Statistics for Terrified Biologists

Helmut F. van Emden
Emeritus Professor of Horticulture
School of Biological Sciences
University of Reading, UK
## Contents

### Preface

#### 1 How to use this book
- Introduction
- The text of the chapters
- What should you do if you run into trouble?
- Elephants
- The numerical examples in the text
- Boxes
- Spare-time activities
- Executive summaries
- Why go to all that bother?
- The bibliography

#### 2 Introduction
- What are statistics?
- Notation
- Notation for calculating the mean

#### 3 Summarizing variation
- Introduction
- Different summaries of variation
  - Range
  - Total deviation
  - Mean deviation
  - Variance
- Why $n - 1$?
- Why the squared deviations?
- The standard deviation
- The next chapter
- Spare-time activities
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>4 When are sums of squares NOT sums of squares?</strong></td>
<td>21</td>
</tr>
<tr>
<td>Introduction</td>
<td>21</td>
</tr>
<tr>
<td>Calculating machines offer a quicker method of calculating sums of squares</td>
<td>21</td>
</tr>
<tr>
<td>Added squares</td>
<td>21</td>
</tr>
<tr>
<td>The correction factor</td>
<td>22</td>
</tr>
<tr>
<td>Avoid being confused by the term “sum of squares”</td>
<td>22</td>
</tr>
<tr>
<td>Summary of the calculator method of calculating down to standard deviation</td>
<td>23</td>
</tr>
<tr>
<td>Spare-time activities</td>
<td>24</td>
</tr>
<tr>
<td><strong>5 The normal distribution</strong></td>
<td>25</td>
</tr>
<tr>
<td>Introduction</td>
<td>25</td>
</tr>
<tr>
<td>Frequency distributions</td>
<td>25</td>
</tr>
<tr>
<td>The normal distribution</td>
<td>26</td>
</tr>
<tr>
<td>What per cent is a standard deviation worth?</td>
<td>27</td>
</tr>
<tr>
<td>Are the percentages always the same as these?</td>
<td>29</td>
</tr>
<tr>
<td>Other similar scales in everyday life</td>
<td>30</td>
</tr>
<tr>
<td>The standard deviation as an estimate of the frequency of a number occurring in a sample</td>
<td>31</td>
</tr>
<tr>
<td>From per cent to probability</td>
<td>31</td>
</tr>
<tr>
<td>Executive summary 1 – The standard deviation</td>
<td>33</td>
</tr>
<tr>
<td><strong>6 The relevance of the normal distribution to biological data</strong></td>
<td>35</td>
</tr>
<tr>
<td>To recap</td>
<td>35</td>
</tr>
<tr>
<td>Is our observed distribution normal?</td>
<td>36</td>
</tr>
<tr>
<td>Checking for normality</td>
<td>37</td>
</tr>
<tr>
<td>What can we do about a distribution that clearly is not normal?</td>
<td>38</td>
</tr>
<tr>
<td>Transformation</td>
<td>38</td>
</tr>
<tr>
<td>Grouping samples</td>
<td>40</td>
</tr>
<tr>
<td>Doing nothing!</td>
<td>40</td>
</tr>
<tr>
<td>How many samples are needed?</td>
<td>40</td>
</tr>
<tr>
<td>Factors affecting how many samples we should take</td>
<td>41</td>
</tr>
<tr>
<td>Calculating how many samples are needed</td>
<td>41</td>
</tr>
<tr>
<td><strong>7 Further calculations from the normal distribution</strong></td>
<td>42</td>
</tr>
<tr>
<td>Introduction</td>
<td>42</td>
</tr>
<tr>
<td>Is “A” bigger than “B”?</td>
<td>42</td>
</tr>
<tr>
<td>The yardstick for deciding</td>
<td>43</td>
</tr>
<tr>
<td>Derivation of the standard error of a difference between two means</td>
<td>45</td>
</tr>
<tr>
<td>Step 1 – from variance of single data to variance of means</td>
<td>45</td>
</tr>
</tbody>
</table>
Step 2 – from variance of single data to "variance of differences" 48
Step 3 – the combination of Steps 1 and 2; the standard error of difference between means (s.e.d.m.) 49
Recap of the calculation of s.e.d.m. from the variance calculated from the individual values 51
The importance of the standard error of differences between means 52
Summary of this chapter 52
Executive summary 2 – Standard error of a difference between two means 56
Spare-time activities 57

