4

Variance, Standard Deviation, Standard Error, Coefficient of Variation

Variance

Even the inbred animals maintained under well controlled animal house conditions may show some variations among the individuals in responding to a treatment in a pharmacology or toxicology study. Though majority of the individual animals respond to the treatment in a similar manner or magnitude, few of them will be too sensitive or resistant to the treatment. There are several factors that may affect the outcome of an animal experimentation, for example factors related to the experimenter. In a nutshell, even a well designed animal experimentation is bound to show some variations in the result and it is important to understand these variations for interpreting the experimental data. We shall work out an example, to make it very clear.

For a pharmacology experiment 5 rats are randomly picked up and placed them in a cage. As all the rats are of similar age and maintained in identical animal house conditions, one would assume that all the animals will have comparable body weight. The body weight of the rats is given in Table 4.1.

It is evident from the Table that the assumption of 'all animals having comparable body weight' is incorrect. In animal experiments, one can seldom get identical animals. There could be several differences (for example difference in water and feed consumption, difference in activity, difference in certain clinical chemistry parameters, etc.) among them. These differences have an important role in determining the outcome of an

Table 4.1. Body weight of rats (g)

Column 1	Column 2	Column 3	Column 4
Rat No.	**Body Weight** (X)	$(X-\overline{X})$	$(X-\overline{X})^2$
1	245	–7.6	57.76
2	254	+1.4	1.96
3	239	–13.6	184.96
4	266	+13.4	179.56
5	259	+6.4	40.96
Number of observations (n)	5	-	-
Sum (Σ)	1263	0	465.2
Mean $(\overline{X})$	252.6	-	-

animal experimentation. Let us try to find an estimate for these differences. In the example given in Table 4.1, the mean body weight is calculated as 252.6 g. Now, calculate the difference of each observation from the mean value $(X–\overline{X})$. A better statistical terminology for the difference is deviation, which is given in column 3 of Table 4.1. One may think that an estimate of the deviations can be obtained easily by summing up $(X–\overline{X})$. By doing so what you get is a zero. You cannot go further with this zero. When $(X–\overline{X})$ given in column 3 is closely examined, one would realize that the sum of the values bearing plus (+) sign is equal to the sum of the values bearing minus (–) sign. That is why a zero is obtained for the sum of $(X–\overline{X})$. This can be easily solved by squaring $(X–\overline{X})$. Squares of $(X–\overline{X})$ are given in column 4 of Table 4.1. Summing up $(X-\overline{X})^2$, a value 465.2, is called as sum of the squares (SS) of deviations is obtained. By dividing 465.2, i.e., the sum of the squares of deviations by n–1, a very important statistical parameter called 'variance' is derived.

Variance = 465.2/(5–1) = 116.3

One may ask why the SS is divided by 4 (n–1), instead of 5 (n). The denominator to calculate the variance is called as 'degrees of freedom'. Degrees of freedom is one less than the total number of observations. Let us try to explain this logically. Five different coloured boxes, say Black, Blue, Green, Red and Yellow are placed on a table. You have the 'freedom' to pick up the boxes in an unbiased manner, one by one. You may think that there are 5 boxes and the number of the 'freedoms' that you can exercise in picking up the boxes is also 5. You exercised your 'freedoms' to pick up the boxes as given in Table 4.2.

Table 4.2. Degrees of freedom exercised in picking up coloured boxes

Boxes picked up	Degrees of freedom exercised	Degrees of freedom left
Red	1	5–1 = 4
Yellow	2	5–2 = 3
Black	3	5–3 = 2
Green	4	5–4 = 1
Blue	This is the last box left out. You cannot exercise any degree of freedom for picking up this box.	

Initially, you thought that you would have had 5 degrees of freedom before picking up any box. Firstly, you picked up the red box and your degrees of freedom is reduced by 1 (5–1). The next time you picked up the yellow box, and now your degrees of freedom is reduced by 2 (5–2). When you picked up the black box, you have only 2 degrees of freedom left. After picking the green box, you have only 1 degree of freedom left. But you cannot exercise any freedom to pick up the blue box. Blue box is the last box left out and you have to pick up this without any choice. Therefore, the actual degrees of freedom that one can exercise is not equal to the total number of observations, but 1 less than the total number of observations.

Standard Deviation (SD)

Standard deviation is the square root of variation:

$$\mathrm{SD} = \sqrt{\mathrm{Variance}} = \sqrt{116.3} = \pm 10.78$$

A ± sign should always be added as a prefix to SD.

Some statisticians are of the opinion that the ± symbol is superfluous (Everett and Benos, 2004). According to them, a standard deviation is a single positive number, the notation of the SD should be: Mean (SD X), where X is the value for SD (for example, body weight of rats = 252.6 g (SD 10.78). We are in favor of prefixing a ± sign to SD as it gives an easily perceivable indication about the lowest and highest values of the sample observations.

Standard deviation is a useful measure to explain the distribution of the sample observations around the mean. SD can also be used to see whether a single observation falls within the normal range (Cumming, 2007). If the observations follow a normal distribution, mean ± 1 SD covers a range of 68% of the observations. About 95% of individuals will have values within 2 standard deviations of the mean (mean ± 2 SD), the other 5%

being equally scattered above and below these limits (Altman and Bland, 1995). Mean ± 3 SD covers a range of 99.7% of the observations.

Standard Error (SE)

SE is the SD of the mean. SE is considered as a measure of the precision of the sample mean (Altman and Bland, 2005). It provides an estimate of the uncertainty of the true value of the population mean (Everett, 2008). In simple words, SE measures the variation in the means of the samples. It can be calculated using the formula:

$$\mathrm{SE} = \mathrm{SD}/\sqrt{n} = 10.78/\sqrt{5} = \pm 4.82$$

Always prefix ± sign to SE.

Coefficient of Variation (CV)

CV is a numerical value where the proportion of the standard deviation in the mean value is shown as a percentage:

$$\mathrm{CV} = \frac{\mathrm{SD}}{\mathrm{Mean}} \times 100 = \frac{10.78}{252.6} \times 100 = 4.27\%$$

CV is an excellent statistical tool that can be used to compare different analytical methods and performance of equipments. Since CV is independent of the scale of measurement, it can be used to compare variables measured on different scales (Daniel, 2007). In a clinical chemistry laboratory, biochemists routinely use the commercially available reagent kits for analyzing clinical chemistry parameters in blood. It is difficult to choose from the plenty of kits available in the market. In such cases, kit with the lowest CV given in the packet insert should be chosen.

CV plays a very important role in determining the significant difference in pharmacology and toxicology experiments. Kobayashi *et al.* (2011) examined 59 parameters from 153 numbers of 28-day repeated dose administration studies conducted in 12 test facilities in order to understand the influence of CV in determining significant difference of quantitative values. CV of electrolytes was comparatively small, whereas enzymes had large CV. A significant difference between the sexes was observed in the CVs of feed consumption, reticulocyte, platelet and leucocyte counts, cholesterol, total protein, albumin, albumin/globulin ratio, alkaline phosphatase, inorganic phosphorus, and pituitary and adrenals weights.

Large differences in CV were observed for major parameters among 7 test facilities. The authors inferred that a statistically significant difference is usually detected if there is a difference of 7% in mean values between the groups and the groups have a CV of about 7%. A parameter with a CV as high as 30% in two groups can be significantly different from each other, if the difference between the two mean values of the groups is about 30% and the number of observation (n) in each group is 10. The authors suggested that it would be ideal to use median value to assess the treatment-related effect, rather than mean, when the CV is very high.

Matsuzawa *et al.* (1993) analyzed historical control data pertaining to clinical pathology of study population covering 14000 rats, 10000 dogs and 1400 monkeys. The authors stated that the serum assay values showed greater variation than the plasma values. Aoyama (2005) suggested that when the number of animals is adjusted, the decentralization of data, like body weight and the organ weight, become comparatively smaller, and a CV of about 10% is obtained. CV for blood levels of various hormones, even in control animals is large. Often, the standard deviation exceeds the mean value by more than 50% for these parameters.

