

# Chapter 7: One-factor Multi-sample Experiments

Introductory Statistics for Engineering Experimentation

Peter R. Nelson, Marie Coffin and Karen A.F. Copeland

Slides by Douglas Bates

7.1 Basic Inference

7.5 Analysis of variance

Multiple comparisons

7.6 The equal variances assumption

7.7 Sample sizes

## Factors with more than two levels

- ▶ In the previous chapter we discussed inference for comparative experiments on two groups.
- ▶ In chapter 3 we showed how to model a continuous response as a function of one or more factors that could have multiple levels. We used the *R* function `aov` to estimate the cell means or the “effects” of the levels of each factor.
- ▶ The statistical assessment of whether or not the effects are significant is usually based on an *analysis of variance*; hence the name `aov` for the model-fitting function and the name `anova` for the extractor function that produces the analysis of variance table.
- ▶ The text book emphasizes a technique called the analysis of means (ANOM). This is not a widely-used technique.
- ▶ We will focus on the analysis of variance and another approach called *multiple comparisons* for follow-up analysis.

## Overview of techniques

- ▶ In chapter 2 we used graphical methods, such as comparative dotplots and comparative density plots, to display a continuous response as it depends on levels of a factor. In section 4.4 we also discussed normal probability plots, which can be used as comparative plots for such data.
- ▶ In chapter 3 we discussed fitting models of the form

$$\mathcal{Y}_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, \dots, I; \quad j = 1, \dots, n_i$$

or

$$\mathcal{Y}_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \dots, I; \quad j = 1, \dots, n_i$$

- ▶ These are the same model; just two different ways of writing it. The first is called the *cell means* form and the second is the *effects* form.
- ▶ We check for differences in the mean response in two stages: first we check if all the means could be equal and, if we reject this hypothesis, we check for which levels of the factor produce significantly different means.

# R functions used in this chapter

# Section 7.1, Basic Inference

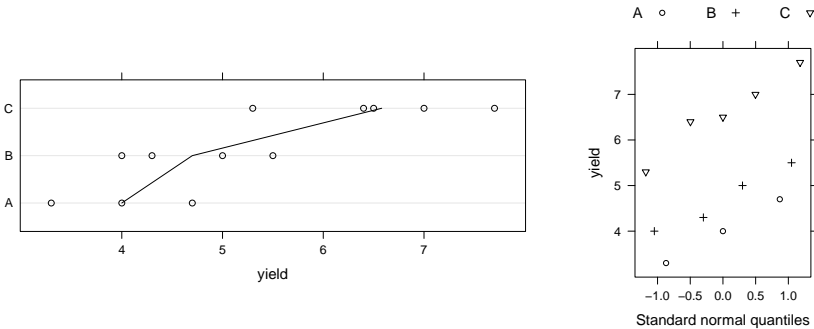
- ▶ Preliminary plots are obtained with `dotplot`, `bwplot`, `densityplot` and `qqmath`, all in the comparative form. Model fits assume that the variances in the groups are more-or-less equal. Hence we check the plots for equal variances as well as equal means.
- ▶ We use `aov` to fit the model, `summary` or `anova` to obtain the analysis of variance table, and `model.tables` to obtain estimates of the cell means or the effects.
- ▶ If we reject the hypothesis  $H_0 : \mu_1 = \mu_2 = \dots = \mu_I$  (or, equivalently,  $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_I = 0$ ) then we use `TukeyHSD` to perform multiple comparisons using Tukey's Honest Significant Difference method.
- ▶ We assess residual plots obtained with `plot(fm, which = 1)` and `plot(fm, which = 2)`

- ▶ We begin by plotting the data, preferably with comparative dotplots or comparative normal probability plots. (The text uses an error-bar chart in figure 7.1 but these are less informative than those mentioned above.)
- ▶ Group means and  $s^2$ , the mean square error, are evaluated by fitting an `aov` model and using `summary` or `anova`. The degrees of freedom for  $s^2$ ,  $n_1 + n_2 + \dots + n_I - I$ , is given in the table.
- ▶ You could use critical values from a  $T_\nu$  distribution to calculate confidence intervals on the individual means (p. 248) but the practice is discouraged.

