Nonparametric Methods for Two Samples

An overview

• In the independent two-sample t-test, we assume normality, independence, and equal variances.

• This t-test is robust against nonnormality, but is sensitive to dependence.

• If \( n_1 \) is close to \( n_2 \), then the test is moderately robust against unequal variance (\( \sigma_1^2 \neq \sigma_2^2 \)). But if \( n_1 \) and \( n_2 \) are quite different (e.g. differ by a ratio of 3 or more), then the test is much less robust.

• How to determine whether the equal variance assumption is appropriate?

• Under normality, we can compare \( \sigma_1^2 \) and \( \sigma_2^2 \) using \( S_1^2 \) and \( S_2^2 \), but such tests are very sensitive to nonnormality. Thus we avoid using them.

• Instead we consider a nonparametric test called Levene’s test for comparing two variances, which does not assume normality while still assuming independence.

• Later on we will also consider nonparametric tests for comparing two means.
Nonparametric Methods for Two Samples

Levene’s test

Consider two independent samples \( Y_1 \) and \( Y_2 \):

Sample 1: 4, 8, 10, 23
Sample 2: 1, 2, 4, 4, 7

Test \( H_0 : \sigma_1^2 = \sigma_2^2 \) vs \( H_A : \sigma_1^2 \neq \sigma_2^2 \).

- Note that \( s_1^2 = 67.58 \), \( s_2^2 = 5.30 \).
- The main idea of Levene’s test is to turn testing for equal variances using the original data into testing for equal means using modified data.
- Suppose normality and independence, if Levene’s test gives a small p-value \((< 0.01)\), then we use an approximate test for \( H_0 : \mu_1 = \mu_2 \) vs \( H_A : \mu_1 \neq \mu_2 \). See Section 10.3.2 of the bluebook.
Nonparametric Methods for Two Samples

Levene’s test

(1) Find the median for each sample. Here \( \tilde{y}_1 = 9, \tilde{y}_2 = 4 \).

(2) Subtract the median from each obs.

Sample 1: -5, -1, 1, 14
Sample 2: -3, -2, 0, 0, 3

(3) Take absolute values of the results.

Sample 1*: 5, 1, 1, 14
Sample 2*: 3, 2, 0, 0, 3

(4) For any sample that has an odd sample size, remove 1 zero.

Sample 1*: 5, 1, 1, 14
Sample 2*: 3, 2, 0, 3

(5) Perform an independent two-sample t-test on the modified samples, denoted as \( Y_1^* \) and \( Y_2^* \). Here \( \bar{y}_1^* = 5.25, \bar{y}_2^* = 2, s_{1^*}^2 = 37.58, s_{2^*}^2 = 2.00 \). Thus \( s_p^2 = 19.79, s_p = 4.45 \) on df = 6 and the observed

\[
t = \frac{5.25 - 2}{4.45 \sqrt{1/4 + 1/4}} = 1.03
\]

on df = 6. The p-value \( 2 \times P(T_6 \geq 1.03) \) is more than 0.20. Do not reject \( H_0 \) at the 5% level.
Nonparametric Methods for Two Samples

Mann-Whitney test

- We consider a nonparametric Mann-Whitney test (aka Wilcoxon test) for independent two samples, although analogous tests are possible for paired two samples.
- We relax the distribution assumption, but continue to assume independence.
- The main idea is to base the test on the ranks of obs.
- Consider two independent samples $Y_1$ and $Y_2$:

  Sample 1: 11, 22, 14, 21
  Sample 2: 20, 9, 12, 10

  Test $H_0: \mu_1 = \mu_2$ vs $H_A: \mu_1 \neq \mu_2$. 
Nonparametric Methods for Two Samples

Mann-Whitney test

(1) Rank the obs

<table>
<thead>
<tr>
<th>rank</th>
<th>obs</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>22</td>
<td>1</td>
</tr>
</tbody>
</table>

(2) Compute the sum of ranks for each sample. Here $RS(1) = 3 + 5 + 7 + 8 = 23$ and $RS(2) = 1 + 2 + 4 + 6 = 13$.

(3) Under $H_0$, the means are equal and thus the rank sums should be about equal. To compute a p-value, we list all possible ordering of 8 obs and find the rank sum of each possibility. Then p-value is $2 \times P(RS(2) \leq 13)$. Here

$$P(RS(2) \leq 13) = P(RS(2) = 10) + P(RS(2) = 11) + P(RS(2) = 12) + P(RS(2) = 13) = \frac{7}{70} = 0.1$$

and thus p-value = 0.2.
Nonparametric Methods for Two Samples

Mann-Whitney test

- If we had observed 10, then p-value = $2 \times \frac{1}{70} = 0.0286$.
- If we had observed 11, then p-value = $2 \times \frac{2}{70} = 0.0571$.
- Thus for this sample size, we can only reject at 5% if the observed rank sum is 10.
- Table A10 gives the cut-off values for different sample sizes. For $n_1 = n_2 = 4$ and $\alpha = 0.05$, we can only reject $H_0$ if the observed rank sum is 10.
Nonparametric Methods for Two Samples

Mann-Whitney test

Recorded below are the longevity of two breeds of dogs.

