16.59	17.96
10	16.57	18.15

Graphical View of the Data

Dot Diagram, Fig. 2-1, pp. 24

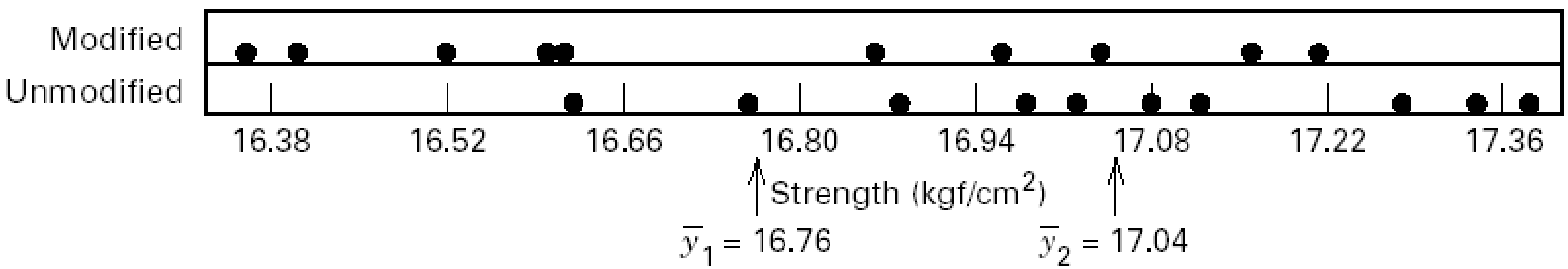

Figure 2-1 Dot diagram for the tension bond strength data in Table 2-1.

Box Plots, Fig. 2-3, pp. 26

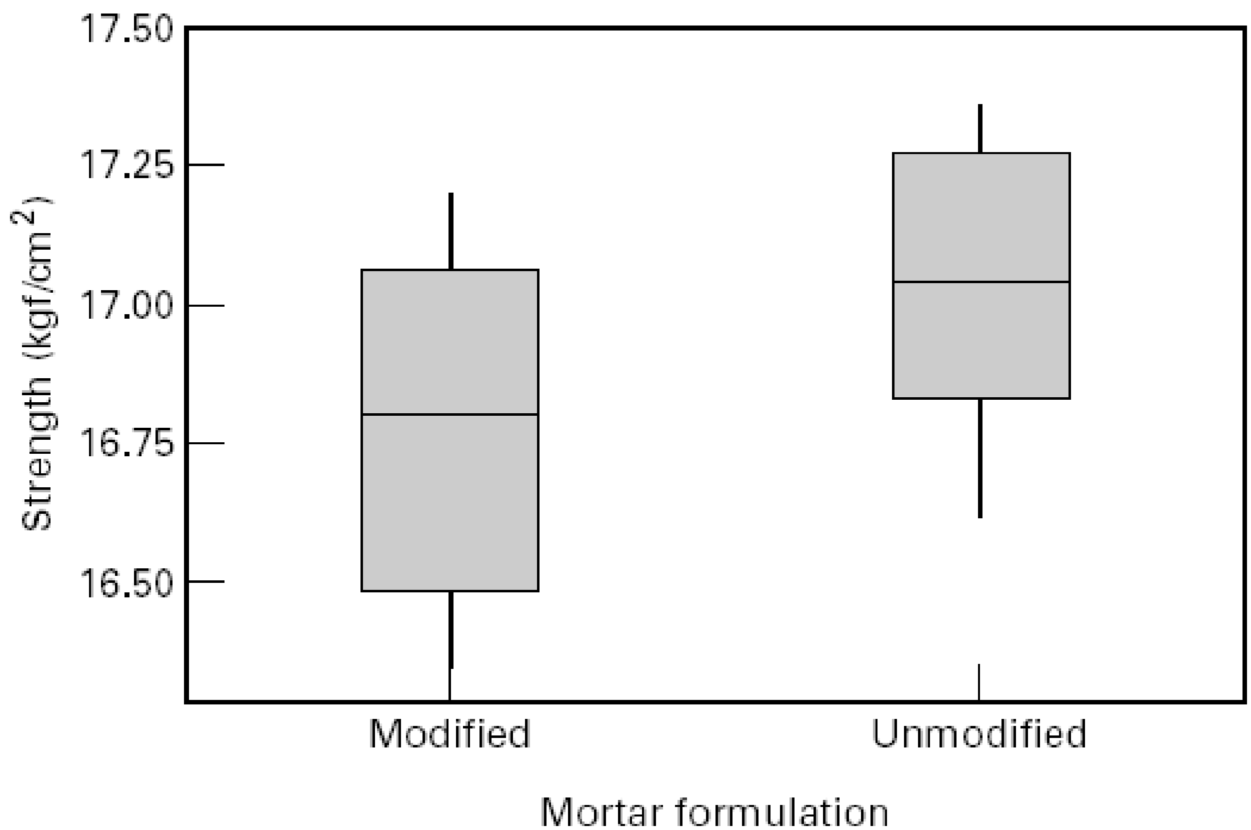

Figure 2-3 Box plots for the portland cement tension bond strength experiment.

