First-year Statistics for Psychology Students through Worked Examples

2. Descriptive Statistics: Mean, Median, Mode and Skewness

Charles McCreery, D.Phil.

Formerly Lecturer in Experimental Psychology Magdalen College Oxford



Oxford Forum Copyright © Charles McCreery, 2018

Acknowledgements

I am grateful to the following for comments and guidance at various points in the evolution of this tutorial: Dr Fabian Wadel, Dr Paul Griffiths, and Professor David Popplewell.

I am also indebted to Andrew Legge for help with the formatting of mathematical formulae and symbols.

Most recently, I have become much indebted to Dr Ed Knorr, who very kindly read through the complete typescript and made numerous suggestions and corrections, both large and small.

Any remaining errors or omissions are my responsibility. I would be pleased to receive information from anyone who spots any error, mathematical or otherwise. I can be contacted via e-mail at:

charles.mccreery@oxford-forum.org

I should also be pleased to hear from anyone who finds this tutorial helpful, either for themselves or for their students.

Charles McCreery

General Introduction¹

There are usually three complementary methods for mastering any new intellectual or artistic task; these are, in ascending order of importance:

- reading books about it
- observing how other people do it
- actually doing it oneself

These tutorials focus on the second of these methods. They are based on handouts that I developed when teaching first-year psychology students at Magdalen College, Oxford. The core of each tutorial is a worked example from an Oxford University Prelims Statistics examination paper. I have therefore placed this section in prime position; however, in teaching the order of events was different, and more nearly corresponded to the three-fold hierarchy of methods given above:

- 1. Students were invited to read one of the chapters on the Recommended Reading list, given at the end of each tutorial. They were also expected to attend a lecture on the topic in question at the Department of Experimental Psychology.
- 2. Students would attend a tutorial, in which we would go through the worked example shown here. They would take away the handouts printed as Appendices at the end of each chapter, which were designed to give structure to the topic and help them when doing an example on their own.
- 3. They would be given another previous examination question to take away and do in their own time, which would be handed in later for marking.

I am strongly in favour of detailed worked examples; following one is the next best thing to attempting a question oneself. Even better than either method is doing a statistical test on data which one has collected oneself, and which therefore has some personal significance to one, but that is not usually practicable in a first-year course.

¹ This is a general introduction to a series of six tutorials available here: <u>http://www.celiagreen.com/charlesmccreery.html</u>

I list three books in the General Bibliography at the end of this tutorial which give worked examples. One of these is Spiegel (1992), in which each chapter has numerous 'solved problems' on the topic in question. These worked problems occupy more than half of each chapter. However, the solutions to the individual problems are not as detailed and discursive as the ones I give here.

Another book which is based on worked examples on each of the topics covered is Greene and D'Oliveira (1982), also listed in the General Bibliography. Their examples are as detailed as those I give here. However, they do not cover probability and Bayes' theorem or Analysis of Variance.

Finally, I strongly recommend the *Introductory Statistics Guide* by Marija Norusis, designed to accompany the statistical package *SPSS-X*, and based on worked examples throughout. Even if the student does not have access to a computer with the *SPSS-X* package on it, this instruction manual contains excellent expositions of all the basic statistical concepts dealt with in my own examples.

Mean, Median, Mode and Standard Deviation

Contents

- 1. The question
- 2. The answer
 - 2.1 Comments on the graph
 - 2.2 Sussex data calculations
 - 2.3 Pittsburgh data calculations
- 3 Concluding comments

Appendix 1: Advantages and disadvantages of the different measures of central tendency

Appendix 2: Summary of four measures of variability or dispersion

Appendix 3: Strategic hints on answering examination questions involving plots

General Bibliography

1. The question²

Explain how you would use the mode, median and mean to determine the symmetry or skewness of a distribution of data.

The data below come from Burrell and Cane (1977) on the patterns of borrowing from libraries. The number of times each book was borrowed in a year was recorded, and this information is presented for those books borrowed at least once in the year. Data are presented for the Hillman Library at the University of Pittsburgh and the long-loan collection at Sussex University.