## Examples 7.1.1 and 7.1.2

```
> reac <-  
+   data.frame(yield = c(3.3, 4.0, 4.7,  
+                       4.0, 5.0, 4.3, 5.5,  
+                       5.3, 6.5, 6.4, 7.0, 7.7),  
+             cat = factor(rep(LETTERS[1:3], c(3,4,5))))  
> summary(fm1 <- aov(yield ~ cat, reac))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
cat	2	14.7012	7.3506	12.099	0.002813
Residuals	9	5.4680	0.6076		



## Section 7.5 Analysis of Variance

- ▶ The F ratio quoted in the analysis of variance table is the ratio of  $MS_{\text{treatment}}$  to the mean square for error,  $MS_e$ . It is a “signal-to-noise” ratio based on the differences the differences between groups versus the differences within groups.
- ▶ Both the numerator and the denominator have degrees of freedom associated with them. In this case they are  $I - 1$  (numerator) and  $N - I$  (denominator) where  $N$  is the total number of observations ( $N = n_1 + n_2 + \dots + n_I$ )
- ▶ The p-value is calculated from a theoretical distribution for this quantity, written  $F_{\nu_1, \nu_2}$ . The `R` functions for this distribution are `df`, `pf`, `qf` and `rf`.
- ▶ The hypothesis being tested is “are all the means the same?” versus “are there any differences?”. Symbolically  $H_0 : \mu_1 = \mu_2 = \dots = \mu_I$ , in the cell means form, or  $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_I = 0$  in the effects form.

## Multiple comparisons

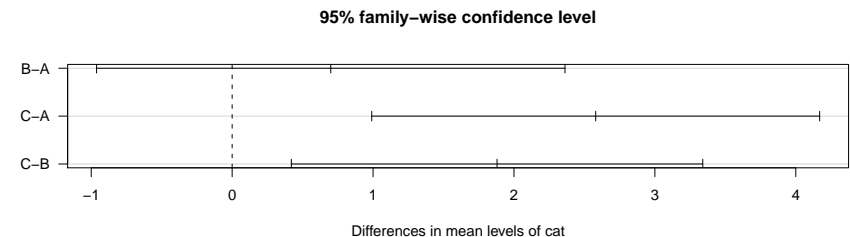
- ▶ If we reject  $H_0$  in the analysis of variance the natural follow-up question is “so which group means are significantly different”.
- ▶ It is tempting to use a series of t-tests to compare each pair of groups but doing so will inflate the probability of a false positive.
- ▶ There are several ways of controlling for this inflated false positive probability. We will use Tukey’s Honest Significant Differences, **TukeyHSD**, which is preferred when the groups are of equal importance. (Other methods are used when we have, say, a control group that we wish to compare with each of several treatments.)

## Example 7.1.1 cont’d

```
> TukeyHSD(fm1)
```

```
Tukey multiple comparisons of means  
95% family-wise confidence level  
Fit: aov(formula = yield ~ cat, data = reac)
```

```
$cat  
      diff      lwr      upr    p adj  
B-A 0.70 -0.9621403 2.362140 0.4954811  
C-A 2.58  0.9906899 4.169310 0.0036432  
C-B 1.88  0.4201254 3.339875 0.0144223
```

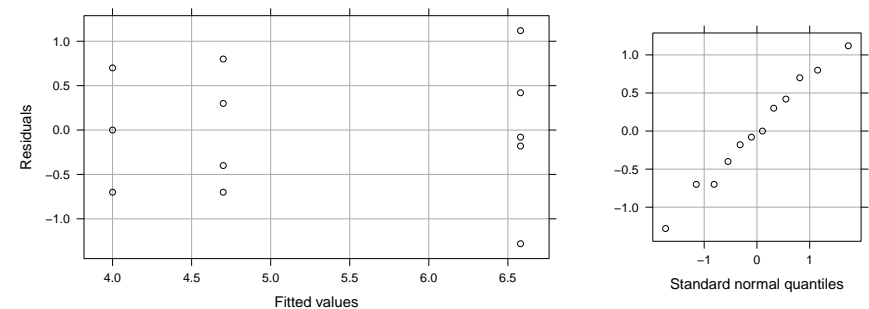


## Section 7.6: The equal variances assumption

- ▶ In the analysis of variance we pool the variance estimates from each group when evaluating the mean square for error.
- ▶ We are implicitly assuming that the variances of the “noise” terms (the  $\epsilon_{ij}$  are equal within and between groups.
- ▶ We should assess this in our data plots and through residual plots.
- ▶ In particular we should check the plot of the residuals versus the fitted values to see if the variability increases as the level of the response increases. If that is the case we consider transformations such as  $\sqrt{y_{ij}}$  or  $\log(y_{ij})$  or  $1/y_{ij}$ .
- ▶ We could do formal statistical tests for equal variances but generally those are not powerful tests. It is better to look for systematic deviations from equal variances.

## Residual plots for model `fm1`

- ▶ The residual plots for model `fm1` fit to the reactor data do not show alarming patterns. There is some slight indication of greater variability with higher mean response by not definitive at all.



# Adding lines to a probability plot

- ▶ Recall that, if the pattern in a normal probability plot is close to a straight line, the slope of the line corresponds to the standard deviation.
- ▶ In a comparative normal probability plot, equal variances corresponds to equal slopes of the lines.
- ▶ We could compare the slopes of lines fit to each group but we do not want to use least squares fits for this purpose. A least squares fit is most sensitive to the data points at the edges and these are the most suspect values.
- ▶ There is a special type of straight line fit (based on quartiles) for `qqmath`. You add such lines to the plot by redefining the `panel` argument.

