

Learning Statistics with JASP

A Tutorial for Psychology Students
and Other Beginners

Danielle J. Navarro
David R. Foxcroft
Thomas J. Faulkenberry



Learning Statistics with JASP: A Tutorial for Psychology Students and Other Beginners

(Version $\frac{1}{\sqrt{2}}$)

Danielle Navarro
University of New South Wales
d.navarro@unsw.edu.au

David Foxcroft
Oxford Brookes University
david.foxcroft@brookes.ac.uk

Thomas J. Faulkenberry
Tarleton State University
faulkenberry@tarleton.edu

<http://www.learnstatswithjasp.com>

Overview

Learning Statistics with JASP covers the contents of an introductory statistics class, as typically taught to undergraduate psychology students. The book discusses how to get started in JASP as well as giving an introduction to data manipulation. From a statistical perspective, the book discusses descriptive statistics and graphing first, followed by chapters on probability theory, sampling and estimation, and null hypothesis testing. After introducing the theory, the book covers the analysis of contingency tables, correlation, *t*-tests, regression, ANOVA and factor analysis. Bayesian statistics is covered at the end of the book.

Citation

Navarro, D.J., Foxcroft, D.R., & Faulkenberry, T.J. (2019). *Learning Statistics with JASP: A Tutorial for Psychology Students and Other Beginners*. (Version $\frac{1}{\sqrt{2}}$).

日本語版について

この本は **JASP ユーザの会** の有志が翻訳を担当しています。日本語の翻訳メンバーは下記の通りです (50 音順)。

- 紀ノ定保礼 (静岡理工科大学)
- 国里愛彦 (専修大学)
- 小杉考司 (専修大学；代表者。連絡先は kosugi@psy.senshu-u.ac.jp です)
- 小林穂波 (関西学院大学)
- 五島光
- 竹林由武 (福岡県立医科歯科大学)
- 徳岡大 (高松大学)
- 難波修史 (国立研究開発法人理化学研究所)
- 北條大樹 (東京大学)
- 平川真 (広島大学)
- 武藤拓之 (京都大学 こころの未来研究センター)
- 山根嵩史 (川崎医療福祉大学)

この本は完全にオープンソースです。つまり、あなたが望む方法で自由に改変することができます (ただし著者に適切なクレジットを与える限りにおいて、です。ライセンス条項を確認してください)。

最新バージョン

この本は、翻訳の進捗に合わせて隨時コンパイルされ、バージョンアップしていきます。最後にコンパイルされたのは 2021 年 11 月 10 日です。

This book is published under a Creative Commons BY-SA license (CC BY-SA) version 4.0. This means that this book can be reused, remixed, retained, revised and redistributed (including commercially) as long as appropriate credit is given to the authors. If you remix, or modify the original version of this open textbook, you must redistribute all versions of this open textbook under the same license - CC BY-SA.

<https://creativecommons.org/licenses/by-sa/4.0/>

The JASP-specific revisions to the original book by Navarro and Foxcroft were made possible by a generous grant to Tom Faulkenberry from the Tarleton State University Center for Instructional Innovation. Also, many thanks to Kristen Bowman for creating the beautiful front and back cover art for the book.

Table of Contents

第 $1/\sqrt{2}$ 版に向けた前書き

素晴らしい姉妹本 “jamovi で学ぶ統計” や “R で学ぶ統計” のアレンジ版, “JASP で学ぶ統計” をご紹介できて嬉しく思っています。このバージョンは Dani Navarro と David Foxcroft による素晴らしい前作の上に成り立っています。前作に投入された努力がなければ、この品質は達成できなかっただでしょう。このアレンジ版を出そうと思ったとき、私はシンプルな目標を持っていました。私は Navarro と Foxcroft のテキストを私自身の授業で使いたかったのですが、直ぐにはそうでないと思ったのは、jamovi ではなく JASP を使っていたからです。どちらも素晴らしいツールなのですが、私は JASP のほうがちょっとばかり好きなのです。というのも、jamovi がプロジェクトとして独立する前から JASP を使っていたからです。ですから、この本を世界の JASP ユーザに提供できることを嬉しく思っています。

2019 年夏、このオープン教育リソース (Open Educational Resource, OER) の執筆ための助成金を与えてくれたタールトン州立大学の Center for Instructional Innovation に感謝します。私の未来の学生 (そして世界中の学生諸君) に向けて、高品質な統計テキストを (おそらく未来永劫) 100/

(誤字脱字などを見つけるなどして) このテキストをよくしてくれる読者を待っています。もし何か貢献できると思ったら私にメールを送ってください (あるいは Githubg ページに参加してフォークしてください)。やっちゃおうぜ!

Thomas J. Faulkenberry

July 12, 2019

バージョン 0.70 に向けた前書き

バージョン 0.65 から今回へのアップデートでは、いくつかの新しい分析が導入されました。ANOVA の章では、反復測定 ANOVA や共分散分析 (ANCOVA) のセクションも追加しました。因子分析やそれに関連する技術の章も導入しました。この新しい要素のスタイルは、本書の他の章と一貫していますが、目の肥えた読者は少し概念的で実用的な説明が強調され、代数的要素が減っていることに気づくかもしれませんね。このことが良いことかどうかわかりませんが、代数については少し後で追加するかもしれません。しかしそれは、私が統計を理解して教えるときの両方のアプローチ、そして私がコース内で教えているとき学生から受け取ったフィードバックを反映したものです。これに合わせて、私も本の残りの部分に目を通し、代数の一部を箱や枠に入れて分離してみました。これらが重要でないとか、役に立たないというのではなく、学生の中にはこれらを読み飛ばしたいと思うかもしれないのに、独立したパートにすることでそうした読者の役に立てばと思うのです。

このバージョンについて、私の学生や同僚、特に Wakefield Morys-Carter から多くのコメントやフィードバックを受けたことに感謝しています。また世界中のみなさんから提案や修正をいただきま

したことに感謝しています！新しい試みの一つとして、この本のサンプルデータファイルが jamovi にアドオンモジュールとして読み込むことができるようになった、というのがあります。Jonathon Love がこれを援助してくれたことに感謝します。

David Foxcroft
February 1st, 2019

バージョン 0.65への前書き

本書は Danielle Navarro による ‘R で学ぶ統計’ の応用で、統計的なソフトウェアや分析例を jamovi に置き換えたものです。R はパワフルな統計プログラミング言語ですが、教師や学生が統計学習の最初に選択するものではありません。教師や学生によっては、ポイントしてクリックするだけで分析できるタイプのソフトウェアを好みますし、それこそ jamovi でできることです。jamovi は R を使う上で二つの側面だけに狙いを定めています。ポイントしてクリックする、グラフィカルユーザインターフェイス (GUI) と、多くの機能を組み合わせた関数を提供しており、SPSS や SAS のような方法を R でプログラミングする方法を提供しています。重要なことは、jamovi はいつもフリーでオープンであること、それが中心的な価値の一つであることです。jamovi は科学コミュニティによって作られ、科学コミュニティのためのものなのですから。

このバージョンでは、多くの人に下書きを読んでもらって、幾つもの提案や訂正をもらいました。特に Dr David Emery and Kirsty Walter に感謝しています。

David Foxcroft
July 1st, 2018

バージョン 0.6への前書き

この本は 2015 年にバージョン 0.5 をリリースしてからそれほど大きく変わってはいません。それでも、前より変わったと言った方がいいと思います。私は 2016 年にアデレードからシドニーに移動し、UNSW での経験はアデレードの頃に比べて変わってしまったので、こちらにきてから取り組む機会がずいぶん減ってしまったのです。実際に振り返ってみると、少し奇妙な感じもします。ちょっとコメントすると…

- 奇妙なことですが、この本では一貫して私の性別を間違えていますが、これについては私自身に責任があると思います(笑)。12ページにこのことについて言及した短い脚注があります。現実の生活では、私はジェンダー多様性を認める活動をしていて、この2年ほどはほとんどshe/herの代名詞を使っています。しかし私は、面倒くさがりなので、この本での文章を訂正しようとは思ってません。
- バージョン0.6にむけて、私はそれほど大きく変更せず、指摘してもらったタイプミスやその他の間違いぐらいの、いくつかのマイナーチェンジだけにしました。ただ、セクション14.4で触れている **lsr** パッケージ(これはもうメンテナンスされていません)に含まれる etaSquared関数に関する問題については注目してもらいたいと思います。この関数は、本書のようなシンプルな例ではうまく機能するのですが、見つけ切ってはないのですけど確かにバグがあるんです!ですから、これについては注意しておいてください。
- 最も大きな変更はライセンスで、私はこれをクリエイティブ・コモンズライセンス(特にCC BY-SA 4.0)のもとでリリースすることにし、誰でも利用できるように全てのソースファイルを GitHub レポジトリに置きました。

おそらく **tidyverse** を使ったバージョンを誰かが書いてくれると思うのですが… これな近々 R にとってもっと重要なトピックになってくるでしょう。:-) では。Danielle Navarro

Preface to Version 0.5

Another year, another update. This time around, the update has focused almost entirely on the theory sections of the book. Chapters 9, 10 and 11 have been rewritten, hopefully for the better. Along the same lines, Chapter 17 is entirely new, and focuses on Bayesian statistics. I think the changes have improved the book a great deal. I've always felt uncomfortable about the fact that all the inferential statistics in the book are presented from an orthodox perspective, even though I almost always present Bayesian data analyses in my own work. Now that I've managed to squeeze Bayesian methods into the book somewhere, I'm starting to feel better about the book as a whole. I wanted to get a few other things done in this update, but as usual I'm running into teaching deadlines, so the update has to go out the way it is!

Dan Navarro

February 16, 2015

Preface to Version 0.4

A year has gone by since I wrote the last preface. The book has changed in a few important ways: Chapters 3 and 4 do a better job of documenting some of the time saving features of Rstudio, Chapters 12 and 13 now make use of new functions in the lsr package for running chi-square tests and t tests, and the discussion of correlations has been adapted to refer to the new functions in the lsr package. The soft copy of 0.4 now has better internal referencing (i.e., actual hyperlinks between sections), though that was introduced in 0.3.1. There's a few tweaks here and there, and many typo corrections (thank you to everyone who pointed out typos!), but overall 0.4 isn't massively different from 0.3.

I wish I'd had more time over the last 12 months to add more content. The absence of any discussion of repeated measures ANOVA and mixed models more generally really does annoy me. My excuse for this lack of progress is that my second child was born at the start of 2013, and so I spent most of last year just trying to keep my head above water. As a consequence, unpaid side projects like this book got sidelined in favour of things that actually pay my salary! Things are a little calmer now, so with any luck version 0.5 will be a bigger step forward.

One thing that has surprised me is the number of downloads the book gets. I finally got some basic tracking information from the website a couple of months ago, and (after excluding obvious robots) the book has been averaging about 90 downloads per day. That's encouraging: there's at least a few people who find the book useful!

Dan Navarro

February 4, 2014

Preface to Version 0.3

There's a part of me that really doesn't want to publish this book. It's not finished.

And when I say that, I mean it. The referencing is spotty at best, the chapter summaries are just lists of section titles, there's no index, there are no exercises for the reader, the organisation is suboptimal, and the coverage of topics is just not comprehensive enough for my liking. Additionally,

there are sections with content that I'm not happy with, figures that really need to be redrawn, and I've had almost no time to hunt down inconsistencies, typos, or errors. In other words, *this book is not finished*. If I didn't have a looming teaching deadline and a baby due in a few weeks, I really wouldn't be making this available at all.

What this means is that if you are an academic looking for teaching materials, a Ph.D. student looking to learn R, or just a member of the general public interested in statistics, I would advise you to be cautious. What you're looking at is a first draft, and it may not serve your purposes. If we were living in the days when publishing was expensive and the internet wasn't around, I would never consider releasing a book in this form. The thought of someone shelling out \$80 for this (which is what a commercial publisher told me it would retail for when they offered to distribute it) makes me feel more than a little uncomfortable. However, it's the 21st century, so I can post the pdf on my website for free, and I can distribute hard copies via a print-on-demand service for less than half what a textbook publisher would charge. And so my guilt is assuaged, and I'm willing to share! With that in mind, you can obtain free soft copies and cheap hard copies online, from the following webpages:

Soft copy: <http://www.compcogscisydney.com/learning-statistics-with-r.html>

Hard copy: www.lulu.com/content/13570633

Even so, the warning still stands: what you are looking at is Version 0.3 of a work in progress. If and when it hits Version 1.0, I would be willing to stand behind the work and say, yes, this is a textbook that I would encourage other people to use. At that point, I'll probably start shamelessly flogging the thing on the internet and generally acting like a tool. But until that day comes, I'd like it to be made clear that I'm really ambivalent about the work as it stands.

