Two-way ANOVA Models

These notes borrow from Neter, Wasserman and Kutner, Chapters 18 and 20, and Fox, Chapter 8.

1 Multifactor Studies

The data for an ANOVA model involving a single factor is sometimes called a *"one-way layout"*, since a table of the data would usually be laid out one way along the factor as a row or column.

When there are two factors, the data is usually laid out as a two dimensional table with the levels of the first factor in rows (say) and levels of the second factor in columns; this is called a *"two-way layout"*, and the corresponding model is called a *two-way ANOVA model.*

In general there may be many factors that a researcher can control (experimental factors, a.k.a. **study factors**) or must take account of (classification factors, a.k.a. *blocking factors*) in a study.

1.1 Some examples

Example 1.

A steel company studied the effects of carbon content and tempering on the strength of steel. Carbon content was investigated at a **high** level and a **low** level. Tempering was also studied at a **high** temperature and a **low** temperature (the precise definitions of carbon and temperature levels are not important here). Thus there were $2 \times 2 = 4$ possible treatments:

Treatment	*Description*
1	High carbon, high temperature
2	High carbon, low temperature
3	Low carbon, high temperature
4	Low carbon, low temperature

The four treatments were then randomly assigned to twelve production batches and the strength of steel in each batch was measured. □

This is an experiment, indeed it is a *completely randomized design.* If the random assignment was done so that exactly three batches were assigned to each treatment, then it is also a *balanced design* (when the sample sizes in the treatments are unequal, the design is called *unbalanced*).

Laid out in a two-way table, the data might look like this:

	Temperature	
Carbon	High	Low
High	60.35	−16.82
	−45.63	−24.02
	76.88	−32.09
Low	−17.00	11.58
	33.48	−96.78
	51.44	63.11

If you think of the design as four individual treatments (as they were listed, 1, 2, 3, 4, above) you could analyze the data using one-way ANOVA. However, when there are two (or more) factors we usually want to analyze the data in special ways that allow us to assess the effects of the two factors, individually and jointly, on the outcome. The individual effects of the factors are called *main effects* and their joint effects are called *interactions*. The separate study of main effects and interactions is why the two-way layout (the second table) and its associated analysis, two-way ANOVA, is important.

Example 2.

An analyst studied the effects of family income (under \$10,000, \$10,000–\$29,999, \$30,000–\$49,999, \$50,000 and more) and stage in the life cycle of the family (1, 2 or 3) on appliance purchases. Thus $4 \times 3 = 12$ treatments are defined:

Treatment	*Description*
1	Under \$10,000, Family stage 1
2	Under \$10,000, Family stage 2
3	Under \$10,000, Family stage 3
4	\$10,000–\$29,999, Family stage 1
⋮	⋮
12	\$50,000 or more, Family stage 3

The analyst selected 20 families whose income and family stage matched each "treatment" above, yielding 240 famlies for the study. □

This is an observational study, because the data were obtained without the analyst assigning "treatments" to each family. It is a balanced design (equal sample sizes for each treatment), and there are two factors, so the data could be displayed in a two-way layout as in the previous example.

Example 3.

A psychologist interested in human-computer interaction wants to know what makes audio menus (like the recorded messages "Press 1 for blah-blah, press 2 for blah-blah" in bank-by-phone and voice mail systems) easy or difficult for people to use. Before beginning the study, he measures subjects' reading span (a measure of the capacity of their short-term memory) and categorizes them as having **high**, **medium**, or **low** reading spans. He then gives each subject a series of tasks to perform with a bank-by-phone system, so that each subject sees each combination of the following task features:

- A **broad menu** system that requires two selections from among nine options each to reach the goal in the menu; versus a **deep menu** system that requires four selections from among three options each to reach the goal;
- A task that is expressible in terms of **numerical** information, like "deposit \$1000", versus one that is expressible in **verbal** terms, such as "change mailing address".
- A task that is **first** in a sequence of two tasks, and a task that is **second** in a sequence of two tasks.

The number of “treatments” in this study is $3 \times 2 \times 2 \times 2 = 24$. Within each reading span group, the psychologist selects four people for each of the eight combinations of task features, for a total of 96 subjects. He then assigns 10 tasks to each person and records the proportion of tasks correctly performed. □

This design is a four-factor, or four-way, design. Sometimes it would be referred to as a 3×2^3 design. It is also balanced—the same number of subjects are assigned to each “treatment.”

1.2 Study Factors vs. Blocking Factors

Study factors are factors that you “designed into the study” because you want to know about them. Estimating their coefficients, or deciding whether they are “significant” in the model, leads to arguments about their relationship (correlational or causal) with the outcome variable.

Blocking factors are factors that you are stuck with, because of the circumstances of data collection. If the blocking factors are not orthogonal to the experimental factors, you should include them even if they are not significant, since otherwise estimates of the experimental effects will be biased.

What is an study factor and what is a blocking factor depends somewhat on what you are interested in, and somewhat on what you can control.

- In the first two examples, both factors (Carbon content and Tempering temperature in **Example 1**; Income and Family Stage in **Example 2**) seem to be “study factors”: we seem to be interested in the effects of both factors on the outcome.
- In **Example 3** there are three “study factors”—the types of menu and task, that the experimenter explicity assigns to subjects—and one “blocking factor”—the reading span of the subjects, which is not determined by the experimenter.

1.3 Complete and Fractional Factorial Designs

The three examples above are *complete factorial designs*: this means, all possible combinations of the factors were studied.

In some studies it may be infeasable or undesirable to study all possible combinations of factors. In **Example 1**, if the steel company wished to study six temperatures, five carbon levels and four methods of cooling the steel, the complete factorial design would have $6 \times 5 \times 4 = 120$ different treatments. It may be too expensive or too time consuming to study three batches of steel under each of these treatments.

A possible alternative is a *fractional factorial design*, in which only a fraction of the 120 possible treatments are studied. If carefully designed, a fractional factorial study can still give us good information about the main effects of the three factors, as well as some of their interactions.

2 Two-Way Layouts

2.1 Example: Balanced two-way layout

A consumer organization studied the effects of age and gender of the seller (car owner) on the the size of a cash offer for a used car. A single, medium-priced, six year old car was offered for sale to 36 car dealers selected at random from all dealers in the region. The 36 dealers were randomly "assigned" in groups of six, to each "owner." The offers, in 100's of dollars, were as follows:

```
   "Owners'"         Dealers'
Age     Gender   Offers (x $100)
------- ------- -----------------
YOUNG   MALE    21 23 19 22 22 23
YOUNG   FEMALE  21 22 20 21 19 25
MIDDLE  MALE    30 29 26 28 27 27
MIDDLE  FEMALE  26 29 27 28 27 29
ELDERLY MALE    25 22 23 21 22 21
ELDERLY FEMALE  23 19 20 21 20 20
```

