Chapter 11

Repeated Measures and Longitudinal Data

In repeated measures designs, there are several individuals (or units) and measurements are taken repeatedly on each individual. When these repeated measurements are taken over time, it is called a *longitudinal* study or, in some applications, a *panel* study. Typically various covariates concerning the individual are recorded and the interest centers on how the response depends on the covariates over time. Often it is reasonable to believe that the response of each individual has several components: a fixed effect, which is a function of the covariates; a random effect, which expresses the variation between individuals; and an error, which is due to measurement or unrecorded variables.

Suppose each individual has response y_i, a vector of length n_i which is modeled conditionally on the random effects γ_i as:

$$y_i|\gamma_i \sim N(X_i\beta + Z_i\gamma_i, \sigma^2\Lambda_i)$$

Notice this is very similar to the model used in the previous chapter with the exception of allowing the errors to have a more general covariance Λ_i. As before, we assume that the random effects $\gamma_i \sim N(0, \sigma^2 D)$ so that:

$$y_i \sim N(X_i\beta, \Sigma_i)$$

where $\Sigma_i = \sigma^2(\Lambda_i + Z_i D Z_i^T)$. Now suppose we have M individuals and we can assume the errors and random effects between individuals are uncorrelated, then we can combine the data as:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_M \end{bmatrix} \qquad X = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_M \end{bmatrix} \qquad \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \dots \\ \gamma_M \end{bmatrix}$$

and $\tilde{D} = diag(D, D, \dots, D)$, $Z = diag(Z_1, Z_2, \dots, Z_M)$, $\Sigma = diag(\Sigma_1, \Sigma_2, \dots, \Sigma_M)$ and $\Lambda = diag(\Lambda_1, \Lambda_2, \dots, \Lambda_M)$. Now we can write the model as

$$y \sim N(X\beta, \Sigma) \qquad \Sigma = \sigma^2(\Lambda + Z\tilde{D}Z^T)$$

The log-likelihood for the data is then computed as previously and estimation, testing, standard errors and confidence intervals all follow using standard likelihood theory as before. There is no strong distinction between the methodology used in this and the previous chapter.

This general structure encompasses a wide range of possible models for different types of data. We explore some of these in the following three examples:

11.1 Longitudinal Data

The Panel Study of Income Dynamics (PSID), begun in 1968, is a longitudinal study of a representative sample of U.S. individuals described in Hill (1992). The study is conducted at the Survey Research Center, Institute for Social Research, University of Michigan, and is still continuing. There are currently 8700 households in the study and many variables are measured. We chose to analyze a random subset of this data, consisting of 85 heads of household who were aged 25–39 in 1968 and had complete data for at least 11 of the years between 1968 and 1990. The variables included were annual income, gender, years of education and age in 1968:

```
data(psid, package="faraway")
head(psid)
```

```
  age educ sex income year person
1  31   12   M   6000   68      1
2  31   12   M   5300   69      1
3  31   12   M   5200   70      1
4  31   12   M   6900   71      1
5  31   12   M   7500   72      1
```

Now plot the data:

```
library(dplyr)
psid20 <- filter(psid, person <= 20)
library(ggplot2)
ggplot(psid20, aes(x=year, y=income))+geom_line()+facet_wrap(~ person)
```