The Hypothesis Testing Framework

- **Statistical hypothesis testing** is a useful framework for many experimental situations
- Origins of the methodology date from the early 1900s
- We will use a procedure known as the **two-sample *t*-test**

The Hypothesis Testing Framework

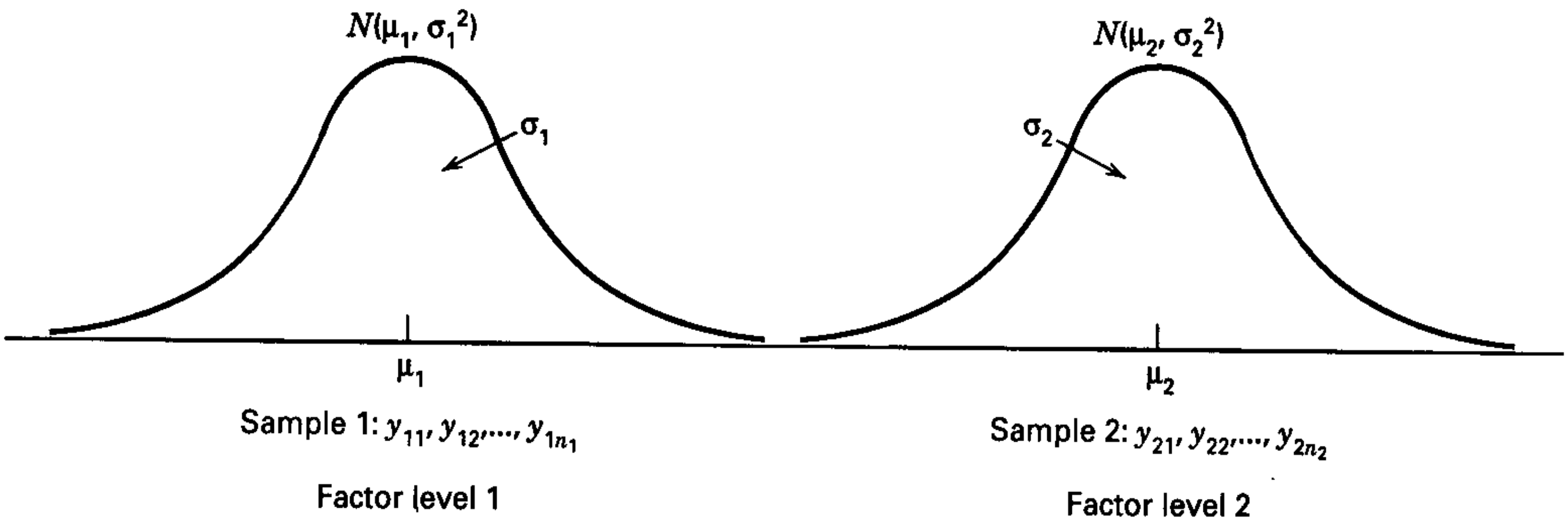

Figure 2-9 The sampling situation for the two-sample *t*-test.

- Sampling from a **normal** distribution
- Statistical hypotheses:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Estimation of Parameters

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \text{ estimates the population mean } \mu$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \text{ estimates the variance } \sigma^2$$

Summary Statistics (pg. 36)

Formulation 1

"New recipe"

$$\bar{y}_1 = 16.76$$

$$S_1^2 = 0.100$$

$$S_1 = 0.316$$

$$n_1 = 10$$

Formulation 2

"Original recipe"

$$\bar{y}_1 = 17.04$$

$$S_1^2 = 0.061$$

$$S_1 = 0.248$$

$$n_1 = 10$$

How the Two-Sample *t*-Test Works:

Use the sample means to draw inferences about the population means

$$\bar{y}_1 - \bar{y}_2 = 16.76 - 17.04 = -0.28$$

$$\frac{\text{Difference in sample means}}{\text{Standard deviation of the difference in sample means}}$$

$$\sigma^2_{\bar{y}} = \frac{\sigma^2}{n}$$

This suggests a statistic:

$$Z_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

How the Two-Sample *t*-Test Works:

Use S_1^2 and S_2^2 to estimate σ_1^2 and σ_2^2

The previous ratio becomes $\dfrac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$

However, we have the case where $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Pool the individual sample variances:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

How the Two-Sample *t*-Test Works:

The test statistic is

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- Values of t_0 that are near zero are consistent with the null hypothesis
- Values of t_0 that are very different from zero are consistent with the alternative hypothesis
- t_0 is a "distance" measure-how far apart the averages are expressed in standard deviation units
- Notice the interpretation of t_0 as a **signal-to-noise** ratio

The Two-Sample (Pooled) *t*-Test

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{9(0.100) + 9(0.061)}{10 + 10 - 2} = 0.081$$

$$S_p = 0.284$$

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{16.76 - 17.04}{0.284\sqrt{\frac{1}{10} + \frac{1}{10}}} = -2.20$$

The two sample means are a little over two standard deviations apart
Is this a "large" difference?

The Two-Sample (Pooled) *t*-Test

- So far, we haven't really done any "statistics"
- We need an **objective** basis for deciding how large the test statistic t_0 really is
- In 1908, W. S. Gosset derived the **reference distribution** for t_0 … called the *t* distribution
- Tables of the *t* distribution - text, page 606

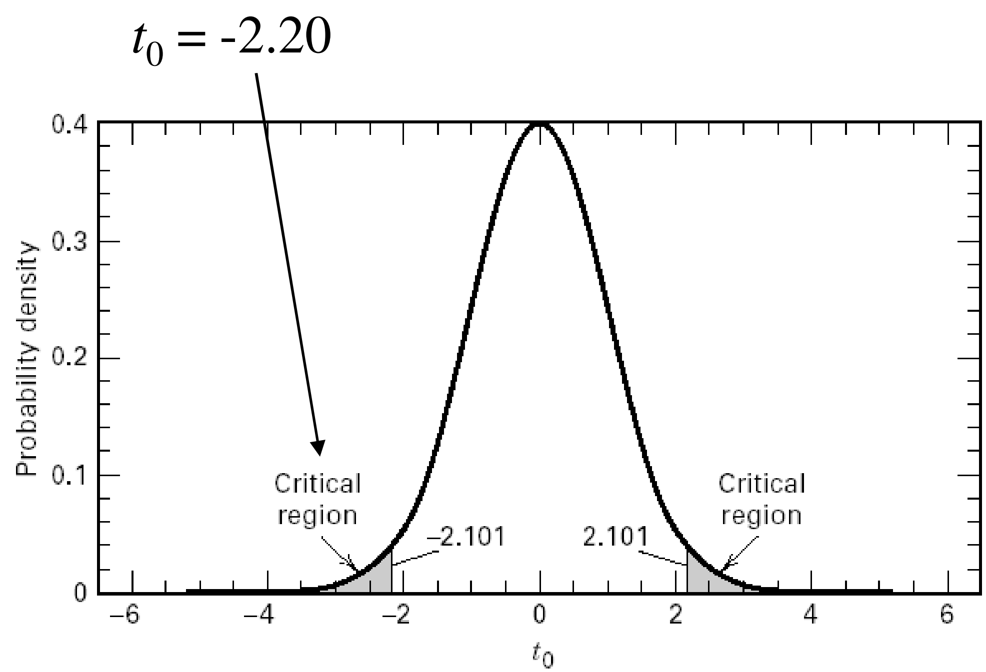

Figure 2-10 The t distribution with 18 degrees of freedom with the critical region $\pm t_{0.025,18} = \pm 2.101$.