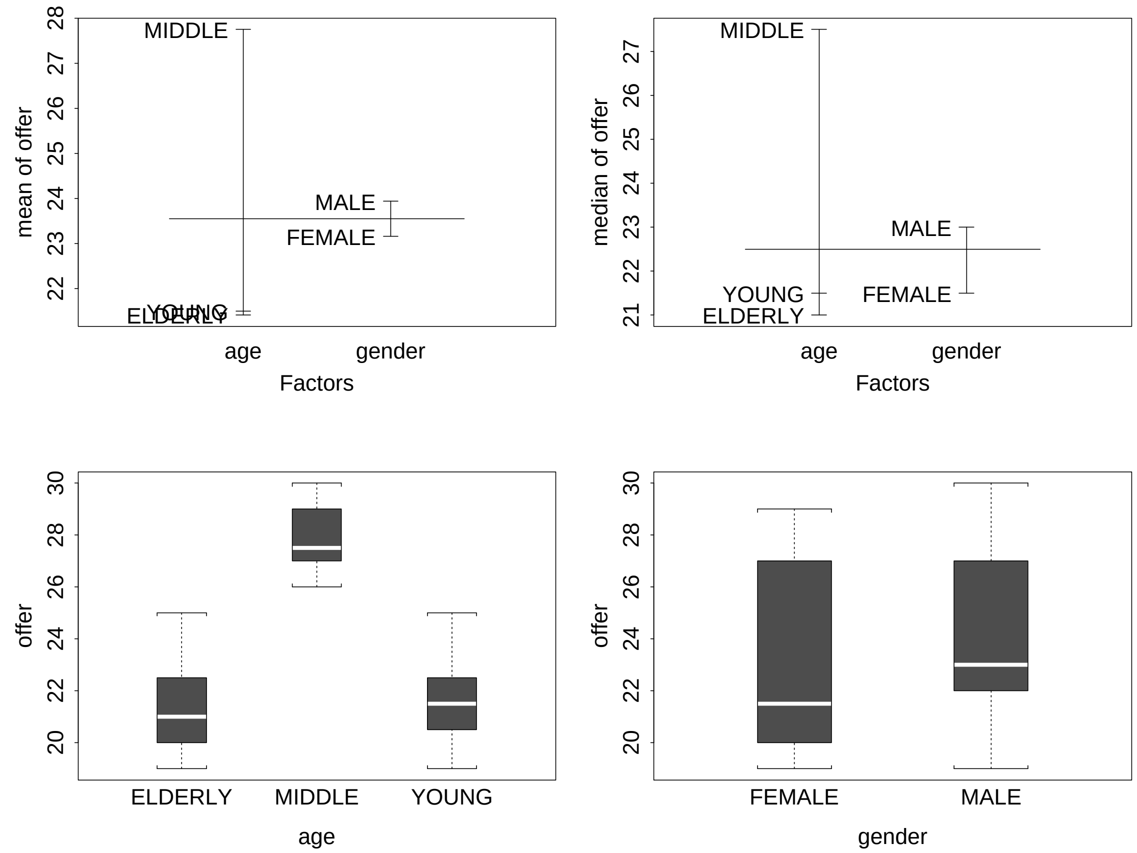
To move the data into Splus and format it in a way that is useful with `aov()` we do the following (there is more than one way to skin this cat, but...):

```
> raw.offers _ read.table("car-offers.txt")
> raw.offers
       V1     V2 V3 V4 V5 V6 V7 V8
1   YOUNG   MALE 21 23 19 22 22 23
2   YOUNG FEMALE 21 22 20 21 19 25
3  MIDDLE   MALE 30 29 26 28 27 27
4  MIDDLE FEMALE 26 29 27 28 27 29
5 ELDERLY   MALE 25 22 23 21 22 21
6 ELDERLY FEMALE 23 19 20 21 20 20
> offer.vector _ c(as.matrix(raw.offers[,3:8]))
> age _ factor(raw.offers[,1])
> gender _ factor(raw.offers[,2])
> car.offers _ data.frame(age=rep(age,6),gender=rep(gender,6),
+ offer=offer.vector)
> rm(age,gender,offer.vector)
> car.offers
       age gender offer
 1   YOUNG   MALE    21
 2   YOUNG FEMALE    21
 3  MIDDLE   MALE    30
 4  MIDDLE FEMALE    26
 5 ELDERLY   MALE    25
 6 ELDERLY FEMALE    23
 7   YOUNG   MALE    23
 8   YOUNG FEMALE    22
```

```
 9  MIDDLE   MALE    29
10  MIDDLE FEMALE    29
11 ELDERLY   MALE    22
12 ELDERLY FEMALE    19
.
.
.
34  MIDDLE FEMALE    29
35 ELDERLY   MALE    21
36 ELDERLY FEMALE    20
```

We can get a general look at the data with `plot.design` and `plot.factor`:

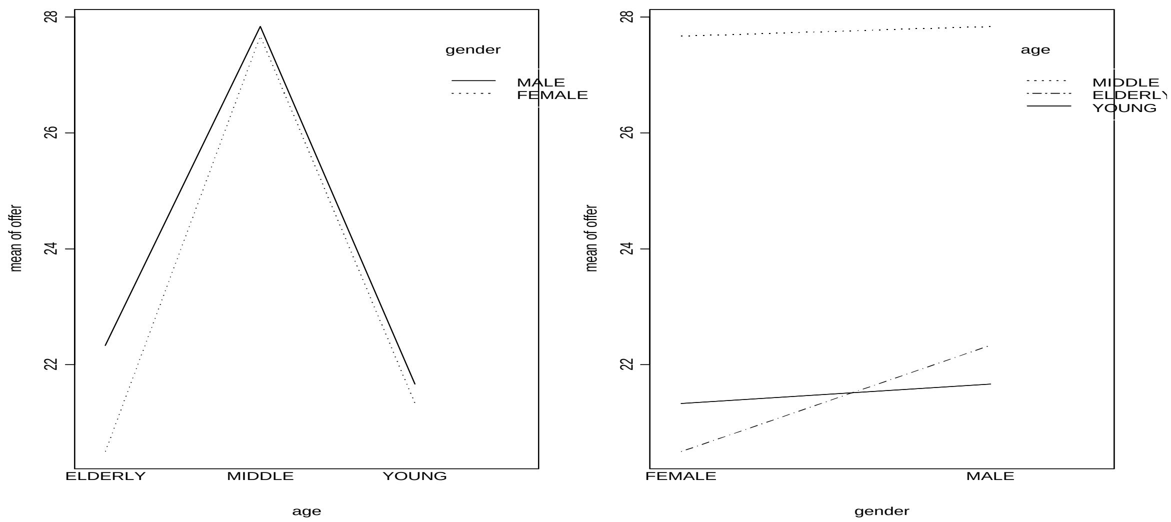
```
> par(mfrow=c(2,2))
> plot.design(car.offers,fun=mean)
> plot.design(car.offers,fun=median)
> plot.factor(car.offers)
```



Some things are already clear from these plots: middle-aged sellers were offered a better price than the young or the old, and although the mean and median offers to men were better than for women, this difference is likely to be swamped by variability in the data. We can also see that the mean and median offers in each treatment were approximately equal, so there is no strong skewing in the data; this impression is confirmed by the boxplots.

Another plot that is interesting to look at with two-way data is the `interaction.plot`.

```
> attach(car.offers)
> interaction.plot(age,gender,offer)
> interaction.plot(gender,age,offer)
> detach()
```



- The interaction plot on the left shows male-female differences in the cell means, at each level of the age factor. If gender and age operate independently of one another to influence the offer made on the car, these differences should be the same at each level of age, i.e. the "curves" in the plot should be parallel.
- The interaction plot on the right shows differences between offers made to each age group, for males and for females. Again, if gender and age operate independently to influence the offer made, these lines should be parallel (the age differences in offer should be the same for male as for female sellers).

Clearly, the two interaction plots contain the same information. But sometimes one plot makes it easier to "see" whether there is an interaction than the other. In this case, the potential for an interaction is hard to see in the plot on the left, but it is easy to see in the plot on the right.

2.1.1 The cell-means model

Please note that the way subscripts are used in these notes may differ from the texbook's usage.

Subscript	Textbook	These Notes
i	Subject	Row
j	Row	Column
k	Column	Subject

By itself the cell-means model is not very interesting, but it gives us some useful notation to use later[1]. The cell-means model for two-way data is

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk}$$

where i is the level of the first (row) factor, j is the level of the second (column) factor, and k indexes observation in treatment cell ij. μ_{ij} is the mean of the observations in treatment cell ij, and ϵ_{ijk} are independent $N(0, \sigma^2)$ errors for each observation. It is useful to think of the data as laid out in rows and columns, like so

	Age			*Row*
Gender	*Young* $(j = 1)$	*Middle* $(j = 2)$	*Old* $(j = 3)$	*Average*
Male $(i = 1)$	μ_{11}	μ_{12}	μ_{13}	$\mu_{1.}$
Female $(i = 2)$	μ_{21}	μ_{22}	μ_{23}	$\mu_{2.}$
Column Average	$\mu_{.1}$	$\mu_{.2}$	$\mu_{.3}$	$\mu_{..}$

The extra parameters in the table are defined as one might expect:

$$\begin{aligned}
\mu_{i.} &= \frac{1}{J}\sum_{j=1}^{J}\mu_{ij} \\
\mu_{.j} &= \frac{1}{I}\sum_{i=1}^{I}\mu_{ij} \\
\mu_{..} &= \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\mu_{ij}
\end{aligned}$$

where I is the number of levels of the row factor and J is the number of levels of the column factor. The cell means are estimated in the natural way, as

$$\hat{\mu}_{ij} = \bar{Y}_{ij} = \frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}} y_{ijk}$$

where the n_{ij} are the number of observations in each cell (when all the n_{ij} are equal, it is a *balanced design*). Note that the interaction plots above are just plots of the rows or columns of the $\bar{Y}_{ij}$'s in the two-way layout. Marginal totals $\mu_{i.}$, $\mu_{.j}$ and $\mu_{..}$ would be estimated by the corresponding averages of $barY$'s, i.e. $\bar{Y}_{i.}$, $\bar{Y}_{.j}$ and $\bar{Y}_{..}$.