<table>
<thead>
<tr>
<th>Breed</th>
<th>A</th>
<th>Breed</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>obs</td>
<td>rank</td>
<td>obs</td>
<td>rank</td>
</tr>
<tr>
<td>12.4</td>
<td>9</td>
<td>11.6</td>
<td>7</td>
</tr>
<tr>
<td>15.9</td>
<td>14</td>
<td>9.7</td>
<td>4</td>
</tr>
<tr>
<td>11.7</td>
<td>8</td>
<td>8.8</td>
<td>3</td>
</tr>
<tr>
<td>14.3</td>
<td>11.5</td>
<td>14.3</td>
<td>11.5</td>
</tr>
<tr>
<td>10.6</td>
<td>6</td>
<td>9.8</td>
<td>5</td>
</tr>
<tr>
<td>8.1</td>
<td>2</td>
<td>7.7</td>
<td>1</td>
</tr>
<tr>
<td>13.2</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.6</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.3</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.1</td>
<td>13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ n_2 = 10 \]
\[ n_1 = 6 \]
\[ T^* = 31.5 \]
Nonparametric Methods for Two Samples

Mann-Whitney test

- Here $n_1$ is the sample size in the smaller group and $n_2$ is the sample size in the larger group.

- $T^*$ is the sum of ranks in the smaller group. Let $T^{**} = n_1(n_1 + n_2 + 1) - T^* = 6 \times 17 - 31.5 = 70.5$.

- Let $T = \min(T^*, T^{**}) = 31.5$ and look up Table A10.

- Since the observed $T$ is between 27 and 32, the p-value is between 0.01 and 0.05. Reject $H_0$ at 5%.

Remarks

- If there are ties, Table A10 gives approximation only.

- The test does not work well if the variances are very different.

- It is not easy to extend the idea to more complex types of data. There is no CI.

- For paired two samples, consider using signed rank test.

- See p.251 of the bluebook for a decision tree.
Nonparametric Methods for Two Samples

Key R commands

> # Levene’s test
> levene.test = function(data1, data2){
+ levene.trans = function(data){
+ a = sort(abs(data-median(data)));
+ if (length(a)%%2)
+ a[a!=0|duplicated(a)]
+ else a
+ }
+ t.test(levene.trans(data1), levene.trans(data2), var.equal=T)
+ }
> y1 = c(4,8,10,23)
> y2 = c(1,2,4,4,7)
> levene.test(y1, y2)

Two Sample t-test

data: levene.trans(data1) and levene.trans(data2)
t = 1.0331, df = 6, p-value = 0.3414
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-4.447408 10.947408
sample estimates:
mean of x mean of y
5.25 2.00

> # Mann-Whitney test example
> samp1 = c(11, 22, 14, 21)
> samp2 = c(20, 9, 12, 10)
> # W = 23-10 = 13
> wilcox.test(samp1, samp2)

Wilcoxon rank sum test

data: samp1 and samp2
W = 13, p-value = 0.2
alternative hypothesis: true mu is not equal to 0
```r
> breedA = c(12.4, 15.9, 11.7, 14.3, 10.6, 8.1, 13.2, 16.6, 19.3, 15.1)
> breedB = c(11.6, 9.7, 8.8, 14.3, 9.8, 7.7)
> # W = 70.5 - 21 = 49.5
> wilcox.test(breedA, breedB)

Wilcoxon rank sum test with continuity correction

data:  breedA and breedB
W = 49.5, p-value = 0.03917
alternative hypothesis: true mu is not equal to 0

Warning message:
  Cannot compute exact p-value with ties in: wilcox.test.default(breedA, breedB)
> 
```
Comparing Two Proportions

Test procedure

Consider two binomial distributions $Y_1 \sim B(n_1, p_1), Y_2 \sim B(n_2, p_2)$, and $Y_1, Y_2$ are independent. We want to test

$$H_0 : p_1 = p_2 \quad \text{vs} \quad H_A : p_1 \neq p_2$$

• Use the point estimator $\hat{p}_1 - \hat{p}_2$, where $\hat{p}_1 = Y_1/n_1, \hat{p}_2 = Y_2/n_2$ are the sample proportions.

• Note that $\mu_{\hat{p}_1 - \hat{p}_2} = E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$ and $\sigma^2_{\hat{p}_1 - \hat{p}_2} = Var(\hat{p}_1 - \hat{p}_2) = p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2$.

• Under $H_0 : p_1 = p_2 = p$, $\mu_{\hat{p}_1 - \hat{p}_2} = 0$ and $\sigma^2_{\hat{p}_1 - \hat{p}_2} = p(1-p)(1/n_1 + 1/n_2)$.

• Under $H_0$, the test statistic is approximately normal,

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{p(1-p)(1/n_1 + 1/n_2)}} \approx N(0, 1)$$

• But we do not know $p$ and thus estimate it by

$$\hat{p} = \frac{Y_1 + Y_2}{n_1 + n_2}$$

• Thus the test statistic is $Z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}} \approx N(0, 1)$ under $H_0$. 

203
Comparing Two Proportions

Potato cure rate example

A plant pathologist is interested in comparing the effectiveness of two fungicide used on infested potato plants. Let $Y_1$ denote the number of plants cured using fungicide A among $n_1$ plants and let $Y_2$ denote the number of plants cured using fungicide B among $n_2$ plants. Assume that $Y_1 \sim B(n_1, p_1)$ and $Y_2 \sim B(n_2, p_2)$, where $p_1$ is the cure rate of fungicide A and $p_2$ is the cure rate of fungicide B. Suppose the obs are $n_1 = 105$, $p_1 = 71/105$ for fungicide A and $n_2 = 87$, $p_2 = 45/87$ for fungicide B. Test $H_0 : p_1 = p_2$ vs $H_A : p_1 \neq p_2$.