Number of	Number of	Number of
times	books	books
borrowed	(Sussex)	(Pittsburgh)
1	9674	63526
2	4351	25653
3	2275	11855
4	1250	6055
5	663	3264
6	355	1727
7	154	931
8	72	497
9	37	275
10	14	124
11	6	68
12	2	28
13	0	13
14	1	6
15	0	9
16	0	4

Plot each set of data on the same graph and comment. For the Sussex data, calculate the mode, median, mean, quartiles and standard deviation. Calculate the same measures for the Pittsburgh library. Describe the similarities and differences between the two sets of data. What can you conclude about the borrowing patterns in the two libraries?

² The question is taken from the Prelims Statistics paper for first-year psychology students at Oxford University, Hilary Term, 1999.

2. The answer



2.1 Comments on the graph

The distributions of the two sets of data look very similar, allowing for the very different sample sizes. Both show an extreme degree of positive (right) skew.

[Note 1: There are less than half the number of values along the x axis of the graph as there are rows in the data table. We have reduced the number of possible times borrowed from 16 to 7. The reason for this is the massive disparity between the numbers for one borrowing and the numbers for 13-16 borrowings, where some of the cells actually have nothing in them.

If we were to plot all 16 possible numbers of borrowings along the x axis, this would mean that the very small numbers would not show up at all on the y axis scale. We therefore choose to collapse all the cells showing borrowing values from 7 to 16 into one cell, i.e., we add all the numbers in these cells together to get a meaningful number which will show up on the graph.

Note 2: For a perfectly symmetrical distribution the mean, median and mode all coincide.

However, if the distribution is skewed to the right (positive skew), mode < median < mean. This is illustrated by the left-hand one of the two distributions illustrated below: it has a longer tail to the right.

If the distribution is skewed to the left (negative skew), mean < median < mode. This is illustrated by the right-hand one of the two distributions below, which has a longer tail to the left.

For more on the mean, median and mode, please refer to Appendix 1 at the end of this worked example.]



Positive skew



Negative skew

X	f	fx	(x- x)	$(\mathbf{x} - \overline{\mathbf{x}})^2$	$f(x-\overline{x})^2$
1	9674	9674	-1.01	1.02	9848.92
2	4351	8702	-0.01	0.00	0.35
3	2275	6825	0.99	0.98	2234.23
4	1250	5000	1.99	3.96	4955.10
5	663	3315	2.99	8.95	5931.25
6	355	2130	3.99	15.93	5654.47
7	154	1078	4.99	24.91	3836.15
8	72	576	5.99	35.89	2584.23
9	37	333	6.99	48.87	1808.34
10	14	140	7.99	63.86	893.99
11	6	66	8.99	80.84	485.03
12	2	24	9.99	99.82	199.64
13	0	0	10.99	120.80	0.00
14	1	14	11.99	143.78	143.78
15	0	0	12.99	168.77	0.00
16	0	0	13.99	195.75	0.00
Totals:	18854	37877			38575.49

2.2 Sussex data—calculations

Mode

The modal value for number of times borrowed is l, because there are more books borrowed this number of times than there are for any of the other number of times.

Median

Middle observation = observation number (N + 1) / 2 = (18,854+1) / 2 = 9427.5.

What value of x (number of times borrowed) corresponds to this middle (i.e., hopefully representative) book? It is 1 because the middle book falls within the first group (9427.5 < 9674). So the median is 1.

$Mean = \overline{x} = \Sigma f x / \Sigma f$

Mean number of borrowings = total number of borrowings for all books (Σfx) divided by total number of books (Σf) = 37,877 / 18,854 = 2.009.³

³ This mean has been calculated on a desktop computer using Microsoft Excel and rounded to three places of decimals. If the reader is using a calculator, rounding to some smaller number of decimal places may be necessary if the calculation is to be practicable. However, this should not significantly affect the result.