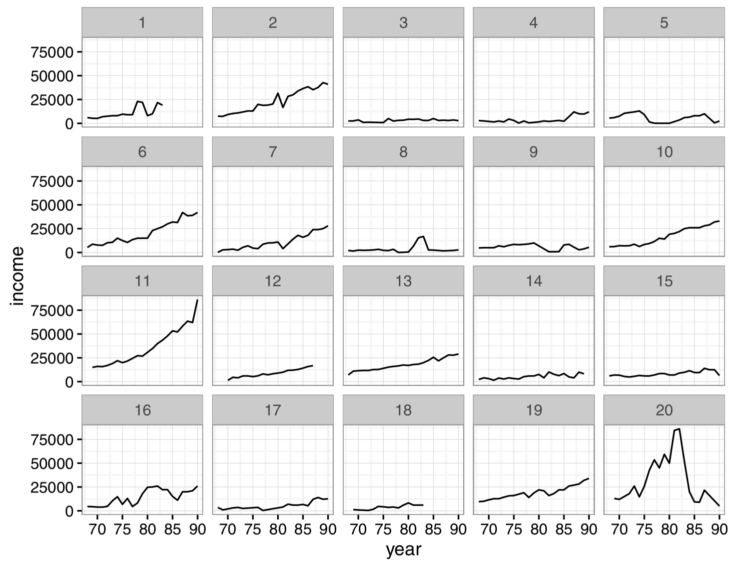

Figure 11.1 *The first 20 subjects in the PSID data. Income is shown over time.*

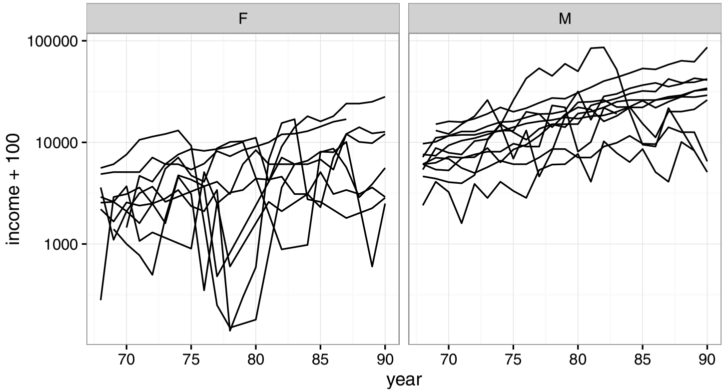

Figure 11.2 *Income change in the PSID data grouped by sex.*

The first 20 subjects are shown in Figure 11.1. We see that some individuals have a slowly increasing income, typical of someone in steady employment in the same job. Other individuals have more erratic incomes. We can also show how the incomes vary by sex. Income is more naturally considered on a log-scale:

```
ggplot(psid20, aes(x=year, y=income+100, group=person)) +geom_line() +
    ↪  facet_wrap(~ sex) + scale_y_log10()
```

See Figure 11.2. We added $100 to the income of each subject to remove the effect of some subjects having very low incomes for short periods of time. These cases distorted the plots without the adjustment. We see that men's incomes are generally higher and less variable while women's incomes are more variable, but are perhaps increasing more quickly. We could fit a line to each subject starting with the first:

```
lmod <- lm(log(income) ~ I(year-78), subset=(person==1), psid)
coef(lmod)
```

```
 (Intercept) I(year - 78)
    9.399957     0.084267
```

We have centered the predictor at the median value so that the intercept will represent the predicted log income in 1978 and not the year 1900 which would be nonsense. We now fit a line for all the subjects and plot the results:

```
library(lme4)
ml <- lmList(log(income) ~ I(year-78) | person, psid)
intercepts <- sapply(ml,coef)[1,]
slopes <- sapply(ml,coef)[2,]
```

The `lmList` command fits a linear model to each group within the data, here specifed by `person`. A list of linear models, one for each group, is returned from which we extract the intercepts and slopes.

```
plot(intercepts,slopes,xlab="Intercept",ylab="Slope")
psex <- psid$sex[match(1:85,psid$person)]
boxplot(split(slopes,psex))
```


In the first panel of Figure 11.3, we see how the slopes relate to the intercepts — there is little correlation. This means we can test incomes and income growths separately. In the second panel, we compare the income growth rates where we see these as higher and more variable for women compared to men. We can test the difference in

