Checking Assumptions

Normal distributions: use probability plot (or quantile-quantile plot); straight line implies normal distribution:

![Normal Q–Q Plot](image)

Sample Quantiles

Theoretical Quantiles

16.4 16.6 16.8 17.0 17.2 17.4

−1.5 −0.5 0.5 1.5

Normal Q–Q Plot

Sample Quantiles
In R, use `qqnorm` to make qq-plots:

```r
qqnorm(cement[, "Unmodified"], datax = TRUE);
qqnorm(cement[, "Modified"], datax = TRUE);
```

Note: by default, `qqnorm` plots the empirical quantiles on the $y$-axis and the theoretical quantiles on the $x$-axis; the `datax` option reverses this, to match the “normal probability plot” as used in the book.

Overlaying qq-plots is a little more work:

```r
qqnorm(cement[, "Unmodified"], pch = 22, ylim = c(16.3, 17.4),
       datax = TRUE);
par(new = TRUE);
qqnorm(cement[, "Modified"], pch = 21, ylim = c(16.3, 17.4),
       datax = TRUE);
```
Sample Size

Recall the hypotheses:

Null hypothesis \( H_0 : \mu_1 = \mu_2 \)
Alternate hypothesis \( H_1 : \mu_1 \neq \mu_2 \).

Two types of error:
- **Type I error**: reject \( H_0 \) when it is true;
- **Type II error**: fail to reject \( H_0 \) when it is false.

\[ P(\text{Type I error}) = \alpha, \quad P(\text{Type II error}) = \beta. \]
We usually choose $\alpha$ at say .05 (or .01, or .10, or ...).

For a given sample size $n$, $\beta$ depends on the difference between the true means, $\delta = |\mu_1 - \mu_2|$.

The *operating characteristic curve*, or O.C. curve, is a graph of $\beta$ against $\delta$.

A graph usually shows the O.C. curves for various sample sizes $n$, which gives guidance about sample size.
Checking Assumptions

Equal variances:
- informally, compare standard deviations;
- also compare slopes in q-q plots;
- formally, use $F$-test for ratio of variances.
If Variances are Unequal

Estimated standard error of $\bar{y}_1 - \bar{y}_2$ is now

$$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},$$

so the test statistic is

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}};$$
Under $H_0 : \mu_1 = \mu_2$, $t_0$ is *approximately* $t$-distributed (Welch’s approximation) with degrees of freedom

$$
\nu = \frac{\left( \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{S_1^2}{n_1} \frac{S_2^2}{n_2} \left( \frac{1}{n_1 - 1} + \frac{1}{n_2 - 1} \right)}.
$$
R command for equal variances

t.test(cement$Modified, cement$Unmodified, var.equal = TRUE)

Output

Two Sample t-test

data:  cement$Modified and cement$Unmodified
t = -2.1869, df = 18, p-value = 0.0422
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  -0.54507339  -0.01092661
sample estimates:
mean of x  mean of y
  16.764       17.042
R command for unequal variances

```r
t.test(cement$Modified, cement$Unmodified, var.equal = FALSE)
# var.equal = FALSE is the default, so it could be omitted
```

Output

```
Welch Two Sample t-test

data:  cement$Modified and cement$Unmodified
t = -2.1869, df = 17.025, p-value = 0.043
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  -0.54617414  -0.00982586
sample estimates:
mean of x  mean of y
   16.764    17.042
```

Note: fewer (non-integer) degrees of freedom; slightly higher \( p \)-value; slightly wider confidence interval.
Paired Comparisons

In some situations, we can manage at least part of the variability in the measurements.

Example: measuring hardness, comparing two instruments (tips).

Making 10 measurements with each tip, we could:

- Choose 20 specimens, divide them into two groups of 10, and test each group with one tip;
- Choose 10 specimens, and test each with both tips: *paired* measurements.
Results for a paired experiment

<table>
<thead>
<tr>
<th>Specimen</th>
<th>Tip 1</th>
<th>Tip 2</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>5</td>
<td>-2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>4</td>
<td>-2</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>5</td>
<td>-1</td>
</tr>
</tbody>
</table>

Mean: 4.8, 4.9, -0.1
S.D.: 2.39, 2.23, 1.20
Differences have less variability than individual measurements, because specimen-to-specimen variations cancel out.

**Statistical model**

\[ y_{i,j} = \mu_i + \beta_j + \epsilon_{i,j}, \quad i = 1, 2, \quad j = 1, 2, \ldots, n. \]

Here:
- \( y_{i,j} \) = measurement for tip \( i \) in pair \( j \);
- \( \mu_i \) = mean strength for tip \( i \), averaged across conditions;
- \( \beta_j \) = deviation from overall mean strength for pair \( j \);
- \( \epsilon_{i,j} \sim N(0, \sigma_i^2) \).
Analysis

- form within-pair differences

\[ d_j = y_{1,j} - y_{2,j} = \mu_1 - \mu_2 + \epsilon_{1,j} - \epsilon_{2,j}; \]

- pair deviation \( \beta_j \) cancels out, so \( d_j \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2 = \sigma_d^2) \);

- standard error of \( \bar{d} \) is \( \sqrt{\sigma_d^2/n} \);

- Sample variance \( S_d^2 \) estimates \( \sigma_d^2 \).

The test statistic:

\[ t_0 = \frac{\bar{d}}{S_d / \sqrt{n}}; \]

Under \( H_0 : \mu_1 = \mu_2 \), \( t_0 \) is \( t \)-distributed with \( n - 1 \) degrees of freedom.
**R command for paired data**

```r
tips <- read.table("data/tips.txt")
t.test(tips$Tip1, tips$Tip2, paired = TRUE)
```

**Output**

Paired t-test

data:  tips$Tip1 and tips$Tip2
t = -0.2641, df = 9, p-value = 0.7976
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.9564389  0.7564389
sample estimates:
mean of the differences
-0.1
Comparison with Unpaired Design

If each run was made on randomly selected materials with no pairing:

\[ y_{i,j} = \mu_i + \beta_{i,j} + \epsilon_{i,j}, \quad i = 1, 2, \quad j = 1, 2, \ldots, n : \]

- now \( \text{Var}(y_{i,j}) = \sigma_{\beta}^2 + \sigma_i^2; \)
- denominator of \( t_0 \) is larger, making test less sensitive;
- degrees of freedom are \( 2(n - 1) \), making test more sensitive;
- on balance, test is usually more sensitive, so pairing is good.

Pairing helps when within-pair variation is much less than among-pair variation.