8 The t-test 58
Introduction 58
The principle of the t-test 58
The t-test in statistical terms 59
Why t? 60
Tables of the t-distribution 61
The standard t-test 64
The procedure 64
The actual t-test 69
t-test for means associated with unequal variances 69
The s.e.d.m. when variances are unequal 70
A worked example of the t-test for means associated with unequal variances 73
The paired t-test 75
Pair when possible 78
Executive summary 3 – The t-test 80
Spare-time activities 82

9 One tail or two? 83
Introduction 83
Why is the analysis of variance F-test one-tailed? 83
The two-tailed F-test 84
How many tails has the t-test? 86
The final conclusion on number of tails 87

10 Analysis of variance – What is it? How does it work? 88
Introduction 88
Sums of squares in the analysis of variance 89
Some "made-up" variation to analyze by Anova 89
The sum of squares table 91
Using Anova to sort out the variation in Table C
  Phase 1 91
  Phase 2 92
SqADS – an important acronym 93
Back to the sum of squares table 96
How well does the analysis reflect the input? 96
End Phase 97
Degrees of freedom in Anova 97
The completion of the End Phase 99
The variance ratio 100
The relationship between “t” and “F” 101
Constraints on the analysis of variance 103
  Adequate size of experiment 103
  Equality of variance between treatments 103
  Testing the homogeneity of variance 104
  The element of chance: randomization 104
Comparison between treatment means in the analysis of variance 107
The least significant difference 108
A caveat about using the LSD 110
Executive summary 4 – The principle of the analysis of variance 111

11 Experimental designs for analysis of variance 115
  Introduction 115
  Fully randomized 116
    Data for analysis of a fully randomized experiment 117
    Prelims 117
    Phase 1 118
    Phase 2 118
    End Phase 120
  Randomized blocks 121
    Data for analysis of a randomized block experiment 123
    Prelims 123
    Phase 1 125
    Phase 2 126
    End Phase 127
  Incomplete blocks 127
  Latin square 130
    Data for the analysis of a Latin square 131
    Prelims 132
    Phase 1 134
    Phase 2 134
12 Introduction to factorial experiments

What is a factorial experiment? 143
Interaction
If there is no interaction 145
What if there is interaction? 147
How about a biological example? 148
Measuring any interaction between factors is often the main/only purpose of an experiment 148
How does a factorial experiment change the form of the analysis of variance? 150
Degrees of freedom for interactions 150
The similarity between the “residual” in Phase 2 and the “interaction” in Phase 3 151
Sums of squares for interactions 152

13 2-Factor factorial experiments

Introduction 154
An example of a 2-factor experiment 154
Analysis of the 2-factor experiment 155
Prelims 155
Phase 1 156
Phase 2 156
End Phase (of Phase 2) 157
Phase 3 158
End Phase (of Phase 3) 162
Two important things to remember about factorials before tackling the next chapter 163
Analysis of factorial experiments with unequal replication 163
Executive summary 6 – Analysis of a 2-factor randomized block experiment 166
Spare-time activity 169

14 Factorial experiments with more than two factors

Introduction 170
Different “orders” of interaction 171
Example of a 4-factor experiment 172
Prelims 173
Phase 1 175
Phase 2 175
Phase 3 176
To the End Phase 183
Addendum – Additional working of sums of squares calculations 186
Spare-time activity 192

15 Factorial experiments with split plots 194
Introduction 194
Deriving the split plot design from the randomized block design 195
Degrees of freedom in a split plot analysis 198
Main plots 198
Sub-plots 198
Numerical example of a split plot experiment and its analysis 201
Calculating the sums of squares 202
End Phase 205
Comparison of split plot and randomized block experiment 206
Uses of split plot designs 209
Spare-time activity 211

16 The t-test in the analysis of variance 213
Introduction 213
Brief recap of relevant earlier sections of this book 214
Least significant difference test 215
Multiple range tests 216
Operating the multiple range test 217
Testing differences between means 222
Suggested “rules” for testing differences between means 222
Presentation of the results of tests of differences between means 223
The results of the experiments analyzed by analysis of variance in Chapters 11–15 225
Spare-time activities 236

17 Linear regression and correlation 238
Introduction 238
Cause and effect 239
Other traps waiting for you to fall into 239
Extrapolating beyond the range of your data 239
Is a straight line appropriate? 239
The distribution of variability 244