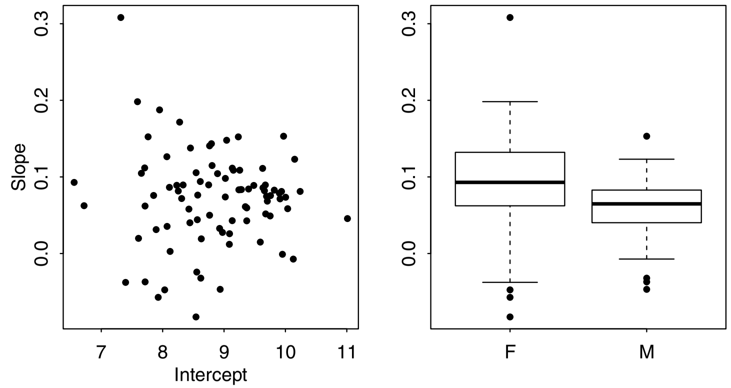

Figure 11.3 *Slopes and intercepts for the individual income growth relationships are shown on the left. A comparison of income growth rates by sex is shown on the right.*

income growth rates for men and women:

```
t.test(slopes[psex=="M"],slopes[psex=="F"])
```

```
        Welch Two Sample t-test

data:  slopes[psex == "M"] and slopes[psex == "F"]
t = -2.3786, df = 56.736, p-value = 0.02077
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.0591687 -0.0050773
sample estimates:
mean of x mean of y
 0.056910  0.089033
```

We see that women have a significantly higher growth rate than men. We can also compare the incomes at the intercept (which is 1978):

```
t.test(intercepts[psex=="M"],intercepts[psex=="F"])
```

```
        Welch Two Sample t-test

data:  intercepts[psex == "M"] and intercepts[psex == "F"]
t = 8.2199, df = 79.719, p-value = 3.065e-12
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.87388 1.43222
sample estimates:
mean of x mean of y
   9.3823    8.2293
```

We see that men have significantly higher incomes.

This is an example of a *response feature* analysis. It requires choosing an important characteristic. We have chosen two here: the slope and the intercept. For many datasets, this is not an easy choice and at least some information is lost by doing this.

Response feature analysis is attractive because of its simplicity. By extracting a univariate response for each individual, we are able to use a wide array of well-known statistical techniques. However, it is not the most efficient use of the data as all the additional information besides the chosen response feature is discarded. Notice that having additional data on each subject would be of limited value.

Suppose that the income change over time can be partly predicted by the subject's age, sex and educational level. We do not expect a perfect fit. The variation may be partitioned into two components. Clearly there are other factors that will affect a subject's income. These factors may cause the income to be generally higher or lower or they may cause the income to grow at a faster or slower rate. We can model this variation with a random intercept and slope, respectively, for each subject. We also expect that there will be some year-to-year variation within each subject. For simplicity, let us initially assume that this error is homogeneous and uncorrelated, that is, $\Lambda_i = I$. We also center the year to aid interpretation as before. We may express these notions in the model:

```
library(lme4)
psid$cyear <- psid$year-78
mmod <- lmer(log(income) ~ cyear*sex +age+educ+(cyear|person),psid)
```

This model can be written as:

$$\begin{aligned}\log(\text{income})_{ij} &= \mu+\beta_y\text{year}_i+\beta_s\text{sex}_j+\beta_{ys}\text{sex}_j\times\text{year}_i+\beta_e\text{educ}_j+\beta_a\text{age}_j\\ &+ \gamma^0_j+\gamma^1_j\text{year}_i+\varepsilon_{ij}\end{aligned}$$

where i indexes the year and j indexes the individual. We have:

$$\begin{pmatrix}\gamma^0_k\\ \gamma^1_k\end{pmatrix}\sim N(0,\sigma^2 D)$$

The model summary is:

```
sumary(mmod, digits=3)
Fixed Effects:
            coef.est coef.se
(Intercept)  6.674    0.543
cyear        0.085    0.009
sexM         1.150    0.121
age          0.011    0.014
educ         0.104    0.021
cyear:sexM  -0.026    0.012

Random Effects:
 Groups   Name        Std.Dev. Corr
 person   (Intercept) 0.531
          cyear       0.049    0.187
 Residual             0.684
---
number of obs: 1661, groups: person, 85
```



```
AIC = 3839.8, DIC = 3751.2
deviance = 3785.5
```

Let's start with the fixed effects. We see that income increases about 10% for each additional year of education. We see that age does not appear to be significant. For females, the reference level in this example, income increases about 8.5% a year, while for men, it increases about $8.5-2.6=5.9\%$ a year. We see that, for this data, the incomes of men are $\exp(1.15)=3.16$ times higher.