- Here $\hat{p}_1 = 71/105 = 0.676$, $\hat{p}_2 = 45/87 = 0.517$, and the pooled estimate of cure rate is
  \[
  \hat{p} = \frac{71 + 45}{105 + 87} = 0.604
  \]

- Thus the observed test statistic is
  \[
  z = \frac{(0.676 - 0.517) - 0}{\sqrt{0.604 \times 0.396 \times (1/105 + 1/87)}} = 2.24
  \]

- Compared to $Z$, the p-value is $2 \times P(Z \geq 2.24) = 0.025$.

- Reject $H_0$ at the 5% level. There is moderate evidence against $H_0$. 

204
Comparing Two Proportions

Remarks

- For constructing a \((1 - \alpha)\) CI for \(p_1 - p_2\), there is no \(H_0\). Since \(Var(\hat{p}_1 - \hat{p}_2) = p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2\), estimate by

\[
\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}
\]

and the CI is

\[
\hat{p}_1 - \hat{p}_2 - z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \leq p_1 - p_2 \\
\leq \hat{p}_1 - \hat{p}_2 + z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}
\]

- In the potato cure rate example, a 95% CI for \(p_1 - p_2\) is

\[
(0.676 - 0.517) \pm 1.96 \times \sqrt{\frac{0.676 \times 0.324}{105} + \frac{0.517 \times 0.483}{87}}
\]

which is \(0.159 \pm 0.138\) or \([0.021, 0.297]\).

- In constructing CI for \(p_1 - p_2\), normal approximation works well if \(n_1\hat{p}_1 \geq 5, n_1(1 - \hat{p}_1) \geq 5, n_2\hat{p}_2 \geq 5, n_2(1 - \hat{p}_2) \geq 5\).

- In testing \(H_0 : p_1 = p_2\), normal approximation works well if \(n_1\hat{p} \geq 5, n_1(1 - \hat{p}) \geq 5, n_2\hat{p} \geq 5, n_2(1 - \hat{p}) \geq 5\).
Comparing Two Proportions

Key R commands

```
> # potato cure rate example
> y1 = 71
> n1 = 105
> y2 = 45
> n2 = 87
> p1 = y1/n1
> p2 = y2/n2
> poolp = (y1+y2)/(n1+n2)
> poolp
[1] 0.6041667
> z.value = (p1-p2)/sqrt(poolp*(1-poolp)*(1/n1+1/n2))
> z.value
[1] 2.241956
> # p-value
> 2*pnorm(z.value, lower.tail=F)
[1] 0.02496419
> # 95% CI
> alpha = 0.05
> qnorm(alpha/2, lower.tail=F)*sqrt(p1*(1-p1)/n1+p2*(1-p2)/n2)
[1] 0.1379716
> c(p1-p2-qnorm(alpha/2, lower.tail=F)*sqrt(p1*(1-p1)/n1+p2*(1-p2)/n2),
+ p1-p2+qnorm(alpha/2, lower.tail=F)*sqrt(p1*(1-p1)/n1+p2*(1-p2)/n2))
[1] 0.02097751 0.29692068
> prop.test(c(71, 45), c(105, 87), correct=F)

2-sample test for equality of proportions without continuity correction

data:  c(71, 45) out of c(105, 87)
X-squared = 5.0264, df = 1, p-value = 0.02496
alternative hypothesis: two.sided
95 percent confidence interval:
 0.02097751 0.29692068
sample estimates:
  prop 1  prop 2
0.6761905 0.5172414
```
> prop.test(c(71, 45), c(105, 87))

    2-sample test for equality of proportions with continuity correction

data:  c(71, 45) out of c(105, 87)
X-squared = 4.3837, df = 1, p-value = 0.03628
alternative hypothesis: two.sided
95 percent confidence interval:
  0.01046848 0.30742971
sample estimates:
  prop 1  prop 2
0.6761905 0.5172414
One-way ANOVA

An overview

• So far we have learned statistical methods for comparing two trts.

• One-way analysis of variance (ANOVA) provides us with a way to compare more than two trts.

• One-way ANOVA can be viewed as an extension of the independent two sample case to independent multiple samples.

• The key idea is to break up the sum of squares

$$\sum (Y_i - \bar{Y})^2$$

• First reconsider the independent two-sample case and then generalize the idea to independent multiple samples.
One-way ANOVA

Independent two samples

- Consider the following independent two samples:
  
  \[ X: 4, 12, 8 \]
  \[ Y: 17, 8, 11 \]

- The summary statistics are

  \[ \bar{x} = 8, \quad s_x^2 = 16, \quad \sum_{i=1}^{3} (x_i - \bar{x})^2 = 32 \]

  \[ \bar{y} = 12, \quad s_y^2 = 21, \quad \sum_{i=1}^{3} (y_i - \bar{y})^2 = 42, \quad s_p^2 = 18.5 \]

- For testing \( H_0 : \mu_1 = \mu_2 \) vs \( H_A : \mu_1 \neq \mu_2 \), use t-test

  \[ t = \frac{(12 - 8) - 0}{\sqrt{18.5\left(\frac{1}{3} + \frac{1}{3}\right)}} = 1.14 \]

  on df = 4. The p-value \( 2 \times P(T_4 \geq 1.14) \) is greater than 0.10. Thus do not reject \( H_0 \) at 5% and there is no evidence against \( H_0 \).