There is a misconception that the variability in the experimental data occurs only in animal experiments. One may think that the instruments used in bioanalytical laboratories are highly sophisticated and automated, hence the results obtained from these instruments show minimum to no variation. This is not true. There is variability in analytical chemistry and the measured values differ from the actual values and 'if the variability of a measurement is not characterized and stated along with the result of the measurement, then the data can only be interpreted in a limited sense' (USP, 2008).

When to Use a Standard Deviation (SD)/Standard Error (SE)?

Pharmacologists and toxicologists ambiguously use SD and SE in their study reports. A confusion in the use of SD and SE is evident in scientific articles published in various journals (Herxheimer, 1988; Nagele, 2003).

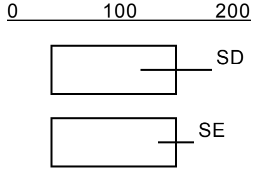

Figure 4.1. SD and SE calculated for human γ-GTP data[a]

[a]Data—42, 60, 26, 48, 56, 31, 30, 80, 79, 93 γ-GTP (IU/l)

Since SE is smaller than the SD (see Figure 4.1), some authors use SE, perhaps intentionally, in order to reduce the variability of their samples (Streiner, 1996; Lang, 1997; Fisher, 2000).

Although SD and SE are related, they give two very different types of information (Carlin and Doyle, 2000). In animal experiments, generally SD is 8–20% of the mean of the measured values, hence, the bar presented by the SD in a graph seems to be well balanced against the mean value. It is not permitted to use SE intentionally just to show a small width of the bar (Matsumoto, 1990). The next question is how precisely mean and SD should be specified? Mean should not be specified with more than one extra decimal place over the raw data but for SD greater precision can be given (Altman and Bland, 1996).

In conclusion, SD gives a fairly good indication about the distribution of the observed values around the mean. SE gives an indication about the variability of the mean. In toxicology experiments, especially with rodents, where the number of animals in a group is usually 10, it would be more ideal to use SD and in pharmacology experiments, where the number of animals in a group is usually <5 it would be more ideal to use SE, though there is no hard and fast rule for these.

References

Altman, D.G. and Bland, J.M. (1995): Statistics notes: The normal distribution. BMJ, 310, 298.

Altman, D.G. and Bland, J.M. (1996): Presentation of numerical data. BMJ, 312, 572.

Altman, D.G. and Bland, J.M. (2005): Standard deviations and standard errors. BMJ, 331, 903.

Aoyama, H. (2005): Applications and limitations of *in vivo* bioassays for detecting endocrine disrupting effects of chemicals on mammalian species of animals. J. Natl. Inst. Public Health, 54(1), 29–34.

Carlin, J.B. and Doyle, L.W. (2000): Basic concepts of statistical reasoning: standard errors and confidence intervals. J. Paediatr. Child Health, 36, 502–505.

Cumming, G. (2007): Error bars in experimental biology. JCB, 177(1), 7–11.

Daniel, W.W. (2007): Biostatistics-A Foundation of Analysis in the Health Sciences. 7th Edition, John Wiley & Sons (Asia) Pte. Ltd., Singapore.

Everett, D.C. (2008): Explorations in statistics: standard deviations and standard errors. Adv. Physiol. Educ., 32, 203–208.

Everett, D.C. and Benos, D.J. (2004): Guidelines for reporting statistics in journals published by the American Physiological Society. Adv. Physiol. Educ., 28, 85–87.

Fisher, D.M. (2000): Research Design and Statistics in Anesthesia. In: Anesthesia, 5th Edition, Vol. 1., Edited by Miller, R.D., Churchill Livingston, Philadelphia, USA.

Herxheimer, A. (1988): Misuse of standard error of the mean. Br. J. Clin. Pharmacol., 26, 197.

Kobayashi, K., Sakuratani, Y., Abe, T., Yamazaki, K., Nishikawa, S., Yamada, J., Hirose, A., Kamata, E. and Hayashi, M. (2011): Influence of coefficient of variation in determining significant difference of quantitative values obtained from 28-day repeated-dose toxicity studies in rats. J. Toxicol. Sci., 36(1), 63–71.

Lang, T.A.S.M. (1997): How to report statistics in medicine: annotated guidelines for authors, editors, and reviewers. American College of Physicians, Philadelphia, USA.

Matsuzawa, T., Nomura, M. and Unno, T. (1993): Clinical pathology reference ranges of laboratory animals. J. Vet. Med. Sci., 55(3), 351–362.

Matsumoto, K. (1990): Japanese Laboratory Animal Engineer Society, No. 6.

Nagele, P. (2003): Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. Br. J. Anaesthesiol., 90, 514–516.

Streiner, D.L. (1996): Maintaining standards: differences between the standard deviation and standard error, and when to use each. Can. J. Psychiatry, 41, 498–502.

USP (2008): The United States Pharmacopeia, The National Formularly, USP 31, NF 26, Asian Edition, Volume1, Port City Press, Baltimore, USA.

5

Analysis of Normality and Homogeneity of Variance

Distribution of Data in Toxicology and Pharmacology Experiments

It is important to know how the data are distributed for selecting a statistical tool for the analysis of the data (Bradlee, 1968). In toxicology and pharmacology experiments, data could be distributed in various patterns. The three commonly seen patterns of data distribution are given in Figure 5.1.

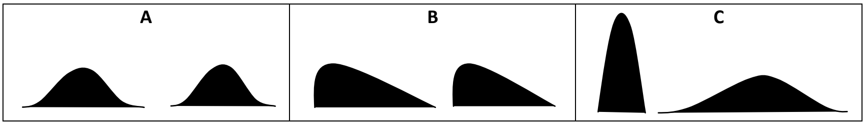

Figure 5.1. Three patterns of data distribution in toxicology and pharmacology experiments

A. Normal distribution and homogeneity of variance, B. Non-normal distribution and homogeneity of variance, C. Non-normal distribution in and heterogeneity in variance.

Analysis of Normality

The two types of non-normal distributions that are generally encountered in statistical analysis are skewness and kurtosis. The mean and median are different in a skewed distribution. Skewness can be positive or negative. The data are positively skewed, when the tail of the distribution curve is extended towards more positive values and the data are negatively skewed, when the tail of the distribution curve is extended towards more negative values (Čisar and Čisar, 2010).

Peakedness of a distribution is depicted by kurtosis. A distribution can be ‘platykurtic’ or ‘leptokurtic’. Platykurtic is more flat-topped and leptokurtic is less flat-topped. Usually platykurtic has long tails, whereas leptokurtic has short tails. In a leptokurtic distribution, the individual measures are concentrated near the mean, whereas in a platykurtic distribution, the individual measures are spread out across their range.

Most of the results obtained from toxicity studies do not follow a normal distribution. When Weil (1982) examined the distribution pattern of hematological and blood chemistry parameters of toxicological studies, skewness and kurtosis were observed in many cases. Kobayashi (2005) examined the measured items of a carcinogenicity/chronic toxicity study in rats. He reported that majority of hematological and biochemical parameters presented a non-normal distribution—mean corpuscular volume, mean corpuscular hemoglobin, platelets, protein, alanine aminotransferase, aspartate aminotransferase, gamma-glutamyl transpeptidase, creatinine phosphokinase, cholesterol and potassium were skewed positively, whereas hematocrit, hemoglobin, red blood cells and mean corpuscular hemoglobin concentration were negatively skewed.

Tests for Analyzing Normal Distribution

Several tests are available for analyzing normal distribution of the data, for example, Kolmogorov-Smirnov (Chakravarti *et al.*, 1967; Park, 2008), Lilliefors (1967), Shapiro-Wilk's *W* (Shapiro and Wilk, 1965) and Chi-distribution using goodness of fit tests (Snedecor and Cochran, 1989).