We know the mean for males and females, but individuals will vary about this. The standard deviation for the intercept and slope are 0.531 and 0.049 ($\sigma\sqrt{D_{11}}$ and $\sigma\sqrt{D_{22}}$), respectively. These have a correlation of 0.189 ($cor(\gamma^0,\gamma^1)$). Finally, there is some additional variation in the measurement not so far accounted for having standard deviation of 0.684 ($sd(\varepsilon_{ijk})$). We see that the variation in increase in income is relatively small while the variation in overall income between individuals is quite large. Furthermore, given the large residual variation, there is a large year-to-year variation in incomes.

We can test the fixed effect terms for significance. We use the Kenward-Roger adjusted F-test:

```
library(pbkrtest)
mmod <- lmer(log(income) ~ cyear*sex +age+educ+(cyear|person),psid,
    ↪ REML=FALSE)
mmodr <- lmer(log(income) ~ cyear + sex +age+educ+(cyear|person),psid,
    ↪  REML=FALSE)
KRmodcomp(mmod,mmodr)
F-test with Kenward-Roger approximation; computing time: 0.30 sec.
large : log(income) ~ cyear + sex + age + educ + (cyear | person) + cyear:sex
small : log(income) ~ cyear + sex + age + educ + (cyear | person)
       stat   ndf   ddf F.scaling p.value
Ftest  4.61  1.00 81.33         1   0.035
```

We have tested the interaction term between year and sex as this is the most complex term in the model. We see that this term is marginally significant so there is no justification to simplify the model by removing this term. Female incomes are increasing faster than male incomes.

We could test the random effect terms using perhaps the parametric bootstrap method. It is less trouble to create confidence intervals for all the parameters:

```
confint(mmod, method="boot")
                                  2.5 %     97.5 %
sd_(Intercept)|person         0.440965  0.6095268
cor_cyear.(Intercept)|person -0.044677  0.4486294
sd_cyear|person               0.039271  0.0582838
sigma                         0.658930  0.7087268
(Intercept)                   5.571034  7.7676101
cyear                         0.067160  0.1027455
sexM                          0.899570  1.3772171
age                          -0.017808  0.0365997
educ                          0.064944  0.1530431
cyear:sexM                   -0.051526 -0.0028899
```

We see that all the standard deviations are clearly well above zero. There might be a case for removing the correlation between the intercept and slope but this term is difficult to interpret and little would be gained from removing it. It is simpler just to leave it in.

There is a wider range of possible diagnostic plots that can be made with longitudinal data than with a standard linear model. In addition to the usual residuals, there are random effects to be examined. We may wish to break the residuals down by sex as seen in the QQ plots in Figure 11.4:

```
diagd <- fortify(mmod)
ggplot(diagd,aes(sample=.resid))+stat_qq()+facet_grid(~sex)
```

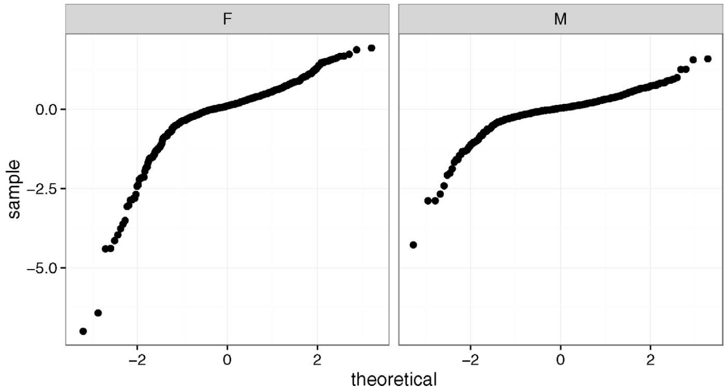

Figure 11.4 *QQ plots by sex.*

We see that the residuals are not normally distributed, but have a long tail for the lower incomes. We should consider changing the log transformation on the response. Furthermore, we see that there is greater variance in the female incomes. This suggests a modification to the model. We can make the same plot broken down by subject although there will be rather too many plots to be useful.

