

2

Examples

This chapter introduces the longitudinal sets of data which will be used throughout the book. The rat data are presented in Section 2.1. The TDO data, studying toenails, are described in Section 2.2. Section 2.3 is devoted to the Baltimore Longitudinal Study of Aging, with two substudies: prostate-specific antigen data (Section 2.3.1) and data on hearing (Section 2.3.2). Section 2.4 introduces the Vorozole study, focusing on quality of life in breast cancer patients. In Section 2.5, we will introduce data, previously analyzed by Goldstein (1979), on the heights of 20 schoolgirls. Section 2.6 presents the growth data of Potthoff and Roy (1964). Mastitis in dairy cattle is the subject of Section 2.7.

To complement the data introduced in this chapter, five case studies, involving additional sets of data, are presented in Chapter 24.

2.1 The Rat Data

In medical science, there has recently been increased interest in the therapeutic use of hormones. However, such drastic therapies require detailed knowledge about their effect on the different aspects of growth. To this respect, an experiment has been set up at the Department of Orthodontics of the Catholic University of Leuven (KUL) in Belgium (see Verdonck *et al.* 1998). The primary aim was to investigate the effect of the inhibition

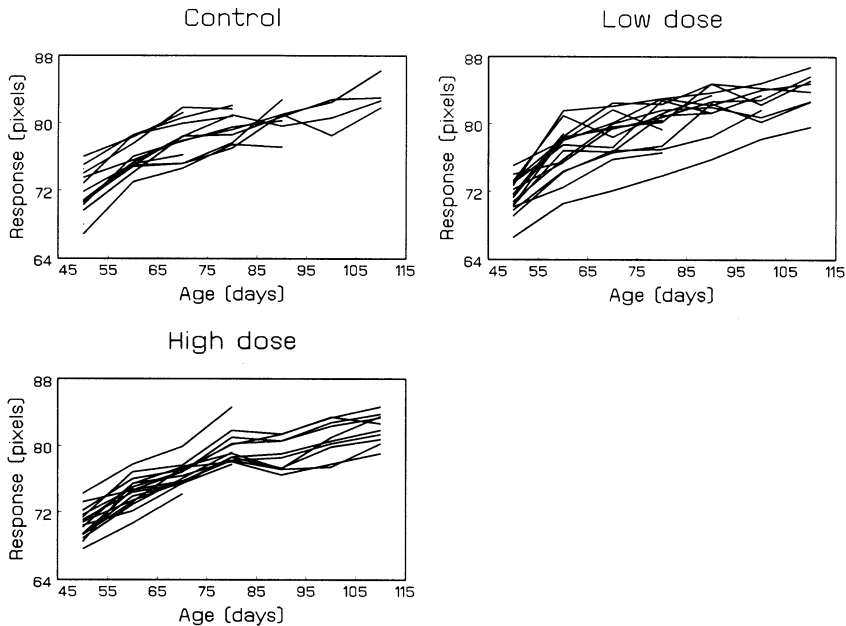


FIGURE 2.1. *Rat Data. Individual profiles for each of the treatment groups in the rat experiment separately.*

of the production of testosterone in male Wistar rats on their craniofacial growth.

A total of 50 male Wistar rats have been randomized to either a control group or one of the two treatment groups where treatment consisted of a low or high dose of the drug Decapeptyl, which is an inhibitor for testosterone production in rats. The treatment started at the age of 45 days, and measurements were taken every 10 days, with the first observation taken at the age of 50 days. The responses of interest are distances (in pixels) between well-defined points on X-ray pictures of the skull of each rat, taken after the rat has been anesthetized. Of primary interest is the estimation of changes over time and testing whether these changes are treatment dependent.

For the purpose of this book, we will consider one of the measurements which can be used to characterize the height of the skull. The individual profiles are shown in Figure 2.1. It is clear that not all rats have measurements up to the age of 110 days. This is due to the fact that many rats do not survive anaesthesia and therefore drop out before the end of the study. Table 2.1 shows the number of rats observed at each occasion. While 50 rats have been randomized at the start of the experiment, only 22 of them survived the 6 first measurements, so measurements on only 22 rats are available in the way anticipated at the design stage. For example, at the

TABLE 2.1. *Rat Data. Summary of the number of observations taken at each occasion in the rat experiment, for each group separately and in total.*

Age (days)	# Observations			
	Control	Low	High	Total
50	15	18	17	50
60	13	17	16	46
70	13	15	15	43
80	10	15	13	38
90	7	12	10	29
100	4	10	10	24
110	4	8	10	22

second occasion (age = 60 days), only 46 rats were available, implying that for 4 rats only 1 measurement could be recorded.