The Two-Sample (Pooled) *t*-Test

- A value of t_0 between –2.101 and 2.101 is consistent with equality of means
- It is possible for the means to be equal and t_0 to exceed either 2.101 or –2.101, but it would be a "**rare event**" … leads to the conclusion that the means are different
- Could also use the ***P*-value** approach

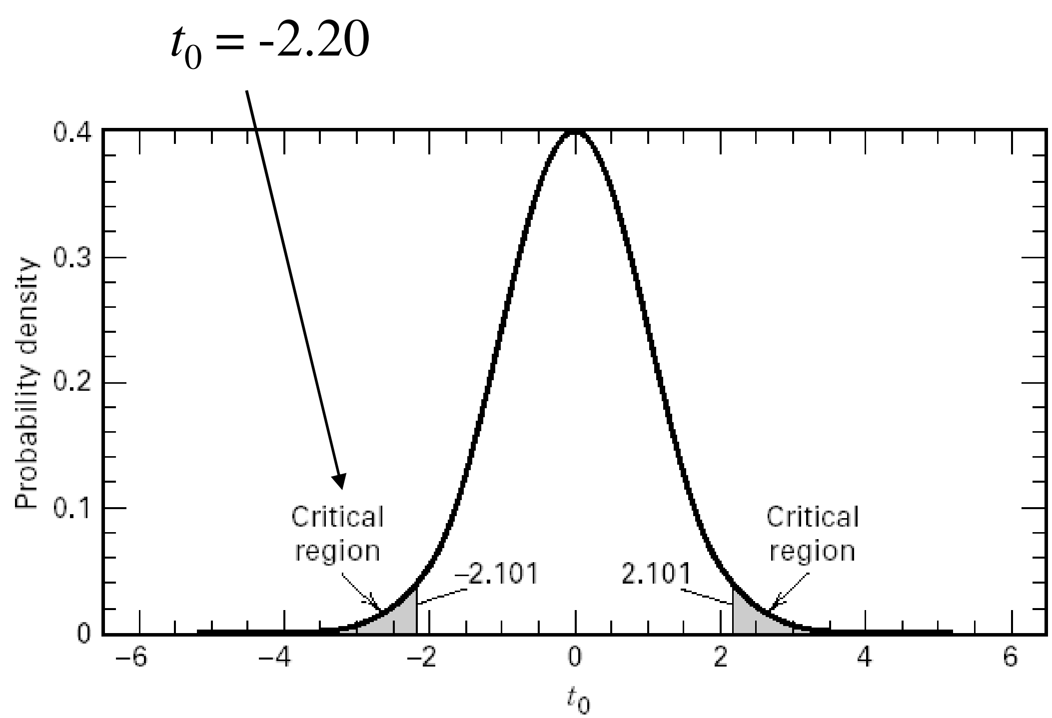

Figure 2-10 The t distribution with 18 degrees of freedom with the critical region $\pm t_{0.025,18} = \pm 2.101$.

The Two-Sample (Pooled) t-Test

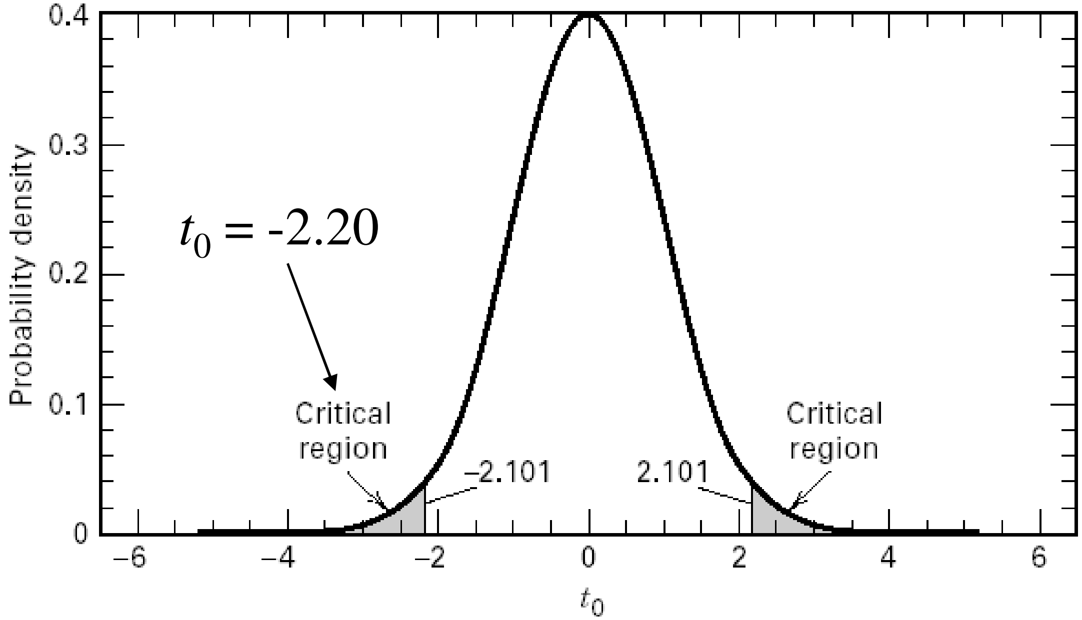

Figure 2-10 The t distribution with 18 degrees of freedom with the critical region $\pm t_{0.025,18} = \pm 2.101$.