All of the above being said, there is one group of people that I can enthusiastically endorse this book to: the psychology students taking our undergraduate research methods classes (DRIP and DRIP:A) in 2013. For you, this book is ideal, because it was written to accompany your stats lectures. If a problem arises due to a shortcoming of these notes, I can and will adapt content on the fly to fix that problem. Effectively, you've got a textbook written specifically for your classes, distributed for free (electronic copy) or at near-cost prices (hard copy). Better yet, the notes have been tested: Version 0.1 of these notes was used in the 2011 class, Version 0.2 was used in the 2012 class, and now you're looking at the new and improved Version 0.3. I'm not saying these notes are titanium plated awesomeness on a stick – though if you wanted to say so on the student evaluation forms, then you're totally welcome to – because they're not. But I am saying that they've been tried out in previous years and they seem to work okay. Besides, there's a group of us around to troubleshoot if any problems come up, and you can

guarantee that at least *one* of your lecturers has read the whole thing cover to cover!

Okay, with all that out of the way, I should say something about what the book aims to be. At its core, it is an introductory statistics textbook pitched primarily at psychology students. As such, it covers the standard topics that you'd expect of such a book: study design, descriptive statistics, the theory of hypothesis testing, *t*-tests, χ^2 tests, ANOVA and regression. However, there are also several chapters devoted to the R statistical package, including a chapter on data manipulation and another one on scripts and programming. Moreover, when you look at the content presented in the book, you'll notice a lot of topics that are traditionally swept under the carpet when teaching statistics to psychology students. The Bayesian/frequentist divide is openly discussed in the probability chapter, and the disagreement between Neyman and Fisher about hypothesis testing makes an appearance. The difference between probability and density is discussed. A detailed treatment of Type I, II and III sums of squares for unbalanced factorial ANOVA is provided. And if you have a look in the Epilogue, it should be clear that my intention is to add a lot more advanced content.

My reasons for pursuing this approach are pretty simple: the students can handle it, and they even seem to enjoy it. Over the last few years I've been pleasantly surprised at just how little difficulty I've had in getting undergraduate psych students to learn R. It's certainly not easy for them, and I've found I need to be a little charitable in setting marking standards, but they do eventually get there. Similarly, they don't seem to have a lot of problems tolerating ambiguity and complexity in presentation of statistical ideas, as long as they are assured that the assessment standards will be set in a fashion that is appropriate for them. So if the students can handle it, why *not* teach it? The potential gains are pretty enticing. If they learn R, the students get access to CRAN, which is perhaps the largest and most comprehensive library of statistical tools in existence. And if they learn about probability theory in detail, it's easier for them to switch from orthodox null hypothesis testing to Bayesian methods if they want to. Better yet, they learn data analysis skills that they can take to an employer without being dependent on expensive and proprietary software.

Sadly, this book isn't the silver bullet that makes all this possible. It's a work in progress, and maybe when it is finished it will be a useful tool. One among many, I would think. There are a number of other books that try to provide a basic introduction to statistics using R, and I'm not arrogant enough to believe that mine is better. Still, I rather like the book, and maybe other people will find it useful, incomplete though it is.

Dan Navarro

January 13, 2013

Part I.

Background

1. Why do we learn statistics?

*"Thou shalt not answer questionnaires
 Or quizzes upon World Affairs,
 Nor with compliance
 Take any test. Thou shalt not sit
 With statisticians nor commit
 A social science"*

– W.H. Auden^{*1}

1.1

On the psychology of statistics

To the surprise of many students, statistics is a fairly significant part of a psychological education. To the surprise of no-one, statistics is very rarely the *favourite* part of one's psychological education. After all, if you really loved the idea of doing statistics, you'd probably be enrolled in a statistics class right now, not a psychology class. So, not surprisingly, there's a pretty large proportion of the student base that isn't happy about the fact that psychology has so much statistics in it. In view of this, I thought that the right place to start might be to answer some of the more common questions that people have about stats.

A big part of this issue at hand relates to the very idea of statistics. What is it? What's it there for? And why are scientists so bloody obsessed with it? These are all good questions, when you

^{*1}The quote comes from Auden's 1946 poem *Under Which Lyre: A Reactionary Tract for the Times*, delivered as part of a commencement address at Harvard University. The history of the poem is kind of interesting: <http://harvardmagazine.com/2007/11/a-poets-warning.html>

think about it. So let's start with the last one. As a group, scientists seem to be bizarrely fixated on running statistical tests on everything. In fact, we use statistics so often that we sometimes forget to explain to people why we do. It's a kind of article of faith among scientists – and especially social scientists – that your findings can't be trusted until you've done some stats. Undergraduate students might be forgiven for thinking that we're all completely mad, because no-one takes the time to answer one very simple question:

Why do you do statistics? Why don't scientists just use common sense?

It's a naive question in some ways, but most good questions are. There's a lot of good answers to it,^{*2} but for my money, the best answer is a really simple one: we don't trust ourselves enough. We worry that we're human, and susceptible to all of the biases, temptations and frailties that humans suffer from. Much of statistics is basically a safeguard. Using "common sense" to evaluate evidence means trusting gut instincts, relying on verbal arguments and on using the raw power of human reason to come up with the right answer. Most scientists don't think this approach is likely to work.

In fact, come to think of it, this sounds a lot like a psychological question to me, and since I do work in a psychology department, it seems like a good idea to dig a little deeper here. Is it really plausible to think that this "common sense" approach is very trustworthy? Verbal arguments have to be constructed in language, and all languages have biases – some things are harder to say than others, and not necessarily because they're false (e.g., quantum electrodynamics is a good theory, but hard to explain in words). The instincts of our "gut" aren't designed to solve scientific problems, they're designed to handle day to day inferences – and given that biological evolution is slower than cultural change, we should say that they're designed to solve the day to day problems for a *different world* than the one we live in. Most fundamentally, reasoning sensibly requires people to engage in "induction", making wise guesses and going beyond the immediate evidence of the senses to make generalisations about the world. If you think that you can do that without being influenced by various distractors, well, I have a bridge in London I'd like to sell you. Heck, as the next section shows, we can't even solve "deductive" problems (ones where no guessing is required) without being influenced by our pre-existing biases.

1.1.1 **The curse of belief bias**

People are mostly pretty smart. We're certainly smarter than the other species that we share

^{*2}Including the suggestion that common sense is in short supply among scientists.

the planet with (though many people might disagree). Our minds are quite amazing things, and we seem to be capable of the most incredible feats of thought and reason. That doesn't make us perfect though. And among the many things that psychologists have shown over the years is that we really do find it hard to be neutral, to evaluate evidence impartially and without being swayed by pre-existing biases. A good example of this is the **belief bias effect** in logical reasoning: if you ask people to decide whether a particular argument is logically valid (i.e., conclusion would be true if the premises were true), we tend to be influenced by the believability of the conclusion, even when we shouldn't. For instance, here's a valid argument where the conclusion is believable:

All cigarettes are expensive (Premise 1)
Some addictive things are inexpensive (Premise 2)
Therefore, some addictive things are not cigarettes (Conclusion)

And here's a valid argument where the conclusion is not believable:

All addictive things are expensive (Premise 1)
Some cigarettes are inexpensive (Premise 2)
Therefore, some cigarettes are not addictive (Conclusion)

The logical *structure* of argument #2 is identical to the structure of argument #1, and they're both valid. However, in the second argument, there are good reasons to think that premise 1 is incorrect, and as a result it's probably the case that the conclusion is also incorrect. But that's entirely irrelevant to the topic at hand; an argument is deductively valid if the conclusion is a logical consequence of the premises. That is, a valid argument doesn't have to involve true statements.

On the other hand, here's an invalid argument that has a believable conclusion:

All addictive things are expensive (Premise 1)
Some cigarettes are inexpensive (Premise 2)
Therefore, some addictive things are not cigarettes (Conclusion)

And finally, an invalid argument with an unbelievable conclusion:

All cigarettes are expensive (Premise 1)
Some addictive things are inexpensive (Premise 2)
Therefore, some cigarettes are not addictive (Conclusion)

Now, suppose that people really are perfectly able to set aside their pre-existing biases about what is true and what isn't, and purely evaluate an argument on its logical merits. We'd expect 100%

of people to say that the valid arguments are valid, and 0% of people to say that the invalid arguments are valid. So if you ran an experiment looking at this, you'd expect to see data like this:

| | conclusion feels true | conclusion feels false |
|---------------------|-----------------------|------------------------|
| argument is valid | 100% say "valid" | 100% say "valid" |
| argument is invalid | 0% say "valid" | 0% say "valid" |

If the psychological data looked like this (or even a good approximation to this), we might feel safe in just trusting our gut instincts. That is, it'd be perfectly okay just to let scientists evaluate data based on their common sense, and not bother with all this murky statistics stuff. However, you guys have taken psych classes, and by now you probably know where this is going.

In a classic study, **Evans1983** ran an experiment looking at exactly this. What they found is that when pre-existing biases (i.e., beliefs) were in agreement with the structure of the data, everything went the way you'd hope:

| | conclusion feels true | conclusion feels false |
|---------------------|-----------------------|------------------------|
| argument is valid | 92% say "valid" | |
| argument is invalid | | 8% say "valid" |

Not perfect, but that's pretty good. But look what happens when our intuitive feelings about the truth of the conclusion run against the logical structure of the argument:

| | conclusion feels true | conclusion feels false |
|---------------------|------------------------|------------------------|
| argument is valid | 92% say "valid" | 46% say "valid" |
| argument is invalid | 92% say "valid" | 8% say "valid" |

Oh dear, that's not as good. Apparently, when people are presented with a strong argument that contradicts our pre-existing beliefs, we find it pretty hard to even perceive it to be a strong argument (people only did so 46% of the time). Even worse, when people are presented with a weak argument that agrees with our pre-existing biases, almost no-one can see that the argument is weak (people got that one wrong 92% of the time!).^{*3}

If you think about it, it's not as if these data are horribly damning. Overall, people did do better

^{*3}In my more cynical moments I feel like this fact alone explains 95% of what I read on the internet.

than chance at compensating for their prior biases, since about 60% of people's judgements were correct (you'd expect 50% by chance). Even so, if you were a professional "evaluator of evidence", and someone came along and offered you a magic tool that improves your chances of making the right decision from 60% to (say) 95%, you'd probably jump at it, right? Of course you would. Thankfully, we actually do have a tool that can do this. But it's not magic, it's statistics. So that's reason #1 why scientists love statistics. It's just *too easy* for us to "believe what we want to believe". So instead, if we want to "believe in the data", we're going to need a bit of help to keep our personal biases under control. That's what statistics does, it helps keep us honest.

1.2

The cautionary tale of Simpson's paradox

The following is a true story (I think!). In 1973, the University of California, Berkeley had some worries about the admissions of students into their postgraduate courses. Specifically, the thing that caused the problem was that the gender breakdown of their admissions, which looked like this:

| | Number of applicants | Percent admitted |
|---------|----------------------|------------------|
| Males | 8442 | 44% |
| Females | 4321 | 35% |

Given this, they were worried about being sued!*⁴ Given that there were nearly 13,000 applicants, a difference of 9% in admission rates between males and females is just way too big to be a coincidence. Pretty compelling data, right? And if I were to say to you that these data *actually* reflect a weak bias in favour of women (sort of!), you'd probably think that I was either crazy or sexist.

Oddly, it's actually sort of true. When people started looking more carefully at the admissions data they told a rather different story (**Bickel1975**). Specifically, when they looked at it on a department by department basis, it turned out that most of the departments actually had a slightly *higher* success rate for female applicants than for male applicants. The table below shows the admission figures for the six largest departments (with the names of the departments removed for privacy reasons):

*⁴Earlier versions of these notes incorrectly suggested that they actually were sued. But that's not true. There's a nice commentary on this here: <https://www.refsmmat.com/posts/2016-05-08-simpsons-paradox-berkeley.html>. A big thank you to Wilfried Van Hirtum for pointing this out to me.

| Department | Males | | Females | |
|------------|------------|------------------|------------|------------------|
| | Applicants | Percent admitted | Applicants | Percent admitted |
| A | 825 | 62% | 108 | 82% |
| B | 560 | 63% | 25 | 68% |
| C | 325 | 37% | 593 | 34% |
| D | 417 | 33% | 375 | 35% |
| E | 191 | 28% | 393 | 24% |
| F | 272 | 6% | 341 | 7% |

Remarkably, most departments had a *higher* rate of admissions for females than for males! Yet the overall rate of admission across the university for females was *lower* than for males. How can this be? How can both of these statements be true at the same time?

Here's what's going on. Firstly, notice that the departments are *not* equal to one another in terms of their admission percentages: some departments (e.g., A, B) tended to admit a high percentage of the qualified applicants, whereas others (e.g., F) tended to reject most of the candidates, even if they were high quality. So, among the six departments shown above, notice that department A is the most generous, followed by B, C, D, E and F in that order. Next, notice that males and females tended to apply to different departments. If we rank the departments in terms of the total number of male applicants, we get **A>B>D>C>F>E** (the "easy" departments are in bold). On the whole, males tended to apply to the departments that had high admission rates. Now compare this to how the female applicants distributed themselves. Ranking the departments in terms of the total number of female applicants produces a quite different ordering **C>E>D>F>**A>B****. In other words, what these data seem to be suggesting is that the female applicants tended to apply to "harder" departments. And in fact, if we look at Figure 1.1 we see that this trend is systematic, and quite striking. This effect is known as **Simpson's paradox**. It's not common, but it does happen in real life, and most people are very surprised by it when they first encounter it, and many people refuse to even believe that it's real. It is very real. And while there are lots of very subtle statistical lessons buried in there, I want to use it to make a much more important point: doing research is hard, and there are *lots* of subtle, counter-intuitive traps lying in wait for the unwary. That's reason #2 why scientists love statistics, and why we teach research methods. Because science is hard, and the truth is sometimes cunningly hidden in the nooks and crannies of complicated data.