The Kolmogorov-Smirnov test is used to analyse continuous distributions. The Lilliefors test is a modified Kolmogorov-Smirnov test. The Shapiro-Wilk *W* test is capable of detecting non-normality for a wide variety of statistical distributions. Owing to this, a lot of attention has been paid to this test in the literature (Sen *et al.*, 2003). The power of Shapiro-Wilk's *W* test for detecting a non-normal distribution is better than other normality tests (Chen, 1971; Liang *et al.*, 2009). The chi-square test is an excellent test to examine whether the data are normally distributed. The major advantage of the chi-square test is that it can be applied to discrete distributions and its disadvantage is that it requires a larger sample size.

Shapiro-Wilk's W test

Let us understand Shapiro-Wilk's *W* test in detail by working out an example given in Table 5.1, body weight of F344 male rats. The data are arranged in an orderly fashion.

Table 5.1. Body weight of F344 male rats

Animal No.	1	2	3	4	5	6	7	8	9	10
Body weight (g)	71	86	92	95	100	102	105	108	118	123
Observation	1	1	2		4				1	1

The data in Table 5.1. is analysed using SAS-JMP and the statistics are given in Tables 5.2. and 5.3. The body weight distribution is given in Figure 5.2.

Table 5.2. Quantiles

100%	**Maximum**	**123.0**
99.5%		123.0
97.5%		123.0
90.0%		122.5
75.0%	**Quartile**	**110.5**
50.0%	**Median**	**101.0**
25.0%	**Quartile**	**90.5**
10.0%		72.5
2.5%		71.0
0.5%		71.0
0.0%	**Minimum**	**71.0**

Note: The term, quantile was introduced by Kendall (1940). Quantiles divide the distributions such that there is a given proportion of observations below the quantile. Quartiles and percentiles are quantiles. Quartile divides the quantile into four equal parts (0–25%, 25–50%, 50–75% and 75–100%). A percentile is the value of a variable below which a certain percent of observations fall. For example, the 10th percentile is that position in a data set which has 90% of data points above it, and 10% below it.

Table 5.3. Estimates

N	10
Sum (Σ)	1000
Mean ($\overline{X}$)	100
Standard error (SE)	4.7981478
Upper 95% mean	110.85416
Lower 95% mean	89.145836
Sum of squares (X-$\overline{X}$)^2	2072
Standard deviation (SD)	15.173076
Variance	230.22222
Coefficient of variation (CV)	15.173076
Skewness	–0.36285
Kurtosis	0.3549171

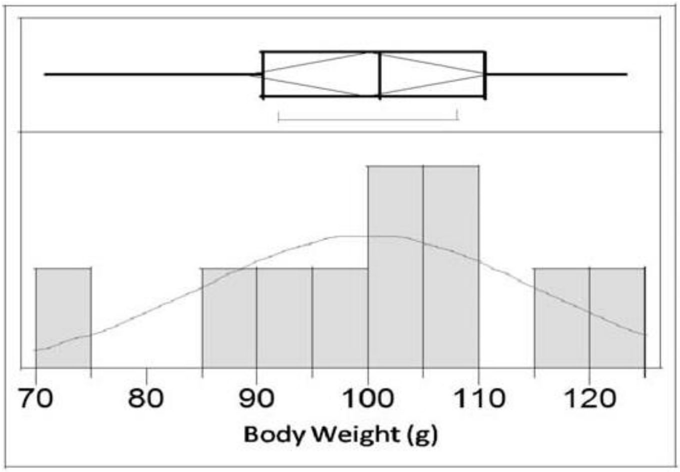

Figure 5.2. Body weight of F344 male rats

Shapiro-Wilk's W test-calculation steps

Step 1: Find the difference between the first set of extreme values (123 and 71 g from Table 5.1). Then find the difference between the second set of extreme values (118 and 86 g). In such a manner find the difference between the extreme values of remaining sets sequentially. If the number of samples is an odd number, ignore the remaining value.

Step 2: Find the Shapiro-Wilk *W* coefficients corresponding to the difference between the extreme values from the Appendix 1. In this example, the number of samples, N=10. The Shapiro-Wilk *W* coefficients corresponding to the difference between the 1st, 2nd, 3rd, 4th and 5th sets of extreme values are 0.5739, 0.3291, 0.2141, 0.1224 and 0.0399, respectively. Calculate the product of difference between extreme values and Shapiro-Wilk *W* coefficients (Table 5.4).

Step 3: Calculate the statistic, *W*, as given below:

$$W = \frac{45.10^2}{2072} = 0.98166$$

Compare *W* (0.98166) with the quantiles of the Shapiro-Wilk *W* test statistic given in Appendix 2. At 10 degrees of freedom the quantiles at 0.95 and 0.98 are 0.978 and 0.983, respectively. Since, the calculated *W* (0.98166) falls between 0.978 and 0.983, it could be concluded that the body weight of all the 10 animals follow a normal distribution. The same is confirmed by Test for goodness of fit:

Test for goodness of fit by Shapiro-Wilk test

W	Prob < *W*
0.981120	0.9673

Since the probability 0.9673<0.981120 (*W*), it is confirmed that the body weight of all the 10 animals follow a normal distribution pattern.

Table 5.4. Product of difference between extreme values and Shapiro-Wilk *W* coefficients

Animal No.	Body weight (g)	Difference between extreme values (D)		Shapiro-Wilk *W* coefficients (C)	Product (DxC)
1	71	First set	123–71=52	0.5739	29.8428
2	86	Second set	118–86=32	0.3291	10.5312
3	92	Third set	108–92=16	0.2141	3.4256
4	95	Fourth set	105–95=10	0.1224	1.2240
5	100	Fifth set	102–100=2	0.0399	0.0798
6	102	-	-	-	-
7	105	-	-	-	-
8	108	-	-	-	-
9	118	-	-	-	-
10	123	-	-	-	-
Sum					45.10

Power of Shapiro-Wilk's W test

Shapiro-Wilk's *W* test can be used in small as well as large sample sizes (Singh, 2009).

However, the power of this test varies with the number of animals in the group. This can be demonstrated with the help of an example of weight of rats on week 13, in a repeated dose administration study. Four situations are simulated in the example:

Situation 1 (Seventeen observations): 70, 80, 85, 90, 94, 99, 101, 102, 104, 105, 108, 111, 112, 114, 121, 125, and 131. The distribution of the observations is given in Figure 5.3a.

Statistics

Mean	103.05882
SD	16.009648
SE	3.8829099
Upper 95% mean	111.29022
Lower 95% mean	94.827422
N	17

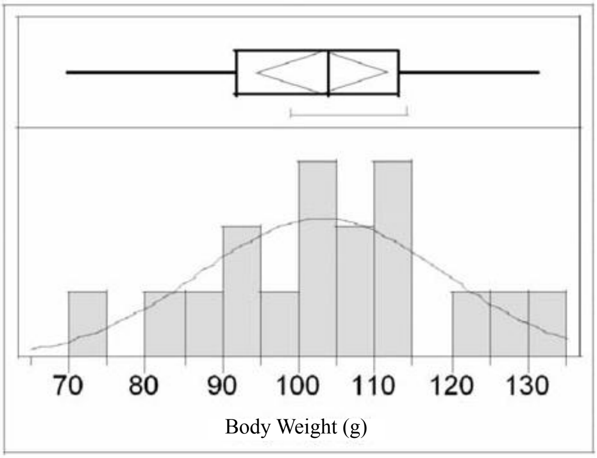

Figure 5.3a. Distribution pattern of body weight (g) of rats—17 observations

Shapiro-Wilk's *W* test

W	Prob <*W*
0.987278	0.9891

Situation 2 (Thirty four observations, the observations of situation 1 are used twice): 70, 80, 85, 90, 94, 99, 101, 102, 104, 105, 108, 111, 112, 114, 121, 125, 131, 70, 80, 85, 90, 94, 99, 101, 102, 104, 105, 108, 111, 112, 114, 121, 125, and 131. The distribution of the observations is given in Figure 5.3b.