- The ***P*-value** is the risk of **wrongly rejecting** the null hypothesis of equal means (it measures rareness of the event)
- The P-value in our problem is $P = 0.042$

Minitab Two-Sample *t*-Test Results

Table 2-2 Two-Sample *t*-Test from Minitab

```
Two-sample T for Modified vs Unmodified

                  N      Mean     StDev    SE Mean
Modified         10    16.764     0.316       0.10
Unmodified       10    17.042     0.248      0.078

Difference = mu (Modified) - mu (Unmodified)
Estimate for difference: -0.278000
95% CI for difference: (-0.545073, -0.010927)
T-Test of difference = 0 (vs not =): T-Value = -2.19
P-Value = 0.042 DF = 18
Both use Pooled StDev = 0.2843
```

Checking Assumptions – The Normal Probability Plot

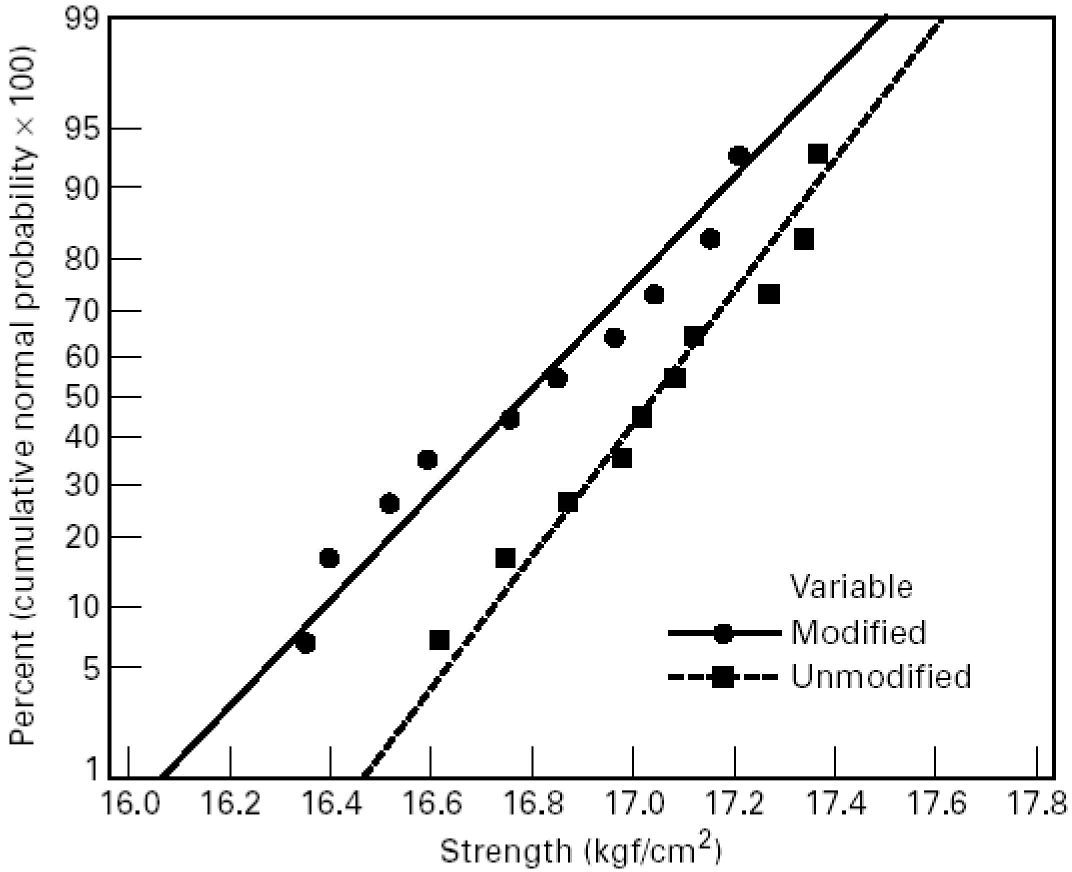

Figure 2-11 Normal probability plots of tension bond strength in the portland cement experiment.

Importance of the *t*-Test

- Provides an **objective** framework for simple comparative experiments
- Could be used to test all relevant hypotheses in a two-level factorial design, because all of these hypotheses involve the mean response at one "side" of the cube versus the mean response at the opposite "side" of the cube

Confidence Intervals (See pg. 43)

- Hypothesis testing gives an objective statement concerning the difference in means, but it doesn't specify "how different" they are
- **General form** of a confidence interval

$$L \leq \theta \leq U \text{ where } P(L \leq \theta \leq U) = 1 - \alpha$$

- The 100(1- α)% **confidence interval** on the difference in two means:

$$\bar{y}_1 - \bar{y}_2 - t_{\alpha/2, n_1+n_2-2} S_p \sqrt{(1/n_1) + (1/n_2)} \leq \mu_1 - \mu_2 \leq$$

$$\bar{y}_1 - \bar{y}_2 + t_{\alpha/2, n_1+n_2-2} S_p \sqrt{(1/n_1) + (1/n_2)}$$

What If There Are More Than Two Factor Levels?

- The *t*-test does not directly apply
- 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 **analysis of variance** (ANOVA) is the appropriate analysis "engine" for these types of experiments – Chapter 3, textbook
- The ANOVA was developed by Fisher in the early 1920s, and initially applied to agricultural experiments
- Used extensively today for industrial experiments

An Example (See pg. 60)

- An engineer is interested in investigating the relationship between the RF power setting and the etch rate for this tool. The objective of an experiment like this is to model the relationship between etch rate and RF power, and to specify the power setting that will give a desired target etch rate.
- The response variable is etch rate.
- She is interested in a particular gas (C2F6) and gap (0.80 cm), and wants to test four levels of RF power: 160W, 180W, 200W, and 220W. She decided to test five wafers at each level of RF power.
- The experimenter chooses 4 **levels** of RF power 160W, 180W, 200W, and 220W
- The experiment is **replicated** 5 times – runs made in random order

An Example (See pg. 62)

Table 3-1 Etch Rate Data (in Å/min) from the Plasma Etching Experiment

Power (W)	Observations 1	2	3	4	5	Totals	Averages
160	575	542	530	539	570	2756	551.2
180	565	593	590	579	610	2937	587.4
200	600	651	610	637	629	3127	625.4
220	725	700	715	685	710	3535	707.0

- Does **changing** the power change the mean etch rate?
- Is there an **optimum** level for power?