Before leaving this topic entirely, I want to point out something else really critical that is often overlooked in a research methods class. Statistics only solves *part* of the problem. Remember that we started all this with the concern that Berkeley's admissions processes might be unfairly biased against female applicants. When we looked at the "aggregated" data, it did seem like the university was discriminating against women, but when we "disaggregate" and looked at the

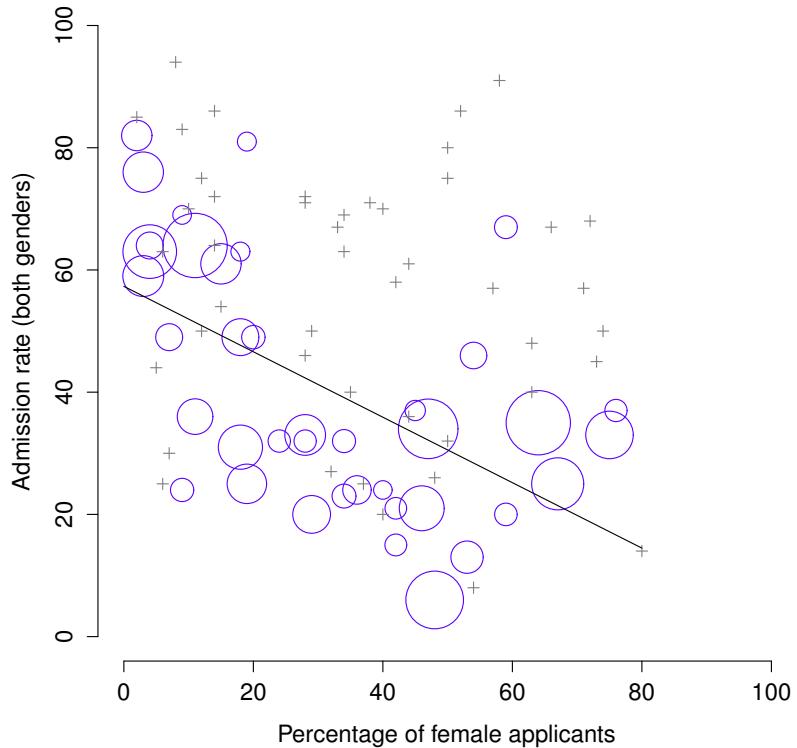


Figure1.1 The Berkeley 1973 college admissions data. This figure plots the admission rate for the 85 departments that had at least one female applicant, as a function of the percentage of applicants that were female. The plot is a redrawing of Figure 1 from **Bickel1975**. Circles plot departments with more than 40 applicants; the area of the circle is proportional to the total number of applicants. The crosses plot departments with fewer than 40 applicants.

.....

individual behaviour of all the departments, it turned out that the actual departments were, if anything, slightly biased in favour of women. The gender bias in total admissions was caused by the fact that women tended to self-select for harder departments. From a legal perspective, that would probably put the university in the clear. Postgraduate admissions are determined at the level of the individual department, and there are good reasons to do that. At the level of individual departments the decisions are more or less unbiased (the weak bias in favour of females at that level is small, and not consistent across departments). Since the university can't dictate which departments people choose to apply to, and the decision making takes place at the level of the department it can hardly be held accountable for any biases that those choices produce.

That was the basis for my somewhat glib remarks earlier, but that's not exactly the whole story, is it? After all, if we're interested in this from a more sociological and psychological perspective, we might want to ask *why* there are such strong gender differences in applications. Why do males tend to apply to engineering more often than females, and why is this reversed for the English department? And why is it the case that the departments that tend to have a female-application bias tend to have lower overall admission rates than those departments that have a male-application bias? Might this not still reflect a gender bias, even though every single department is itself unbiased? It might. Suppose, hypothetically, that males preferred to apply to "hard sciences" and females prefer "humanities". And suppose further that the reason for why the humanities departments have low admission rates is because the government doesn't want to fund the humanities (Ph.D. places, for instance, are often tied to government funded research projects). Does that constitute a gender bias? Or just an unenlightened view of the value of the humanities? What if someone at a high level in the government cut the humanities funds because they felt that the humanities are "useless chick stuff". That seems pretty *blatantly* gender biased. None of this falls within the purview of statistics, but it matters to the research project. If you're interested in the overall structural effects of subtle gender biases, then you probably want to look at *both* the aggregated and disaggregated data. If you're interested in the decision making process at Berkeley itself then you're probably only interested in the disaggregated data.

In short there are a lot of critical questions that you can't answer with statistics, but the answers to those questions will have a huge impact on how you analyse and interpret data. And this is the reason why you should always think of statistics as a *tool* to help you learn about your data. No more and no less. It's a powerful tool to that end, but there's no substitute for careful thought.

1.3 _____

Statistics in psychology

I hope that the discussion above helped explain why science in general is so focused on statistics. But I'm guessing that you have a lot more questions about what role statistics plays in psychology, and specifically why psychology classes always devote so many lectures to stats. So here's my attempt to answer a few of them...

- **Why does psychology have so much statistics?**

To be perfectly honest, there's a few different reasons, some of which are better than others. The most important reason is that psychology is a statistical science. What I mean

by that is that the “things” that we study are *people*. Real, complicated, gloriously messy, infuriatingly perverse people. The “things” of physics include objects like electrons, and while there are all sorts of complexities that arise in physics, electrons don’t have minds of their own. They don’t have opinions, they don’t differ from each other in weird and arbitrary ways, they don’t get bored in the middle of an experiment, and they don’t get angry at the experimenter and then deliberately try to sabotage the data set (not that I’ve ever done that!). At a fundamental level psychology is harder than physics.^{*5}

Basically, we teach statistics to you as psychologists because you need to be better at stats than physicists. There’s actually a saying used sometimes in physics, to the effect that “if your experiment needs statistics, you should have done a better experiment”. They have the luxury of being able to say that because their objects of study are pathetically simple in comparison to the vast mess that confronts social scientists. And it’s not just psychology. Most social sciences are desperately reliant on statistics. Not because we’re bad experimenters, but because we’ve picked a harder problem to solve. We teach you stats because you really, really need it.

- **Can’t someone else do the statistics?**

To some extent, but not completely. It’s true that you don’t need to become a fully trained statistician just to do psychology, but you do need to reach a certain level of statistical competence. In my view, there’s three reasons that every psychological researcher ought to be able to do basic statistics:

- Firstly, there’s the fundamental reason: statistics is deeply intertwined with research design. If you want to be good at designing psychological studies, you need to at the very least understand the basics of stats.
- Secondly, if you want to be good at the psychological side of the research, then you need to be able to understand the psychological literature, right? But almost every paper in the psychological literature reports the results of statistical analyses. So if you really want to understand the psychology, you need to be able to understand what other people did with their data. And that means understanding a certain amount of statistics.
- Thirdly, there’s a big practical problem with being dependent on other people to do all your statistics: statistical analysis is *expensive*. If you ever get bored and want to look up how much the Australian government charges for university fees, you’ll notice something interesting: statistics is designated as a “national priority” category, and so the fees are much, much lower than for any other area of study. This is because

^{*5}Which might explain why physics is just a teensy bit further advanced as a science than we are.

there's a massive shortage of statisticians out there. So, from your perspective as a psychological researcher, the laws of supply and demand aren't exactly on your side here! As a result, in almost any real life situation where you want to do psychological research, the cruel facts will be that you don't have enough money to afford a statistician. So the economics of the situation mean that you have to be pretty self-sufficient.

Note that a lot of these reasons generalise beyond researchers. If you want to be a practicing psychologist and stay on top of the field, it helps to be able to read the scientific literature, which relies pretty heavily on statistics.

- **I don't care about jobs, research, or clinical work. Do I need statistics?**

Okay, now you're just messing with me. Still, I think it should matter to you too. Statistics should matter to you in the same way that statistics should matter to *everyone*. We live in the 21st century, and data are *everywhere*. Frankly, given the world in which we live these days, a basic knowledge of statistics is pretty damn close to a survival tool! Which is the topic of the next section.

1.4

Statistics in everyday life

*"We are drowning in information,
but we are starved for knowledge"*

– Various authors, original probably John Naisbitt

When I started writing up my lecture notes I took the 20 most recent news articles posted to the ABC news website. Of those 20 articles, it turned out that 8 of them involved a discussion of something that I would call a statistical topic and 6 of those made a mistake. The most common error, if you're curious, was failing to report baseline data (e.g., the article mentions that 5% of people in situation X have some characteristic Y, but doesn't say how common the characteristic is for everyone else!). The point I'm trying to make here isn't that journalists are bad at statistics (though they almost always are), it's that a basic knowledge of statistics is very helpful for trying to figure out when someone else is either making a mistake or even lying to you. In fact, one of the biggest things that a knowledge of statistics does to you is cause you to get angry at the newspaper or the internet on a far more frequent basis. You can find a good example of this in Section 3.1.5. In later versions of this book I'll try to include more anecdotes along those lines.

There's more to research methods than statistics

So far, most of what I've talked about is statistics, and so you'd be forgiven for thinking that statistics is all I care about in life. To be fair, you wouldn't be far wrong, but research methodology is a broader concept than statistics. So most research methods courses will cover a lot of topics that relate much more to the pragmatics of research design, and in particular the issues that you encounter when trying to do research with humans. However, about 99% of student *fears* relate to the statistics part of the course, so I've focused on the stats in this discussion, and hopefully I've convinced you that statistics matters, and more importantly, that it's not to be feared. That being said, it's pretty typical for introductory research methods classes to be very stats-heavy. This is not (usually) because the lecturers are evil people. Quite the contrary, in fact. Introductory classes focus a lot on the statistics because you almost always find yourself needing statistics before you need the other research methods training. Why? Because almost all of your assignments in other classes will rely on statistical training, to a much greater extent than they rely on other methodological tools. It's not common for undergraduate assignments to require you to design your own study from the ground up (in which case you would need to know a lot about research design), but it *is* common for assignments to ask you to analyse and interpret data that were collected in a study that someone else designed (in which case you need statistics). In that sense, from the perspective of allowing you to do well in all your other classes, the statistics is more urgent.

But note that "urgent" is different from "important" – they both matter. I really do want to stress that research design is just as important as data analysis, and this book does spend a fair amount of time on it. However, while statistics has a kind of universality, and provides a set of core tools that are useful for most types of psychological research, the research methods side isn't quite so universal. There are some general principles that everyone should think about, but a lot of research design is very idiosyncratic, and is specific to the area of research that you want to engage in. To the extent that it's the details that matter, those details don't usually show up in an introductory stats and research methods class.

Part II.

Describing and displaying data with JASP

2. JASP 入門

ロボットは良く働く。

—Roger Zelazny^{*1}

この章では、JASP の入門方法について説明します。JASP をダウンロードしてインストールする方法について簡単に説明しますが、この章のほとんどでは、JASP ユーザーインターフェースの使用方法の入門に焦点を当てます。この章の目標は、統計の概念を学ぶことではありません。そうではなく、JASP の仕組みとソフトと快適にやり取りする方法について学びます。これを行うために、データセットと変数を検討することに時間を費やします。そうすることで、JASP での作業がどのようなものかを少し感じることができます。

ただし、詳細に進む前に、JASP を使用する理由について少し説明することには少なからず価値があります。本書を読んでいるということは、あなたにはすでに JASP を使用する理由があるのでしょうただ、その理由が「統計の授業で使用しているから」である場合、なぜ教授が授業で JASP を使用することを選択したのかについて少し説明する価値があります。もちろん、他の人々がなぜ JASP を選択するのかは本当のところ知らないので、私が使う理由について話します。

- 当たり前のことですが、手動で行うよりもコンピューターで統計を行うことは、速く、簡単で、強力であるということは述べる価値があります。コンピューターは頭を使わない反復作業に優れており、統計計算の多くは頭を使わない反復作業です。ほとんどの人にとって、鉛筆と紙で統計計算を行う唯一の理由は、学習のためです（新しい概念を学ぶ時は専門家でさえこれを行います）。私の授業では、そのようにいくつかの計算を行うことを時々提案しますが、その唯一の真の価値は教育です。自分でいくつか計算することは、統計の「感覚」を得るために役立ちますので、一度行う価値があります。しかし、一度だけです！
- 従来のスプレッドシート（例えば、Microsoft Excel）で統計を行うことは、一般的には長期的に見ると良くない考えです。多くの人はそれらに馴染みがあると感じるかもしれません、スプレッドシートでは、分析できる範囲が非常に限られています。スプレッドシートを使用し

^{*1}Source: *Dismal Light* (1968).