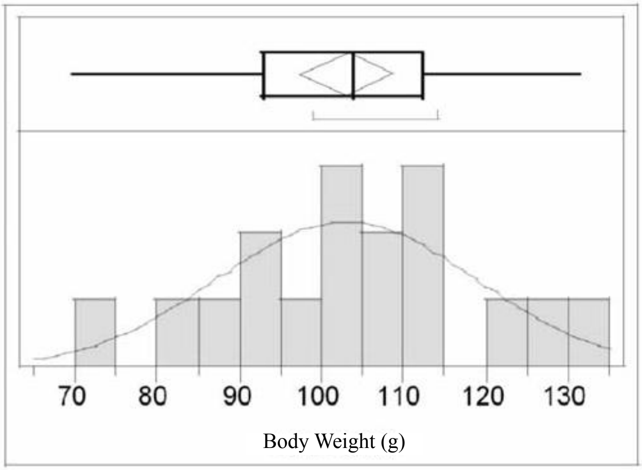

Figure 5.3b. Distribution pattern of body weight (g) of rats—34 observations

Statistics

Mean	103.05882
SD	15.765211
SE	2.7037114
Upper 95% mean	108.55957
Lower 95% mean	97.558081
N	34

Shapiro-Wilk's *W* test

W	Prob <*W*
0.968746	0.5017

Situation 3 (Fifty one observations, the observations of situation 1 are used thrice): 70, 80, 85, 90, 94, 99, 101, 102, 104, 105, 108, 111, 112, 114, 121, 125, 131, 70, 80, 85, 90, 94, 99, 101, 102, 104, 105, 108, 111, 112, 114, 121, 125, 131, 70, 80, 85, 90, 94, 99, 101, 102, 104, 105, 108, 111, 112, 114, 121, 125, and 131. The distribution of the observations is given in Figure 5.3c.

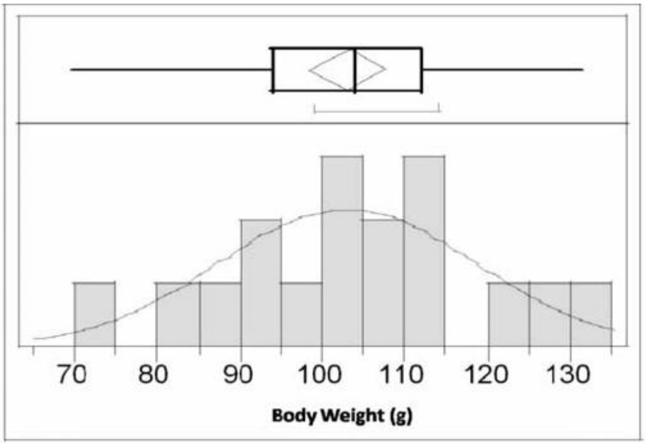

Figure 5.3c. Distribution pattern of body weight (g) of rats—51 observations

Statistics

Mean	103.05882
SD	15.686187
SE	2.1965056
Upper 95% mean	107.47063
Lower 95% mean	98.647012
N	51

Test for goodness of fit, Shapiro-Wilk's *W* test

W	Prob <*W*
0.959888	0.1486

Situation 4 (Sixty eight observations, the observations of situation 1 are used four times): 70, 80, 85, 90, 94, 99, 101, 102, 104, 105, 108, 111, 112, 114, 121, 125, 131, 70, 80, 85, 90, 94, 99, 101, 102, 104, 105, 108, 111, 112, 114, 121, 125, 131, 70, 80, 85, 90, 94, 99, 101, 102, 104, 105, 108, 111, 112, 114, 121, 125, 131, 70, 80, 85, 90, 94, 99, 101, 102, 104, 105, 108, 111, 112, 114, 121, 125, and 131. The distribution of the observations is given in Figure 5.3d.

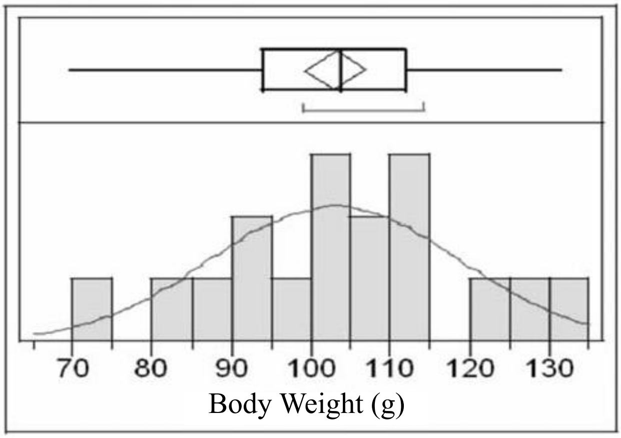

Figure 5.3d. Distribution pattern of body weight (g) of rats—68 observations

Statistics

Mean	103.05882
SD	15.647118
SE	1.8974918
Upper 95% mean	106.84623
Lower 95% mean	99.271414
N	68

Shapiro-Wilk's *W* test

W	Prob <*W*
0.954862	0.0383

The statistics given in Figure 5.3a–Figure 5.3d are consolidated in Table 5.5. Shapiro-Wilk's *W* test revealed a significant *P*, when the number of animals was 68, indicating a non-normal distribution.

Table 5.5. Change in power of Shapiro-Wilk's *W* test with the change in number of animals

N	Mean	Coefficient of variance (%)	Shapiro-Wilk's *W* test	
			W	*P*
17	103	15.5	0.987278	0.9891 (NS)
34		15.3	0.968746	0.5017 (NS)
51		15.2	0.959888	0.1486 (NS)
68		15.2	0.954862	**0.0383** (S)

NS-Not significant (normal distribution); S-Significant (non-normal distribution)

Parametric and Non-parametric Analyses

The two basic assumptions for any statistical analysis are the distribution of the data (normal or non-normal) and homogeneity of variance (homogeneous or heterogeneous). If the variances of the groups are heterogeneous and or the data are non-normally distributed, the choice of the statistical tools is non-parametric (Kobayashi *et al.*, 2011a). Non-parametric tests are also called as 'distribution-free tests'. A parametric test is always based on the assumption that the data follow a normal distribution and variances of the groups are homogeneous.

Analysis of Homogeneity of Variance

One of the assumptions of parametric analysis is that variances of the observations in the individual groups are equal (the other assumption is that the data are normally distributed). When the variances of the groups are equal, the situation is referred to as homogeneity of variance (also called as homoscedasticity of variance). When the variances of the groups are different (not homogeneous), the situation is called as heteroscedasticity.

Bartlett's homogeneity test

In most of the pharmacological and toxicological studies, Bartlett's test is commonly used to examine the data for homogeneity of variance (Bartlett, 1937). However, according to Finney (1995) "Bartlett's test is notorious for its unwanted sensitivity to non-normality of error distribution, and is an untrustworthy instrument for classifying some data sets as homogeneous in variance, other as heterogeneous."

Homogeneity of variance by Bartlett's test is calculated using the below given formula:

$$X^2 cal = 2.3026 \times \frac{\{(\text{Sum of N} - \text{Number of group}) \times \log V - \text{N of each group} - 1 \times \log \text{Sum of Variance}\ \}}{1 + \dfrac{\dfrac{1}{\text{Sum of (N of each group} - 1)} - \dfrac{1}{\text{Sum of total number} - \text{Number of group}}}{3 \times (\text{Numbe of group} - 3)}}$$

where,

$$V = \frac{(\text{Variance of each group} \times \text{Sum of } \ (N - 1))}{(\text{Sum of } N) - \text{Number of group}}$$

X^2 *cal* (chi square calculated) is compared with the value given in chi square Table (N=number of groups-1) at 5% probability level. If the computed value is less than the table value, it is interpreted that the variances of the groups are similar (no heterogeneity). It may be noted that Bartlett's test is not suitable for detecting a heterogeneity when the number of animals in a group very small.

Levene's homogeneity test

Another test used to examine the data for homogeneity of variance is Levene's test (Levene, 1960; Nichols, 1994), which has less sensitivity to non-normality of error distribution. Interestingly, compared to Bartlett's test, Levene's test is less commonly used to analyse the data obtained from toxicological and pharmacological experiments.