```
> qqmath(~ wear, tennis, groups = type, type = c("g","p"),  
+       panel = function(...){  
+         panel.qqmath(...)  
+         panel.qqmathline(...)}  
+       )
```

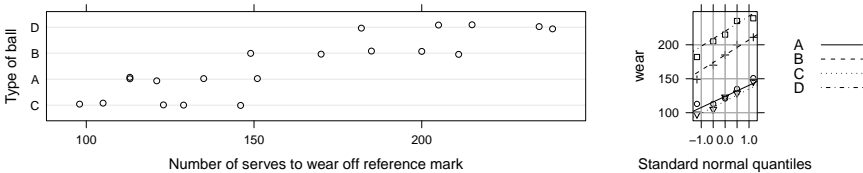
# Transformation to stabilize the variance

- ▶ When a change in variance is systematic, typically the variance increasing as the level of the response increases, and the response must be positive and the dynamic range (ratio of largest to smallest observation) is large then we consider nonlinear transformation of the response.
- ▶ The most common transformation is the logarithm. Powers, such as the square root and the inverse are also used.
- ▶ Box and Cox formulated a family of transformations, indexed by a parameter  $\lambda$  that includes powers and the logarithm. This family is written

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(y) & \lambda = 0 \end{cases}$$

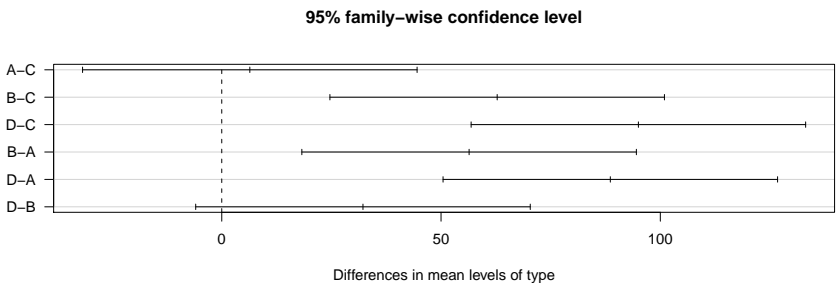
- ▶ There are ways to find the optimal value of  $\lambda$  for a particular data set and model (see `help(boxcox, package = "MASS")`) but frequently we consider a sequence of possible transformations in the order,  $\sqrt{y}$ ,  $\log(y)$  and  $1/y$ .

# Tennis ball wear



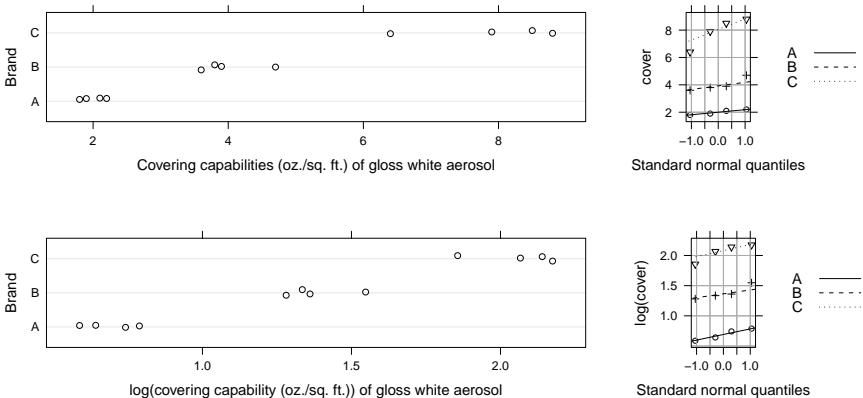
```
> summary(fm2 <- aov(wear ~ type, tennis))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
type	3	31347.0	10449.0	23.518	4.161e-06
Residuals	16	7108.8	444.3		



# Example 7.6.4

```
> aerosol <- data.frame(cover = c(2.1,1.9,1.8,2.2,  
+                               4.7,3.6,3.9,3.8, 6.4,8.5,7.9,8.8),  
+                       brand = gl(3, 4, labels = LETTERS[1:3]))
```



