

An Example of ANOVA using R

by EV Nordheim, MK Clayton & BS Yandell, September 20, 2004

In class we handed out "An Example of ANOVA". Below we redo the example using R. There are three groups with seven observations per group. We denote group i values by y_i :

```
> 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)
```

Now we combine them into one long vector, with a second vector, `group`, identifying group membership:

```
> y = c(y1, y2, y3)
> n = rep(7, 3)
> n
```

```
[1] 7 7 7
```

```
> group = rep(1:3, n)
> group
```

```
[1] 1 1 1 1 1 1 1 2 2 2 2 2 2 2 3 3 3 3 3 3 3
```

Descriptive Summaries

Here are summaries by group and for the combined data. First we show stem-leaf diagrams.

```
> tmp = tapply(y, group, stem)
```

The decimal point is at the |

```
16 | 8
17 | 6
18 | 28
19 | 17
20 | 1
```

The decimal point is at the |

```
15 | 9
16 | 4
17 | 47
```

```
18 | 47
19 | 1
```

The decimal point is at the |

```
15 | 29
16 | 57
17 | 17
18 | 8
```

```
> stem(y)
```

The decimal point is at the |

```
15 | 299
16 | 4578
17 | 14677
18 | 24788
19 | 117
20 | 1
```

Now we show summary statistics by group and overall. We locally define a temporary function, `tmpfn`, to make this easier.

```
> tmpfn = function(x) c(sum = sum(x), mean = mean(x), var = var(x),
+   n = length(x))
> tapply(y, group, tmpfn)
```

```
 $"1"
      sum      mean      var      n
130.300000 18.614286  1.358095  7.000000
```

```
 $"2"
      sum      mean      var      n
123.600000 17.657143  1.409524  7.000000
```

```
 $"3"
      sum      mean      var      n
117.900000 16.842857  1.392857  7.000000
```

```
> tmpfn(y)
      sum      mean      var      n
371.800000 17.704762  1.798476 21.000000
```

ANOVA Table

While we could show you how to use R to mimic the computation of SS by hand, it is more natural to go directly to the ANOVA table. See Appendix 11 for other examples of the use of R commands for ANOVA.

```
> data = data.frame(y = y, group = factor(group))
> fit = lm(y ~ group, data)
> anova(fit)
```

Analysis of Variance Table

```
Response: y
      Df  Sum Sq Mean Sq F value Pr(>F)
group   2 11.0067  5.5033  3.9683 0.03735 *
Residuals 18 24.9629  1.3868
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The `anova(fit)` object can be used for other computations on the handout and in class. For instance, the tabled F values can be found by the following. First we extract the treatment and error degrees of freedom. Then we use `qt` to get the tabled F values.

```
> df = anova(fit)[, "Df"]
> names(df) = c("trt", "err")
> df

trt err
  2  18

> alpha = c(0.05, 0.01)
> qf(alpha, df["trt"], df["err"], lower.tail = FALSE)

[1] 3.554557 6.012905
```

Confidence Interval for Variance

A confidence interval on the pooled variance can be computed as well using the `anova(fit)` object. First we get the residual sum of squares, `SSTrt`, then we divide by the appropriate chi-square tabled values.

```
> anova(fit)["Residuals", "Sum Sq"]

24.96286

> anova(fit)["Residuals", "Sum Sq"]/qchisq(c(0.025, 0.975), 18,
+     lower.tail = FALSE)

[1] 0.7918086 3.0328790
```

Comparison of Means

Chapter 12 concerns comparing means after conducting an analysis of variance overall F-test. Here is a way to conduct pairwise t-tests.

```
> pairwise.t.test(y, group)
```

```
$method
```

```
[1] "t tests with pooled SD"
```

```
$data.name
```

```
[1] "y and group"
```

```
$p.value
```

```
          1          2
2 0.29149074          NA
3 0.03444923 0.2914907
```

```
$p.adjust.method
```

```
[1] "holm"
```

```
attr(,"class")
```

```
[1] "pairwise.htest"
```