て実際のデータ分析を行う習慣を身につけた場合、非常に深い穴に掘り込まれることになります。

- プロプライエタリ・ソフトウェア^{*2}を避けることは、とても良い考えです。購入できる商用パッケージはたくさんあります。私が好きなものもあれば、そうでないものもあります。通常、商用パッケージは外観の体裁がとても良く、一般に非常に強力です（スプレッドシートよりもはるかに強力です）。しかし、非常に高価です。通常、企業は「学生版」（本物の一部が使えない版）を非常に安く販売し、その後、びっくりするような価格で完全版の「教育版」を販売しています。また、驚愕するほど高い値段で、商用ライセンスを販売しています。ここでのビジネスモデルは、学生時代にあなたを引き込んで、現実の世界に出かけるときに彼らのツールに依存したままにすることです。しゃくにさわるからといって彼らを責めるのは難しいですが、個人的には、避けることができるなら、何千ドルも払いたくはありません。そして、あなたはそれを避けることができます。JASP のような、オープンソースで無料のパッケージを利用すれば、法外なライセンス料を支払う必要がなくなります。

これらが JASP を使用する主な理由です。ただし、欠点がないわけではありません。JASP は、比較的新しいため^{*3}、それをサポートする教科書やその他のリソースがあまりありません。私たちがよく陥ってしまういくつかの迷惑な癖がありますが、全体的には長所が短所を上回っていると思います。これまでに出会った他のどの選択肢よりもそうです。

2.1 _____

JASP のインストール

さて、セールストークは十分でしょう。始めましょう。他のソフトウェアと同じように、JASP はコンピューターにインストールする必要があります。幸いなことに、JASP はオンラインで無料で配布されており、JASP ホームページからダウンロードできます。

<https://jasp-stats.org/>

ページの上方で、「ダウンロード」という見出しをクリックします。次に、Windows ユーザー、Mac ユーザー、および Linux ユーザー用の個別のリンクが表示されます。関連リンクをたどると、読んで字のとおりのオンラインの説明があります。この原稿の執筆時点では、JASP の現在のバージョンは 0.9.2.0 ですが、通常は数か月ごとに更新されるので、おそらく新しいバージョンが必要になります。

^{*2} 訳注 ソフトウェアの配布者が、ソフトの使用・改変・複製などを制限しているソフトウェア

^{*3} これが執筆された 2019 年 5 月

ます。^{*4}

2.1.1 JASP の起動

いずれにせよ、使用しているオペレーティングシステムに関係なく、JASP を開いて、起動させましょう。JASP の初回起動時に、図 2.1 のようなユーザーインターフェイスが表示されます。

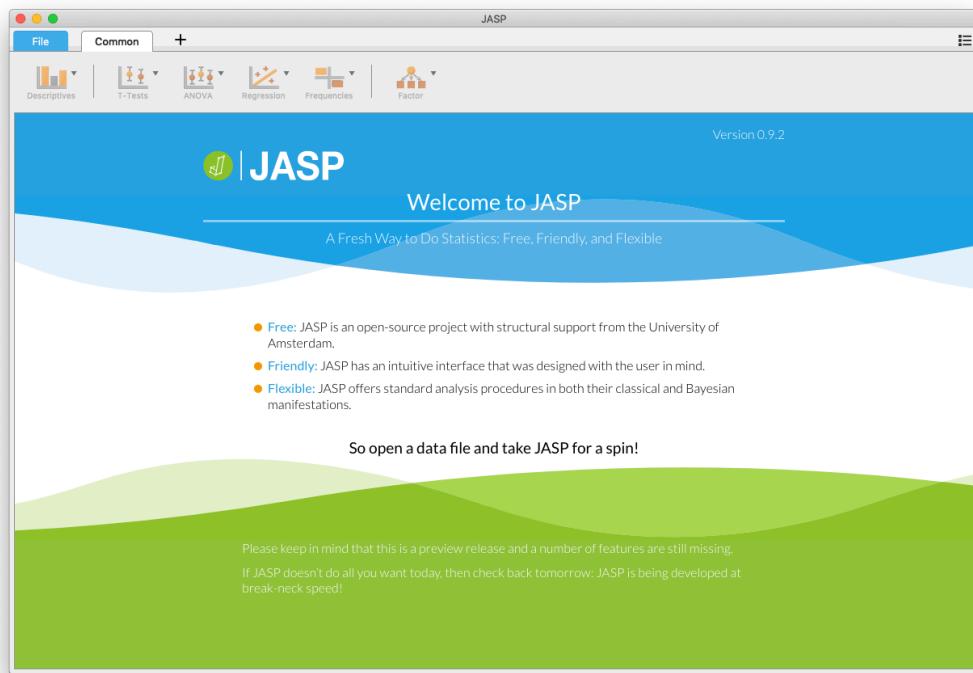


Figure2.1 起動時の JASP

他の統計ソフトウェアの使用経験がある場合、データの入力を開始する場所がないことに少しがっかりするかもしれません。これは JASP 開発者側の意図的な決定です。彼らの哲学は、ユーザーが最も快適なエディターを使用できるようにすることです^{*5}。したがって、JASP にデータを読み込むため上で推奨される方法は、CSV ファイル (.csv) を読み込むことです。CSV ファイルは、スプレッドシートプログラムで作成（と開くことが）できるテキストベースのデータ形式です。これについて

^{*4}この本でやる作業とは違って JASP は頻繁に更新されます。実際、この本の執筆中に何度かアップグレードがありましたが、この本の内容に大きな違いはありませんでした。

^{*5}この重要な問題についての議論が、<https://jasp-stats.org/2018/05/15/data-editing-in-jasp/> にあるので、参照ください

の詳細は、このあとすぐに説明します。

2.2

分析

分析は、上にあるいくつかのボタンから選択できます。分析を選択すると、特定の分析のための「options panel」が表示されます。あなたは、分析のさまざまな部分にさまざまな変数を割り当てたり、さまざまなオプションを選択できます。同時に、分析結果は右側の「Results panel」に表示され、オプションを変更するとリアルタイムで更新されます。

分析を正しく設定したら、オプションパネルの右上にある「OK」ボタンをクリックして、分析オプションを閉じることができます。これらのオプションに戻りたい場合は、結果をクリックすることができます。このようにして、あなた（または同僚）が以前に作成した分析に戻ることができます。

特定の分析が不要になった場合は、結果のコンテキストメニューで削除できます。特定の結果のヘッダー（もしくは、▼）をクリックしてメニューを表示して、「Remove Analysis」を選ぶと、分析を削除できます。しかし、これについては後で詳しく説明します。まず、JASP にいくつかのデータを入れてみましょう。

2.3

JASPへのデータ読み込み

データ分析を行う時に、私たちに関係があると思われるファイル形式がいくつかあります。この本の観点から特に重要なのは 2 つです:

- *jasp files* は、拡張子が.jasp のファイルです。これは、JASP がデータ、変数、および分析を保存するために使用する標準的なファイル形式です。
- コンマ区切り (CSV) ファイルは、拡張子が.csv のファイルです。これは、一般的な古いテキストファイルであり、さまざまなソフトウェアプログラムで開くことができます。csv ファイルは非常にシンプルなので、csv ファイルにデータを保存するのにかなりよく使われます。

2.3.1 CSV ファイルからデータをインポートする

かなり広く使用されているデータ形式の 1 つは、地味な「カンマ区切り」ファイルです。CSV ファ

The screenshot shows a Microsoft Excel spreadsheet titled 'booksales'. On the left, there is a table with four columns: 'Month' (A1:A14), 'Days' (B1:B14), 'Sales' (C1:C14), and 'Stock.Levels' (D1:D14). The first row contains column headers. The data rows show values for each month from January to December. On the right, a separate window displays the raw CSV text:

```

Month,Days,Sales,Stock.Levels
January,31,0,high
February,28,100,high
March,31,200,low
April,30,50,out
May,31,0,out
June,30,0,high
July,31,0,high
August,31,0,high
September,30,0,high
October,31,0,high
November,30,0,high
December,31,0,high

```

Figure2.2 booksales.csv のデータファイル。左側は、スプレッドシートソフトを使用してファイルを開きました。ファイルが基本的にテーブルであることを示しています。右側は、同じファイルが標準のテキストエディター（Mac のTextEdit）で開きました。ファイルがどのようにフォーマットされているか示しています。テーブルへの記入は、コンマで区切られます。

イルとも呼ばれ、通常は拡張子.csvを持ちます。CSV ファイルは、昔からある単にシンプルなテキストファイルであり、保存されるのは基本的に单なるデータのテーブルです。これを Figure 2.2 に示します。この図は、私が作成した booksales.csv というファイルを示しています。ご覧のとおり、各行は 1 ヶ月間の書籍販売データを表します。最初の行には実際のデータは含まれませんが、変数の名前があります。

CSV ファイル（あなたが作成したファイルか、誰かが提供したファイル）があれば、左上隅にある「File」タブをクリックして「Open」を選択し、表示されたオプションから選択をすることで、JASP でファイルを開けます。最も一般的には、「Computer」を選択してから「Browse」を選択します。これにより、あなたが使っているオペレーティングシステムに特有のファイルブラウザが開きます。Mac を使用している場合は、ファイルの選択に使用する通常の Finder ウィンドウのように見えるでしょう。Windows では、エクスプローラーウィンドウのように見えます。Mac での表示例は、Figure

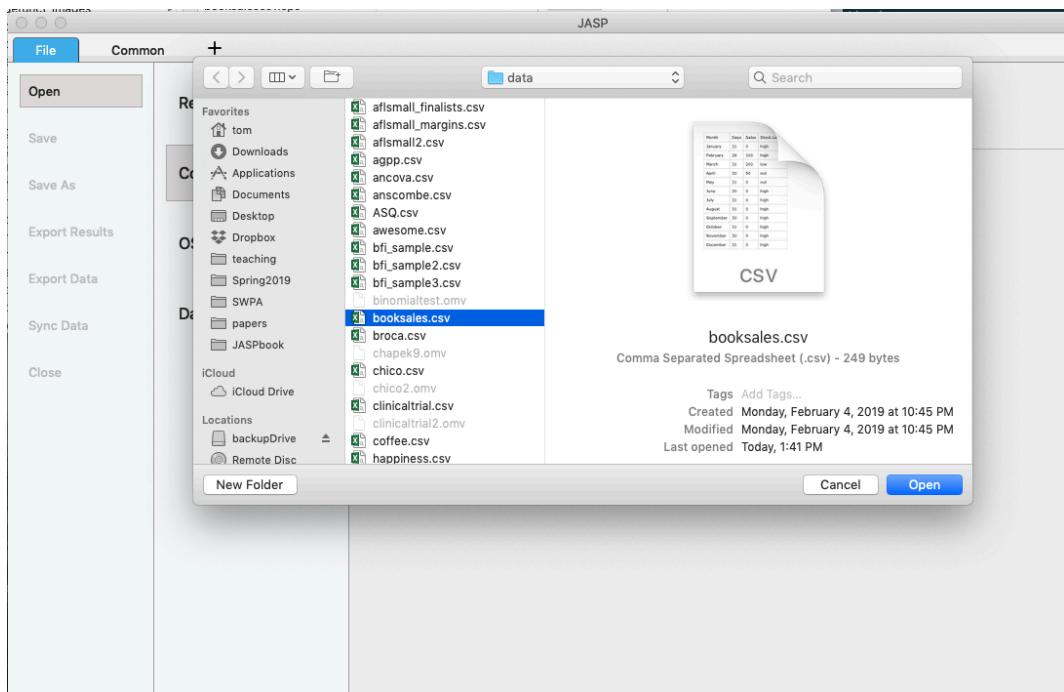


Figure2.3 JASP がインポートする CSV ファイルを選択するように求める Mac 上のダイアログボックス。Mac ユーザーはこれをすぐに理解すると思います。これは、Mac があなたにファイルを探す時に要求する一般的な方法です。Windows ユーザーにはこれは表示されませんが、代わりに、ファイルを選択するときに Windows がいつも出してくる通常のエクスプローラーウィンドウが表示されます。

~??に示されています。あなたはきっと自分のコンピュータに慣れているでしょうから、インポートしたい csv ファイルを見つけるのに問題はないはずです！ 必要なものを見つけて、「Open」ボタンをクリックしてください。

2.4

The spreadsheet

Once loaded into JASP, data is represented in a spreadsheet with each column representing a ‘variable’ and each row representing a ‘case’ or ‘participant’ .

2.4.1 Variables

The most commonly used variables in JASP are ‘Data Variables’ , which contain data loaded from a CSV file. Data variables can be one of three measurement levels, which are designated by the symbol in the header of the variable’ s column.

Nominal variables are for categorical variables which are text labels, for example a column called Gender with the values Male and Female would be nominal. So would a person’ s name. Nominal variable values can also have a numeric value. These variables are used most often when importing data which codes values with numbers rather than text. For example, a column in a dataset may contain the values 1 for males, and 2 for females. It is possible to add nice ‘human-readable’ labels to these values with the variable editor (more on this later).