Power of Bartlett's and Levene's homogeneity tests

Bartlett's test is used for testing the homogeneity of variance of the data that follow a normal distribution. Bartlett's test is very sensitive to the data that are non-normal to the slightest extent. According to Finney (1995), Bartlett's test is not necessarily to be carried out for examining homogeneity of variance before ANOVA (Analysis of variance, an important statistical tool for comparing more than two groups; you will learn more about ANOVA in Chapter 11). The reason for this is that the power of the Bartlett's homogeneity test is too strong for examining homogeneity of variance, as mentioned above. Toxicity studies using Bartlett's test for testing homogeneity of variance at 1% probability level, which is not so conventional, have been reported (Hayashi *et al.*, 1994; Katsumi *et al.*, 1999; Kudo *et al.*, 2000; Mochizuki *et al.*, 2009; Ishii *et al.*, 2009). The reason for setting a 1% probability level for detecting a significant difference probably could be: if a significant difference is detected by Bartlett's test at the conventional 5% probability level, then the data should be analysed using the non-parametric Dunnett type rank sum test (joint type) (Yamazaki *et al.*, 1981) and/or Dunn test (Hollander and Wolfe, 1973), which have low detection power. Therefore, when the probability level is set at 1%, it is unlikely that the data show a heteroscedacity in variance by Bartlett's test. The reason for this is that to detect a significant difference at 1% probability level, the chi square value has to be larger than that of the 5% probability level.

Do We Need to Examine the Data for Both Normality and Homogeneity?

Kobayashi *et al.* (2011b) made an attempt to compare the statistical tools used to analyse the data of repeated dose administration studies with rodents conducted in 45 countries, with that of Japan. They found that the statistical techniques used for testing the above data for homogeneity of variance are similar in Japan and other countries. In most of the countries, including Japan, the data are generally not tested for normality.

Kobayashi *et al.* (2008; 2011b) suggested that the data may be examined for both homogeneity of variance and normal distribution. However, in bioequivalence clinical trials, because of the limited sample size a reliable determination of the distribution of the data set is not required (EMEA, 2006).

Which Test to be Used for Examining Homogeneity of Variance?

In pharmacological and toxicological experiments, treatments that lower mean values often decrease variance in the treated groups, substantially (Colquhoun, 1971). In these cases, statistical analyses based on the assumption of normal distribution and homogeneity of variance are inappropriate (Spector and Vesell, 2006).

Water consumption of B6C3F1 female mice during the week 13 of a repeated dose administration study is given in Table 5.6. There were four groups and each group consisted of 10 mice. Homogeneity of variances among the groups was analysed using Brown-Forsythe's (Brown and Forsythe, 1974), O'Brien's, Levene's and Bartlett's tests.

Table 5.6. Water consumption (g/week)of B6C3F1 female mice during the week 13 of a repeated dose administration study

Groups	N	Mean± S.D.	*P*			
			O'Brien	Brown-Foresythe	Levene	Bartlett
1	10	43.8 ± 9.0	0.0459	0.0340	0.0014	<0.0001
2	10	35.4 ± 3.4				
3	10	31.9 ± 1.5				
4	10	30.7 ± 2.1				

It is clear from the table that the sensitivity of Bartlett's test is higher, followed by Levene's test. O'Brien's and Brown-Forsythe's tests have very low sensitivity.

Brown-Forsythe's test is a modified Levene's test. Both Brown-Forsythe's and Levene's tests use transformed values (Maxwell and Delaney, 2004). It is more appropriate to use the Levene's, Brown-Forsythe's or O'Brien's tests (O'Brien, 1979; 1981) for testing the homogeneity of variance of the data that follow a non-normal distribution (SAS, 1996). Kobayashi *et al.* (1999) suggested Levene's test for examining homogeneity of variance of the data obtained from toxicity studies.

References

Bartlett, M.S. (1937): Properties of sufficiency and statistical tests. Proceedings of the Royal Statistical Society Series A, 160, 268–282.

Bradlee, J.V. (1968): Distribution-Free Statistical Tests. Prentice-Hall, Englewood Cliffs, New Jersey, USA.

Brown, M.B. and Forsythe, A.B. (1974): Robust tests for equality of variances. J. Am. Stat. Assoc., 69, 364–367.

Chakravarti, I.M, Laha, R.G. and Roy, J. (1967): Handbook of Methods of Applied Statistics, Volume I, John Wiley and Sons, New York, USA.

Chen, E.H. (1971): The power of Shapiro-Wilk *W* test for normality in samples from contaminated normal distribution. J. Am. Stat. Assoc., 66(336), 760–762.

Čisar, P. and Čisar, S.M. (2010): Skewness and kurtosis in function of selection of network traffic distribution. Acta Polytech. Hung., 7(2), 95–106.

Colquhoun, D. (1971): Lecture on Biostatistics. Clarendon Press, Oxford, UK.

EMEA (2006): European Medicines Agency. Biostatistical Methodology in Clinical Trials. ICH Topic E 9—Statistical Principles for Clinical Trials, CPMP/ICH/363/96, London, UK.

Finney, D.J. (1995): Thoughts suggested by a recent paper: Questions on non-parametric analysis of quantitative data (letter to editor). J. Toxicol. Sci., 20(2), 165–170.

Hayashi, T., Yada, H., Auletta, C.S., Daly, I.W., Knezevich, A.L. and Cockrell, B.Y. (1994): A six-month interperitoneal repeated dose toxicity study of tazobactam/piperacillin and tazobactam in rats. J. Toxicol. Sci., 19, Suppl. II, 155–176.

Hollander, M. and Wolf, D.A. (1973): Nonparametric Statistical Methods, John Wiley and Sons, New York, USA.

Ishii, S., Ube, M., Okada, M., Adachi, T., Sugimoto, J., Inoue, Y., Uno, Y. and Mutai, M. (2009): Collaborative work on evaluation of ovarian toxicity (17). J. Toxicol. Sci., 34, SP175–SP188.

Kendall, M.G. (1940): Note on the distribution of quantiles for large samples. Supp. J. Royal Stat. Soc., 7(1), 83–85.

Kobayashi, K. (2005): Analysis of quantitative data obtained from toxicity studies showing non-normal distribution. J. Toxicol. Sci., 30(2), 127–134.

Kobayashi, K., Kitajima, S., Miura, D., Inoue, H., Ohori, K., Takeuchi, H. and Takasaki, K. (1999): Characteristics of quantitative data obtained in toxicity rodents—The necessity of Bartlett's test for homogeneity of variance to introduce a rank test. J. Environ. Biol., 20, 3748.

Kobayashi, K., Pillai, K.S., Suzuki, M. and Wang, J. (2008): Do we need to examine the quantitative data obtained from toxicity studies for both normality and homogeneity of variance? J. Environ. Biol., 29, 47–52.

Kobayashi, K., Sadasivan Pillai, K., Soma Guhatakurta, Cherian, K.M., and Ohnishi, M. (2011b): Statistical tools for analysing the data obtained from repeated dose toxicity studies with rodents: A comparison of the statistical tools used in Japan with that of used in other countries. J. Environ. Biol., 32: 11–16.

Kobayashi, K., Sakuratani, Y., Abe, T., Yamazaki, K., Nishikawa, S., Yamada, J., Hirose, A., Kamata, E. and Hayashi, M. (2011a): Influence of coefficient of variation in determining significant difference of quantitative values obtained from 28-day repeated-dose toxicity studies in rats. J.Toxicol. Sci., 36(1), 63–71.

Kudo, S., Tanase, H., Yamasaki, M., Nakao, M., Miyata, Y., Tsuru, K. and Imai, S. (2000): Collaborative work to evaluate toxicity on male reproductive organs by repeated dose studies in rats (23). J. Toxicol. Sci., 25, SP223–SP232.

Levene, H. (1960): Robust tests for the equality of variances. In: Contributions to Probability and Statistics, Edited Olkin, I. Stanford University Press, USA.

Liang, J., Tang, M.L. and Chan, P.S. (2009): A generalized Shapiro-Wilk's *W* statistic for testing high dimensional normality. Comp. Stat. Data Anal., 53(11), 3883–3891.

Lilliefors, H. (1967): On the Kolmogorov–Smirnov test for normality with mean and variance unknown. J. Am. Stat. Assoc., 62, 399–402.