[1] ...and it is the right place to start when trying to relate ANOVA models back to plain vanilla linear regression models, as we have seen before

The fitted values are $\hat{y}_{ij} = \bar{Y}_{ij}$, and the residuals are $r_{ijk} = y_{ijk} - \hat{y}_{ij}$. We could partition the total sum of squares, as for one-way ANOVA, into between- and within- sums of squares,

$$\begin{aligned} \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{n_{ij}}(\bar{Y}_{..} - y_{ijk})^2 &= \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}(\bar{Y}_{..} - \hat{y}_{ij})^2 + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{n_{ij}}(\hat{y}_{ij} - y_{ijk})^2, \text{ or} \\ TSS &= RegSS + RSS \end{aligned}$$

and this gives rise to the usual[2] ANOVA table, as before.

Source	*SS*	*df*	*MS*
Regression	RegSS	$IJ-1$	RegMS=$RegSS/(IJ-1)$
Error	RSS	$n-IJ$	RMS=$RSS/(n-IJ)$
Total	TSS	$n-1$	TMS=$TSS/(n-1)$

which we can examine immediately in SPLUS

```
> car.aov _ aov( offer ~ gender:age,data=car.offers)
> anova(car.aov)
Analysis of Variance Table

Response: offer

Terms added sequentially (first to last)
           Df Sum of Sq  Mean Sq  F Value         Pr(F)
gender:age  5  327.2222 65.44444 27.39535 2.451114e-010
 Residuals 30   71.6667  2.38889
```

Note the use of ":" to indicate the row and column variables (more generally ":" is used to construct interactions, but more about that later). Unsurprisingly (given the plots we did before), something about the factors matters, but this analysis does not reveal clearly what it is.

2.1.2 Main Effects and Interactions

To learn more about which of the two factors are important and how the factors may or may not interact, we now define and examine the *main effects* and *interactions* of the two factors for this problem.

- Main row (first factor) effects:

$$\alpha_i = \mu_{i.} - \mu_{..};$$

[2]The degrees of freedom look mysterious in this table, but they are not. If you were doing this as a one-way ANOVA, you would have IJ indicators, and you know you must drop one term from the model to alleviate collinearity problems. Hence the df for the model (total number of parameters to be estimated) is $IJ - 1$.

- Main column (second factor) effects:

$$\beta_j = \mu_{.j} - \mu_{..};$$

It is easy to verify that $\sum_i \alpha_i = 0$ and $\sum_j \beta_j = 0$, and it is easy to see how to estimate α_i and β_i in terms of the $\bar{Y}$'s.

To see whether the first factor "matters" we want to construct a test of $H_0 : \alpha_i \equiv 0$ vs $H_1 :$ *not all* 0; and similarly to see whether the second factor "matters" we should test $H_0 : \beta_j \equiv 0$ vs $H_1 :$ *not all* 0.

In one-way ANOVA we say that the cell means model and the grand means model were related by the equation $\mu_i = \mu_. + \alpha_i$. The corresponding equation here,

$$\begin{aligned} \mu_{ij} &\stackrel{?}{=} \mu_{..} + \alpha_i + \beta_j \\ &= \mu_{i.} + \mu_{.j} - \mu_{..} \end{aligned}$$

is not true in general. The differences between the left hand sides of this "equation" and the right hand sides are called the *interactions*,

- Interactions[3]:

$$\begin{aligned} \gamma_{ij} &= \mu_{ij} - (\alpha_i + \beta_i + \mu_{..}) \\ &= \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..} \end{aligned}$$

It is also easy to see that $\sum_i \sum_j \gamma_{ij} = 0$. Now we have equality:

$$\begin{aligned} \mu_{ij} &= \mu + \alpha_i + \beta_j + \gamma_{ij} \\ &= \mu + (\mu_{i.} - \mu) + (\mu_{.j} - \mu) + (\mu_{ij} - \mu_{i.} - \mu_{.j} + \mu), \end{aligned}$$

where $\mu = \mu_{..}$ is the grand mean; and this gives rise to the "grand mean plus effects" model for two-way ANOVA, namely

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

subject to the conditions that the α_i's sum to zero, the β_j's sum to zero, and the γ_{ij}'s sum to zero.

[3]Some texts use the notation $(\alpha\beta)_{ij}$ instead of γ_{ij} for interactions, to emphasize the joint influence of the two factors on the outcome.

Recall our general ANOVA table for sums of squares[4] in linear regression and ANOVA models,

Source	*SS*	*df*	*MS*
α	$SS(\alpha)$	df_α	$SS(\alpha)/df_\alpha$
$\beta\vert\alpha$	$SS(\beta\vert\alpha)$	df_β	$SS(\beta\vert\alpha)/df_\beta$
$\gamma\vert\alpha,\beta$	$SS(\gamma\vert\alpha,\beta)$	df_γ	$SS(\gamma\vert\alpha,\beta)/df_\gamma$
Regression	$SS(\alpha,\beta,\gamma)$	k	$SS(\alpha,\beta,\gamma)/k$
Error	$RSS(\alpha,\beta,\gamma)$	$n-(k+1)$	$RSS(\alpha,\beta,\gamma)/[n-(k+1)]$
Total	TSS	$n-1$	

From the definitions of $SS(\cdot|\cdots)$ as differences in $RegSS$'s for the corresponding nested model[5], we know that the SS's listed above always add up to the total $RegSS = SS(\alpha,\beta,\gamma)$, no matter what order the terms α,β,γ are added to the model,

$$\begin{aligned} SS(\alpha,\beta,\gamma) &= SS(\alpha) + SS(\beta|\alpha) + SS(\gamma|\alpha,\beta) \\ &= SS(\beta) + SS(\gamma|\beta) + SS(\alpha|\gamma,\beta) \\ &= etc\ldots \end{aligned}$$

but clearly the values of the individual summands on the right will vary depending on the order the terms are added to the model.

Something nicer happens in balanced (equal number of observations in each cell) ANOVA problems. To see why, we begin by noting that in any balanced two-way ANOVA with the grand mean plus interactions parameterization, it turns out that the group of main effects for the first factor (i.e. the row effects) is orthogonal to the group of main effects for the second factor (the column effects), and both of these are orthogonal to the group of effects for the interactions.

For example, let us return to the 2×3, two-way layout

	Age			*Row*
Gender	*Young* $(j=1)$	*Middle* $(j=2)$	*Old* $(j=3)$	*Average*
Male $(i=1)$	μ_{11}	μ_{12}	μ_{13}	$\mu_{1.}$
Female $(i=2)$	μ_{21}	μ_{22}	μ_{23}	$\mu_{2.}$
Column Average	$\mu_{.1}$	$\mu_{.2}$	$\mu_{.3}$	$\mu_{..}$

(where now for example, $\mu_{11} = \mu + \alpha_1 + \beta_1 + \gamma_{11}$, $\mu_{21} = \mu + \alpha_2 + \beta_1 + \gamma_{21}$, and the α's, β's and γ's satisfy the sum-to-zero constraints) and suppose that it is perfectly balanced (same number of

[4]The set of sums of squares displayed in this table are called *sequential sums of squares* because their values depend on the order, or sequence, in which terms are added to the model. The set of sums of squares $SS(\alpha|\beta,\gamma)$, $SS(\beta|\alpha,\gamma)$, and $SS(\gamma|\alpha,\beta)$, are called *partial sums of squares*, and correspond to the numerators of the squared t statistics for adding each term *last* in the model. I think SAS's "Type III" sums of squares are exactly the partial sums of squares. SAS also displays "Type II" sums of squares, which I believe are basically partial sums of squares for adding each main effect last, but ignoring interactions. Type II SS's seem less useful to me.

[5]E.g., $SS(\beta|\alpha) = SS(\alpha,\beta) - SS(\alpha)$.

observations for each cell in the table). Then, following our development of the $0/+1/-1$-coded "deviation regressors" $D_?$ for the grand mean plus effects parametrization, as differences of 0/1-coded indicator variables, one can show that the "deviation regressors" for the two-way layout will be

Cell		D_1	D_2	D_3	D_4	D_5
		α_1	β_1	β_2	γ_{11}	γ_{12}
Row	*Column*	R_1	C_1	C_2	R_1C_1	R_1C_2
1	1	1	1	0	1	0
1	2	1	0	1	0	1
1	3	1	-1	-1	-1	-1
2	1	-1	1	0	-1	0
2	2	-1	0	1	0	-1
2	3	-1	-1	-1	1	1

where each row should be replicated as many times as there are observations in each cell of the layout. Because each row is replicated the same number of times, the deviation regressors will clearly be orthogonal (dot product or any two columns equal to zero).