Quartiles

First quartile comes after 18,854 / 4 = 4713.5 cases. This point comes within the first group because there are 9,674 cases within this group; so first quartile = 1.

Second quartile = median = 1.

Third quartile comes after $18,854 \times 3/4$ cases = 14,140.5 cases. This point comes in the third group: (9,674 + 4,351) < 14,140.5 < (9,674 + 4,351 + 2,275). So third quartile = 3.

Standard Deviation

SD = V(Variance)Variance = $\sigma^2 = \Sigma f(x - \overline{x})^2 / (\Sigma f - 1) = 38,575.49 / (18,854 - 1) = 2.046$. So SD = V2.046 = 1.430.

[NB1: The relative sizes of the mean, median and mode are consistent with the illustration at the start of the answer, allowing for the fact that the mode and the median cannot be separated in this case.

NB2: The divisor in the formula for Variance is $(\Sigma f - 1)$, not Σf . The reason for subtracting 1 from the number of borrowings (18,854) when calculating the variance is that we are dealing here with only a sample from the underlying population of all possible observations of this kind. The correction of *n* to (n - 1) is designed to reduce the probability of bias being introduced by using a sample rather than the complete population.⁴]

⁴ This correction is known as the Bessel correction, after the German astronomer and mathematician Friedrich Bessel (1784-1846).

2.3 Pittsburgh data—calculations

X	f	fx	(x - x)	$(\mathbf{x} - \overline{\mathbf{x}})^2$	$f(x - \overline{x})^2$
1	63526	63526	-0.90	0.81	51456.06
2	25653	51306	0.10	0.01	256.53
3	11855	35565	1.10	1.21	14344.55
4	6055	24220	2.10	4.41	26702.55
5	3264	16320	3.10	9.60	31367.04
6	1727	10362	4.10	16.80	29030.87
7	931	6517	5.10	26.00	24215.31 ⁻
8	497	3976	6.10	37.20	18493.37
9	275	2475	7.10	50.40	13862.75
10	124	1240	8.10	65.59	8135.64
11	68	478	9.10	82.79	5631.08
12	28	336	10.10	101.99	2856.28
13	13	169	11.10	123.19	1601.73
14	6	84	12.10	146.39	878.46
15	9	135	13.10	171.58	1544.49
16	4	64	14.10	198.78	795.24
Totals:	114035	216773			231171.31

Mode

The modal value for number of times borrowed is 1.

Median

Median observation = (N + 1) / 2 = (114,035 + 1) / 2 = 57,018. This observation (book) falls in the first group (borrowed just once) because 57,108 < 63,526. So median = 1.

$Mean = \Sigma f x / \Sigma f$

Mean number of borrowings = total number of borrowings for all books / total number of books = 216,773 / 114,035 = 1.901

Quartiles

First quartile comes after 114,035 / 4 cases = 28,508.75 cases. This point comes within the first group, so first quartile = 1.

Second quartile = median = 1.

Third quartile comes after $114,035 \times 3/4$ cases = 85,526.25 cases. This point comes in the second group (63,526 < 85,526.25 < (63,526 + 25,653)). So third quartile = 2.

Standard Deviation

SD = $\sqrt{(\text{Variance})}$ Variance = $\sigma^2 = \Sigma f(x - \overline{x})^2 / (\Sigma f - 1) = 231,171.31 / (114,035 - 1) = 2.027$. So SD = $\sqrt{2.027} = 1.424$.

3. Concluding comments

The means for the two data sets are very similar (2.009 and 1.901), while the other measures of central tendency, the median and mode, are identical. The main measure of dispersion, the SD, is somewhat larger in the Sussex than the Pittsburgh data; this difference is reflected in the interquartile range (third quartile minus first quartile) which is greater in the Sussex data.

The great disparity between the sizes of the two data sets suggests that the Pittsburgh library is either much larger than that at Sussex, or more used. Despite this disparity, the borrowing pattern seems remarkably similar in the two libraries.