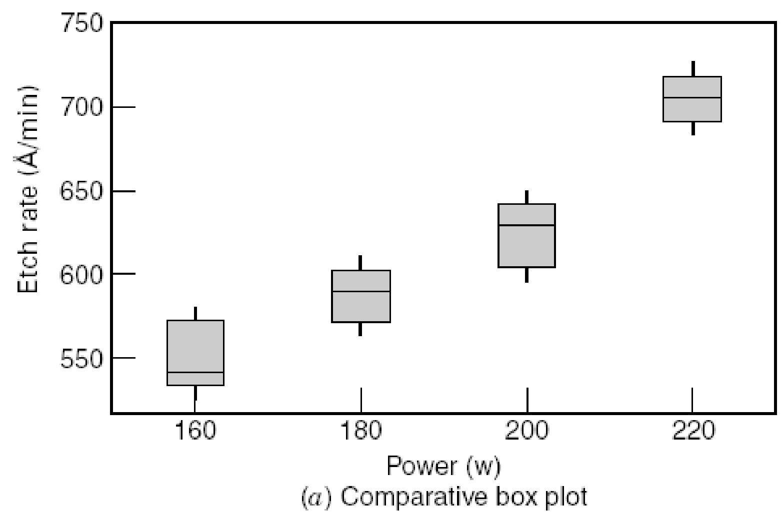

Figure 3-2 Box plots and scatter diagram of the etch rate data.

The Analysis of Variance (Sec. 3-2, pg. 63)

Table 3-2 Typical Data for a Single-Factor Experiment

Treatment (level)	Observations				Totals	Averages
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2.}$	$\bar{y}_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$y_{a.}$	$\bar{y}_{a.}$
					$y_{..}$	$\bar{y}_{..}$

- In general, there will be *a* **levels** of the factor, or *a* **treatments, and** *n* **replicates** of the experiment, run in **random order…**a completely randomized design **(CRD)**
- $N = an$ total runs
- We consider the **fixed effects** case…the **random effects** case will be discussed later
- Objective is to test hypotheses about the equality of the *a* treatment means

The Analysis of Variance

- 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

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

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

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

Models for the Data

There are several 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

The Analysis of Variance

- **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:

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

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

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

The Analysis of Variance

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

- 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:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_a$$
$$H_1 : \text{ At least one mean is different}$$

The Analysis of Variance

- 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, the treatment and error mean squares will be (theoretically) equal.
- If treatment means differ, the treatment mean square will be larger than the error mean square.

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	$SS_{\text{Treatments}} = n \sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^2$	$a - 1$	$MS_{\text{Treatments}}$	$F_0 = \frac{MS_{\text{Treatments}}}{MS_E}$
Error (within treatments)	$SS_E = SS_T - SS_{\text{Treatments}}$	$N - a$	MS_E	
Total	$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2$	$N - 1$		

- Computing…see text, pp 66-70
- The **reference distribution** for F_0 is the $F_{a\text{-}1,\, a(n\text{-}1)}$ distribution
- **Reject** the null hypothesis (equal treatment means) if

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

ANOVA Table
Example 3-1

Table 3-4 ANOVA for the Plasma Etching Experiment

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
RF Power	66,870.55	3	22,290.18	$F_0 = 66.80$	<0.01
Error	5339.20	16	333.70		
Total	72,209.75	19			

The Reference Distribution:

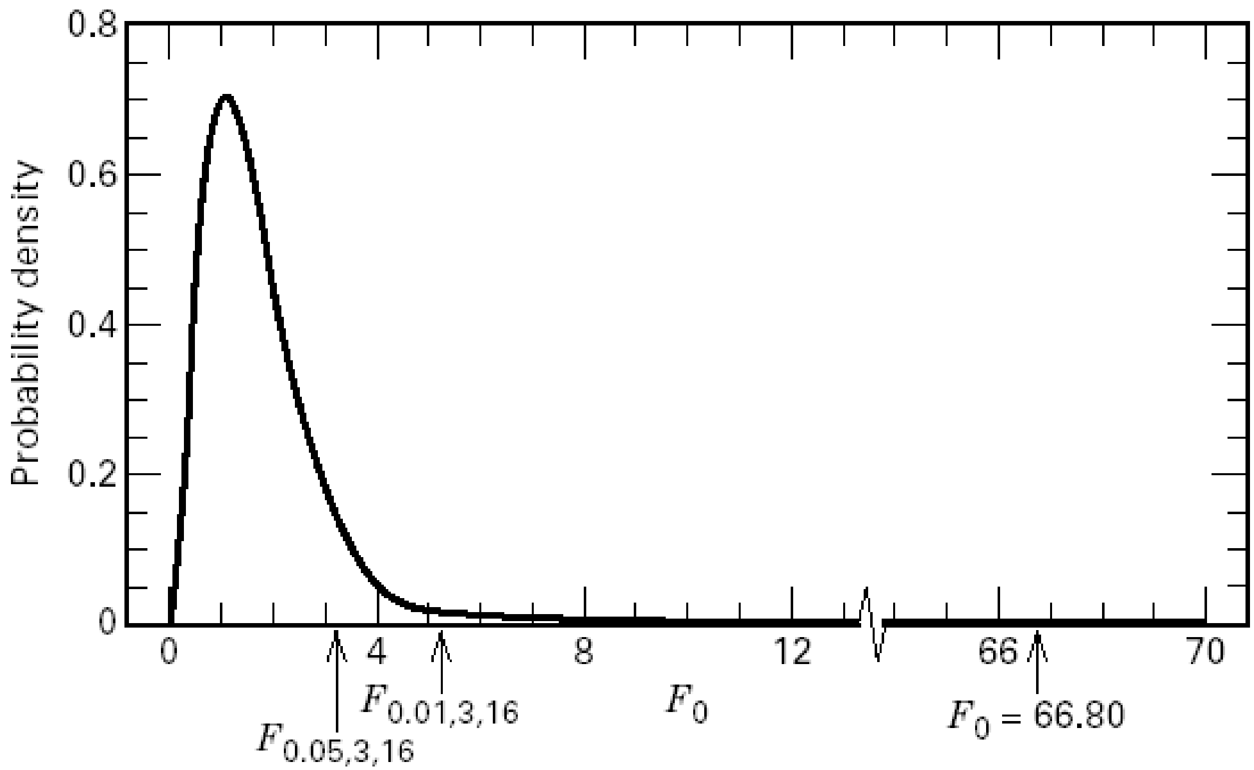

Figure 3-3 The reference distribution ($F_{3,16}$) for the test statistic F_0 in Example 3-1.

ANOVA calculations are usually done via computer

- Text exhibits sample calculations from two very popular software packages, Design-Expert and Minitab
- See page 99 for Design-Expert, page 100 for Minitab
- Text discusses some of the summary statistics provided by these packages

Model Adequacy Checking in the ANOVA
Text reference, Section 3-4, pg. 75

- **Checking assumptions** is important
- Normality
- Constant variance
- Independence
- Have we fit the right model?
- Later we will talk about what to do if some of these assumptions are **violated**

Model Adequacy Checking in the ANOVA

- Examination of **residuals** (see text, Sec. 3-4, pg. 75)

$$e_{ij} = y_{ij} - \hat{y}_{ij}$$
$$= y_{ij} - \bar{y}_{i.}$$

- Design-Expert generates the residuals
- **Residual plots** are very useful
- **Normal probability plot** of residuals

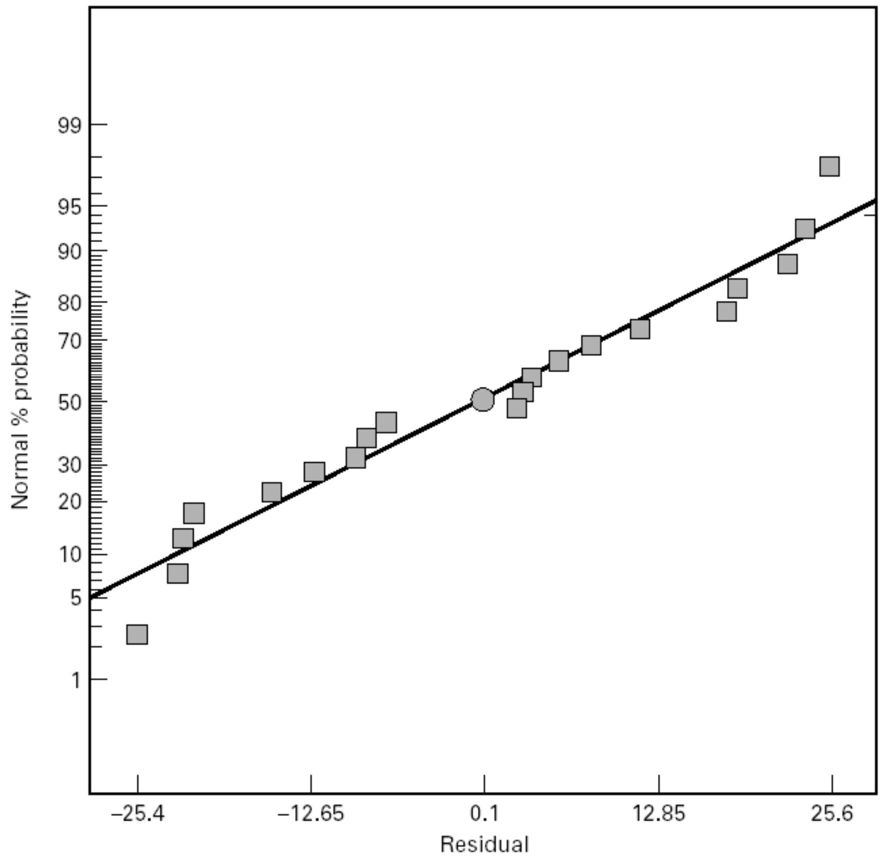

Figure 3-4 Normal probability plot of residuals for Example 3-1.

Other Important Residual Plots

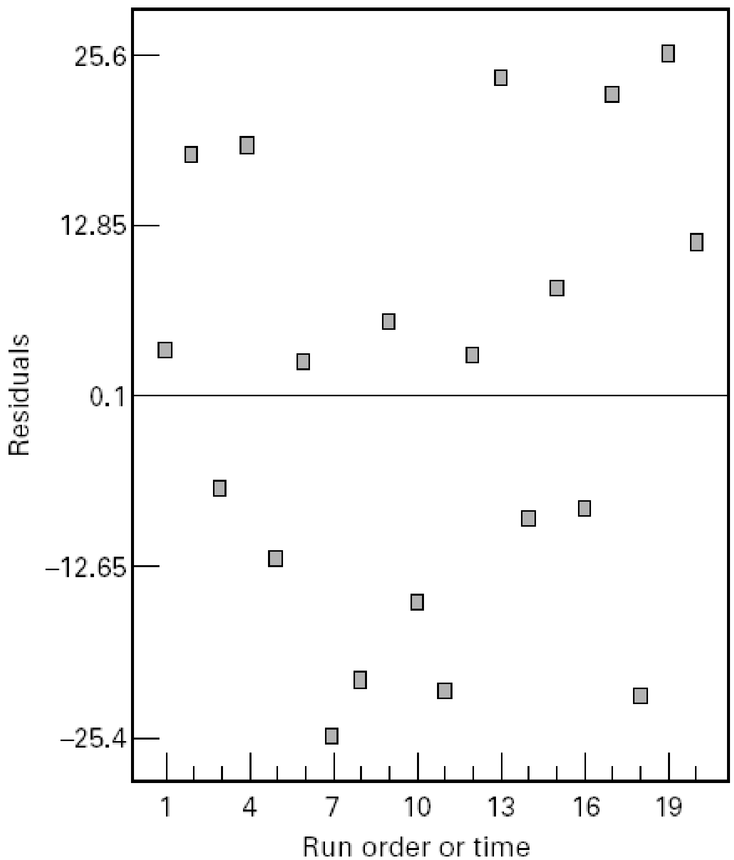

Figure 3-5 Plot of residuals versus run order or time.