A consequence of this orthogonality (we will see the proof later) is that

$$\begin{aligned} SS(\alpha|\beta,\gamma) &= SS(\alpha|\beta) = SS(\alpha) \\ SS(\beta|\alpha,\gamma) &= SS(\beta|\alpha) = SS(\beta) \\ SS(\gamma|\alpha,\beta) &= SS(\gamma) \end{aligned}$$

This means we can partition $SS(\alpha,\beta,\gamma)$ in the ANOVA table above into contributions from each set of main effects and the interactions, that do not depend on the order in which the terms were added into the model:

$$\begin{aligned} SS(\alpha,\beta,\gamma) &= SS(\alpha) + SS(\beta|\alpha) + SS(\gamma|\alpha,\beta) \\ &= SS(\alpha) + SS(\beta) + SS(\gamma) \end{aligned}$$

You can verify directly for example[6] that

$$\begin{aligned} SS(\alpha,\beta,\gamma) &= \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}(\bar{Y}_{..} - \hat{y}_{ij})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}(\bar{Y}_{..} - \bar{Y}_{ij})^2 \\ &= \sum_{i=1}^{I} n_{i.}(\bar{Y}_{..} - \bar{Y}_{i.})^2 + \sum_{j=1}^{J} n_{.j}(\bar{Y}_{..} - \bar{Y}_{.j})^2 + n\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{Y}_{..} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{ij})^2 \\ &= SS(\alpha) + SS(\beta) + SS(\gamma) \end{aligned}$$

and this greatly simplifies computation of the ANOVA model and interpreation of the F-tests one can build from it.

Let us say $n_{ij} \equiv m$ for all cells in the balanced design; then the expanded ANOVA table looks like this:

[6] Though I don't recommend it unless you have a lot of time to kill.

Source	*SS*	*df*	*MS*
Main effects (α)	SS(α)	$I-1$	MS(α)=SS(α)/$(I-1)$
Main effects (β)	SS(β)	$J-1$	MS(β)=SS(β)/$(J-1)$
Interactions	SS(γ)	$(I-1)(J-1)$	MS(γ)=SS(γ)/$((I-1)(J-1))$
Regression	SS(α,β,γ)	$IJ-1$	MS(α,β,γ)=SS(α,β,γ)/$(IJ-1)$
Error	RSS(α,β,γ)	$IJ(m-1)$	MSE=RSS(α,β,γ)/$(IJ(m-1))$
Total	TSS	$mIJ-1$	TMS=$TSS/(mIJ-1)$

This is also easy to obtain in SPLUS:

```
> car.aov _ aov( offer ~ gender*age,data=car.offers)
> anova(car.aov)
Analysis of Variance Table

Response: offer

Terms added sequentially (first to last)
           Df Sum of Sq  Mean Sq  F Value    Pr(F)
    gender  1    5.4444   5.4444  2.27907 0.141592
       age  2  316.7222 158.3611 66.29070 0.000000
gender:age  2    5.0556   2.5278  1.05814 0.359700
 Residuals 30   71.6667   2.3889
```

Note the use of the "*" to indicate that main effects `gender`, `age` as well as the interactions `gender:age` should be included in the model. Scanning the F-tests here, it seems clear that age matters and gender does not (since the gender and age effects are orthogonal it does not matter in which order we examine them, as can be seen from the alternate order for the factors below).

```
> car.aov _ aov( offer ~ age*gender,data=car.offers)
> summary(car.aov)
           Df Sum of Sq  Mean Sq  F Value    Pr(F)
       age  2  316.7222 158.3611 66.29070 0.000000
    gender  1    5.4444   5.4444  2.27907 0.141592
age:gender  2    5.0556   2.5278  1.05814 0.359700
 Residuals 30   71.6667   2.3889
```

Moreover, the interaction doesn't seem to matter. When we can safely set the interaction terms to zero in a two-way ANOVA, we arrive at what is called the *additive model*,

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

In the additive model it is true that $\mu_{ij} = \mu + \alpha_i + \beta_j$, so differences in cell means in one factor are the same, no matter the value of the second factor:

$$\mu_{ij} - \mu_{i'j} = \mu_{ij'} - \mu_{i'j'}$$

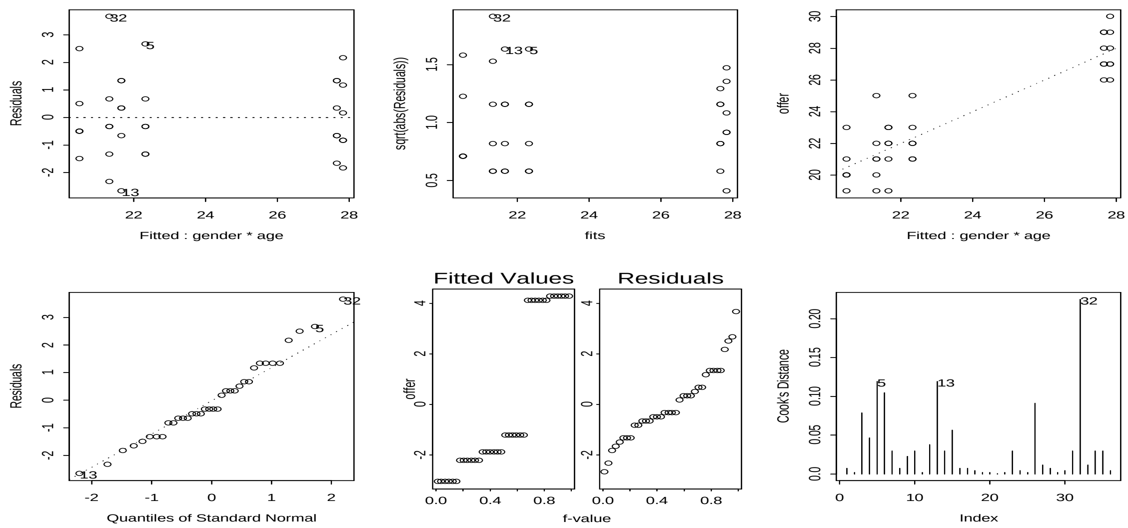
This is why the curves in an interaction plot should be parallel when there is no interaction between the variables.

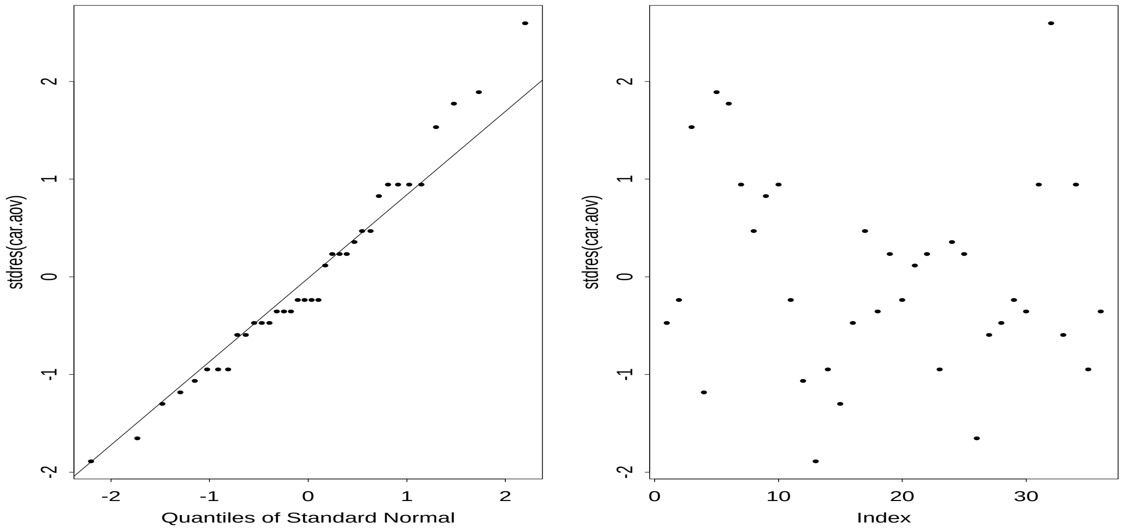
2.1.3 Model Selection and Diagnostics

Usually, it is only sensible to include or exclude entire factors or interactions in an ANOVA model, rather than looking at individual factor levels. In the car offer problem, we probably would not include the interaction `gender:age`, and we would probably also exclude the gender effects.