- Now we will examine this using the idea of breaking up sums of squares.
One-way ANOVA

Sums of squares (SS)

- Total SS: Pretend that all obs are from a single population. The overall mean is
  \[
  \frac{4 + 12 + 8 + 17 + 8 + 11}{6} = 10
  \]
  and the SS Total is
  \[
  (4-10)^2 + (12-10)^2 + (8-10)^2 + (17-10)^2 + (8-10)^2 + (11-10)^2 = 98
  \]
  on df = 5.

- Treatment SS: How much of the total SS can be attributed to the differences between the two trt groups? Replace each obs by its group mean.

  X: 8, 8, 8
  Y: 12, 12, 12

  The overall mean here is
  \[
  \frac{8 + 8 + 8 + 12 + 12 + 12}{6} = 10
  \]
  and the SS Trt is
  \[
  (8-10)^2 + (8-10)^2 + (8-10)^2 + (12-10)^2 + (12-10)^2 + (12-10)^2 = 24
  \]
  on df = 1.
One-way ANOVA

Sums of squares (SS)

- Error SS: How much of the total SS can be attributed to the differences within each trt group? The SS Error is

\[(4 - 8)^2 + (12 - 8)^2 + (8 - 8)^2 + (17 - 12)^2 + (8 - 12)^2 + (11 - 12)^2 = 74\]

on df = 4.

- Note that SSError/df = 74/4 = 18.5 = \(s_p^2\).

- Note also that

\[
\text{SS Total} = \text{SS Trt} + \text{SS Error} \quad (98 = 24 + 74)
\]

\[
\text{df Total} = \text{df Trt} + \text{df Error} \quad (5 = 1 + 4)
\]

- An ANOVA table summarizes the information.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trt</td>
<td>1</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Error</td>
<td>4</td>
<td>74</td>
<td>18.5</td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>98</td>
<td></td>
</tr>
</tbody>
</table>

- Here MS = SS/df.
One-way ANOVA

F-test

- $H_0 : \mu_1 = \mu_2 \text{ vs } H_A : \mu_1 \neq \mu_2$
- A useful fact is that, under $H_0$, the test statistic is:
  \[ F = \frac{\text{MSTtrt}}{\text{MSError}} \sim F_{\text{dfTrt}, \text{dfError}} \]
- In the example, the observed $f = 24/18.5 = 1.30$.
- Compare this to an F-distribution with 1 df in the numerator and 4 df in the denominator using Table D. The (one-sided) p-value $P(F_{1,4} \geq 1.30)$ is greater than 0.10. Do not reject $H_0$ at the 10% level. There is no evidence against $H_0$.
- Note that a small difference between the two trt means relative to variability is associated with a small $f$, a large p-value, and accepting $H_0$, whereas a large difference between the two trt means relative to variability is associated with a large $f$, a small p-value, and rejecting $H_0$.
- Note that $f = 1.30 = (1.14)^2 = t^2$. That is $f = t^2$, but only when the df in the numerator is 1.
- Note that the p-value is one-tailed, even though $H_A$ is two-sided.
One-way ANOVA

A recap

In the simple example above, there are 2 trts and 3 obs/trt. The overall mean is 10,

\[
\text{SSTotal} = \sum_{i=1}^{3} (x_i - 10)^2 + \sum_{i=1}^{3} (y_i - 10)^2 = 98
\]

\[
\text{SSTrt} = 3 \times (\bar{x} - 10)^2 + 3 \times (\bar{y} - 10)^2 = 24
\]

\[
\text{SSErr} = \sum_{i=1}^{3} (x_i - 8)^2 + \sum_{i=1}^{3} (y_i - 12)^2 = 74
\]

with df = 5, 1, and 4, respectively.
One-way ANOVA

Generalization to $k$ independent samples

- Consider $k$ trts and $n_i$ obs for the $i^{th}$ trt.
- Let $y_{ij}$ denote the $j^{th}$ obs in the $i^{th}$ trt group.
- Tabulate the obs as follows.

<table>
<thead>
<tr>
<th>Trt</th>
<th>1</th>
<th>2</th>
<th>⋮</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs</td>
<td>$y_{11}$</td>
<td>$y_{21}$</td>
<td>⋮</td>
<td>$y_{k1}$</td>
</tr>
<tr>
<td></td>
<td>$y_{12}$</td>
<td>$y_{22}$</td>
<td>⋮</td>
<td>$y_{k2}$</td>
</tr>
<tr>
<td></td>
<td>⋮</td>
<td>⋮</td>
<td>⋮</td>
<td>⋮</td>
</tr>
<tr>
<td></td>
<td>$y_{1n_1}$</td>
<td>$y_{2n_2}$</td>
<td>⋮</td>
<td>$y_{kn_k}$</td>
</tr>
<tr>
<td>Sum</td>
<td>$y_1.$</td>
<td>$y_2.$</td>
<td>⋮</td>
<td>$y_k.$</td>
</tr>
<tr>
<td>Mean</td>
<td>$\bar{y}_1.$</td>
<td>$\bar{y}_2.$</td>
<td>⋮</td>
<td>$\bar{y}_k.$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trt</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>9</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>37</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>Mean</td>
<td>9.25</td>
<td>9</td>
<td>5.6</td>
</tr>
</tbody>
</table>