2.2 The Toenail Data (TDO)

The data introduced in this section were obtained from a randomized, double-blind, parallel group, multicenter study for the comparison of two oral treatments (in the sequel coded as *A* and *B*) for toenail dermatophyte onychomycosis (TDO), described in full detail by De Backer *et al.* (1996). TDO is a common toenail infection, difficult to treat, affecting more than 2 out of 100 persons (Roberts 1992). Antifungal compounds, classically used for treatment of TDO, need to be taken until the whole nail has grown out healthy. The development of new such compounds, however, has reduced the treatment duration to 3 months. The aim of the present study was to compare the efficacy and safety of 12 weeks of continuous therapy with treatment *A* or with treatment *B*.

In total, 2×189 patients were randomized, distributed over 36 centers. Subjects were followed during 12 weeks (3 months) of treatment and followed further, up to a total of 48 weeks (12 months). Measurements were taken at baseline, every month during treatment, and every 3 months afterward, resulting in a maximum of seven measurements per subject. For our purposes, we will only consider one of the secondary endpoints, unaffected nail length, which is measured as follows. At the first occasion, the treating physician indicates one of the affected toenails as the target nail, the nail which will be followed over time. At each occasion, the unaffected nail length (measured from the nail bed to the infected part of the nail,

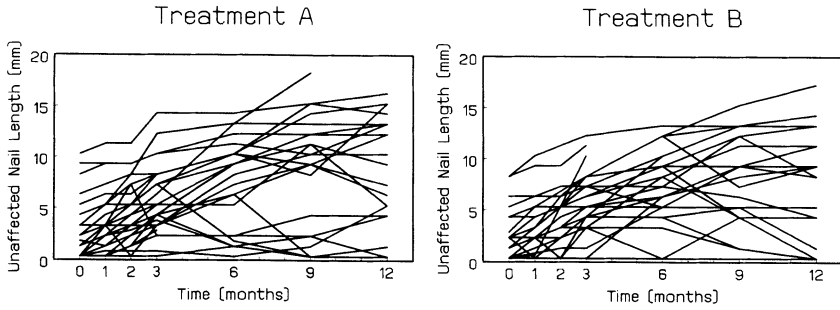


FIGURE 2.2. *Toenail Data. Individual profiles of 30 randomly selected subjects in each of the treatment groups in the toenail experiment.*

which is always at the free end of the nail) of the target nail is measured in millimeters. Obviously, this response will be related to the toe size. Therefore, we will only include here those patients for which the target nail was one of the two big toenails. This reduces our sample under consideration to 150 and 148 subjects, respectively. Figure 2.2 shows the observed profiles of 30 randomly selected subjects from treatment group *A* and treatment group *B*, respectively.

Due to a variety of reasons, 72 (24%) out of the 298 participants left the study prematurely. Table 2.2 summarizes the number of subjects still in the study at each occasion, for both treatment groups separately. Although the comparison of the average evolutions in both treatment groups was of primary interest, there was also some interest in studying the relationship between the dropout process and the actual outcome. For example, are patients who drop out doing better or worse than patients who do not drop out from the study ?

2.3 The Baltimore Longitudinal Study of Aging (BLSA)

The Baltimore Longitudinal Study of Aging (BLSA) is an ongoing multi-disciplinary observational study, which started in 1958, and with the study of normal human aging as primary objective (Shock *et al.* 1984). Participants in the BLSA are volunteers who return approximately every 2 years for 3 days of biomedical and psychological examinations. They are predominantly white (95%), well educated (over 75% have bachelor's degrees), and financially comfortable (82%). So far, over 1400 men with an average of almost 7 visits and 16 years of follow-up have participated in the study since its inception in 1958. Later on, females have been included in the study as well.

TABLE 2.2. *Toenail Data. Summary of the number of observations taken at each occasion in the TDO study, for each group separately and in total.*

Time (months)	# Observations		
	Treatment A	Treatment B	Total
0	150	148	298
1	149	142	291
2	146	138	284
3	140	131	271
6	131	124	255
9	120	109	229
12	118	108	226

The BLSA (Pearson *et al.* 1994) is a unique resource for rapidly evaluating longitudinal hypotheses because of the availability of data from repeated clinical examinations and a bank of frozen blood samples from the same individuals over 30 years of follow-up (where new studies would require many years to conduct). On the other hand, the observational aspect of the study poses additional complications on the statistical analysis. For example, although repeated visits are scheduled every 2 years, some subjects may have more than one visit within 1 year of time, while others have over 10 years between two successive visits. Also, longitudinal evolutions may be highly influenced by many covariates which may or may not be recorded in the study.