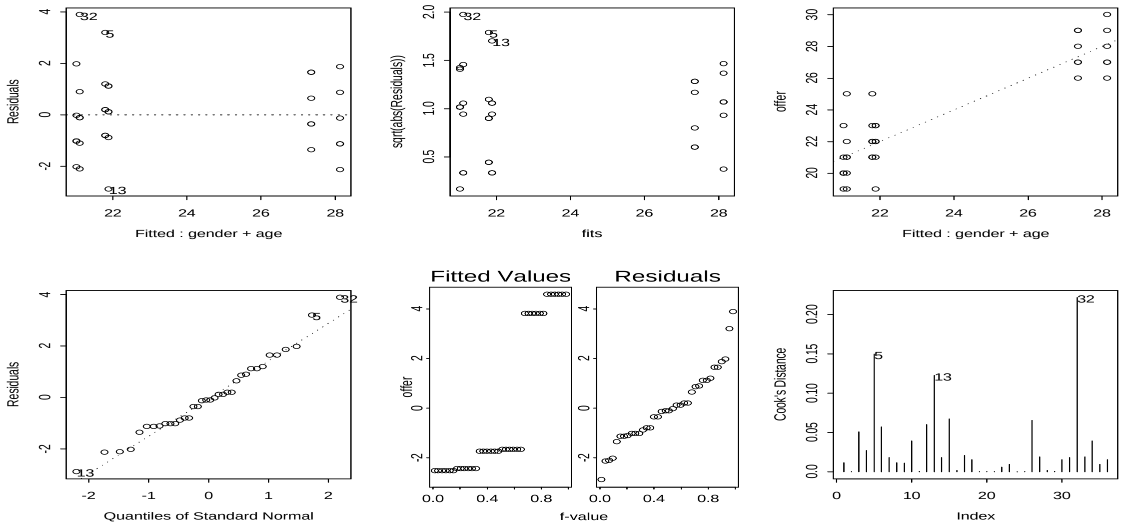
To illustrate diagnostics we will examine the diagnostic plots of all three models under consideration.

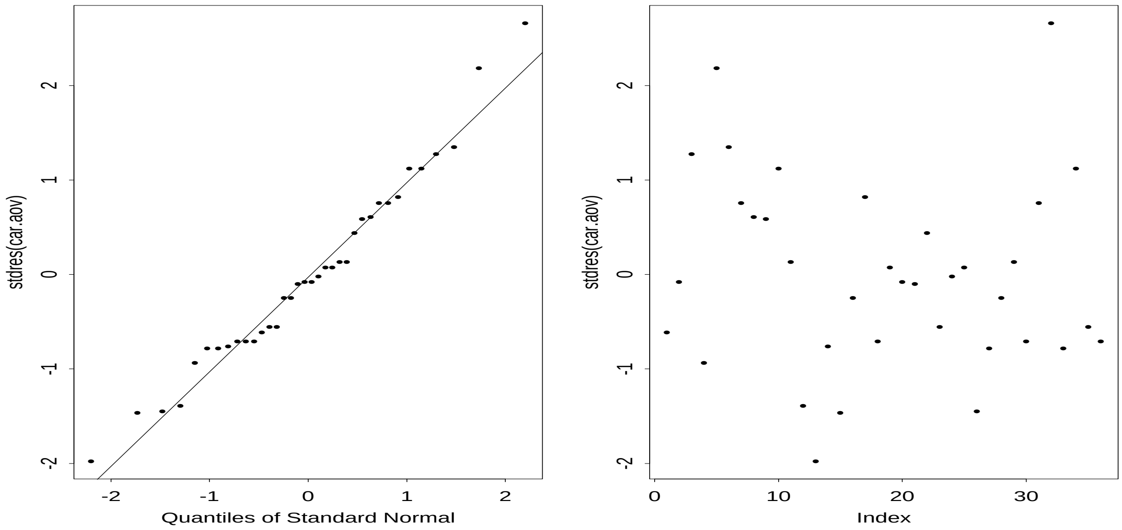
```
> car.aov _ aov( offer ~ gender*age,data=car.offers)
> par(mfrow=c(2,3)); plot(car.aov)
> par(mfrow=c(1,2))
> source("stdres.splus") # function from week04
> qqnorm(stdres(car.aov)); qqline(stdres(car.aov)); plot(stdres(car.aov))
```



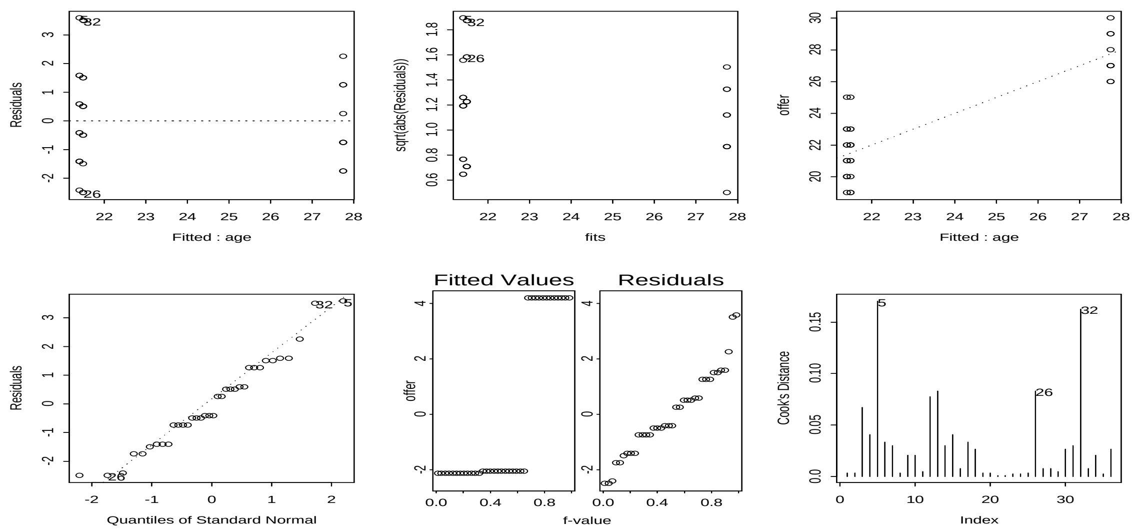


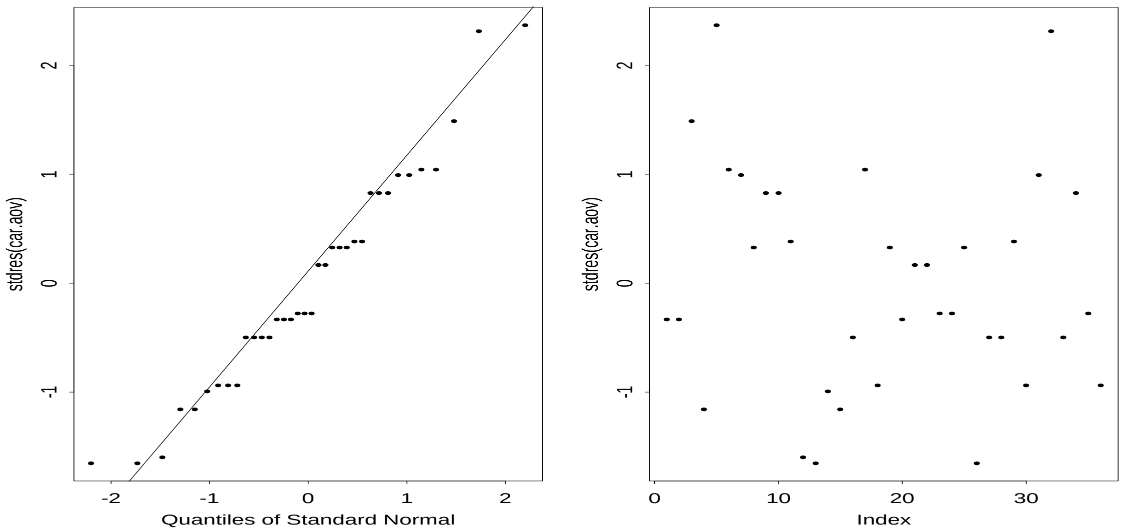

```
> car.aov _ aov( offer ~ gender + age,data=car.offers)
> anova(car.aov)
          Df Sum of Sq  Mean Sq  F Value     Pr(F)
   gender  1    5.4444   5.4444  2.27082 0.1416397
      age  2  316.7222 158.3611 66.05069 0.0000000
Residuals 32   76.7222   2.3976
> par(mfrow=c(2,3))
> plot(car.aov)
> par(mfrow=c(1,2))
> qqnorm(stdres(car.aov)); qqline(stdres(car.aov)); plot(stdres(car.aov))
```






```
> car.aov _ aov( offer ~ age,data=car.offers)
> anova(car.aov)
          Df Sum of Sq  Mean Sq  F Value          Pr(F)
      age  2  316.7222 158.3611 63.60142 4.768963e-012
Residuals 33   82.1667   2.4899
> par(mfrow=c(2,3))
> plot(car.aov)
> par(mfrow=c(1,2))
> qqnorm(stdres(car.aov)); qqline(stdres(car.aov)); plot(stdres(car.aov))
```





Other diagnostics, for example examining the relationship between within-cell variances and estimated cell means to find a variance-stabilizing transformation, proceed in much the same way as one-way ANOVA.