Ordinal variables are like Nominal variables, except the values have a specific order. An example is a Likert scale with 3 being ‘strongly agree’ and -3 being ‘strongly disagree’ .

Scale variables are variables which exist on a continuous scale. Examples might be height or weight. This is also referred to as ‘Interval’ or ‘Ratio scale’ .

Note that when opening a data file JASP will try and guess the variable type from the data in each column. In both cases this automatic approach may not be correct, and it may be necessary to manually specify the variable type with the variable editor.

2.4.2 Computed variables

Computed Variables are those which take their value by performing a computation on other variables. Computed Variables can be used for a range of purposes, including log transforms, z-scores, sum-scores, negative scoring and means.

Computed variables can be added to the data set with the ‘+’ button in the header row of the data spreadsheet. This will produce a dialog box where you can specify the formula using either R code or a drag-and-drop interface. At this point, I simply want you to know that the capability exists, but describing how to do it is a little beyond our scope right now. More later!

2.4.3 Copy and Paste

As a final note, we will mention that JASP produces nice American Psychological Association (APA) formatted tables and attractive plots. It is often useful to be able to copy and paste these, perhaps into a Word document, or into an email to a colleague. To copy results, click on the header of the object of interest and from the menu select exactly what you want to copy.

Selecting “copy” copies the content to the clipboard and this can be pasted into other programs in the usual way. You can practice this later on when we do some analyses. Also, if you use the \LaTeX document preparation system, you can select “Copy special” and “ \LaTeX code”; doing so will place the \LaTeX syntax into your clipboard.

2.5

Changing data from one measurement scale to another

Sometimes you want to change the variable level. This can happen for all sorts of reasons. Sometimes when you import data from files, it can come to you in the wrong format. Numbers sometimes get imported as nominal, text values. Dates may get imported as text. ParticipantID values can sometimes be read as continuous: nominal values can sometimes be read as ordinal or even continuous. There’s a good chance that sometimes you’ll want to convert a variable from one measurement level into another one. Or, to use the correct term, you want to **coerce** the variable from one class into another.

In 2.4 we saw how to specify different variable levels, and if you want to change a variable’s measurement level then you can do this in the JASP data view for that variable. Just click the check box for the measurement level you want – continuous, ordinal, or nominal.

2.6

Quitting JASP

There’s one last thing I should cover in this chapter: how to quit JASP. It’s not hard, just close the program the same way you would any other program. However, what you might want to do before you quit is save your work! There are two parts to this: saving any changes to the data set, and saving the analyses that you ran.

It is good practice to save any changes to the data set as a *new* data set. That way you can always go back to the original data. To save any changes in JASP, select ‘Export Data’ from the ‘File’ tab, click ‘Browse’ and navigate to the directory location in which you want to save the file, and create a new file name for the changed data set.

Alternatively, you can save *both* the changed data and any analyses you have undertaken by saving as a .jasp file. To do this, from the ‘File’ tab select ‘Save as’, click ‘Browse’ to navigate to the directory location in which you want to save the file, and type in a file name for this .jasp

file. Remember to save the file in a location where you can find it again later. I usually create a new folder for specific data sets and analyses.

2.7

Summary

Every book that tries to teach a new statistical software program to novices has to cover roughly the same topics, and in roughly the same order. Ours is no exception, and so in the grand tradition of doing it just the same way everyone else did it, this chapter covered the following topics:

- Section 2.1. We downloaded and installed JASP, and started it up.
- Section 2.2. We very briefly oriented to the part of JASP where analyses are done and results appear, but then deferred this until later in the book.
- Section 2.3. We saw how to load data files (formatted as .csv files) in JASP.
- Section 2.4. We spent more time looking at the spreadsheet part of JASP, and considered different variable types, and briefly mentioned how to compute new variables.
- Section 2.5. And saw that sometimes we need to coerce data from one type to another.
- Section 2.6. Finally, we looked at good practice in terms of saving your data set and analyses when you have finished and are about to quit JASP.

We still haven't arrived at anything that resembles data analysis. Maybe the next Chapter will get us a bit closer!

3. 記述統計

新しいデータを手に入れたときはいつでも、最初にやるべきことの一つは、データを簡単にまとめ、その傾向を理解しやすくする方法を見つけることです。これこそ記述統計の全てです（この反対は推測統計です）。実際、多くの人が“統計”という言葉を、記述統計の同義語だと思っています。この章で話そうとしているのがそれなのですが、詳細に入る前に、なぜ記述統計が必要なのかという感覚を掴んでもらいたいと思います。そうするためにまず、`aflsmall_margins` ファイルを開いて、ファイルの中にある変数を見てみましょう。

このように、一つの変数 `afl.margins` しかありません。この章ではこの変数に注目しますので、これが何なのか少し説明します。この本に含まれるデータセットとは違って、これは実際に得たデータであり、オーストラリアのフットボールリーグ (AFL) に関するデータです^{*1} 変数 `afl.margins` は、2010 年シーズンのホームゲーム、アウェイゲーム含めた全 176 ゲームの得点差 (獲得点数) が含まれています。

このアウトプットから、このデータが何を言おうとしているのか掴み取るのは簡単ではありません。“データを眺めている”だけでは、データを理解するのに全く効果的ではないのです。このデータが何を言おうとしているか、それを掴み取るために、記述統計を計算する必要があり（この章で扱います）、わかりやすい図を描くことです（第 4 章で扱います）。二つのやり方のうち、記述統計の方がより簡単なのですが、私たちが見ようとしているデータがどんなものなのかのイメージを掴むために、この `afl.margins` データのヒストグラムをお見せすることにしましょう。図 3.2 を見てください。どうやってヒストグラムを描くかについては、セクション 4.1 で説明しますから。今は、ヒストグラムを見てそれが `afl.margins` データを正しく理解する方法であることがわかってもらえば結構です。

^{*1} オーストラリア人ではない人にむけた注意：AFL はオーストラリアのルールで行われるフットボール競技です。この章を読むためにオーストラリアのルールを調べる必要は全くありません。

| | afl.margins | |
|----|-------------|--|
| 1 | 56 | |
| 2 | 31 | |
| 3 | 56 | |
| 4 | 8 | |
| 5 | 32 | |
| 6 | 14 | |
| 7 | 36 | |
| 8 | 56 | |
| 9 | 19 | |
| 10 | 1 | |
| 11 | 3 | |
| 12 | 104 | |
| 13 | 43 | |
| 14 | 44 | |

Figure3.1 JASP が aflsmall_margins.csv ファイルを開いて変数を見せて いるスクリーンショット

.....

3.1 _____

傾向の測定

図 3.2 で示したようなデータの絵を描くというのは、データがどうなっているのかの“要点”をもたらす優れた方法です。データをいくつかの単純な“集約された”統計量に凝縮してみることが、特に便利です。いろんな場面で、まず計算してもらいたいのは**中心傾向**についての測定です。すなわち、あなたのデータの“平均”や“真ん中”がどのあたりにあるんのかを捉えて欲しいのです。最もよく使われる三つの数字は、平均値、中央値、最頻値です。これを順番に説明していくので、その後でそれぞれがどういうときに便利なのかをみていきましょう。

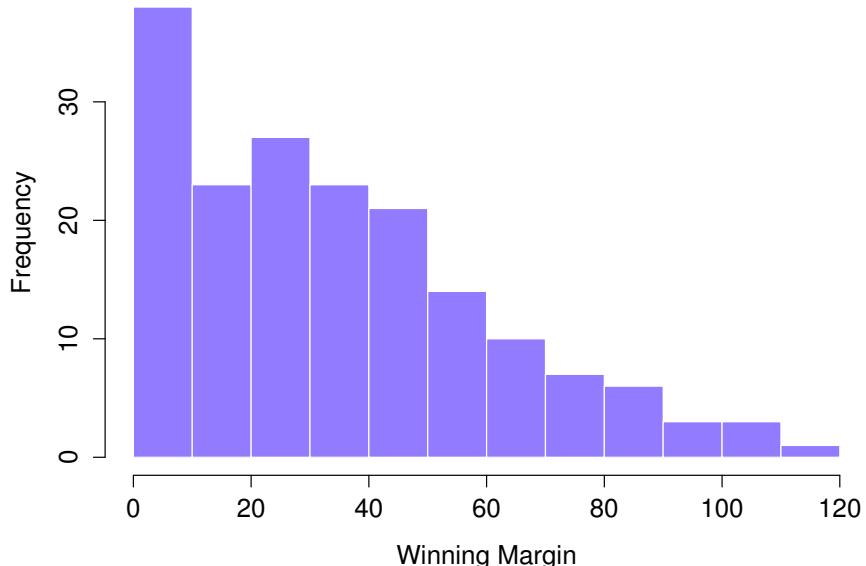


Figure3.2 2020 年の AFL 得点差データ (変数 `afl.margins`) のヒストグラム。ご想像の通り、より大差がつくゲームはより少ないので見て取れます。

3.1.1 平均値

観測値のセットの**平均値**は、普通の、昔ながらの平均値です。全ての値を足し上げて、足した値の数で割ります。最初の 5 つの AFL の得点差は、56,31,56,8,32 ですが、これらの平均値を計算するには単に次のようにするだけです。

$$\frac{56 + 31 + 56 + 8 + 32}{5} = \frac{183}{5} = 36.60$$

もちろん、この平均の定義は誰にとっても新しいものではないでしょう。アベレージ (すなわち平均値) は、日常生活でもよく使われていますから、みなさんにとってもなじみ深い物でしょう。平均の概念についてはみなさん理解しているでしょうから、この計算を表記するために統計学者が使う数学的表記法について説明する機会とさせてもらって、その後で JASP でどのように計算するか紹介することにしましょう。

最初に導入する表記法は N です。これは平均するときの観測度数の数を表すのに使います (今回の場合は $N = 5$ です)。つぎに、観測値そのものについてのラベルをつけます。これには伝統的に X が用いられ、具体的にそのどれを指し示すのかについて、添字を使います。つまり、 X_1 とすれば最

初の観測値, X_2 とすれば 2 番目の観測値, 以下同様に X_N までいきます。あるいは, 同じことをもう少し抽象的に表現するために, X_i で i 番目の観測値を指すことにします。表記法についてはっきりさせるために, 以下の表では `afl.margins` 変数にある 5 つの観測について, 数学的表記法と対応する実際の値の関係をリストアップしています。

| the observation | its symbol | the observed value |
|------------------------|------------|--------------------|
| winning margin, game 1 | X_1 | 56 points |
| winning margin, game 2 | X_2 | 31 points |
| winning margin, game 3 | X_3 | 56 points |
| winning margin, game 4 | X_4 | 8 points |
| winning margin, game 5 | X_5 | 32 points |

オウケイ、では平均の式を書いてみましょう。伝統的に、平均を表すのに \bar{X} を使います。平均の計算は以下の式で表現できます。

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_{N-1} + X_N}{N}$$

この式はまったく正しいのですが、ちょっとばかり長ったらしいので、総和記号である Σ を導入してこれを短縮しましょう^a ここでは最初の 5 つの観測について足しあわせをしたいわけですから、長い書き方ですと $X_1 + X_2 + X_3 + X_4 + X_5$ となります。ここで総和の記号を使ってこれを次のように短縮します。

$$\sum_{i=1}^5 X_i$$

文字通り、これは「1 から 5 までの全ての i について、 X_i の値を足し合わせる」と読みます。しかしその意味は基本的に「最初の 5 つの観測値を足す」、です。どちらにせよ、これは平均を使うための記号として使われ、次のように書きます。

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

正直なところ、この数学的な表記法が平均の概念を明確にするのに役立つとは思えません。実際には、私が言葉で言ったのと同じことを書き出しているだけです。すなわち、全ての数字を足しあわせて、足した項の数で割る、です。しかし詳細に書き込んだ本当の理由はこれではありません。私のゴールは、誰もがこの本を読むときに使われるであろう記号について、はっきりと理解しておいてもらうことがあります。 \bar{X} は平均、 Σ は総和、 X_i は i th 番目の観測値で N は観測の総数、ということをね。これらの記号は再利用されるので、みなさんがこれを使った式を「読む」ことができるよう、さらに「多くのものを足しあわせて別のもので割る」と言えるように理解してもらうことが重要なのです。

^a 総和に対して Σ を使うのは、勝手に決めたわけではありません。これはギリシア文字シグマの大文字で、アルファベットで言う S のアナロジーだからです。同様に、全ての積を示すための記号もあって、それは “products”(総積) と呼ばれるので文字としては Π を使います(ギリシアのパイの大文字で、これはアルファベットの P のアナロジーだからです)。

3.1.2 JASP での平均の計算

数学の話はここまで。計算してくれる魔法の箱はどうやって手に入れたらいいでしょうか？ 観測値の数が大きな数字になったら、コンピュータを使って計算させるのが何より簡単です。全てのデータを使って平均の計算をするために、JASP を使いましょう。最初のステップは ‘記述’ のボタンをクリックして、次に ‘記述統計’ をクリックしてください。それから変数 `afl.margins` をハイライト