In this book, two of the many responses measured in the BLSA will be used to illustrate the statistical methodology. In Section 2.3.1, it will be discussed how data from the BLSA can be used to study the natural history of prostate disease. Afterward, in Section 2.3.2, the hearing data will be presented.

2.3.1 *The Prostate Data*

During the last 10 years, many papers have been published on the natural history of prostate disease; see, for example, Carter *et al.* (1992a, 1992b) and Pearson *et al.* (1991, 1994). According to Carter and Coffey (1990), prostate disease is one of the most common and most costly medical problems in the United States, and prostate cancer has become the second leading cause of male cancer deaths. It is therefore very important to look for markers which can detect the disease at an early stage. The prostate-specific antigen (PSA) is such a marker. PSA is an enzyme produced by

TABLE 2.3. *Prostate Data. Description of subjects included in the prostate data set, by diagnostic group. The cancer cases are subdivided into local/regional (L/R) and metastatic (M) cancer cases.*

	Controls	BPH cases	Cancer Cases	
			L/R	M
Number of participants	16	20	14	4
Age at diagnosis (years)				
Median	66	75.9	73.8	72.1
Range	56.7-80.5	64.6-86.7	63.6-85.4	62.7-82.8
Years of follow-up				
Median	15.1	14.3	17.2	17.4
Range	9.4-16.8	6.9-24.1	10.6-24.9	10-25.3
Time between measurements (years)				
Median	2	2	1.7	1.7
Range	1.1-11.7	0.9-8.3	0.9-10.8	0.9-4.8
Number of measurements per individual				
Median	8	8	11	9.5
Range	4-10	5-11	7-15	7-12

both normal and cancerous prostate cells, and its level is related to the volume of prostate tissue. Still, an elevated PSA level is not necessarily an indicator of prostate cancer because patients with benign prostatic hyperplasia (BPH) also have an enlarged volume of prostate tissue and therefore also an increased PSA level. This overlap of the distribution of PSA values in patients with prostate cancer and BPH has limited the usefulness of a single PSA value as a screening tool since, according to Pearson *et al.* (1991), up to 60% of BPH patients may be falsely identified as potential cancer cases based on a single PSA value.

Based on clinical practice, researchers have hypothesized that the rate of change in PSA level might be a more accurate method of detecting prostate cancer in the early stages of the disease. This has been extensively investigated by Pearson *et al.* (1994), who analyzed repeated PSA measures from the Baltimore Longitudinal Study of Aging (BLSA), using linear mixed models.

A retrospective case-control study was undertaken that utilized frozen serum samples from 18 BLSA participants identified as prostate cancer cases, 20 cases of BPH, and 16 controls with no clinical signs of prostate disease. In order to be eligible for the analyses, men had to meet several criteria:

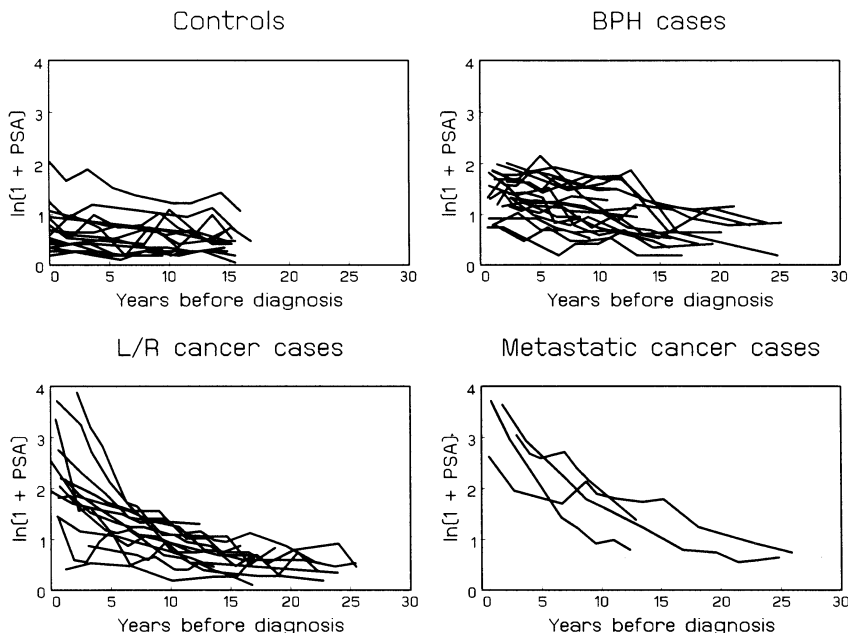


FIGURE 2.3. *Prostate Data. Longitudinal trends in PSA in men with prostate cancer, benign prostatic hyperplasia, or no evidence of prostate disease.*