2.1.4 Fitted Values, Parameter Estimates

Fitted values are easy to obtain in SPLUS from the `aov` object, using `fitted()`, or by computing cell means by hand.

- If interactions are *absent*, it is usually most interesting to obtain estimates of the marginal means $\mu_{i.}$ and $\mu_{.j}$.
- If interactions are *present*, then one usually must focus on the cell means μ_{ij} themselves.

Parameter estimates are a little tricky, again because SPLUS's default ANOVA parameterization gives estimates of the Helmert contrasts instead of something sensible like the factor effects and interactions.

If one wants estimates of the factor effects and interactions, one can obtain them much as with one-way ANOVA. Here is how to do it with the `C( , )` function; for example `C(age,sum)` says to parameterize the model so that age is estimated with factor effects that sum to zero.

```
> car.aov _ aov(offer ~ C(gender,sum) + C(age,sum), data=car.offers)
> summary.lm(car.aov)

Call: aov(formula = offer ~ C(gender, sum) + C(age, sum), data = car.offers)
Residuals:
    Min      1Q  Median     3Q   Max
 -2.889 -1.028 -0.1111 0.9444 3.889

Coefficients:
```

```
                 Value Std. Error  t value Pr(>|t|)
   (Intercept)  23.5556   0.2581    91.2765   0.0000
C(gender, sum)  -0.3889   0.2581    -1.5069   0.1416
  C(age, sum)1  -2.1389   0.3650    -5.8606   0.0000
  C(age, sum)2   4.1944   0.3650    11.4928   0.0000

Residual standard error: 1.548 on 32 degrees of freedom
Multiple R-Squared: 0.8077
F-statistic: 44.79 on 3 and 32 degrees of freedom, the p-value is 1.466e-011

Correlation of Coefficients:
               (Intercept) C(gender, sum) C(age, sum)1
C(gender, sum)  0.0
  C(age, sum)1  0.0         0.0
  C(age, sum)2  0.0         0.0           -0.5
```

2.2 Example: One observation per cell

The data below (used by Box, Hunter, and Hunter (1978)) were collected to determine the effect of treatments A, B, C, and D on the yield of penicillin in a penicillin manufacturing process. The values of the response variable `yield` are the numbers in the table, and the columns of the table correspond to the levels A, B, C, and D of the treatment factor.

There was a second factor, namely the blend factor, since a separate blend of the corn-steep liquor had to be made for each application of the treatments. Your main interest is in determining whether the treatment factor affects yield. The blend factor is of only secondary interest; it is a *blocking* variable introduced to increase the sensitivity of the inference for treatments. The order of the treatments within blocks was chosen at random.

This is a *randomized blocks* experiment: four penicillin cultures within each blend were randomly assigned to the four treatments.

```
Effect of four treatments on penicillin yield

         Treatment
Block    A  B  C  D
Blend-1 89 88 97 94
Blend-2 84 77 92 79
Blend-3 81 87 87 85
Blend-4 87 92 89 84
Blend-5 79 81 80 88
```

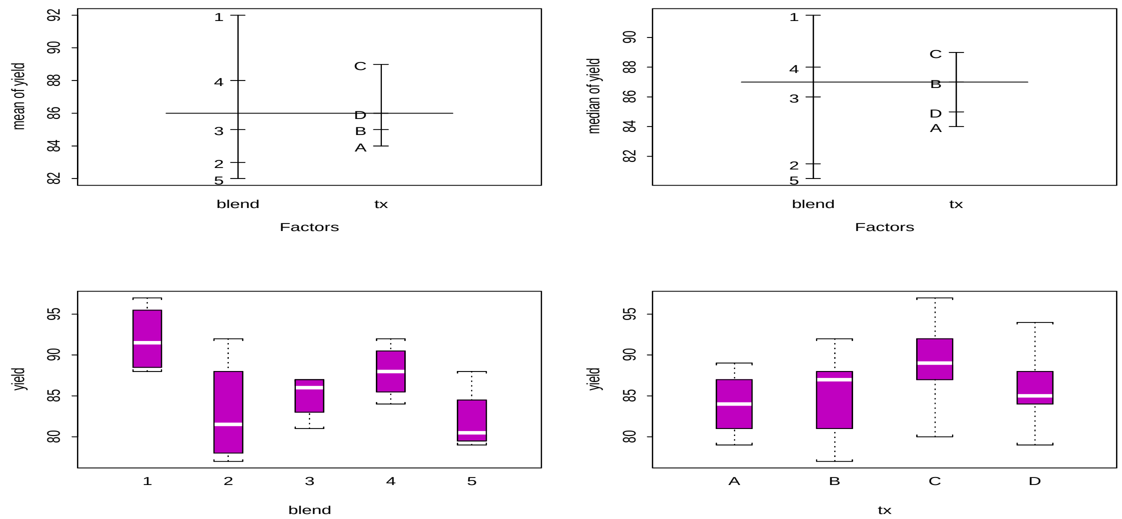
Here is one way to convert this to a form SPLUS will like

```
> pen.raw _ read.table("penicillin.txt",header=T)
> pen.raw _ c(as.matrix(pen.raw))
> pen.raw
 [1] 89 84 81 87 79 88 77 87 92 81 97 92 87 89 80 94 79 85 84 88
```

```
> pendata _ data.frame(blend=factor(rep(1:5,4)),
+ tx=rep(LETTERS[1:4],rep(5,4)), yield=pen.raw)
> pendata
   blend tx yield
 1     1  A    89
 2     2  A    84
 3     3  A    81
 4     4  A    87
 5     5  A    79
 .
 .
 .
20     5  D    88
```

We begin with the usual plots

```
> par(mfrow=c(2,2))
> plot.design(pendata)
> plot.design(pendata,fun=median)
> plot.factor(pendata)
```



It turns out that you can't fit an ANOVA model with full, unconstrained interactions to a data sets that has only one observation per cell, if you want any degrees of freedom left over for the error variance.

To see this, recall the ANOVA table for balanced designs ($n_{ij} \equiv m$ for all cells),

Source	*SS*	*df*	*MS*
Main effects (α)	SS(α)	$I-1$	MS(α)=SS(α)/$(I-1)$
Main effects (β)	SS(β)	$J-1$	MS(β)=SS(β)/$(J-1)$
Interactions	SS(γ)	$(I-1)(J-1)$	MS(γ)=SS(γ)/$((I-1)(J-1))$
Regression	SS(α,β,γ)	$IJ-1$	MS(α,β,γ)=SS(α,β,γ)/$(IJ-1)$
Error	RSS(α,β,γ)	$IJ(m-1)$	MSE=RSS(α,β,γ)/$(IJ(m-1))$
Total	TSS	$mIJ-1$	TMS=$TSS/(mIJ-1)$

we see that, when $m=1$, if we include the interaction terms, there will be no degrees of freedom left for calculating the RSS and RMS. We can see this directly in SPLUS:

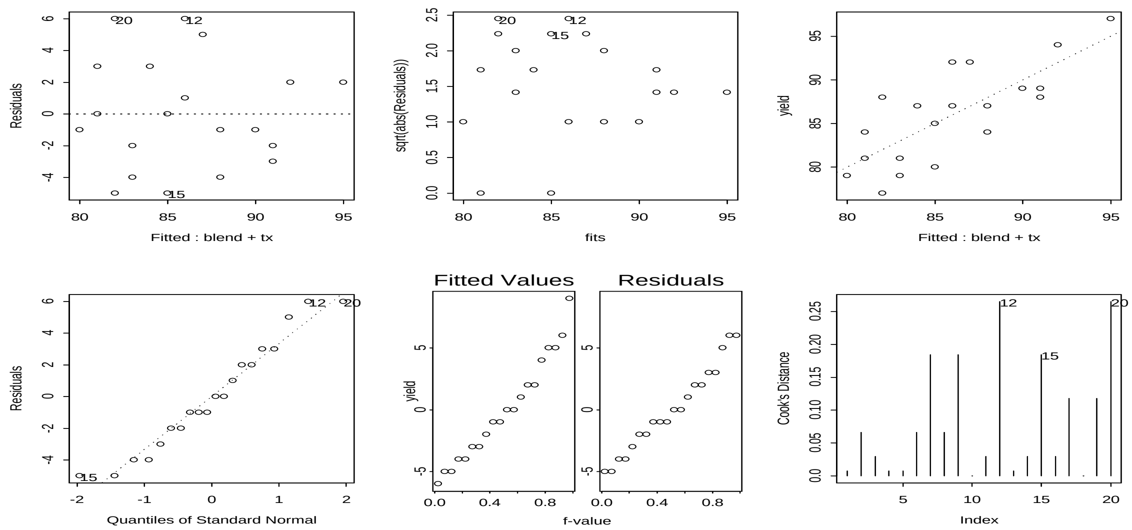
```
> pen.aov _ aov(yield ~ blend*tx,data=pendata)
> summary(pen.aov)
         Df Sum of Sq  Mean Sq
   blend  4       264 66.00000
      tx  3        70 23.33333
blend:tx 12       226 18.83333
```