The fact that in both data sets the mean is approximately equal to the variance is compatible with the idea that the borrowing pattern follows a Poisson distribution in both libraries, since by definition mean = variance in the Poisson.

(I use the phrase 'is compatible with' rather than 'proves', because there might be other reasons for the similarity, including coincidence.)

Appendix 1

Advantages and disadvantages of the different measures of central tendency

	Advantages	Disadvantages
Mean	Can be manipulated	Affected by outliers
	algebraically	
		Affected by skewness
	Most stable between samples,	
	hence best predictor of	
	population mean	
Median	Not affected by outliers	
	Does not require interval scale	
Mode	Not affected by outliers	May not be representative
	Applicable to nominal data	

Appendix 2

Summary of four measures of variability or dispersion

Measure	Definition	Potential problems
Range	The distance from the lowest to the highest score	Is affected by outliers
Interquartile range	The range of the middle 50% of the observations	
Variance	A measure of the degree of spread of a distribution	Is sensitive to outliers
Standard deviation	The positive square root of the variance	Is sensitive to outliers

Rule of thumb:

Given a large sample, two standard deviations (one on each side of the mean) usually includes about 68% of the observations. Four SDs (two on each side) usually contains about 95%.

Appendix 3

Strategic hints on answering examination questions involving plots

1. Collapsing categories

Consider whether any categories have to be collapsed (amalgamated) to make the plot a meaningful visual representation of the data.

2. Grouping and gaps

In the case of bar charts, should the bars be grouped? If so, leave a space between each group.

3. Scaling

Always consider whether there is any good reason why the scales on the x and y axes should not start at nought (regardless of where the first data point comes). Starting at nought is less likely to be misleading, other things being equal.

NB All of these points may be illustrated by the question about book-borrowings answered above.

General Bibliography

Textbooks of the kind listed below are usually updated every few years. If the reader finds there is an edition later than the one listed here, he or she is recommended to buy the latest version.

Greene, Judith and D'Oliveira, Manuela (1982). *Learning to Use Statistical Tests in Psychology*. Milton Keynes: Open University Press.

Hays, William L. (1994). *Statistics* (5th edition). Orlando, Florida: Harcourt Brace.

Hoel, Paul G. (1976). *Elementary Statistics* (4th edition). New York: Wiley.

Howell, David C. (1997). *Statistical Methods for Psychology* (4th edition). London: Duxbury Press.

Norusis, Marija J. (1988). SPSS-X Introductory Statistics Guide, for SPSS-X Release 3. Chicago, Illinois: SPSS Inc.

Spiegel, Murray R. (1992). *Schaum's Outline of Theory and Problems of Statistics* (2nd edition). New York: McGraw-Hill.

Tabachnick, Barbara G. and Fidell, Linda S. (1983). *Using Multivariate Statistics*. London: Pearson Education Ltd.

Charles McCreery is a Research Director at Oxford Forum, an independent association of academics, set up to research and publish in currently neglected areas of psychology, theoretical physics, philosophy and economics.

If you feel have derived some benefit from this tutorial, please consider making a donation to Oxford Forum's work, via the PayPal button on the following webpage: https://celiagreen.blogspot.co.uk/

Some other publications by members of Oxford Forum are described in the following pages. They are available from <u>Amazon UK</u> and <u>Amazon USA</u>.

Lucid Dreams Foreword by Professor H.H. Price, FBA



The original pioneering study of lucid dreams: dreams in which the subject is aware that he or she is dreaming, and sometimes able to control the course of the dream.

"A fascinating subject together with a wealth of equally fascinating examples."

J.B. Priestley

"the author should be congratulated on her choice and treatment of a subject on which so very little previous work has been done." *Times Literary Supplement*

"This fascinating book raises interesting questions which will doubtless form the basis of experimentation."