Maxwell, S.E. and Delaney, H.D. (2004): Designing Experiments and Analysing Data—A Model Comparison Perspective. 2nd Ed., Lawrence Erlbaum Associates, Inc., New Jersey, USA.

Mochizuki, M., Shimizu, S., Urasoko, Y., Umeshita, K., Kamata, T., Kitazawa, T. Nakamura, D., Nishihata, Y., Ohishi, T. and Edamoto, H. (2009): Carbon tetrachloride-induced hepatotoxicity in pregnant and lactating rats. J. Toxicol. Sci., 34(2), 175–181.

Nichols, D. (1994): Levene test, SPSS Inc., nichols@spss.com

O'Brien, R.G. (1979): A general ANOVA method for robust test of additive models for variance. J. American Stat. Asso., 74, 877–880.

O'Brien, R.G. (1981): A simple test for variance effects in experimental designs. Psych. Bull., 89, 570–574.

Park, H.M. (2008): Univariate analysis and normality test using SAS, Stata and SPSS. Univ. Inf. Tech. Serv., Centre Stat. Math. Comp., Indiana Univ., Bloomington, USA.

SAS (1996): JMP Start Statistics. SAS Institute, USA.

Sen, P.K., Jureckov, J. and Picek, J. (2003): Goodness-of-Fit Test of Shapiro-Wilk Type with Nuisance Regression and Scale. Aust. J. Stat., 32(1&2), 163–167.

Shapiro, S.S. and Wilk, M.B. (1965): An analysis of variance test for normality (complete samples). Biometrika, 52(3-4), 591–611.

Singh, K. (2009): Quantitative Social Research Methods. Sage Publication Pvt. Ltd., New Delhi, India.

Snedecor, G.W. and Cochran, W.G. (1989): Statistical Methods, 8th Edition, Iowa State University Press, Ames, USA.

Spector, R. and Vesell, E.S. (2006): Pharmacology and statistics: Recommendations to strengthen a productive partnership. Pharmacology, 78, 113–122.

Weil, C.S. (1982): Statistical analysis and normality of selected hematologic and clinical chemistry measurements used in toxicologic studies. Arch. Toxicol. Suppl., 5, 237–253.

Yamazaki, M., Noguchi, Y., Tanda, M. and Shintani, S. (1981): Statistical method appropriate for general toxicological studies in rats. J. Takeda Res. Lab., 40(3/4), 163–187.

6

Transformation of Data and Outliers

Transformation of Data

There are situations in pharmacological and toxicological experiments that the data show heterogeneous variance across the groups of animals. Using parametric tests to analyse such data may give rise to Type I error. One way to overcome this situation is to transform the data (Wallenstein *et al.,* 1980). It is most likely that the variance of the transformed data show homogeneity.

In Table 6.1, transformed values of alanine aminotransferase activity of Wistar rats of the control group in a 14-day repeated dose administration study is given.

Table 6.1. Alanine aminotransferase activity (U/L) of Wistar rats of the control group in a 14-day repeated dose administration study

45.3, 63.8, 82, 42, 40.8, 38.2, 35.9, 37.9, 39.1, 35.5 (N=10)		
Transformation	Mean±SD	CV (%)
None	46 ± 15	32.7
Logarithm	1.6 ± 0.12	7.2
Square root	6.7 ± 1.0	15.0
Reciprocal	0.02 ± 0.005	22.8

For the non-transformed data, the CV was 32.7%, which substantially decreased, when the data were transformed to logarithms. CV also decreased when the data were transformed to square roots and reciprocals, but in a lesser magnitude than the logarithmic-transformed data.

Concentrations of blood constituents usually show a non-normal distribution (Flynn *et al*., 1974). Therefore, statistical analysis is usually carried out with the transformed values of blood constituents (Niewczas *et al.*, 2009). According to Lew (2007), in pharmacology, the data may be transformed to their logarithms in order to eliminate heterogeneity in variation. For example, plasma/serum concentration of drug and/or its metabolites in drug metabolism and pharmacokinetic studies (DMPK) in laboratory animals (Girard *et al.*, 1992; Steinke *et al.*, 2000; Zheng *et al.*, 2010) and bioavailability/bioequivalence (BA/BE) studies in volunteers (Dubey *et al.*, 2009) are usually analysed in their logarithmic-transformed values. FDA (2003) and EMEA (2006) recommend logarithmic-transformation of exposure measures before statistical analysis in BA/BE studies. It should be borne in mind that the data showing a non-normal distribution may also display other patterns of uneven variation that cannot be easily eliminated (Keppel and Wickens, 2004).

Statistical analysis using transformed values are not the same as using measured values. Therefore, interpreting the transformed values may be difficult (Jenifer, 2010). In the words of Finney (1995), "When a scientist measures a quantity such as concentration of a chemical compound in a body fluid, his interest usually lies in the scale, perhaps mg/ml, that he has used; he is less likely to be interested in a summary of results relating to a transformed quantity such as the logarithm of blood concentration. If he analyzes in terms of logarithm, encouraged perhaps by an elementary but uncritical statistical textbook or by a convenient software package, he may find significant differences but to express his conclusions in meaningful numbers may be impossible. I do not assert that a scientist should never transform data before analysis; I urge that data should be transformed only after careful consideration of all consequences. Textbook implications that; 'In certain specified circumstance, data must be transformed' should not be unthinkingly accepted. Remember that any transformation is likely to increase the difficulty of interpreting results in relation to the original measurements." Therefore, when a significant difference is obtained for transformed values, following a statistical analysis, it is necessary to describe that the significant difference obtained is for the transformed values.

Outliers

Data obtained from pharmacological and toxicological studies are not free from outliers. An outlier can be defined as 'an observation which deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism' (Hawkins, 1980). Outliers can have deleterious effects on statistical analyses (Rasmussen, 1988; Schwager and Margolin, 1982). Outliers increase error rates and distort statistical estimates when using either parametric or nonparametric tests (Zimmerman, 1995; 1998). Outliers arise from two sources—from errors in the data and from the inherent variability of the data (Anscombe, 1960). According to Barnett and Lewis (1994), ‘not all outliers are illegitimate contaminants, and not all illegitimate scores show up as outliers’.

Hypoglycemic property of a drug was evaluated in alloxan-induced hyperglycemic rats. These rats were divided into two groups (5 rats/group), Groups 1 and 2. Group 1 (control) was treated with vehicle and Group 2 was treated with the drug. Following the administration of vehicle or drug, blood glucose was determined in individual rat (Table 6.2.).

Table 6.2. Blood glucose (mg/dl) in alloxan-treated rats following administration of drug

Group 1 (Vehicle treated)	Group 2 (Drug treated)
189, 195, 169, 206, 175	138, 161, 156, 171, ***259***
Mean ± SD = 186.8 ± 14.9 (n=5)	Mean ± SD = 177.0 ± 47.4 (n=5)
	Mean ± SD = 156.5 ± 13.4 (n=4)

The blood glucose level of the vehicle treated group was 186.8 ± 14.9 mg/dl (mean ± SD), whereas the drug treated group was 177.0 ± 47.4 mg/dl (mean ± SD). Though a decrease in blood glucose level was observed in the drug treated animals, it was statistically insignificant by Aspin Welch’s *t*-test using one-sided (we used Aspin Welch’s *t*-test because the variance of the groups is different. You will read more about this test in Chapter 8). The SD of drug treated group exploded considerably, indicating a large variance. Close observation of the individual values of the drug treated animals shows that all the values in this group are close to each other, except the value, 259. Let us recompute the mean and SD of this group, after removing 259 from the data. The revised mean ± SD is 156.5 ± 13.4 (n=4). We are comfortable with this SD, as this is very close to the SD of the vehicle treated group, indicating a homogeneity of variance between the vehicle treated and drug treated animals. The blood glucose of drug treated animals (after removing the value, 259) is statistically different from the vehicle treated animals by Student’s *t*-test (we used the Student’s *t*-test because the variance of the groups is not different. You will read more about this test in Chapter 8). In this example, the value 259 is an outlier, as it clearly stands out of other values, but in many pharmacological and toxicological experiments it is not easy to spot an outlier. A simple method to identify an outlier mentioned in several books on statistics is given below (Hogan and Evalenko, 2006):

Lower outlier = 25th percentile – (1.5 x IQR)
Upper outlier = 75th percentile + (1.5 x IQR)

Readers may go back to Chapter 2 and refresh their memory on box-and-whisker plot and IQR (inter-quartile range or hinge spread).