- Sum for the $i^{th}$ trt: $y_{i.} = \sum_{j=1}^{n_i} y_{ij}$
- Mean for the $i^{th}$ trt: $\bar{y}_{i.} = y_{i.}/n_i$
- Grand sum: $y_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij} = \sum_{i=1}^{k} y_{i.}$
- Grand mean: $\bar{y}_{..} = y_{..}/N$ where the total # of obs is:

$$N = \sum_{i=1}^{k} n_i = n_1 + n_2 + \cdots + n_k.$$
One-way ANOVA

Basic partition of SS

\[ SS \text{ Total} = SS \text{ Trt} + SS \text{ Error} \]
\[ df \text{ Total} = df \text{ Trt} + df \text{ Error} \]

where

\[
SS \text{ Total} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N}
\]
\[ df \text{ Total} = N - 1 \]
\[
SS \text{ Trt} = \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y}_{..})^2 = \sum_{i=1}^{k} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N}
\]
\[ df \text{ Trt} = k - 1 \]
\[
SS \text{ Error} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2
\]
\[ = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2 \]
\[ df \text{ Error} = N - k = (n_1 - 1) + \cdots + (n_k - 1) \]

or simply \( SS \text{ Error} = SS \text{ Total} - SS \text{ Trt} \) and \( df \text{ Error} = df \text{ Total} - df \text{ Trt} \).
One-way ANOVA

Fish length example

- Consider the length of fish (in inch) that are subject to one of three types of diet, with seven observations per diet group. The raw data are:

<table>
<thead>
<tr>
<th></th>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$Y_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.2</td>
<td>20.1</td>
<td>17.6</td>
<td>16.8</td>
</tr>
<tr>
<td>17.4</td>
<td>18.7</td>
<td>19.1</td>
<td>16.4</td>
</tr>
<tr>
<td>15.2</td>
<td>18.8</td>
<td>17.7</td>
<td>16.5</td>
</tr>
</tbody>
</table>

- A stem and leaf display of these data looks like:

<table>
<thead>
<tr>
<th>Y_1</th>
<th>Y_2</th>
<th>Y_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.9</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>16.8</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>17.6</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>16.4</td>
<td>74</td>
<td>8</td>
</tr>
<tr>
<td>15.9</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>19.7</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

- Summary statistics are:

\[
\begin{align*}
y_1. &= 130.3 \quad \bar{y}_1. = 18.61 \quad s^2_1 = 1.358 \quad n_1 = 7 \\
y_2. &= 123.6 \quad \bar{y}_2. = 17.66 \quad s^2_2 = 1.410 \quad n_2 = 7 \\
y_3. &= 117.9 \quad \bar{y}_3. = 16.84 \quad s^2_3 = 1.393 \quad n_3 = 7 \\
y_\cdot & = 371.8 \quad \bar{y}_\cdot = 17.70 \quad N = 21
\end{align*}
\]
One-way ANOVA

Fish length example

• The sums of squares are:

\[
\text{SSTotal} = \sum_{i=1}^{3} \sum_{j=1}^{7} y_{ij}^2 - \frac{(y..)^2}{N} \\
= 6618.60 - 6582.63 = 35.97
\]

\[
\text{SSTrt} = \sum_{i=1}^{3} \frac{(y_i.)^2}{n_i} - \frac{(y..)^2}{N} \\
= \frac{1}{7}[(130.3)^2 + (123.6)^2 + (117.9)^2] - 6582.63 \\
= 11.01
\]

\[
\text{SSErr} = \text{SSTot} - \text{SSTrt} = 35.97 - 11.01 = 24.96
\]

• Or \( \text{SSErr} = 6s_1^2 + 6s_2^2 + 6s_3^2 = 24.96 \)

• The corresponding ANOVA table is:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trt</td>
<td>2</td>
<td>11.01</td>
<td>5.505</td>
</tr>
<tr>
<td>Error</td>
<td>18</td>
<td>24.96</td>
<td>1.387</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>35.97</td>
<td></td>
</tr>
</tbody>
</table>

217
One-way ANOVA

Fish length example

• Note that the MS for Error computed above is the same as the pooled estimate of variance, $s_p^2$.

• The null hypothesis $H_0$: “all population means are equal” versus the alternative hypothesis $H_A$: “not all population means are equal”.

• The observed test statistic is:

$$f = \frac{\text{MSTrt}}{\text{MSErr}} = \frac{5.505}{1.387} = 3.97$$

• Compare this with $F_{2,18}$ from Table D: at 5% $f_{2,18} = 3.55$, and at 1% $f_{2,18} = 6.01$, so for our data $0.01 < p\text{-value} < 0.05$.

• Reject $H_0$ at the 5% level. There is moderate evidence against $H_0$. That is, there is moderate evidence that there is a diet effect on the fish length.
One-way ANOVA

Assumptions

1. For each trt, a random sample \( Y_{ij} \sim N(\mu_i, \sigma^2_i) \).
2. Equal variances \( \sigma^2_1 = \sigma^2_2 = \cdots = \sigma^2_k \).
3. Independent samples across trts.

That is, independence, normality, and equal variances.