させて、‘右矢印’をクリックしてそれを‘変数ボックス’に移します。するとすぐに画面の右側に表が現れます。そこには‘記述’についての情報があります。図 3.3 を見てください。

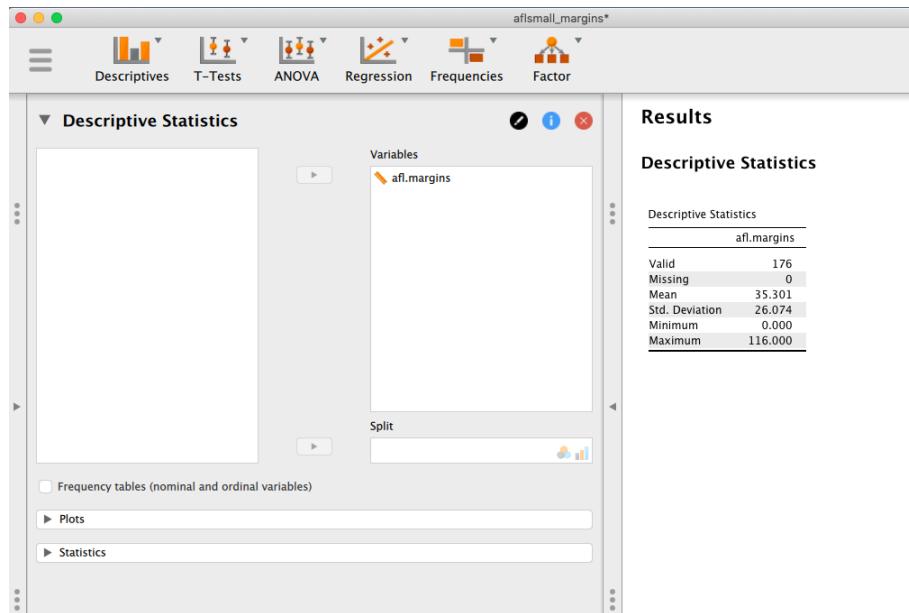


Figure3.3 AFL における 2010 年得点差データ (変数 `afl.margins`) のデフォルトで示される記述統計

図 3.3 に見て取れるように、変数 `afl.margins` の平均値は 35.301 です。他の情報として、観測度数の総数 ($N=176$) や、欠損値の数 (ありません)、変数の中央値、最小値、最大値も含まれています。

3.1.3 中央値

中心化傾向の二つ目の測度としてよく使われるのは、**中央値**です。この説明は平均よりも簡単です。変数セットの中央値というのは、ちょうど真ん中の値という意味です。AFL データの最初の 5 つの値、56,31,56,8,32 に興味があると思ってください。これらの数字の中央値を探すために、これを昇順に並べ替えます。

8, 31, **32**, 56, 56

見てみると、これら 5 つの観測値の中央値は 32 ですね。並べ替えたリストの真ん中にあるからです (より分かりやすくするために、太字にしました)。簡単なことです。でも 5 つでなくて 6 つの観測値に興味があったらどうしましょう? シーズン 6 番目のゲームが得点差 14 点だったとすると、並べ替

えリストは今や次のようにになります。

8, 14, **31**, **32**, 56, 56

そして真ん中の数字はふたつあって、31と32になります。中央値は、この二つの数字の平均値として定義されるので、31.5になります。前と同じで、数字がもっとたくさんあると人の手でやるのはとても難しくなります。実際には、もちろん、誰も真ん中の値を探すためにデータを並べ替えるなんてことはしません。コンピュータを使って、この面倒な作業をやらせるのです。JASPはお願いしたら中央値を出してくれます；単に‘統計’をクリックして、ドロップダウンメニューから‘中心化傾向’メニューの‘中央値’を選んでください。結果は自動的に中央値を含むものにアップデートされ、JASPは[afl.margins](#)変数の中央値が30.500であるとレポートしてくれます。

3.1.4 平均値か中央値か？その違いは？

平均値と中央値の計算方法を知ることは、このお話の一部に過ぎません。あなたはそれがデータの何についてものを言い、それらを使うときに何が仄めかされることになるのかを理解する必要があります。図3.4にそれを描いてみました。平均は、データセットの“重心”的なもので、中央値はデータの“真ん中の値”です。これが意味することは、あなたがこれらのどちらかを使うときに、データの種類が何であって、それで何をやろうとしているのかに関わってきます。ざっくりいうと、

- データが名義尺度水準であれば、平均値も中央値も使うべきではありません。平均値も中央値も数字が割り当てられた値に意味がある、という考え方には依存しているからです。
- データが順序尺度水準であれば、平均値よりも中央値を使う方が良いでしょう。中央値はあなたのデータの順序情報（すなわち、どの数字が大きいか）にだけ関わり、正確な数字には依存しないからです。これこそあなたのデータが順序尺度水準である状況でしょう。それに対して平均は、正確な量的値が観測対象に割り当てられているときに使われる所以、順序尺度データには適していないのです。
- 間隔尺度あるいは比率尺度水準のデータであれば、どちらでも一般的に受け入れられます。どちらを選ぶかは、あなたが何をしたいかによります。平均値はデータの全ての情報を使用します（あなたが大量のデータを持っているときには便利です）。が、極端な、外れ値には敏感です。

最後のパートを少し拡張しましょう。一つの結論として、平均値と中央値の間の体系的な違いは、ヒストグラムが非対称であるとき（歪んでいるとき；セクション3.3を参照）に現れます。これは図3.4に描かれています。中央値は（右図）、ヒストグラムの“ボディ”近くにありますが、平均値（左図）は“尻尾”（極端な値があるところ）に引っ張られています。わかりやすい例を示すために、ボブ（年収\$50,000）、ケイト（年収\$60,000）、ジェーン（年収\$65,000）が席についていると思ってください。テーブルの平均値は\$58,333で、中央値は\$60,000です。ここにビルが座ります。彼の年収は（\$100,000,000）です。年収の平均値は\$25,043,750に跳ね上がりますが、中央値は\$62,500にあが

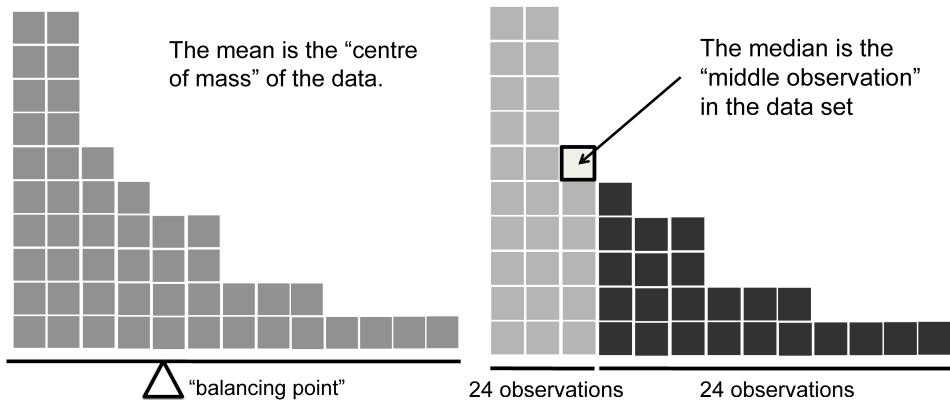


Figure 3.4 平均値と中央値の違いをどう解釈するかについてのイラスト。平均値は基本的にデータセットの“重心”です。データのヒストグラムが固体物だと考えたら、そのバランスを取る点(シーソーみたいに)が平均値です。それに対して、中央値は真ん中の観測で、それより小さいデータが半分、それより大きいデータが半分あるということです。

るだけです。席についている人の全体的な年収に興味があるおなら、平均が正しい答えになるでしょう。しかし典型的な年収の人が知りたいのであれば、中央値がより良い選択肢になるのです。

3.1.5 現実的な例

平均値と中央値の違いについて、何故注意を払うべきなのかの感覚を得るために、現実生活での例で考えてみましょう。私は科学的・統計的知識の足りないジャーナリストを馬鹿にする傾向があるのですが、信頼すべきところは信頼すべきだと思っています。これは2010年9月24日のABCニュース^{*2}になった、ある素晴らしい論文です。

コモンウェルス銀行の上級幹部がこの数週間、世界各地を訪問し、オーストラリアの住宅価格と所得に対する主要な価格の比率が、類似国と比較してどのように優れているかを示すプレゼンテーションを行いました。“住宅価格はこの5.6年、実質的に横ばい状態である”と銀行トレーディング部門のチーフエコノミスト Carig James は言っています。

これはおそらく、住宅ローンを抱えている人や、住宅ローンを希望している人、家賃を払っている人、オーストラリアの住宅市場でここ数年続いていることに全く気がついていない人にとっては、大きな驚きではないでしょうか。元の論文に戻ってみましょう。

CBA(コモンウェルス銀行のこと)は、グラフ、数字、国際比較などで住宅の運命が決まると信じている人と戦ってきました。プレゼンテーションの中には、オーストラリアの家賃は収入に比べて割高であるという議論を、銀行が否定しているとされています。オーストラリアにおいて、世帯主の価格に対する住宅価格は大都市において5.6、全国的には4.3であり、他の多くの先進国と同じぐらいであるとしています。また、サンフランシスコとニューヨークではこの比率は7,

^{*2}www.abc.net.au/news/stories/2010/09/24/3021480.htm

オークランドでは 6.7, バンクーバーでは 9.3 にもなります。

もっとびっくりなニュースです! だけど、この論文は次のように見立てています。

アナリストの多くは、これは銀行によってミスリーディングな図、比較がなされたからだと言います。CBA の資料 4 ページ目をみて、グラフや表の下に書いてある情報ソースをみたら、国際比較の追加的なソースがあることに気づくでしょう—人口動態学についての。コモンウェルス銀行が人口動態学の情報を使ってオーストラリアの住宅価格・収入比率の分析をしていたとすると、その実態は 5.6 とか 4.3 ではなく 9 近くになります。

うーむ、かなりの違いがありますね。一方では 9 といい、他方では 4-5 だ、と言っています。この違いを区分して、本当の値はこの間にあるんだとでもしたほうがよいでしょうか? 全く違います! 正しい答えと、間違った答えがあるような状態なのです。人口動態学は正しく、コモンウェルス銀行は間違っています。論文では次のように指摘しています。:

コモンウェルス銀行の住宅価格対収入の図には明らかな問題があり、平均年収と住宅価格の中央値を比較しているのです(人口動態学の図は収入の中央値と価格の中央値の比較をしているのに)。中央値は真ん中にある点で、極端に高いあるいは低い値を効率よくカットしますが、平均値は年収や資産価値については高所得者が含まれるので高くなる傾向があります。別の言い方をすれば、コモンウェルス銀行の図は Ralph Norris の数百万ドルにも及ぶ給料を収入が話に入れ、かれの(間違いなく)高価な住宅価格は図の中に入れないようにしているので、住宅価格はオーストラリアの中級ぐらいの年収と比較することになります。

これ以上いうことはありません。人口動態学的に計算した比率の方が正しいのです。銀行がやったやり方は間違っています。なぜ数字に得意なはずの銀行がこのような基本的なミスをしたのかというと… 彼らが何を考えていたのかは分からないので、ここまでにしましょう。しかしこの論文が以下の事実についての注意を促しています。関係があるかどうかわかりませんが。

オーストラリア最大の住宅業界牽引者であるコモンウェルス銀行は、住宅価格の上昇については最大級の興味を持っています。住宅ローンや多くの中小企業向けローンの担保として、オーストラリアの住宅の大部分を事実上所有しています。

むにやむにや。

3.1.6 最頻値

サンプルの最頻値は、とても単純です。それは最も頻度が多い値、なのです。AFL の別の変数を使ってこれを説明してみましょう。決勝で最も多くプレーしている選手は誰でしょう? `aflsmall_finalists` ファイルを開いて、`afl.finalists` 変数をみてみましょう。図 3.5 がそれです。この変数には全 400 チームの、1987 年から 2010 年までの間に開催された 200 回の決勝戦情報が載っています。

我々がやるべきことは、全 400 試合を読み通して、決勝戦リストに出てくるチームの名前を数え上げ、**度数分布表**を作ることです。しかしこれは頭を使わない退屈な作業で、まさにコンピュータが得意とするような作業ですね。だから JASP を使いましょう。「記述」の下にある「記述統計」の、`afl.finalists` 変数を選び「変数」ボックスに移し、「度数分布表」と書かれた小さなチェックボック

The screenshot shows the JASP interface with the title bar 'aflsmall_finalists'. The menu bar includes Descriptives, T-Tests, ANOVA, Regression, Frequencies, and Factor. A toolbar below the menu has icons for Descriptives, T-Tests, ANOVA, Regression, Frequencies, and Factor. The main window displays a table with 13 rows and 2 columns. The first column contains row numbers 1 through 13. The second column contains team names: Hawthorn, Melbourne, Carlton, Melbourne, Hawthorn, Carlton, Melbourne, Carlton, Hawthorn, Melbourne, Melbourne, Hawthorn, and Melbourne.