There are several statistical tools available for detecting an outlier. Among them, the Dixon test and Grubb test are widely used (Verma and Ruiz, 2006) and these tests are suggested by ASTM (2008). Outlier tests suggested in USP (2008) are ESD test, Dixon-Type test and Hampel's rule.

We shall discuss 3 outlier tests in detail:

1. Masuyama's Rejection Limit Test (Shibata, 1970)

Let us examine whether the value 259 of the example given in Table 6.2 is an outlier. Masuyama's rejection limit test is calculated using the following equation:

$$\overline{X} \pm \left(Sx \cdot \sqrt{\frac{n+1}{n}} \cdot t_{(n-1)\,0.05} \right), \text{ where}$$

Sx: Standard deviation; $t_{(n\text{-}1)0.05}$ is *t* value at 5% probability level (n–1 degrees of freedom).

The mean and SD of the data (138, 161, 156, 171, ***259***) given in Table 6.2 are;

Mean = 177.0; SD = 47.4 (n=5)

$t_{(5-1)0.05} = 2.776$ [from *t* Table by two-tailed test]

$$\text{Rejection limits} = 177 \pm (47.4 \times \sqrt{\frac{5+1}{5}} \times 2.776) = 177 \pm 157.89$$

$\therefore$ 19.11~334.89

As indicated above, Masuyama's rejection limit test gives the rejection limits in a wider range. Masuyama's rejection test is not sensitive in detecting an outlier. Hence, use of this test should be done in toxicology/pharmacology with a little caution.

2. Thompson's Rejection Test (Thompson, 1935)

Let us again work out the example of blood glucose levels of drug-treated rats given in Table 6.2. The values are 138, 161, 156, 171, ***259*** mg/dl. We shall apply Thompson's rejection test to examine whether the value, ***259*** mg/dl is an outlier.

$\Sigma X = 885$, $\overline{X} = 177$, Sum of squares (SS), $\Sigma(X-\overline{X})^2 = 8978$

$\therefore \delta = 177 – 259 = –82$

$$Sn = \sqrt{\frac{8978}{5}} = \sqrt{1795.6} = 42.37$$

$$\therefore \tau = \frac{-82}{42.37} = -1.94$$

When you substitute these calculations for the expression of *t*:

$$t_{(5-2)} = \frac{-1.94\sqrt{5-2}}{\sqrt{5-1-\left(-1.94^2\right)}}$$

$$\therefore t_{(3)} = 14.2$$

The Table value for *t* at 0.001 probability level (Table 6.3) for three degrees of freedom, is 12.923. Since the calculated *t* value is greater than the table value, we consider the blood glucose value, ***259*** mg/dl is an outlier.

Table 6.3. *t* test critical values (Yoshimura, 1987)

df\2α	0.2	0.1	0.05	0.02	0.01	0.002	0.001
df\α	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
2	1.885	2.919	4.302	6.964	9.924	22.327	31.59
3	1.637	2.353	3.182	4.540	5.840	10.214	**12.923**
4	1.533	2.131	2.776	3.746	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	2.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587

α=One-sided, 2α=Two-sided test.

3. Smirnov-Grubbs' Rejection Test (Grubbs, 1969)

Smirnov-Grubbs' rejection test is one of the tests for outliers used widely in various fields of biology (Sunaga *et al.*, 2006; Kawano *et al.*, 2007; Ishikawa *et al.*, 2010; Okubo *et al.*, 2010).

In animal experiments, the Smirnov-Grubbs' test is used more frequently than the Thompson's rejection test. Smirnov-Grubbs' test has a high power when the outlier is only one observation. However, when outliers are two or more observations, power of this test decreases due to the masking effect of one outlier to the other.

The calculation procedure of Smirnov- Grubbs' test is very simple. We can use the same example that we used for Thompson's rejection test.

First, calculate T_n.

$$T_n = \frac{(X_1 - \overline{X})}{\sqrt{V}},$$

Where n = Number of samples; X_1= The outlier.

Blood glucose level of drug treated rats are 138, 161, 156, 171, ***259*** mg/dl.

ΣX = 885, $\overline{X}$ = 177, Sum of squares (SS), $\Sigma(X-\overline{X})^2$ = 8978, Variance (*V*) = 1795.6.

$$T_5 = \frac{|259 - 177\ |}{\sqrt{1795.6}} = \frac{82}{42.37} = 1.94$$

The Table value for Smirnov- Grubbs at 0.01 probability level (Table 6.4) for 5 degrees of freedom, is 1.749. Since the calculated value (1.94) is greater than the table value (1.749), the test confirms that the blood glucose value ***259*** mg/dl is an outlier.

A Cautionary Note

Though human and other errors are major contributing factors for outliers, a positive outcome from an outlier test should be investigated (Ellison *et al.*, 2009). Before discarding an outlier, one has to confirm that the value discarded as an outlier is not a genuine data point. Hubrecht and Kirkwood (2010) suggested that one way to deal with an outlier is to carry out the statistical analysis with and without it. If the analytical results provide

Table 6.4. Smirnov-Grubbs' Table[a] (Aoki, 2002; 2006)

N	0.1	0.05	0.025	0.01
3	1.148	1.153	1.154	1.155
4	1.425	1.462	1.481	1.493
5	1.602	1.671	1.715	1.749
6	1.729	1.822	1.887	1.944
7	1.828	1.938	2.020	2.097
8	1.909	2.032	2.127	2.221
9	1.977	2.110	2.215	2.323
10	2.036	2.176	2.290	2.410
11	2.088	2.234	2.355	2.484
12	2.134	2.285	2.412	2.549
13	2.176	2.331	2.462	2.607
14	2.213	2.372	2.507	2.658
15	2.248	2.409	2.548	2.705
16	2.279	2.443	2.586	2.747
17	2.309	2.475	2.620	2.785
18	2.336	2.504	2.652	2.821
19	2.361	2.531	2.681	2.853
20	2.385	2.557	2.708	2.884

[a]One-sided table.

similar interpretation, the outlier should not be discarded. By merely not falling in the 'expected' range should not be the only reason for considering a data point as an outlier and discarding it (Petrie and Sabin, 2009). Let us examine the data on hemoglobin concentration of F344 male rats on week 104 in a repeated dose administration study given in Figure 6.1.

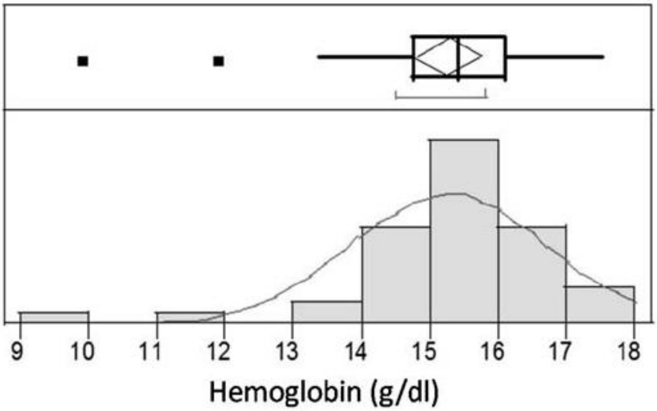

Figure 6.1. Hemoglobin concentration (g/dl) of F344 male rats on week 104

The data between 9 and 13 g/dl, appear to be outliers. Box-and-Whisker plot given in the upper section of the Figure provides useful information on the spread of the data and two outlier data points. It may be also possible that an outlier test done on the data of the Figure 6.1 confirms this view. But the values lower than 13 g/dl should not be considered as outliers, since this is how hemoglobin is distributed in the rat population of the study, which is non-normal. However, according to Ye (2003), an outlier is valid if it represents an accurate measurement and still falls well outside range of majority of values.