1. seven or more years of follow-up prior to diagnosis of prostate cancer, simple prostatectomy for BPH, or exclusion of prostate disease by a urologist,
2. confirmation of the pathological diagnosis, and
3. no prostate surgery prior to diagnosis.

To the extent possible, age at diagnosis and years of follow-up were matched for the control, BPH, and cancer groups. However, due to the high prevalence of BPH in men over age 50, it was difficult to find age-matched controls with no evidence of prostate disease. In fact, the control group remained significantly younger at first visit and at diagnosis, compared to the BPH group. For this reason, our analyses of this data set will always correct for age differences at the time of the diagnosis.

A description of the data, differentiating between local/regional (L/R) cancer cases and metastatic cancer cases, is given in Table 2.3. The number of repeated PSA measurements per individual varies between 4 and 15, and the follow-up period ranges from 6.9 to 25.3 years. Since it was anticipated that PSA values would increase exponentially in prostate cancer cases, the responses were transformed to $\ln(\text{PSA} + 1)$. These transformed individual profiles are shown in Figure 2.3.

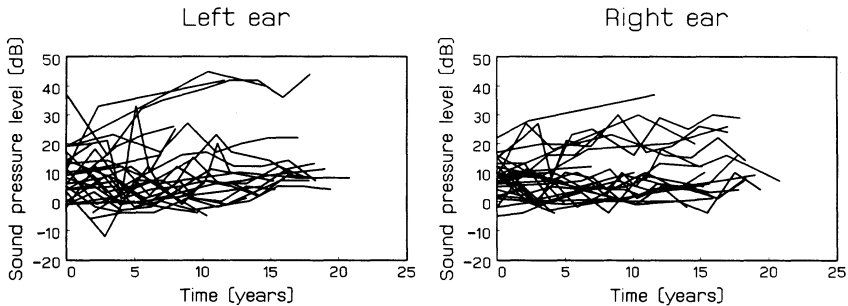


FIGURE 2.4. *Hearing Data. Individual profiles of 30 randomly selected subjects in the hearing data set, for the left and the right ear separately.*

2.3.2 The Hearing Data

Also recorded in the BLSA study are hearing threshold sound pressure levels (SPLs in dB), measured at 11 different frequencies [varying from 125 to 8000 hertz (Hz)] on both ears, yielding a maximum of 22 observations per visit. This was done by means of a sound proof chamber and a Bekesy audiometer. Using these data, Brant and Fozard (1990) have shown that the relationship between hearing threshold level and frequency can be well described by a quadratic function of the logarithm of frequency, the parameters of which depend on age and are highly subject-specific. Morrell and Brant (1991) and Brant and Pearson (1994) considered the data of 268 elderly male participants whose first visit occurred at about 70 years of age or older. They studied how hearing thresholds change over time and how these evolutions depend on age and on the frequency under consideration.

For our purposes, we now consider all available hearing thresholds for 500 Hz, from male BLSA participants only, without otologic disease, unilateral hearing loss, or evidence of noise-induced hearing loss. Individual profiles on the left and right ear separately are shown in Figure 2.4 for 30 randomly selected subjects.

In total, we have 6170 observations (3089 on the left ear and 3081 on the right ear), from 681 males. Their age at the first visit ranged from 17.2 to 90.5 years, with median value equal to 53 years. The number of visits per subject varied from 1 to 15, and some of the participants were followed for over 22 years (median 7.5 years).