Plots of residuals and fitted values are also valuable. We have broken education into three levels: less than high school, high school or more than high school:

```
diagd$edulevel <- cut(psid$educ,c(0,8.5,12.5,20), labels=c("lessHS","
    ↪ HS","moreHS"))
ggplot(diagd, aes(x=.fitted,y=.resid)) + geom_point(alpha=0.3) + geom_
    ↪ hline(yintercept=0) + facet_grid(~ edulevel) + xlab("Fitted") +
    ↪  ylab("Residuals")
```

See Figure 11.5. Again, we can see evidence that a different response transformation should be considered. Plots of the random effects would also be useful here.

11.2 Repeated Measures

The acuity of vision for seven subjects was tested. The response is the lag in milliseconds between a light flash and a response in the cortex of the eye. Each eye is tested at four different powers of lens. An object at the distance of the second number appears to be at distance of the first number. The data is given in Table 11.1. The data comes from Crowder and Hand (1990) and was also analyzed by Lindsey (1999).

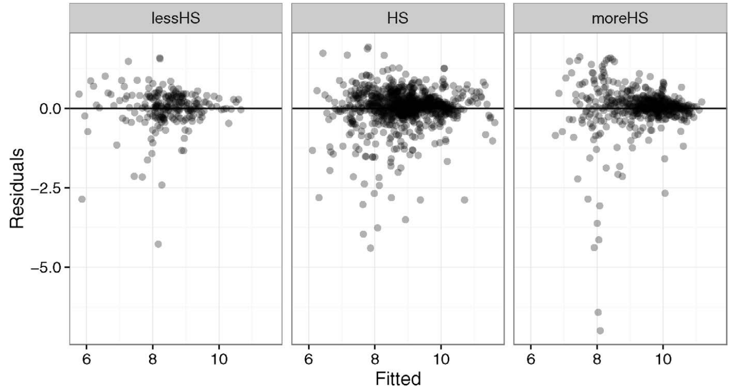

Figure 11.5 *Residuals vs. fitted plots for three levels of education: less than high school on the left, high school in the middle and more than high school on the right.*

Power							
6/6	6/18	6/36	6/60	6/6	6/18	6/36	6/60
Left				Right			
116	119	116	124	120	117	114	122
110	110	114	115	106	112	110	110
117	118	120	120	120	120	120	124
112	116	115	113	115	116	116	119
113	114	114	118	114	117	116	112
119	115	94	116	100	99	94	97
110	110	105	118	105	105	115	115

Table 11.1 *Visual acuity of seven subjects measured in milliseconds of lag in responding to a light flash. The power of the lens causes an object six feet in distance to appear at a distance of 6, 18, 36 or 60 feet.*

We start by making some plots of the data. We create a numerical variable representing the power to complement the existing factor so that we can see how the acuity changes with increasing power:

```
data(vision, package="faraway")
vision$npower <- rep(1:4,14)
ggplot(vision, aes(y=acuity, x=npower, linetype=eye)) + geom_line() +
    ↪ facet_wrap(~ subject, ncol=4) + scale_x_continuous("Power",
    ↪ breaks=1:4,labels=c("6/6","6/18","6/36","6/60"))
```

See Figure 11.6. There is no apparent trend or difference between right and left eyes.

However, individual #6 appears anomalous with a large difference between the eyes. It also seems likely that the third measurement on the left eye is in error for this individual.