Non-normal distribution of several parameters is normally seen in biological experiments. In a non-normal distribution, the data points that fall outside the range of majority of the values should not be considered as outliers. It is worth mentioning here that in bioequivalence trials the regulatory agencies may permit exclusion of outliers from the statistical analysis if they are caused by product or process failure but the regulatory agencies may not permit exclusion of outliers from the statistical analysis if they are caused by subject-by-treatment interaction (Schall *et al.*, 2010).

References

Anscombe, F.J. (1960): Rejection of outliers. Technometrics, 2, 123–147.

Aoki, S. (2002): http://aoki2.si.gunma-u.ac.jp/lecture/Grubbs/Grubbs-table.html

Aoki, S. (2006): http://aoki2.si.gunma-u.ac.jp/lecture/Grubbs/Grubbs.html

ASTM (2008): American Society for Testing Materials. Standard Practice for Dealing With Outlying Observations, ASTM E178-08, ASTM International Philadelphia, USA.

Barnett, V. and Lewis, T. (1994): Outliers in Statistical Data, 3rd Edition. Wiley, New York, USA.

Dubey, S.K., Patni, A., Khuroo, A., Thudi, N.R., Reyar, S., Arun Kumar, Tomar, M.S., Jain, R., Nand Kumar and Monif, T. (2009): A quantitative analysis of memantine in human plasma using ultra performance liquid chromatography/Tandem mass spectrometry. E- J. Chem., 6(4), 1063–1070.

Ellison, S.L.R., Barwick, V.J. and Farrant, T.J.D. (2009): Practical Statistics for the Analytical Scientist—A Bench Guide. 2nd Edition, The Royal Society of Chemistry, Cambridge, U.K.

EMEA (2006): European Medicines Agency. Biostatistical Methodology in Clinical Trials. ICH Topic E 9—Statistical Principles for Clinical Trials, CPMP/ICH/363/96, London, UK.

FDA (2003): Food and Drug Administration. Guidance for Industry Bioavailability and Bioequivalence Studies for Orally Administered Drug Products—General Considerations. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Rockville, USA.

Finney, D.J. (1995): Thoughts suggested by a recent paper: Questions on non-parametric analysis of quantitative data (letter to editor). J. Toxicol. Sci., 20(2), 165–170.

Flynn, F.V., Piper, K.A.J., Garcia-Webb, P., McPherson, K. and Healy, M.J.R. (1974): The frequency distributions of commonly determined blood constituents in healthy blood donors. Clin. Chim. Acta, 52,163–171.

Girard, D., Gootz, T.D. and Mcguirk, P.R. (1992): Studies of CP-74667, a new quinolone, in laboratory animals. Antimicrobial Agents and Chemotherapy, 36(8), 11671–1676.

Grubbs, F.E. (1969): Procedures for detecting outlying observations in samples. Technometrics, 11, 1–21.

Hawkins, D.M. (1980): Identification of Outliers. Chapman and Hall Ltd., New York, USA.

Hogan, T.P. and Evalenko, K. (2006): The elusive definition of outliers in introductory statistics text books. Teaching of Psychology, 33, 252–256.

Hubrecht, R. and Kirkwood, J. (2010): The UFAW Handbook on the Care and Management of Laboratory and Other Research Animals. John Wiley and Sons, West Sussex, UK.

Ishikawa, Y., Kiyoi, H., Watanabe, K., Miyamura, K., Nakano, Y., Kitamura, K., Kohno, A., Sugiura, I., Yokozawa, T., Hanamura, A., Yamamoto, K., Iida, H., Emi, N., Suzuki, R., Ohnishi, K. and Naoe, T. (2010): Trough plasma concentration of imatinib reflects BCR-ABL kinase inhibitory activity and clinical response in chronic-phase chronic myeloid leukemia: a report from the BINGO study. Cancer Sci., 101(10), 2186–2192.

Jenifer, L.H. (2010): S Guide to Doing Statistics in Second Language Research Using SPSS. Taylor & Francis, New York, USA.

Kawano, N., Egashira, Y. and Sanada, H. (2007): Effect of dietary fiber in edible seaweeds on the development of D-galactosamine-induced hepatopathy in rats. J. Nutr. Sci. Vitaminol., 53(5), 446–450.

Keppel, G. and Wickens, T.D. (2004): Design and Analysis, a Researcher's Handbook. 4th Edition, Pearson Prentice Hall, New Jersey, USA.

Lew, M. (2007): Good statistical practice in pharmacology Problem 1. Br. J. Pharmacol., 152(3), 295–298.

Niewczas, M.A., Ficociello, L.H., Johnson, A.C., Walker, W., Rosolowsky, E.T., Roshan, B., Warram, J.H. and Krolewski, A.S. (2009): Pathways and renal function in nonproteinuric patients with type 1 diabetes. Clin. J. Am. Soc. Nephrol., 4, 62–70.

Okubo, Y., Kaneoka, K., Imai, A., Shiina, I., Tatsumura, M., Izumi, S. and Miyakawa, S. (2010): Comparison of the activities of the deep trunk muscles measured using intramuscular and surface electromyography. J. Mech. Med. Biol., 10(4), 611–620.

Petrie, A. and Sabin, C. (2009): Medical Statistics at a Glance. 3rd Edition. Wiley-Blackwell, Chichester, UK.

Rasmussen, J.L. (1988): Evaluating outlier identification tests: Mahalanobis D Squared and Comrey D. Multivariate Behavioral Res., 23(2), 189–202.

Schall, R., Endrenyi, L. and Ring, A. (2010): Residuals and outliers in replicate design crossover studies J. Biopharm. Stat., 20(4), 835–849.

Schwager, S.J. and Margolin, B.H. (1982): Detection of multivariate outliers. Ann. Stat., 10, 943–954.

Shibata, K. (1970): Biostatistics, Tokyo University of Agriculture, Tokyo, Japan.

Steinke, W., Archimbaud, Y., Becka, M., Binder, R., Busch, U., Dupont, P. and Maas, J. (2000): Quantitative distribution studies in animals: Cross-validation of radioluminography versus liquid-scintillation measurement. Reg. Toxicol. Pharmacol., 31, S33–S43.

Sunaga, H., Kaneko, M. and Amaki, Y. (2006): The efficacy of intratracheal administration of vecuronium in rats, compared with intravenous and intramuscular administration. Int. Anesthesia Res. Soc., 103(3), 601–607.

Thompson, W.R. (1935): On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation, Ann. Math. Stat., 6, 215–219.

USP (2008): The United States Pharmacopeia, The National Formularly, USP 31, NF 26, Asian Edition, Volume1, Port City Press, Baltimore, USA.

Verma, S.P. and Ruiz, A.Q. (2006): Critical values for six Dixon tests for outliers in normal samples up to sizes 100, and applications in science and engineering. Revista Mexicana de Ciencias Geológicas, 23(2), 133–161.

Wallenstein, S., Zucker, C.L. and Fleiss, J.L. (1980): Some statistical methods useful in circulation research. Circ. Res., 47, 1–9.

Ye, N. (2003): The Handbook of Data Mining. Lawrence Elbaum Associate Inc., New Jersey, USA.

Yoshimura, I. (1987): Statistical Analysis of Toxicological Data. Scientist Press, Tokyo, Japan.

Zheng, Y., Liu, H., Ma, G., Yang, P., Zhang, L., Gu, Y., Zhu, Q., Shao, T., Zhang, P., Zhu, Y., and Cai, W. (2010): Determination of S-propargyl-cysteine in rat plasma by mixed-mode reversed-phase and cation-exchange HPLC–MS/MS method and its application to pharmacokinetic studies. J. Pharm. Biomed. Anal., 54(5), 1187–1191.

Zimmerman, D.W. (1995). Increasing the power of nonparametric tests by detecting and downweighting outliers. J. Exp. Edu., 64(1), 71–78.

Zimmerman, D.W. (1998). Invalidation of parametric and nonparametric statistical tests by concurrent violation of two assumptions. J. Exp. Edu., 67(1), 55–68.