TABLE 2.4. *Heights of Schoolgirls. Classification of 20 preadolescent school girls in three groups, according to their mother's height*

	Mothers height	Children numbers
Small mothers	< 155 cm	1 → 6
Medium mothers	[155cm; 164cm]	7 → 13
Tall mothers	> 164 cm	14 → 20

2.4 The Vorozole Study

This study was an open-label, multicenter, parallel group design conducted at 67 North American centers. Patients were randomized to either the new drug Vorozole (2.5 mg taken once daily) or the standard drug megestrol acetate (40 mg four times daily). The patient population consisted of post-menopausal patients with histologically confirmed estrogen-receptor positive metastatic breast carcinoma. All 452 randomized patients were followed until disease progression or death. The main objective was to compare the treatment groups with respect to response rate, whereas secondary objectives included a comparison relative to duration of response, time to progression, survival, safety, pain relief, performance status, and quality of life. Full details of this study are reported in Goss *et al.* (1999). In this book, we will focus on overall quality of life, measured by the total Functional Living Index: Cancer (FLIC; Schipper *et al.* 1984). Precisely, a higher FLIC score is the more desirable outcome. Even though this outcome is, strictly speaking, of the ordinal type, the total number of categories encountered exceeds 70, justifying the use of continuous-outcome methods.

Patients underwent screening and for those deemed eligible, a detailed examination at baseline (occasion 0) took place. Further measurement occasions were months 1, then from months 2 at bimonthly intervals until month 44.

Goss *et al.* (1999) analyzed FLIC using a two-way ANOVA model with effects for treatment, disease status, as well as their interaction. No significant difference was found. Apart from treatment, important covariates are dominant site of the disease as well as clinical stage.

This example will be used, for example, to introduce exploratory tools in Chapter 4.

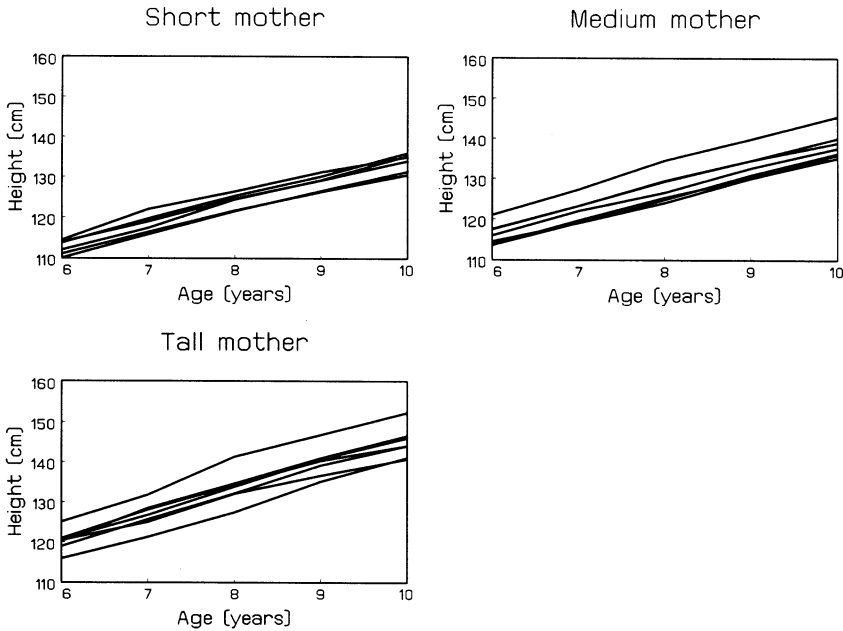


FIGURE 2.5. *Heights of Schoolgirls.* Growth curves of 20 school girls from age 6 to 10, for girls with small, medium, or tall mothers.

2.5 Heights of Schoolgirls

Goldstein (1979, Table 4.3, p. 101) reports growth curves of 20 preadolescent girls, measured on a yearly basis from age 6 to 10. The girls were classified according to the height of their mother, which was discretized as in Table 2.4. The individual profiles are shown in Figure 2.5, for each group separately. The measurements are given at exact years of age, some having been previously adjusted to these. The values Goldstein reports for the fifth girl in the first group are 114.5, 112, 126.4, 131.2, and 135.0. This suggests that the second measurement is incorrect. We therefore replaced it by 122. An extensive analysis of this data set can be found in Section 4.2 of Verbeke and Molenberghs (1997). Of primary interest is to test whether the growth of these schoolgirls is related to the height of their mothers.