Professor W.H Sprott, The Listener

"[Lucid dream research] rests almost entirely on the meticulous descriptions and classifications of types and subtypes put forward by Green in her initial publications ... All of us 'second generation' researchers have found ourselves continuously in her debt." **Professor Harry Hunt, Brock University**

Hamish Hamilton, reissued by Institute of Psychophysical Research

ISBN 978 09000760 08 (hardback)

Charles McCreery

Dreams and Psychosis A new look at an old hypothesis

DREAMS AND PSYCHOSIS A New Look at an Old Hypothesis
Charles McCreery
Psychological Paper No. 2008-1

This paper proposes a theory of psychosis based on a link between sleep and hyperarousal. It is argued that the phenomenological similarities between psychosis and dreams arise from the fact that sleep can occur, not only in states of deafferentation and low arousal, but also in states of hyperarousal resulting from extreme stress.

It is proposed that both schizophrenic and manic-depressive patients are people who are prone to episodes of hyperarousal. Various sorts of electrophysiological evidence are adduced for this proposition, drawn from the fields of electroencephalography, studies of the galvanic skin response and studies of smooth pursuit eye movements. In addition, it is suggested that a key finding is the apparently paradoxical one that catatonic patients can be aroused from their seeming stupor by the administration of sedatives rather than stimulants.

It is proposed that a tendency to hyperarousal leaves certain individuals vulnerable to 'micro-sleeps' in everyday life, with the attendant phenomena of hallucination and other sorts of reality-distortion. Delusional thinking may follow as an attempt to rationalise these intrusions of dream-phenomena into daylight hours.

Oxford Forum

Price £4.95; 34 pages ISBN 978 09536772 83

Out-of-the-Body Experiences Foreword by Professor H.H. Price, FBA



An analysis of four hundred first-hand case histories in which people seemed to leave their body and see it from outside.

"While there had been stories of out-of-body experiences for centuries, Green was the first to systematically examine a large number of first-hand accounts, from more than four hundred people ..."

Professor Oliver Sacks, <u>Hallucinations</u>

"The present volume is the first in which contemporary instances are collected, collated and studied [...] the results are extraordinarily interesting, stimulating and well worth examining by the reader." *Times Literary Supplement*

"With this rich lode of material at her disposal [Celia Green] has been able to make an exciting advance in the clarification of her subject."

The Tablet

Hamish Hamilton, reissued by Institute of Psychophysical Research

ISBN 978 09000760 15 (hardback)

Celia Green and Charles McCreery

Apparitions



An analysis of eighteen hundred first-hand accounts of experiences in which people saw, heard or sometimes even seemed to touch people or things that were not really there.

The authors advance the hypothesis that some, or even all, apparitional experiences are 'metachoric', meaning the whole of the visual field is hallucinatory.

"An excellent piece of documentation, soberly treated, and well worth reading.""

Anthony Powell, Daily Telegraph

"Enthralling" Manchester Evening News

"A fascinating and thought-provoking book." Coventry Evening Telegraph

Hamish Hamilton, reissued by Institute of Psychophysical Research

ISBN 978 09000760 91 (hardback)

The Human Evasion Foreword by R.H. Ward



An attack on the way of thought of contemporary man, revealing the patterns of prejudice which underlie his most cherished and sacrosanct opinions.

For all its seriousness, the book is written with sustained wit and intellectual audacity. Surveying the whole field of modern thought, the author reveals the same disease at work in modern Christianity as in theoretical physics.

"A subtle and sustained attack on contemporary ways of thought." *Times Literary Supplement*

"Few books, long or short, are great ones; this book is short and among those few." **R.H. Ward**

Hamish Hamilton, reissued by Institute of Psychophysical Research

ISBN 978 09000760 91 (hardback)

The Decline and Fall of Science



An attack on the attitudes of the contemporary scientific and intellectual establishment, including a vigorous and controversial defense of capitalism and private incomes.

"The central thesis is absolutely valid." *New Society*

"None can fail to acknowledge the brilliance of the author's writing." *Hampstead and Highgate Express*

Hamish Hamilton, reissued by Oxford Forum

ISBN 978 09536772 52 (hardback)

Advice to Clever Children



Reflections on education, religion and the human predicament.