(and no error line is given in the ANOVA table). However, if we omit the interaction term, then we can use the interaction sum of squares as the error sum of squares:

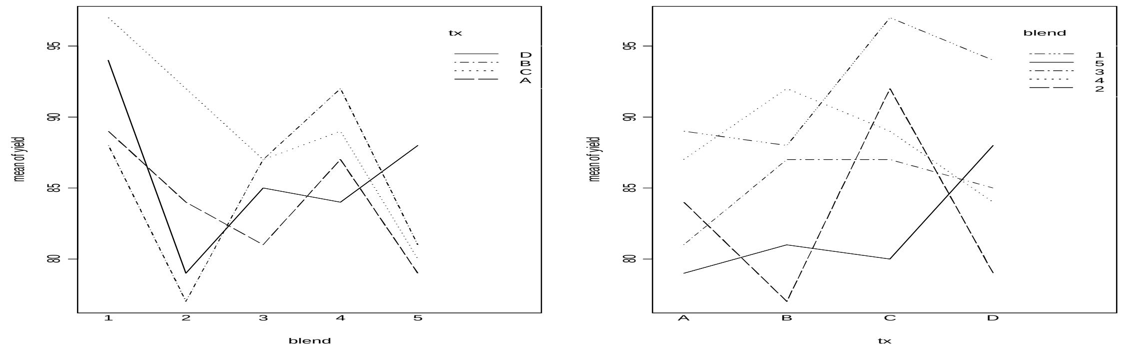
```
> pen.aov _ aov(yield ~ blend + tx, data=pendata)
> summary(pen.aov)
          Df Sum of Sq  Mean Sq  F Value     Pr(F)
    blend  4       264 66.00000 3.504425 0.0407462
       tx  3        70 23.33333 1.238938 0.3386581
Residuals 12       226 18.83333
```

Unfortunately, the treatment does not seem to be having much effect on the penicillin yield. On the other hand, we seem to have accidentally discovered that the blocking factor, the blend of the corn-steep liquor used, is marginally significant!

The diagnostic plots do not yield any substantial difficulties with the regression. A time series plot of the residuals, not reproduced here, also does not reveal any substantial problems.



One might also ask whether the interaction plots can be used to assess whether there "might be" an interaction that the ANOVA simply can't pick up. We can maket the plots, but they are very difficult to interpret. For example the interaction plots for this data set are quite jumbled and nonparallel, yet there is no strong indication of interaction in them.



Chapter 13 of the SPLUS Guide to Statistics has more information about testing and estimating specially-constrained interaction terms; however for many purposes these interaction terms would be hard to interpret.

Finally, it might be interesting to compare these results with the results one would get if one ignored the blocking factor and did a one-way ANOVA on the four treatments:

```
> pen.aov _ aov(yield ~ tx, data = pendata)
> summary(pen.aov)
           Df Sum of Sq  Mean Sq   F Value     Pr(F)
       tx  3        70 23.33333 0.7619048 0.5317827
Residuals 16       490 30.62500
```

Note that the p-value is even higher for `tx`, the treatment, than it was under the two-way, randomized blocks analysis. A more careful inspection of the two ANOVA tables explains why: the error SS in the one-way analysis is the sum of the error variance in the two-way analysis plus the SS due to the blocks: in other words, the variation due to differing blocks (differing corn-liquor blends) has been lumped into the error variance of the one-way analysis. This is why blocking is such an effective idea in experimental design.

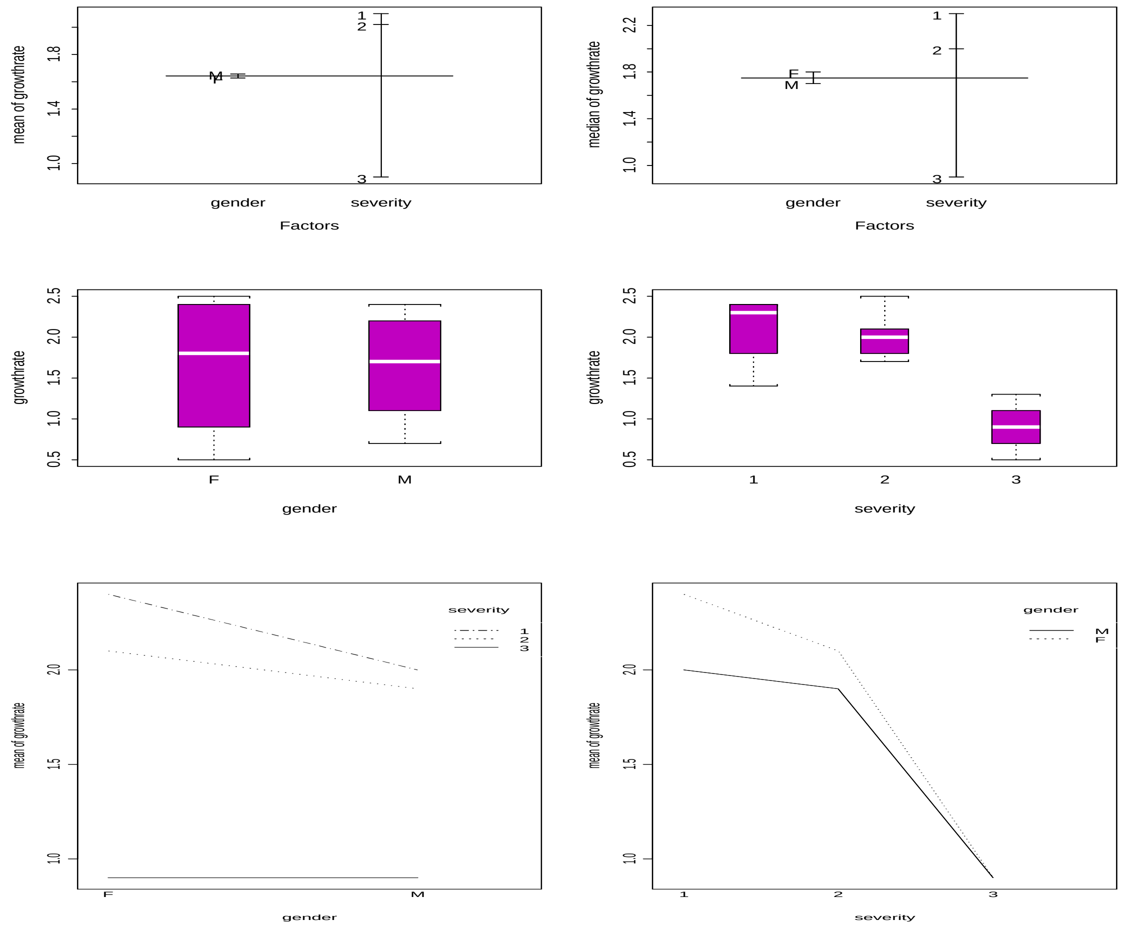
2.3 Example: Unbalanced two-way layout

Human growth hormone was administered at a clinical research center to growth-hormone-deficient, short children who had not yet reached puberty. The investigator was interested in the effects of gender (factor A: male or female) and bone development (factor B: severely (1), moderately (2), or mildly (3), depressed) on the rate of growth induced by the hormone.

```
M 1 1.4
M 1 2.4
M 1 2.2
M 2 2.1
M 2 1.7
M 3 0.7
M 3 1.1
F 1 2.4
F 2 2.5
F 2 1.8
F 2 2.0
F 3 0.5
F 3 0.9
F 3 1.3
```

Reading the data in this format is straightforward, so we do this and then produce what we hope are useful initial plots.

```
> hormone.data _ read.table("hormone.txt")
> names(hormone.data)
[1] "V1" "V2" "V3"
> names(hormone.data)  _ c("gender","severity","growthrate")
> hormone.data$severity _ factor(hormone.data$severity)
> par(mfrow=c(2,2))
> plot.design(hormone.data)
> plot.design(hormone.data, fun=median)
> plot.factor(hormone.data)
> attach(hormone.data)
> par(mfrow=c(1,2))
> interaction.plot(gender,severity,growthrate)
> interaction.plot(severity,gender,growthrate)
> detach()
```



It appears that severity will matter, but gender will probably not. There does not appear to be much evidence of interactions in the interaction plots.