2.6 Growth Data

These data, introduced by Potthoff and Roy (1964), contain growth measurements for 11 girls and 16 boys. For each subject, the distance from the center of the pituitary to the maxillary fissure was recorded at ages 8,

TABLE 2.5. *Growth Data for 11 Girls and 16 Boys. Measurements marked with * were deleted by Little and Rubin (1987).*

Girl	Age (in years)				Boy	Age (in years)			
	8	10	12	14		8	10	12	14
1	21.0	20.0	21.5	23.0	1	26.0	25.0	29.0	31.0
2	21.0	21.5	24.0	25.5	2	21.5	22.5*	23.0	26.5
3	20.5	24.0*	24.5	26.0	3	23.0	22.5	24.0	27.5
4	23.5	24.5	25.0	26.5	4	25.5	27.5	26.5	27.0
5	21.5	23.0	22.5	23.5	5	20.0	23.5*	22.5	26.0
6	20.0	21.0*	21.0	22.5	6	24.5	25.5	27.0	28.5
7	21.5	22.5	23.0	25.0	7	22.0	22.0	24.5	26.5
8	23.0	23.0	23.5	24.0	8	24.0	21.5	24.5	25.5
9	20.0	21.0*	22.0	21.5	9	23.0	20.5	31.0	26.0
10	16.5	19.0*	19.0	19.5	10	27.5	28.0	31.0	31.5
11	24.5	25.0	28.0	28.0	11	23.0	23.0	23.5	25.0
					12	21.5	23.5*	24.0	28.0
					13	17.0	24.5*	26.0	29.5
					14	22.5	25.5	25.5	26.0
					15	23.0	24.5	26.0	30.0
					16	22.0	21.5*	23.5	25.0

Source: Pothoff and Roy (1964), Jennrich and Schluchter (1986).

10, 12, and 14. The data were used by Jennrich and Schluchter (1986) to illustrate estimation methods for unbalanced data, where unbalancedness is now to be interpreted in the sense of an unequal number of boys and girls.

Little and Rubin (1987) deleted 9 of the $[(11 + 16) \times 4]$ measurements, rendering 9 incomplete subjects. Deletion is confined to the age 10 measurements. Little and Rubin (1987) describe the mechanism to be such that subjects with a low value at age 8 are more likely to have a missing value at age 10. The data are presented in Table 2.5. The measurements that were deleted are marked with an asterisk. In Section 17.4.1, the complete data will be analyzed in some detail. Sections 17.4.2 and 17.4.3 are devoted to frequentist and likelihood-based ignorable analyses of the incomplete version of the data, respectively. Section 17.4.4 is devoted to insight in the missingness mechanism.

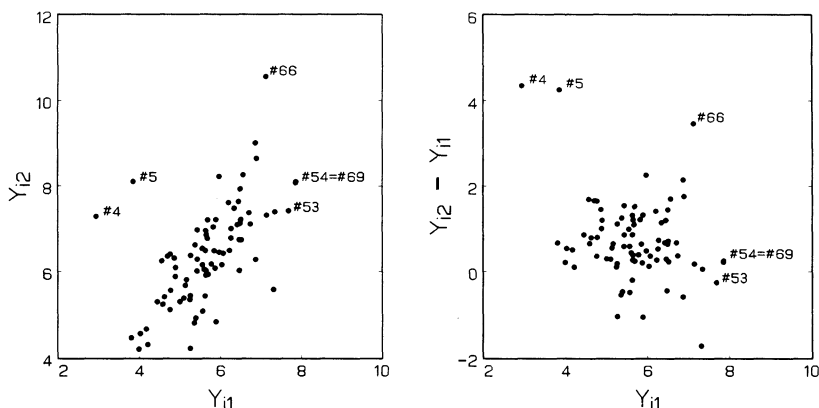


FIGURE 2.6. *Mastitis in Dairy Cattle.* The first panel shows a scatter plot of the second measurement versus the first measurement. The second panel shows a scatter plot of the change versus the baseline measurement.

2.7 Mastitis in Dairy Cattle

This example, concerning the occurrence of the infectious disease mastitis in dairy cows, was introduced in Diggle and Kenward (1994) and reanalyzed in Kenward (1998). Data were available of the milk yields in thousands of liters of 107 dairy cows from a single herd in 2 consecutive years: Y_{ij} ($i = 1, \dots, 107; j = 1, 2$). In the first year, all animals were supposedly free of mastitis; in the second year, 27 became infected. Mastitis typically reduces milk yield, and the question of scientific interest is whether the probability of occurrence of mastitis is related to the yield that would have been observed had mastitis not occurred. A graphical representation of the complete data is given in Figure 2.6.

<http://www.springer.com/978-1-4419-0299-3>

Linear Mixed Models for Longitudinal Data

Verbeke, G.; Molenberghs, G.

2000, XXII, 570 p. 128 illus., Softcover

ISBN: 978-1-4419-0299-3