A unified model

\[ Y_{ij} = \mu_i + e_{ij} \]

where \( e_{ij} \) are iid \( N(0, \sigma^2) \). Let

\[ \mu = \frac{1}{k} \sum_{i=1}^{k} \mu_i, \quad \alpha_i = \mu_i - \mu. \]

Then equivalently the model is:

\[ Y_{ij} = \mu + \alpha_i + e_{ij} \]

where \( e_{ij} \) are iid \( N(0, \sigma^2) \).
One-way ANOVA

Hypotheses

\[ H_0: \mu_1 = \mu_2 = \cdots = \mu_k \text{ vs. } H_A: \text{Not all } \mu_i \text{'s are equal.} \]

Equivalently

\[ H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0 \text{ vs. } H_A: \text{Not all } \alpha_i \text{'s are zero.} \]

F-test

Under \( H_0 \), the test statistic is

\[ F = \frac{\text{MSTrt}}{\text{MSErr}} \sim F_{\text{dfTrt,dfError}} \]

Parameter estimation

- Estimate \( \sigma^2 \) by \( S_p^2 \).
- Estimate \( \mu_i \) by \( \bar{Y}_i \).
- Or estimate \( \mu \) by \( \bar{Y}_.. \) and estimate \( \alpha_i \) by \( \bar{Y}_i - \bar{Y}_.. \).
- We will discuss inference of parameters later on.
One-way ANOVA

A brief review

<table>
<thead>
<tr>
<th>Dist’n</th>
<th>One-Sample Inference</th>
<th>Two-Sample Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>$H_0: \mu = \mu_0$</td>
<td>Paired $H_0: \mu_D = 0$, CI for $\mu_D$ ($Z$ or $T_{n-1}$)</td>
</tr>
<tr>
<td></td>
<td>CI for $\mu$</td>
<td>2 ind samples $H_0: \mu_1 = \mu_2$, CI for $\mu_1 - \mu_2$ ($T_{n_1+n_2-2}$)</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$ is known ($Z$) or unknown ($T_{n-1}$)</td>
<td>$k$ ind samples $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ ($F_{k-1,N-k}$)</td>
</tr>
<tr>
<td></td>
<td>$H_0: \sigma^2 = \sigma^2_0$, CI for $\sigma^2$ ($\chi^2_{n-1}$)</td>
<td>$H_0: \sigma^2 = \sigma^2_0$, CI for $\sigma^2$ ($\chi^2_{N-k}$)</td>
</tr>
<tr>
<td>Arbitrary</td>
<td>$H_0: \mu = \mu_0$, CI for $\mu$ (CLT $Z$)</td>
<td>Paired $H_0: \mu_D = 0$ (Signed rank)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 ind samples $H_0: \mu_1 = \mu_2$ (Mann-Whitney)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 ind samples $H_0: \sigma^2_1 = \sigma^2_2$ (Levene’s)</td>
</tr>
<tr>
<td>Binomial</td>
<td>$H_0: p = p_0$ (Binomial $Y \sim B(n,p)$)</td>
<td>2 ind samples $H_0: p_1 = p_2$, CI for $p_1 - p_2$ (CLT $Z$)</td>
</tr>
<tr>
<td></td>
<td>$H_0: p = p_0$, CI for $p$ (CLT $Z$)</td>
<td></td>
</tr>
</tbody>
</table>

- For testing or CI, address model assumptions (e.g. normality, independence, equal variance) via detection, correction, and robustness.

- In hypothesis testing, $H_0$, $H_A$ (1-sided or 2-sided), test statistic and its distribution, p-value, interpretation, rejection region, $\alpha$, $\beta$, power, sample size determination.

- For paired t-test, the assumptions are $D \sim iid N(\mu_D, \sigma^2_D)$ where $D = Y_1 - Y_2$. $Y_1, Y_2$ need not be normal. $Y_1$ and $Y_2$ need not be independent.
One-way ANOVA

More on assumptions

<table>
<thead>
<tr>
<th>Assumptions</th>
<th>Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normality</td>
<td>Stem-and-leaf plot; normal scores plot</td>
</tr>
<tr>
<td>Independence</td>
<td>Study design</td>
</tr>
<tr>
<td>Equal variance</td>
<td>Levene’s test</td>
</tr>
<tr>
<td>Correct model</td>
<td>More later</td>
</tr>
</tbody>
</table>

Detect unequal variance

- Plot trt standard deviation vs trt mean.
- Or use an extension of Levene’s test for

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2.$$  

The main idea remains the same, except that a one-way ANOVA is used instead of a two-sample t-test.
One-way ANOVA

Levene’s test

For example, consider $k = 3$ groups of data.

Sample 1: 2, 5, 7, 10
Sample 2: 4, 8, 19
Sample 3: 1, 2, 4, 4, 7

(1) Find the median for each sample. Here $\tilde{y}_1 = 6$, $\tilde{y}_2 = 8$, $\tilde{y}_3 = 4$.

(2) Subtract the median from each obs and take absolute values.

Sample 1*: 4, 1, 1, 4
Sample 2*: 4, 0, 11
Sample 3*: 3, 2, 0, 0, 3

(3) For any sample that has an odd sample size, remove 1 zero.