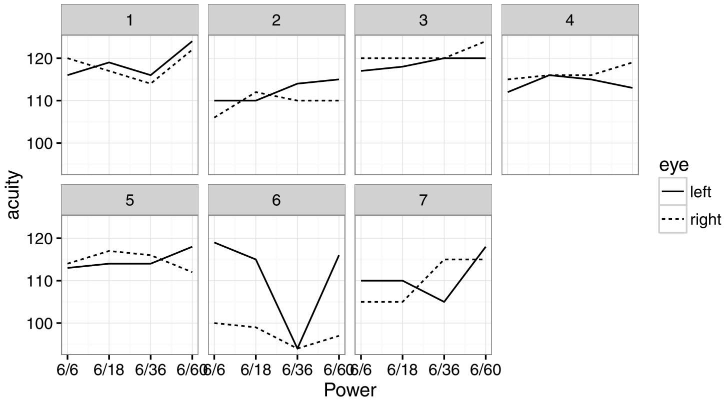

Figure 11.6 *Visual acuity profiles. The left eye is shown as a solid line and the right as a dashed line. The four powers of lens displayed are 6/6, 6/18, 6/36 and 6/60.*

We must now decide how to model the data. The power is a fixed effect. In the model below, we have treated it as a nominal factor, but we could try fitting it in a quantitative manner. The subjects should be treated as random effects. Since we do not believe there is any consistent right-left eye difference between individuals, we should treat the eye factor as nested within subjects. We start with this model:

```
mmod <- lmer(acuity~power + (1|subject) + (1|subject:eye),vision)
```

Note that if we did believe there was a consistent left vs. right eye effect, we would have used a fixed effect, putting `eye` in place of `(1|subject:eye)`.

We can write this (nested) model as:

$$y_{ijk} = \mu + p_j + s_i + e_{ik} + \varepsilon_{ijk}$$

where $i = 1,\ldots,7$ runs over individuals, $j = 1,\ldots,4$ runs over power and $k = 1,2$ runs over eyes. The p_j term is a fixed effect, but the remaining terms are random. Let $s_i \sim N(0,\sigma_s^2)$, $e_{ik} \sim N(0,\sigma_e^2)$ and $\varepsilon_{ijk} \sim N(0,\sigma^2\Sigma)$ where we take $\Sigma = I$. The summary output is:

```
sumary(mmod)
Fixed Effects:
            coef.est coef.se
(Intercept) 112.64     2.23
power6/18     0.79     1.54
power6/36    -1.00     1.54
power6/60     3.29     1.54
```



```
Random Effects:
 Groups      Name        Std.Dev.
 subject:eye (Intercept) 3.21
 subject     (Intercept) 4.64
 Residual                4.07
---
number of obs: 56, groups: subject:eye, 14; subject, 7
AIC = 342.7, DIC = 349.6
deviance = 339.2
```

We see that the estimated standard deviation for subjects is 4.64 and that for eyes for a given subject is 3.21. The residual standard deviation is 4.07. The random effects structure we have used here induces a correlation between measurements on the same subject and another between measurements on the same eye. We can compute these two correlations, respectively, as:

```
4.64^2/(4.64^2+3.21^2+4.07^2)
[1] 0.44484
(4.64^2+3.21^2)/(4.64^2+3.21^2+4.07^2)
[1] 0.65774
```

As we might expect, there is a stronger correlation between observations on the same eye than between the left and right eyes of the same individual.