# Example 7.6.4 (cont'd)

```
> summary(fm3 <- aov(cover ~ brand, aerosol))

            Df Sum Sq Mean Sq F value    Pr(>F)
brand          2  72.027   36.013   76.806 2.208e-06
Residuals      9    4.220    0.469

> summary(fm3a <- aov(log(cover) ~ brand, aerosol))

            Df Sum Sq Mean Sq F value    Pr(>F)
brand          2   3.751    1.8755   133.39 2.049e-07
Residuals      9   0.1265    0.0141

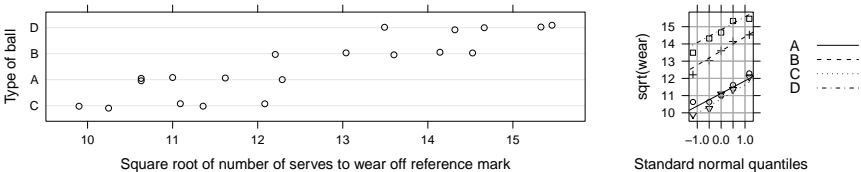
> TukeyHSD(fm3a)

    Tukey multiple comparisons of means
      95% family-wise confidence level
Fit: aov(formula = log(cover) ~ brand, data = aerosol)

$brand
      diff      lwr      upr    p adj
B-A 0.6911097 0.4570052 0.9252141 4.64e-05
C-A 1.3694858 1.1353814 1.6035903 2.00e-07
C-B 0.6783762 0.4442717 0.9124806 5.38e-05
```

# Transformation of tennis example

- ▶ Data representing counts, such as the tennis example, are often transformed by taking the square root. This is because we expect counts to have approximately a Poisson distribution and the “variance stabilizing” transformation for the Poisson is the square root.

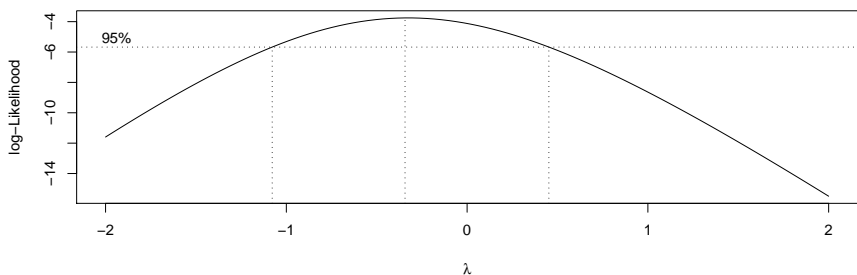


```
> summary(fm2a <- aov(sqrt(wear) ~ type, tennis))

            Df Sum Sq Mean Sq F value    Pr(>F)
type          3  48.314   16.105   23.334 4.377e-06
Residuals     16  11.043    0.690
```

# Box-Cox analysis of aerosol (not required)

```
> library(MASS)
> boxcox(fm3)
```



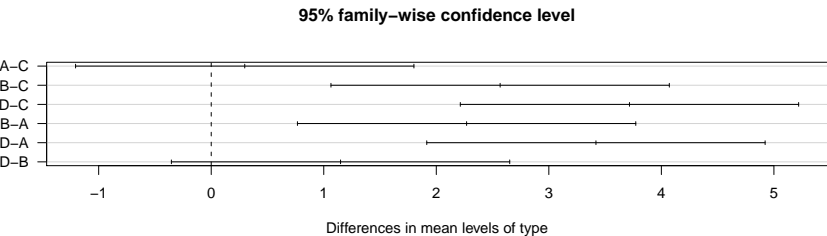
- ▶ This plot shows that the optimal value of  $\lambda$  is about -0.25, corresponding to the inverse fourth root but that  $\lambda = 0$  (the logarithm) and  $\lambda = -1$  (the inverse) are both reasonable.  $\lambda = 1$  (original scale) is not in the 95% confidence interval.

# Multiple comparisons on the transformed scale

```
> TukeyHSD(fm2a, ordered = TRUE)

    Tukey multiple comparisons of means
      95% family-wise confidence level
    factor levels have been ordered
Fit: aov(formula = sqrt(wear) ~ type, data = tennis)

$type
      diff      lwr      upr    p adj
A-C 0.2979205 -1.2053478 1.801189 0.9404445
B-C 2.5673121  1.0640438 4.070580 0.0008549
D-C 3.7165853  2.2133170 5.219854 0.0000143
B-A 2.2693916  0.7661233 3.772660 0.0026768
D-A 3.4186648  1.9153965 4.921933 0.0000391
D-B 1.1492732 -0.3539951 2.652541 0.1689889
```



## Section 7.7: sample sizes

- ▶ Sample sizes or power calculations for analysis of variance are obtained with the `power.anova.test` function. It is based on working values of the “between” and “within” variances (only their ratio is important).
- ▶ The methods described in the text are based on the maximum difference in the means, standardized by the underlying  $\sigma$ . This is the TukeyHSD comparison statistic. Some tables are given in Appendix B9 but they are rather sparse.