Sample 1*: 4, 1, 1, 4
Sample 2*: 4, 11
Sample 3*: 3, 2, 0, 3

(4) Perform a one-way ANOVA f-test on the final results.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>2</td>
<td>44.6</td>
<td>22.30</td>
<td>3.95</td>
<td>0.05 &lt; p &lt; 0.10</td>
</tr>
<tr>
<td>Error</td>
<td>7</td>
<td>39.5</td>
<td>5.64</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>84.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
One-way ANOVA

Key R commands

> # Fish length example
> y1 = c(18.2,20.1,17.6,16.8,18.8,19.7,19.1)
> y2 = c(17.4,18.7,19.1,16.4,15.9,18.4,17.7)
> y3 = c(15.2,18.8,17.7,16.5,15.9,17.1,16.7)
> y = c(y1, y2, y3)
> n1 = length(y1)
> n2 = length(y2)
> n3 = length(y3)
> trt = c(rep(1,n1),rep(2,n2),rep(3,n3))
> oneway.test(y~factor(trt), var.equal=T)

One-way analysis of means

data:  y and factor(trt)
F = 3.9683, num df = 2, denom df = 18, p-value = 0.03735

> fit.lm = lm(y~factor(trt))
> anova(fit.lm)
Analysis of Variance Table

Response: y

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>factor(trt)</td>
<td>2</td>
<td>11.0067</td>
<td>5.5033</td>
<td>3.9683</td>
</tr>
<tr>
<td>Residuals</td>
<td>18</td>
<td>24.9629</td>
<td>1.3868</td>
<td></td>
</tr>
</tbody>
</table>

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> # Alternatively use data frame
> eg = data.frame(y=y, trt=factor(trt))
> eg

    y trt
1   18.2 1
2   20.1 1
3   17.6 1
4   16.8 1
5   18.8 1
6   19.7 1

224
eg.lm = lm(y~trt, eg)
> anova(eg.lm)
Analysis of Variance Table

Response: y

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>2</td>
<td>11.0067</td>
<td>5.5033</td>
<td>3.9683</td>
<td>0.03735*</td>
</tr>
<tr>
<td>Residuals</td>
<td>18</td>
<td>24.9629</td>
<td>1.3868</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> # Kruskal-Wallis rank sum test
> kruskal.test(y~trt)

        Kruskal-Wallis rank sum test

data:  y by trt
Kruskal-Wallis chi-squared = 5.7645, df = 2, p-value = 0.05601
Comparisons among Means

An overview

• In one-way ANOVA, if we reject $H_0$, then we know that not all trt means are the same.

• But this may not be informative enough. We now consider particular comparisons of trt means.

• We will consider contrasts and all pairwise comparisons.
Comparisons among Means

Fish length example continued

Recall the example with $k = 3$ trts and $n = 7$ obs/trt. Test $H_0: \mu_1 = \mu_3$ vs $H_A: \mu_1 \neq \mu_3$.

- $\bar{y}_1 = 18.61, \bar{y}_3 = 16.84, n_1 = n_3 = 7, s_p = 1.387$ on df = 18.

- The observed test statistic is

\[ t = \frac{\bar{y}_1 - \bar{y}_3}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_3}}} = \frac{18.61 - 16.84}{\sqrt{1.387 \times \frac{2}{7}}} = 2.81 \]

on df = 18. The p-value $2 \times P(T_{18} \geq 2.81)$ is between 0.01 and 0.02.

- We may also construct a $(1 - \alpha)$ CI for $\mu_1 - \mu_3$:

\[ (\bar{y}_1 - \bar{y}_3) \pm t_{df, \alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_3}} \]

- Suppose $\alpha = 0.05$. Thus $t_{18,0.025} = 2.101$ and a 95% CI for $\mu_1 - \mu_3$ is

\[ (18.61 - 16.84) \pm 2.101 \times \sqrt{1.387 \times \frac{2}{7}} \]

which is $[0.45, 3.09]$ or $1.77 \pm 1.32$. 

227
Comparisons among Means

Fish length example continued

Now test $H_0 : \mu_1 - \frac{1}{2}(\mu_2 + \mu_3) = 0$ vs $H_A : \mu_1 - \frac{1}{2}(\mu_2 + \mu_3) \neq 0$.

- Estimate $\mu_1 - \frac{1}{2}(\mu_2 + \mu_3)$ by $\bar{Y}_1 - \frac{1}{2}(\bar{Y}_2 + \bar{Y}_3)$.
- The test statistic is
  \[
  T = \frac{\bar{Y}_1 - \frac{1}{2}(\bar{Y}_2 + \bar{Y}_3) - \mu \bar{Y}_1 - \frac{1}{2}(\bar{Y}_2 + \bar{Y}_3)}{S_{\bar{Y}_1 - \frac{1}{2}(\bar{Y}_2 + \bar{Y}_3)}}
  \]

- We will see that
  \[
  \mu \bar{Y}_1 - \frac{1}{2}(\bar{Y}_2 + \bar{Y}_3) = \mu_1 - \frac{1}{2}(\mu_2 + \mu_3)
  \]
  and
  \[
  S_{\bar{Y}_1 - \frac{1}{2}(\bar{Y}_2 + \bar{Y}_3)} = S_p \sqrt{\frac{1}{n_1} + \frac{1}{4n_2} + \frac{1}{4n_3}}
  \]
- Thus a $(1 - \alpha)$ CI for $\mu_1 - \frac{1}{2}(\mu_2 + \mu_3)$ is
  \[
  \bar{y}_1 - \frac{1}{2}(\bar{y}_2 + \bar{y}_3) \pm s_p \sqrt{\frac{1}{n_1} + \frac{1}{4n_2} + \frac{1}{4n_3}}
  \]
- But first we will generalize this situation.
Comparisons among Means

Contrast

- A *contrast* is a quantity of the form
  \[ \sum_{i=1}^{k} \lambda_i \mu_i \]
  where \( k \) is the # of trts, \( \mu_i \) is the \( i^{th} \) trt mean, and \( \lambda_i \) is the \( i^{th} \) contrast coefficient.