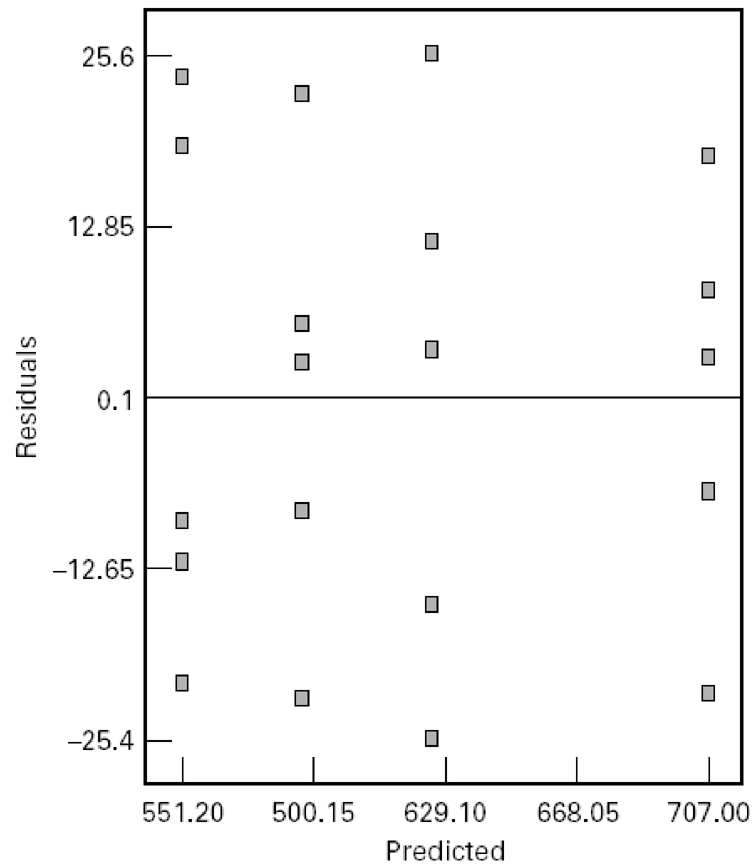

Figure 3-6 Plot of residuals versus fitted values.

Post-ANOVA Comparison of Means

- The analysis of variance tests the hypothesis of equal treatment means
- Assume that residual analysis is satisfactory
- If that hypothesis is rejected, we don't know **which specific means** are different
- Determining which specific means differ following an ANOVA is called the **multiple comparisons problem**
- There are **lots** of ways to do this…see text, Section 3-5, pg. 87
- We will use pairwise *t*-tests on means…sometimes called Fisher's Least Significant Difference (or Fisher's **LSD**) Method

Design-Expert Output

Treatment Means (Adjusted, If Necessary)

	Estimated Mean	Standard Error
1–160	551.20	8.17
2-180	587.40	8.17
3-200	625.40	8.17
4-220	707.00	8.17

Treatment	Mean Difference	DF	Standard Error	t for H_0 Coeff=0	Prob > |t|
1 vs 2	–36.20	1	11.55	–3.13	0.0064
1 vs 3	–74.20	1	11.55	–6.42	<0.0001
1 vs 4	–155.80	1	11.55	–13.49	<0.0001
2 vs 3	–38.00	1	11.55	–3.29	0.0046
2 vs 4	–119.60	1	11.55	–10.35	<0.0001
3 vs 4	–81.60	1	11.55	–7.06	<0.0001

Values of "Prob > |t|" less than 0.0500 indicate the difference in the treatment means is significant.
Values of "Prob > |t|" greater than 0.1000 indicate the difference in the two treatment means is not significant.

Graphical Comparison of Means
Text, pg. 89

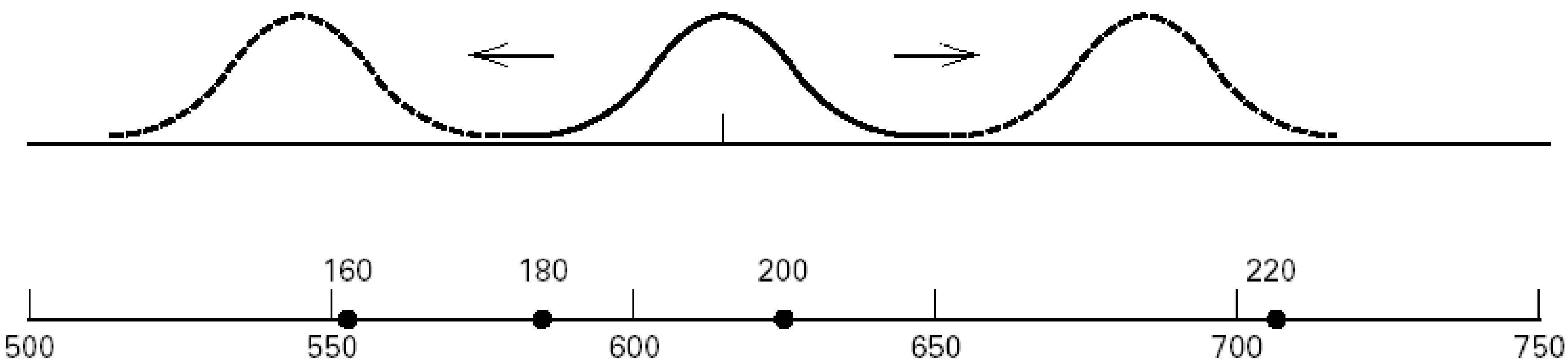

Figure 3-11 Etch rate averages from Example 3-1 in relation to a t distribution with scale factor $\sqrt{MS_E/n} = \sqrt{330.70/5} = 8.13$.