Clearly the data are unbalanced; the cell sample sizes are

```
Cell  n_{ij}
M 1     3
M 2     2
M 3     2
F 1     1
F 2     3
F 3     3
```

In principle there is little difference between doing an ANOVA with unbalanced designs and doing ANOVA with balanced designs. Calculation of sums of squares, and presentation of the ANOVA table, is more difficult, because the nice order-independent equation

$$SS(\alpha,\beta,\gamma) \stackrel{?}{=} SS(\alpha)+SS(\beta)+SS(\gamma)$$

no longer holds—the factors are no longer orthogonal—and we must return to calculations involving the order-dependent (and harder to calculate) sequential sums of squares,

$$SS(\alpha, \beta, \gamma) \quad = \quad SS(\alpha) + SS(\beta|\alpha) + SS(\gamma|\alpha, \beta)$$

are no longer orthogonoal—these equalities no longer hold. Instead, sequential sums of squares as in regression must be calculated. In fact, in many computer packages, the simplest way to get correct parameter estimates and sums of squares for unbalanced designs is to analyze the data using dummy variables in a regression package.

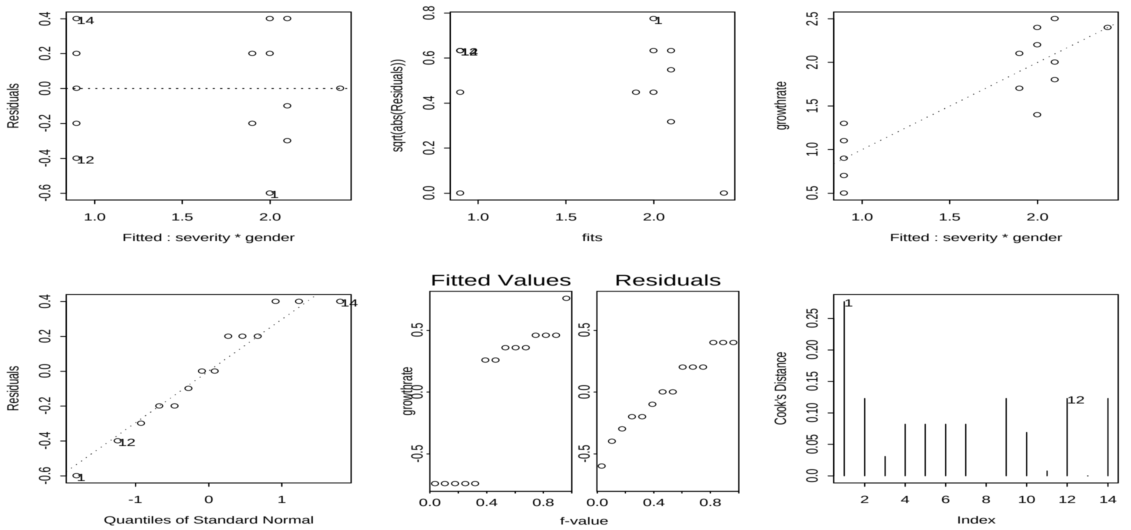
These problems can be illustrated in SPLUS without resorting to regression, because `aov()` does properly handle unbalanced ANOVA problems.

```
> hormone.aov _ aov(growthrate ~ gender*severity,data=hormone.data)
> summary(hormone.aov)
                Df Sum of Sq  Mean Sq  F Value     Pr(F)
         gender  1  0.002857 0.002857  0.01758 0.8977853
       severity  2  4.396000 2.198000 13.52615 0.0027133
gender:severity  2  0.075429 0.037714  0.23209 0.7980337
      Residuals  8  1.300000 0.162500


> hormone.aov _ aov(growthrate ~ severity*gender,data=hormone.data)
> summary(hormone.aov)
                Df Sum of Sq  Mean Sq  F Value     Pr(F)
       severity  2  4.306286 2.153143 13.25011 0.0028912
         gender  1  0.092571 0.092571  0.56967 0.4720219
severity:gender  2  0.075429 0.037714  0.23209 0.7980337
      Residuals  8  1.300000 0.162500

> par(mfrow=c(2,3))
> plot(hormone.aov)
```

Either way, it appears that only severity is an important influence on the growth rate, as we suspected from the boxplots.



The diagnostic residual plots look reasonably good here too. One thing to watch out for when using these plots with unbalanced data is that if you are judging the "equal variance" assumption by eye, be careful to note that even if the variances in all cells *are* the same, cells with fewer observations will appear to look less variable.

Defining the grand mean and the marginal means in an unbalanced two-way layout can be done several different ways, just as for unbalanced one-way problems. The two main choices are

- *Unweighted means.* The definitions of μ and of $\mu_{i.}$ and $\mu_{.j}$ are exactly as for the balanced case.
- *Sample size weighting.* The definitions are changed to be

$$\begin{aligned}\mu_{i.} &= \frac{1}{n_{i.}}\sum_{j=1}^{J} n_{ij}\mu_{ij} \\ \mu_{.j} &= \frac{1}{n_{.j}}\sum_{i=1}^{I} n_{ij}\mu_{ij} \\ \mu_{..} &= \frac{1}{n_{..}}\sum_{i}\sum_{j} n_{ij}\mu_{ij}\end{aligned}$$

  and corresponding changes in the "sum to zero" constraints on the α's, β's, etc.

In SPLUS, the `dummy.coef()` function standardizes factor and treatment effects so that they sum to zero, corresponding to unweighted means.

```
> dummy.coef(hormone.aov)
$"(Intercept)":
 (Intercept)
         1.7

$severity:
```

```
   1   2    3
 0.5 0.3 -0.8

$gender:
   F    M
 0.1 -0.1

$"severity:gender":
  1F             2F   3F   1M             2M  3M
 0.1 -2.081668e-017 -0.1 -0.1 2.081668e-017 0.1
```

The `model.tables()` function produces factor effect estimates that correspond to sample-size weighting.

```
> model.tables(hormone.aov,se=T)
Refitting model to allow projection
Standard error information not returned as design is unbalanced.
Standard errors can be obtained through se.contrast.
Tables of effects

 severity
         1      2       3
    0.4571 0.3771 -0.7429
rep 4.0000 5.0000  5.0000

 gender
          F        M
    0.07714 -0.07714
rep 7.00000  7.00000

 severity:gender
Dim 1 : severity
Dim 2 : gender
         F      M
  1  0.171 -0.057
rep  1.000  3.000
  2  0.011 -0.017
rep  3.000  2.000
  3 -0.069  0.103
rep  3.000  2.000
```

Note that `model.tables()` would not produce standard errors. In the unbalanced case, the standard errors for comparing two means depend on the replication factors, hence it could be very complex to tabulate all combinations. Instead, they can be computed directly with *se.contrast.* For instance, to compare the first and third bone severity levels for boys in our hormone study:

```
> se.contrast(hormone.aov,contrast=list(gender=="M",severity==1,
+ gender=="M",severity==3),data=hormone.data)
Refitting model to allow projection
[1] 0.1703463
```

By default, the standard error of the difference of the means specified by contrast is computed. The contrast itself can be computed using `model.tables()` again:

```
> model.tables(hormone.aov,type="means")
Refitting model to allow projection
Tables of means
Grand mean

 1.643

 severity
      1    2   3
    2.1 2.02 0.9
rep 4.0 5.00 5.0

 gender
       F     M
    1.72 1.566
rep 7.00 7.000

 severity:gender
Dim 1 : severity
Dim 2 : gender
      F   M
  1 2.4 2.0
rep 1.0 3.0
  2 2.1 1.9
rep 3.0 2.0
  3 0.9 0.9
rep 3.0 2.0
```

Thus, an approximate 95% confidence interval for the difference in growthrate for the first and third bone severity levels among boys is $(2.0 - 0.9) \pm qt(0.975, 3) * (0.1703463)$, which runs from 0.557882 to 1.642118 (`qt(0.975,3) = 3.182446`).

Obtaining SE's for other contrasts, using `se.contrast()`, is explained in Chapter 14 of the SPLUS Guide to Statistics.