- For comparison, we require that \( \sum_{i=1}^{k} \lambda_i = 0 \).
- For example, we have seen two contrasts already.
- \( \mu_1 - \mu_3 \) is a contrast with \( \lambda_1 = 1, \lambda_2 = 0, \lambda_3 = -1 \):
  \[ \sum_{i=1}^{k} \lambda_i \mu_i = 1 \times \mu_1 + 0 \times \mu_2 + (-1) \times \mu_3. \]

- \( \mu_1 - \frac{1}{2}(\mu_2 + \mu_3) \) is a contrast with \( \lambda_1 = 1, \lambda_2 = -1/2, \lambda_3 = -1/2 \):
  \[ \sum_{i=1}^{k} \lambda_i \mu_i = 1 \times \mu_1 + (-1/2) \times \mu_2 + (-1/2) \times \mu_3. \]
Comparisons among Means

Contrast

• Estimate \( \sum_{i=1}^{k} \lambda_i \mu_i \) by \( X = \sum_{i=1}^{k} \lambda_i \bar{Y}_i \).

• Consider the distribution of

\[
T = \frac{X - \mu_X}{S_X}
\]

• Here \( \mu_X = \sum_{i=1}^{k} \lambda_i \mu_i \), because

\[
\mu_X = E(\sum_{i=1}^{k} \lambda_i \bar{Y}_i) = \sum_{i=1}^{k} \lambda_i E(\bar{Y}_i) = \sum_{i=1}^{k} \lambda_i \mu_i.
\]

• For \( S_X \), consider variance first.

\[
Var(\sum_{i=1}^{k} \lambda_i \bar{Y}_i) = \sum_{i=1}^{k} \lambda_i^2 Var(\bar{Y}_i) = \sum_{i=1}^{k} \lambda_i^2 \frac{\sigma^2}{n_i} = \sigma^2 \sum_{i=1}^{k} \lambda_i^2 \frac{\sigma^2}{n_i}.
\]

• Estimate \( Var(\sum_{i=1}^{k} \lambda_i \bar{Y}_i) \) by \( S_p^2 \sum_{i=1}^{k} \frac{\lambda_i^2}{n_i} \) and

\[
S_X = S_p \sqrt{\sum_{i=1}^{k} \frac{\lambda_i^2}{n_i}}
\]
Comparisons among Means

Fish length example continued

- For the first contrast, $\lambda_1 = 1$, $\lambda_2 = 0$, $\lambda_3 = -1$,

$$S_X = S_p \sqrt{\frac{1}{7} + \frac{0}{7} + \frac{1}{7}} = S_p \sqrt{\frac{2}{7}}$$

as before.

- For the second contrast, $\lambda_1 = 1$, $\lambda_2 = -1/2$, $\lambda_3 = -1/2$,

$$S_X = S_p \sqrt{\frac{1}{7} + \frac{1/4}{7} + \frac{1/4}{7}} = S_p \sqrt{\frac{3}{14}}$$

- Thus for testing $H_0 : \mu_1 - \frac{1}{2}(\mu_2 + \mu_3) = 0$, the observed test statistic is

$$t = \frac{\bar{y}_1 - \frac{1}{2}(\bar{y}_2 + \bar{y}_3)}{s_p \sqrt{\frac{3}{14}}} = \frac{18.61 - (17.66 + 16.84)/2}{\sqrt{1.387 \times \frac{3}{14}}} = 2.49$$

on df = 18. The p-value $2 \times P(T_{18} \geq 2.49)$ is between 0.02 and 0.05.
Comparisons among Means

Fish length example continued

• We may also construct a 95% CI for \( \mu_1 - \frac{1}{2}(\mu_2 + \mu_3) \):

\[
\bar{y}_1 - \frac{1}{2}(\bar{y}_2 + \bar{y}_3) \pm t_{18.0.025} s_p \sqrt{\frac{3}{14}}
\]

• A 95% CI for \( \mu_1 - \frac{1}{2}(\mu_2 + \mu_3) \) is

\[
18.61 - \frac{1}{2}(17.66 + 16.84) \pm 2.101 \times \sqrt{1.387 \times \frac{3}{14}}
\]

which is [0.21, 2.51] or 1.36 ± 1.15.
Comparisons among Means

Remarks

• If all $n_i = n$, then

$$Var(\sum_{i=1}^{k} \lambda_i \bar{Y}_i) = \frac{\sigma^2}{n} \sum_{i=1}^{k} \lambda_i^2.$$

This is called a balanced case.

• Single sample $S \bar{Y} = S \sqrt{\frac{1}{n}}$

• Two samples $S_{\bar{Y}_1 - \bar{Y}_2} = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

• Multiple samples

$$S_{\sum_{i=1}^{k} \lambda_i \bar{Y}_i} = S_p \sqrt{\sum_{i=1}^{k} \frac{\lambda_i^2}{n_i}}$$