| | afl.finalists |
|----|---------------|
| 1 | Hawthorn |
| 2 | Melbourne |
| 3 | Carlton |
| 4 | Melbourne |
| 5 | Hawthorn |
| 6 | Carlton |
| 7 | Melbourne |
| 8 | Carlton |
| 9 | Hawthorn |
| 10 | Melbourne |
| 11 | Melbourne |
| 12 | Hawthorn |
| 13 | Melbourne |

Figure3.5 aflsmall_finalists.csv ファイルに修められた変数の JASP スクリーンショット

スをクリックします。すると図 3.6 のようなものが得られるでしょう。

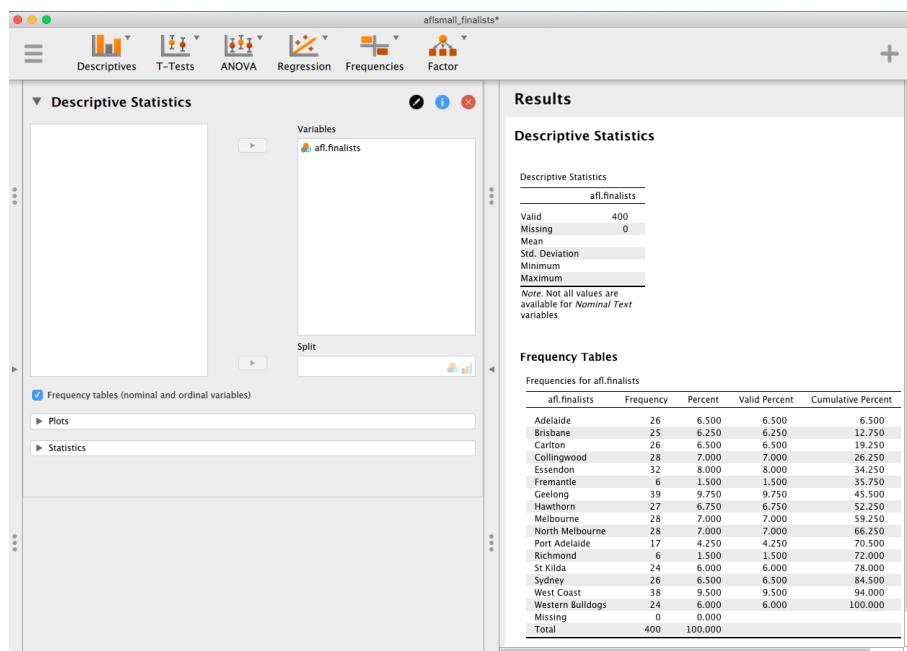


Figure3.6 afl.finalists 変数の度数分布表を示した JASP スクリーンショット

さて度数分布表を入手したわけですが、これをみると 24 年間ずっと、Geelong が他のどのチームよ

りも多く決勝に進んでいることがわかります。ですから `afl.finalists` データの最頻値は "Geelong" だということになります。Geelong(39 回決勝進出) が 1987 年から 2010 年の間で他のどのチームよりも多く決勝に進んでいるのです。また, '記述統計' の表では平均値, 中央値, 最大値, 最小値が計算されていないのも注目です。なぜなら `afl.finalists` 変数は名義的な文字変数であって, これらの値を計算する意味がないからです。

最後に最頻値に関するポイントをもう一つ。名義尺度のデータを持っていたら最頻値を計算するのが最もよくあるケースです。というのも, 平均や最頻値はこの種の変数には向いていないからですが, 順序, 間隔, 比率尺度水準の変数の最頻値を知りたいという時もあります。例えば, `afl.margins` 変数にもどってみましょう。この変数は明らかに比率尺度水準(もしピンとこないのなら, もう一度セクション ?? を読んでみてください)であり, あなたが知りたいのはこの中心に関する測度であれば平均値や中央値を求めるところです。しかしこんなことを考えてみてください: あなたの友達が賭けようぜと言ってきて, ランダムにフットボールのゲームを選ぶとします。誰がプレイするのかを知らずに, 正確な得失点差を推測しないといけないです。正しく当てられたら 50 ドルもらえます。でなければ 1 ドル失います。ほとんど正解に近かった, という残念賞はないものとします。正確に点差を推測しなければならないのです。この賭けをする時, 平均や中央値は全くあなたの役に立ちません。最頻値にかけるべきです。`afl.margins` 変数の最頻値を JASP で計算するには, データセットに戻って '記述' - '記述統計' 画面から, '統計量' と書いてあるセクションを拡大してください。'最頻値' のチェックボックスをクリックして, '記述統計量' テーブルにある最頻値をみます。図 3.7 にあるやつです。そうすると, 2010 年のデータでは 3 点差に賭けるべきだということがわかります。

3.2

変動性の指標

ここまで話してきた統計の話は, 中心化傾向に関するものでした。つまり, そこでの話はデータの "真ん中" とか "代表的な" 値についてでした。しかし, 中心化傾向は計算したい要約統計量の唯一の種類, というわけではありません。計算したい第二のものとして, データの **変動性** があります。つまり, どれぐらいデータが "散らばっているか"? とか, どれぐらい平均や中央値から観測値が "遠くにある" 傾向があるか? というものです。ここでは, データが間隔あるいは比率尺度水準で得られていると考えますから, `afl.margins` データを例に使い続けましょう。このデータを使うことで, 散らばりの指標としていくつかのものを示すことにし, その長所と短所も見ていくことにしましょう。

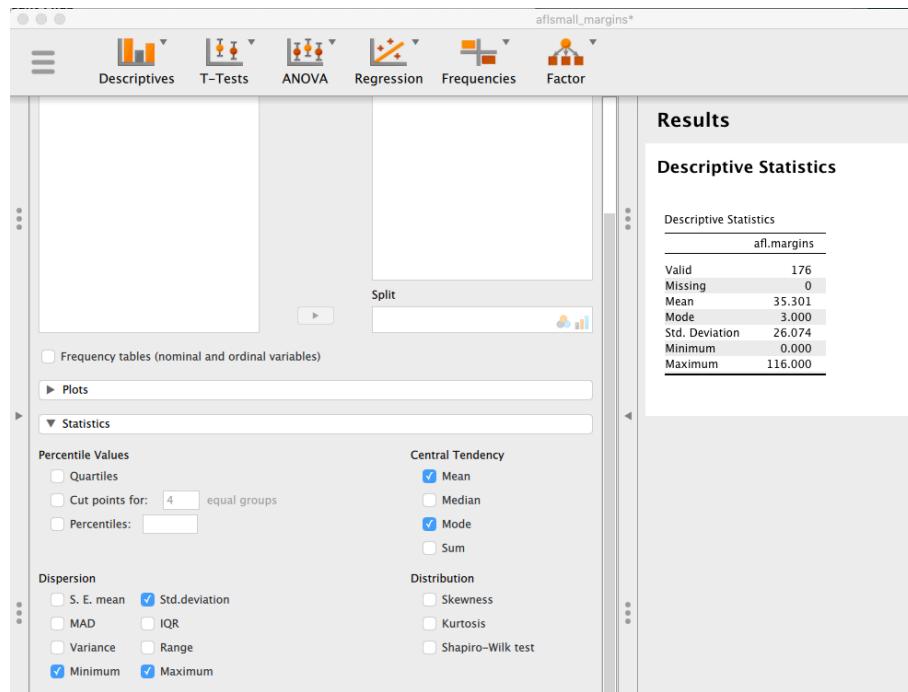


Figure3.7 afl.margins 変数の中央値を示す JASP 画面

3.2.1 範囲

変数の範囲はとてもシンプルなものです。最大値から最小値を引いたもののこと指します。AFL 得点差データの最大値は 116 で最小値は 0 でした。“変動”を表す量として範囲は最も単純なものです、最も悪いものもあります。要約統計量を頑健なものにするために、平均について議論していたことを思い出してください。もしデータセットの中に一つ二つ変な値があると、我々の統計量はそ うしたデータに角に影響されないようにしたいところです。

例えば、変数が極端な外れ値を持っていたとします。

$$-100, 2, 3, 4, 5, 6, 7, 8, 9, 10$$

範囲が頑健な値ではないことは明らかですよね。変数の範囲は 110 になりますが、外れ値を除くとたったの 8 になります。

3.2.2 四分位範囲

四分位範囲 (interquartile range,IQR) は範囲に似ていますが、最大値と最小値の差を使うのではなく、25 パーセンタイルと 75 パーセンタイルの差を使います。パーセンタイルをまだ知らないかも

されませんが、データの 10 パーセンタイルというのはある点 x よりも小さいのがデータの 10% になるような点 x のこと、という意味です。実は、既にこの考え方は出てきています。データの中央値とは、50 パーセンタイルのことですから！JASP では、簡単に 25,50,75 パーセンタイルを見つけることができます。‘記述’の‘記述統計’から‘統計量’の画面にある‘四分位’チェックボックスをクリックするだけです。

Descriptive Statistics

| Descriptive Statistics | |
|-------------------------|---------|
| <hr/> afl.margins <hr/> | |
| Valid | 176 |
| Missing | 0 |
| Mean | 35.301 |
| Mode | 3.000 |
| Std. Deviation | 26.074 |
| Minimum | 0.000 |
| Maximum | 116.000 |
| 25th percentile | 12.250 |
| 50th percentile | 30.500 |
| 75th percentile | 51.500 |

Figure3.8 afl.margins 変数の四分位を示す JASP のスクリーンショット

驚くには值しませんが、図 3.8において 50 パーセンタイルは中央値と同じになっています。そして、 $50.50 - 12.75 = 37.75$ ですから、2010 年の AFL 得点差データの四分位範囲は 37.75 ということになります。範囲の解釈は明らかですが、IQR の解釈の仕方はそこまで明らかだというわけではないですね。これは次のように考えるのが最も単純な方法でしょう。すなわち、四分位範囲はデータの“中半分”の範囲だというものです。つまり、データの一つの四分位が 25 パーセンタイル点で、もう一つの点が 75 パーセンタイル点ですから、この二つの間にデータの“中半分”が位置していることになります。IQR はこの中半分をカバーする範囲なのです。

3.2.3 平均絶対偏差

二つの尺度、範囲と四分位範囲をみてきましたが、どちらもデータのパーセンタイルをみて、データの散らばりを測ろうとするアイデアに基づいています。しかし、これだけがこの問題唯一の解決策ではありません。別のアプローチとして、意味のある参照点（ふつう平均値や中央値ですが）を選び、その参照点からの“典型的な”偏差を報告する、ということがあります。“典型的な”偏

差、というのは何を意味しているでしょう？普通これは偏差の平均値や中央値を指します。実際、ここからは二つの尺度が導かれます。“平均絶対偏差”(平均値からの)と，“中央値絶対偏差”(中央値からの)，です。私がこれまでみてきたところ、中央値に基づく尺度が統計的に使われているようで、そちらの方が優れているようです。しかし正直に言って、心理学でこれらが使われてきたのをあまりみたことがありません。平均に基づく尺度の方が、心理学ではよく出てきます。このセッションでは前者について最初説明しますが、その後で2番目についても触れていきます。

前のパラグラフではちょっと抽象的だったかもしれません、平均からの**平均絶対偏差**についてもう少しゆっくりみていきましょう。この尺度が便利なことの一つに、この名前が実際にどうやって計算するのかを表している、というのがあります。AFLの得点差データについて、もう一度最初の5ゲームをみてみると、得点差は56, 31, 56, 8, 32でしたね。ここで計算はある参照点(今回は平均)からの偏差を見るものですから、最初にするべきことは平均つまり \bar{X} を計算することです。最初の5ケースでは、平均は $\bar{X} = 36.6$ になりました。次のステップは各観測値、 X_i を偏差のスコアに変換することです。これは観測値 X_i と平均 \bar{X} の差を計算することができます。つまり、偏差スコアの定義は $X_i - \bar{X}$ となるのです。今回のサンプルにおける最初の観測値は、 $56 - 36.6 = 19.4$ になります。オーケイ、十分シンプルですね。このプロセス、次のステップはこれらの偏差を絶対偏差にすることです。これは負の値を正の値にすることでできます。数学的には-3の絶対値を $|-3|$ と書き、 $|-3| = 3$ とします。この絶対値を使うのは、平均よりも高かったのか低かったのかを気にしないということであり、興味は平均にどれくらい近かったのかというだけだということです。このプロセスができるだけ明白にするために、下の表では、5つの観測値すべてについてこれらの計算を示しています。

| 用語: | どのゲームで | 値 | 平均偏差 | 絶対偏差 |
|-----|--------|-------|-----------------|-------------------|
| 表記: | i | X_i | $X_i - \bar{X}$ | $ X_i - \bar{X} $ |
| | 1 | 56 | 19.4 | 19.4 |
| | 2 | 31 | -5.6 | 5.6 |
| | 3 | 56 | 19.4 | 19.4 |
| | 4 | 8 | -28.6 | 28.6 |
| | 5 | 32 | -4.6 | 4.6 |