We can check for a power effect using a Kenward-Roger adjusted F-test:

```
library(pbkrtest)
mmod <- lmer(acuity~power+(1|subject)+(1|subject:eye),vision,REML=
    ↪ FALSE)
nmod <- lmer(acuity~1+(1|subject)+(1|subject:eye),vision,REML=FALSE)
KRmodcomp(mmod, nmod)
F-test with Kenward-Roger approximation; computing time: 0.16 sec.
large : acuity ~ power + (1 | subject) + (1 | subject:eye)
small : acuity ~ 1 + (1 | subject) + (1 | subject:eye)
       stat   ndf   ddf F.scaling p.value
Ftest  2.83  3.00 39.00         1   0.051
```

We see the result is just above the 5% level. We might expect some trend in acuity with power, but the estimated effects do not fit with this trend. While acuity is greatest at the highest power, 6/60, it is lowest for the second highest power, 6/36. A look at the data makes one suspect the measurement made on the left eye of the sixth subject at this power. If we omit this observation and refit the model, we find:

```
mmodr <- lmer(acuity~power+(1|subject)+(1|subject:eye),vision,REML=
    ↪ FALSE, subset=-43)
nmodr <- lmer(acuity~1+(1|subject)+(1|subject:eye),vision,REML=FALSE,
    ↪ subset=-43)
KRmodcomp(mmodr, nmodr)
F-test with Kenward-Roger approximation; computing time: 0.15 sec.
large : acuity ~ power + (1 | subject) + (1 | subject:eye)
small : acuity ~ 1 + (1 | subject) + (1 | subject:eye)
      stat  ndf  ddf F.scaling p.value
Ftest  3.6  3.0 38.0         1   0.022
```

Now the power effect is significant, but it appears this is due to an effect at the highest power only. We can check that the highest power has a higher acuity than the average of the first three levels by using Helmert contrasts:

```
op <- options(contrasts=c("contr.helmert", "contr.poly"))
mmodr <- lmer(acuity~power+(1|subject)+(1|subject:eye),vision,subset
    ↪ =-43)
```

```
sumary(mmodr)
Fixed Effects:
            coef.est coef.se
(Intercept) 113.79     1.76
power1        0.39     0.54
power2        0.04     0.32
power3        0.71     0.22
```

By looking at the standard errors relative to the effect sizes, we can see that only the third contrast is of significance. We remember to reset the contrasts back to the default or subsequent output will be surprising:

```
options(op)
```

The Helmert contrast matrix is

```
contr.helmert(4)
  [,1] [,2] [,3]
1   -1   -1   -1
2    1   -1   -1
3    0    2   -1
4    0    0    3
```

We can see that the third contrast (column) represents the difference between the average of the first three levels and the fourth level, scaled by a factor of three. In the output, we can see that this is significant while the other two contrasts are not.

We finish with some diagnostic plots. The residuals and fitted values and the QQ plot of random effects for the eyes are shown in Figure 11.7:

```
plot(resid(mmodr) ~ fitted(mmodr),xlab="Fitted",ylab="Residuals")
abline(h=0)
qqnorm(ranef(mmodr)$"subject:eye"[[1]],main="")
```

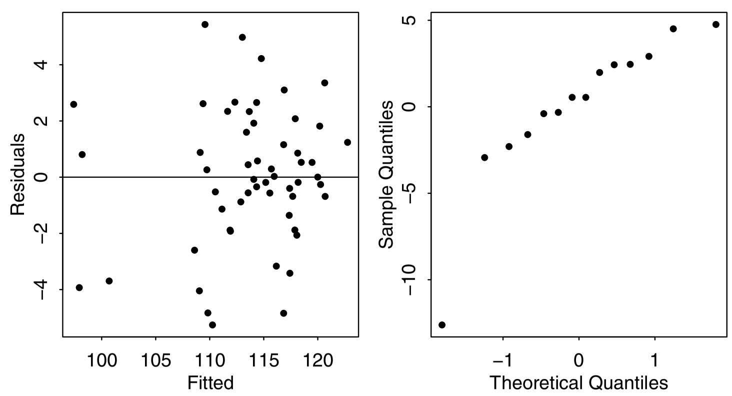

Figure 11.7 *Residuals vs. fitted plot is shown on the left and a QQ plot of the random effects for the eyes is shown on the right.*

The outlier corresponds to the right eye of subject #6. For further analysis, we should consider dropping subject #6. There are only seven subjects altogether, so we would