The Regression Model

$$\hat{y} = 137.62 + 2.527x \qquad \hat{y} = 1147.77 - 8.2555\,x + 0.028375\,x^2$$

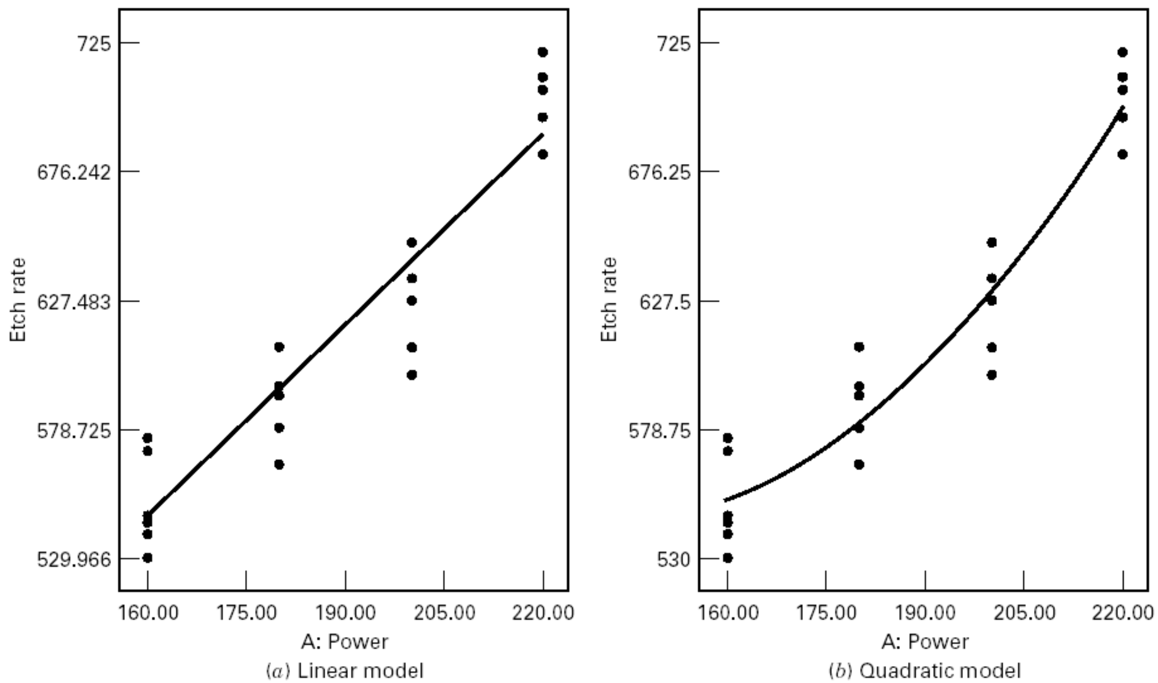

Figure 3-10 Scatter diagrams and regression models for the etch rate data of Example 3-1.

Why Does the ANOVA Work?

We are sampling from normal populations, so

$$\frac{SS_{Treamtents}}{\sigma^2} \sim \chi^2_{a-1} \text{ if } H_0 \text{ is true, and } \frac{SS_E}{\sigma^2} \sim \chi^2_{a(n-1)}$$

Cochran's theorem gives the independence of these two chi-square random variables

$$\text{So } F_0 = \frac{SS_{Treatments}/(a-1)}{SS_E/[a(n-1)]} \sim \frac{\chi^2_{a-1}/(a-1)}{\chi^2_{a(n-1)}/[a(n-1)]} \sim F_{a-1,a(n-1)}$$

$$\text{Finally, } E(MS_{Treatments}) = \sigma^2 + \frac{n\sum_{i=1}^{n}\tau_i^2}{a-1} \text{ and } E(MS_E) = \sigma^2$$

Therefore an upper-tail F test is appropriate.

Sample Size Determination
Text, Section 3-7, pg. 101

- **FAQ** in designed experiments
- Answer depends on lots of things; including what type of experiment is being contemplated, how it will be conducted, resources, and desired **sensitivity**
- Sensitivity refers to the **difference in means** that the experimenter wishes to detect
- Generally, **increasing** the number of **replications** **increases** the **sensitivity** or it makes it easier to detect small differences in means

Sample Size Determination
Fixed Effects Case

- Can choose the sample size to detect a specific difference in means and achieve desired values of **type I and type II errors**
- Type I error – reject H_0 when it is true (α)
- Type II error – fail to reject H_0 when it is false (β)
- **Power** = 1 - β
- **Operating characteristic curves** plot β against a parameter Φ where

$$\Phi^2 = \frac{n\sum_{i=1}^{a}\tau_i^2}{a\sigma^2}$$

Sample Size Determination
Fixed Effects Case---use of OC Curves

- The **OC curves** for the fixed effects model are in the Appendix, Table V, pg. 613
- A very common way to use these charts is to define a difference in two means *D* of interest, then the minimum value of Φ^2 is

$$\Phi^2 = \frac{nD^2}{2a\sigma^2}$$

- Typically work in term of the ratio of D/σ and try values of *n* until the **desired power** is achieved
- Minitab will perform power and sample size calculations – see page 103
- There are some other methods discussed in the text

Power and sample size calculations from Minitab (Page 103)

```
Power and Sample Size

One-way ANOVA

Alpha = 0.01 Assumed standard deviation = 25
Number of Levels = 4

              Sample                      Maximum
SS Means        Size       Power       Difference
  2812.5           5    0.804838               75

The sample size is for each level.

Power and Sample Size

One-way ANOVA

Alpha = 0.01 Assumed standard deviation = 25
Number of Levels = 4

           Sample  Target                          Maximum
SS Means     Size   Power  Actual Power         Difference
  2812.5        6     0.9      0.915384                 75

The sample size is for each level.
```