"Celia Green has written an important and mentally stimulating book which goes far beyond its title." Lord St. John of Fawsley

"What this aggressive, stimulating book does is to make us face up to some of the polite fictions we have come to accept." **Ralph Yarrow, Phoenix**

Institute of Psychophysical Research, reissued by Oxford Forum

ISBN 978 09536772 21 (hardback)

Fabian Tassano

The Power of Life or Death: Medical Coercion and the Euthanasia Debate



A book which argues against medical paternalism and suggests that the increasing power given to doctors to give or withhold treatment represents a dangerous infringement of individual liberty.

"A terse, clear, incisive, intellectually first-class study of the growing power of doctors and of the lack of effective checks upon the too easily concealed but surely numerous abuses of that power." **Professor Antony Flew**

"I would not recommend this book as comfortable bedtime reading. ...If you like an intellectual challenge this one is for you." **British Medical Journal**

"His view goes straight to the medical jugular." *Nature*

"Tassano presents hair-raising case studies ... his book is a timely polemic."

Literary Review

Duckworths, reissued by Oxford Forum

The Lost Cause Causation and the Mind-Body Problem

Foreword by Professor Howard Robinson



"Celia Green has succeeded in bringing together considerations from a wide range of disciplines: philosophy, obviously, but also psychology, neuroscience and fundamental physics, making skilful use of her own empirical investigations ... most impressive." **Dr Michael Lockwood, University of Oxford**

"A worthwhile reminder of the various problems which surround the physicalism which currently dominates the philosophy of mind. ...Green does a good job of exposing the dogmatic underpinnings of current materialism, adherence to which makes mental causation seem deeply problematic."

The Human Nature Review

Oxford Forum

Charles McCreery

Perception and Hallucination The case for continuity



An analysis of empirical arguments for representationalism.

"I think the present paper is a very lucid and useful article. [...] This is the best case I know of, of an attempt to make an empirical – as opposed to a 'philosophical' – argument against direct realism." **Howard Robinson, PhD**

Professor of Philosophy, Central European University

Oxford Forum

Price £4.95; 32 pages

Letters from Exile Observations on a Culture in Decline



A collection of letters and essays written by Celia Green during the period 1990 to 1999, containing trenchant analyses of education, collectivised medicine, and modern ethics. The final section of the book introduces a provocative and original distinction between tribal and territorial morality.

"I have no doubt she is a genius." **Professor Antony Flew**

Oxford Forum

Fabian Tassano

Mediocracy *Inversions and Deceptions in an Egalitarian Culture*



A devil's dictionary for the twenty-first century, and a guide to analysing the ideology often hidden behind contemporary culture.

"Delightfully dissects the language of modern egalitarianism and political correctness. Witty, biting and definitely not to be read by New Labour."

Professor Patrick Minford

"Read this book and gain important insights into the way that the cultural elite's language works to disorient public debate." **Professor Frank Furedi**

"A witty exposure of left wing foibles." Sir Samuel Brittan

"A marvellous counterblast against the psychobabble emanating from the politically correct pseudo-intellectuals who now infest British academia."

Frederick Forsyth

Oxford Forum

Charles McCreery

The Abolition of Genius Foreword by Professor H.J. Eysenck, PhD, DSc

The Abelition of
The Adoition of
Genius
Charles McCreery
13
OXFORD FORUM
Ô
Ψ
www.celiagreen.com

An analysis of the relationship between genius and money. Dr McCreery puts forward the controversial thesis that the possession of a private income, either by the genius or by his or her patron, has been a necessary condition of the productivity of the great majority of geniuses throughout history.

"This is a courageous, well-argued and timely book [...]" H.J. Eysenck

Oxford Forum

ISBN 978 09536772 69

Free online PDF at:

http://www.celiagreen.com/abolition6.pdf