さてデータセットの各観測値について絶対偏差を計算できたので、これらのスコアの平均を計算しましょう。次のようになります。

$$\frac{19.4 + 5.6 + 19.4 + 28.6 + 4.6}{5} = 15.52$$

はいおしまい。これら5つのスコアについて、平均絶対偏差は15.52でした。

ところで、この簡単な例はこれでおしまいですが、少し話が残っています。まず、数学的な定式化をしておくべきです。しかしこれをしようとするとき、平均絶対偏差についての数学的表記が必要です。腹立たしいことに、“平均絶対偏差”と“中央値絶対偏差”はどちらも同じ頭文字 (MAD) ので、曖昧になってしまいますから、平均絶対偏差に何か別の表現を考えないといけないというやれやれ。*average absolute deviation* を短くして、AAD とすることにしましょう。これでもいくらか曖昧な表記ですが、計算は次のように書くことができます。

$$\text{aad}(X) = \frac{1}{N} \sum_{i=1}^N |X_i - \bar{X}|$$

3.2.4 分散

平均絶対偏差は使いでありますが、変動の尺度として最適というわけではありません。純粋に数学的な観点からは、絶対偏差よりも二乗した偏差の方が好ましい理由があります。これを使うと分散とよばれる尺度を手に入れることになります。それは本当にステキな統計的特徴を持っているのですが、それは横に置いておくとして^{*3}、今から取り上げるとても大きな心理学的欠陥も持っていることを説明します。データセット X の分散は $\text{Var}(X)$ と表記されますが、もっと一般的には s^2 と書きます（その理由はすぐにわかります）。

観測されたデータセットの分散を計算する式は次の通りです。

$$\text{Var}(X) = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2$$

ご覧の通り、基本的には平均絶対偏差で使ったものと同じ形をしていますが、違うのは“絶対偏差”のかわりに“偏差平方”を使っているところです。このため、分散は“平均偏差平方”とも言われます。

さて、基本的な概念を手に入れましたので、具体例でみてみましょう。もう一度、AFL ゲームの最初の 5 つのデータを使います。前回同じアプローチをした時に習って、次のような表にしてみました。

^{*3}えーっと、ちょっとだけ何が最高にクールなのか、“クール”の定義をしてから説明してみましょう。分散は加算的なのです。その意味はこんな感じです。私が二つの変数 X と Y を持っていて、それらの分散がそれぞれ $\text{Var}(X)$ と $\text{Var}(Y)$ だとしましょう。ここで新しい変数 Z を、二つの和、 $Z = X + Y$ で定義したとします。そうすると、 Z の分散は $\text{Var}(X) + \text{Var}(Y)$ になるのです。これがとても便利な特徴なのですが、このセクションで私が説明しようとする他の尺度にはないものなのです。

| 用語: 表記: | どのゲームで 表記: <i>i</i> | 値 X_i | 平均偏差 $X_i - \bar{X}$ | 偏差平方 $(X_i - \bar{X})^2$ |
|------------|---------------------------|------------|-------------------------|-----------------------------|
| | 1 | 56 | 19.4 | 376.36 |
| | 2 | 31 | -5.6 | 31.36 |
| | 3 | 56 | 19.4 | 376.36 |
| | 4 | 8 | -28.6 | 817.96 |
| | 5 | 32 | -4.6 | 21.16 |

最後の列には全ての偏差平方が入っていますので、この平均を取れば良いのです。手計算する、つまり電卓を使うと、この分散の値が 324.64 であることがわかります。興奮してきたでしょう？このとき、多分あなたの考えに火がついた問題（すなわち、324.64 の分散って本当に平均なのか？）は横に置いて、JASP でこれをどう計算するかみてみましょう。というのも、これで奇妙なことが明らかになるからです。

まず最初の 5 行だけを含んだ新しいデータを読み込みます。ファイル `aflsmall_margins_first5.csv` を読み込んでください。次に‘統計’メニューの‘記述’-‘記述統計’をクリックし、‘分散’チェックボックスをクリックします（‘ばらつき’グループの中にあるのがわかると思います）。手計算した値（324.64）と同じ数字になりましたか？いや、ちょっと待って、あなたは全く違う答えを手にしたではありませんか（405.800）！おかしいなあ。JASP は壊れてるの？タイプですか？何が起こってる？

起こった通りのこと、答えは no です。タイプではなく、JASP が間違っているわけでもありません。現に、JASP がここで何をしているのかを説明するのはとても簡単なのですが、JASP がなぜそれをしたのか、というのはちょっと説明に苦労します。ですから“何が起ったのか”から始めましょう。JASP は上で示したのとは少し違う数式を評価したのです。偏差平方の平均を計算したのではありません。平均はデータ点の数 N で割りますが、JASP は $N - 1$ で割ったのです。

言い換えると、JASP は次の式を使って計算したのです。

$$\frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$$

これが何をやったかです。本当に知りたいのは、なぜ JASP が N ではなく $N - 1$ で割ったのか、ですよね。結局のところ、分散は偏差平方の平均なのですよね？だったら N で割るべきじゃないか、サンプルの実際の観測数でね。全くその通りです。しかし、第 ?? 章で論じるように、“サンプルを記述する”ことと“サンプルのもとになった母集団を推測すること”とのあいだにはちょっとした違いがあるのです。ここまででは、この差の区別をしてきませんでした。あなたが表現したいのがサンプルなのか、母集団の推測するものなのかどうかにかかわらず、平均は同じように計算できたのです。しかし分散や標準偏差、そのほかの尺度ではそうならないのです。私が最初に説明したこと（つまり、

N で割ることによる実際の平均)は、標本の分散を計算することを想定したものでした。しかしほとんどの場合、標本そのものに興味をもってるわけではないでしょう。むしろ、その標本は世界について何かを伝えるために存在しているはずです。そうであれば、あなたが実際に計算したいのは“標本統計量”ではなくて、“母集団の母数”を推定するためのものになるはずです。しかしこの話は、少し先走りすぎています。今は、JASP がすることをただ信じて、第 ?? 章で推定について論じるときまでこの問題をおいておくことにしましょう。

最後にもう一つ。このセクションはちょっとした推理小説のようになっていました。先ほど分散の式を示し、JASP では “ $N - 1$ ” でやっていること、そしてなぜそうするのかのヒントを書きましたが、最も大事なことは触れていなかったのです。みなさんは分散をどういうものだと理解していますか? 記述統計は記述することだけを目的としていますが、今のところ分散は意味不明な数字しかありません。残念なことに、分散の解釈について人間味のない説明しかできない理由は、それがそもそも人間味のないものだからです。これが分散について最も深刻な問題点です。分散は本当は変動を表現する基本的な量であるというある種の美しい数学的特性はあるのですが、現実的に他者との会話に使いたいと思うときには全く役に立たないです。分散は元の変数に関しては全く意味のない数字になります! 全ての数字は二乗されてしまうので、それは何も意味しないことになるのです。これは大問題だ。例えば、以前示した表について言うと、ゲーム 1 における点差は “376.36 ポイントの二乗分、平均より高い” と言うことになります。これはまったく馬鹿馬鹿しい表現ではないですか。計算した分散の 324.64 の時も同じことがいえます。多くのフットボールゲームを見てきましたが、誰も “ポイントの二乗分” なんて言ってるのを聞いたことがありません。これは測定の実際の単位ではなく、分散は意味のない単位を持っているので、人間にとて全く意味のないことになるのです。

3.2.5 標準偏差

オーケイ、分散を使う理由は分かってもらえたとしましょう。説明はしませんが、分散は数学的に良い特性持っていますからね。でもあなたが人間で、ロボットでないなら、データと同じ単位を持っている(つまり二乗した値ではないもの)尺度を使う方がいいと思うでしょう。じゃあどうしましょう? 答えは簡単です! 分散の平方根を取れば良くって、これは**標準偏差**として知られています。“偏差平方平均の根”，つまり RMSD とも呼ばれます。これで問題がスッキリ解決しました。だれも “分散は 324.68 ポイントの二乗” ということの意味を理解することはできませんが，“標準偏差 18.01 ポイント” は簡単に理解できます。元の単位で表現されているんですから。

標準偏差は分散の正の平方根に等しいので、次の式を見ても驚かないと思います。

$$s = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2}$$

JASP では、‘分散’のチェックボックスと同じセクションに‘標準偏差’のチェックボックスもあります。図 3.8 をみると、JASP は `afl.margins` の標準偏差を `26.074` と答えてくれています。標準偏差はとてもよく使われる所以、チェックするのがデフォルトになっていますが、あなた自身で選んでみてください!!

しかし、分散についての議論でお気づきかもしれません、JASP は実際にはこれとちょっと違ったやり方で計算します。分散を見るだけなら、JASP は N ではなく $N - 1$ で割る方で計算するのです。

第 ?? 章で再びこのトピックに触れると意味がわかると思いますが、この新しい量を $\hat{\sigma}$ (“シグマ・ハット”と読みます) とし、次のように定式化します。

$$\hat{\sigma} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2}$$

標準偏差を解釈するのも少し複雑です。標準偏差は分散から導出されています。そして分散は人にとてあまり意味のない量になっていますから、標準偏差は単純な解釈では済みません。結果的に、私たちのほとんどはちょっとした経験則を用いています。一般的に、平均から標準偏差 1 つぶん離れたところにデータの 68% が含まれ、データの 95% は平均から標準偏差 2 つ分離れたところに 99.7% が、平均から標準偏差 3 つ分離れたところに含まれる、ということが期待できます。このルールはほとんどの場合うまく当てはまりますが、多少の例外はあります。これがちゃんと計算できるのはヒストグラムが対称的で“ベル型”になっているという仮定に基づいています^{*4}。図 3.2 にある AFL の得点差ヒストグラムを見ると、この経験則は私たちのデータに合っているとは思えません! しかし大まかに合っているのです。AFL データの 65.3% が実際に平均から 1 標準偏差の範囲にあります。This is shown visually in Figure 3.9. このことは、図 3.9 に視覚的に示されています。

3.2.6 どの尺度を使いましょうか?

いくつかの範囲についての尺度を紹介してきました。範囲、IQR、平均絶対偏差、分散、標準偏差

^{*4}厳密にいうと、この仮定はデータが正規分布にしたがっているということで、この重要な概念については第?? 章で議論することになります。またこのことは本書で何度も何度も出てきます。

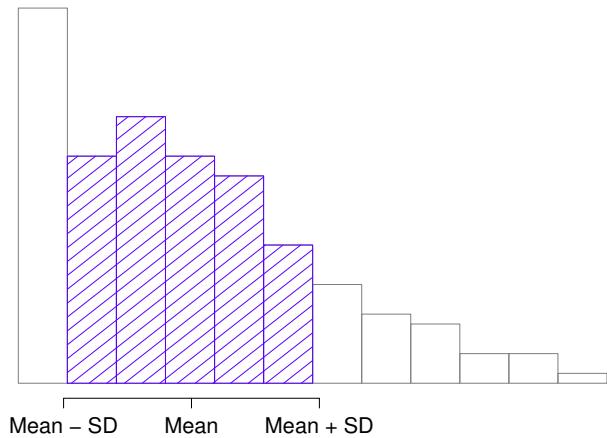


Figure3.9 AFL 得点差データについての標準偏差を描いたもの。色がついているヒストグラムの箇所は平均から 1 標準偏差のなかに入ったデータの数を表しています。今回は 65.3% のデータセットがこの範囲内に入り、次のメイントピックスである“約 68%”のルールに近い結果になっています。

です。そしてその長所と短所についてもみてきました。簡単にまとめておきましょう。

- 範囲 データのちらばり全体を見ます。外れ値に弱く、データの極端な部分を見たいという理由がない場合はあまり使われることはありません。
- 四分位範囲 データの“真ん中あたり”がある場所を教えてくれます。多少、外れ値に強くて中央値を含んでいます。これはよく使われます。
- 平均絶対偏差平均から観測度数が“平均的に”どれくらい離れているかを教えてくれます。解釈しやすいのですが、いくつかの小さな問題点があって（ここでは触れていませんが）、そのせいで統計家は標準偏差ほど魅力を感じていません。時々使われますが、それほど頻度はありません。
- 分散 平均偏差の二乗の平均です。数学的にはエレガントで、平均周りの散らばりを描写するにはたぶん“正しい”方法なのですが、データと同じ単位を使っていないので意味不明な数字になります。数学的なツール以外の用途はほとんどありませんが、非常に多くの統計技法の中に“埋もれて”います。
- 標準偏差分散の平方根です。これは数学的にも非常にエレガントで、データと同じ単位で表現されていますから、解釈も簡単です。平均が中心化傾向の尺度として使われる時は、これが基本です。散らばりの尺度の中で最もポピュラーなものになります。

まとめると、IQR と標準偏差が簡単で、データのばらつきを報告するのに最もよく使われる二大尺度、ということになります。しかし他のものが使われることもあります。この本に載せたのは、わずかではありますがみなさんがどこかで出会うかもしれませんからです。

3.3

歪度と尖度

みなさんが心理学の文献で見かけるかもしれませんる記述統計量が、あと二つあります。歪度と尖度です。実践上はどちらもこれまで話してきた中心化傾向や変動性の尺度ほど、使われるものではありません。歪度はちょっと大事なので見かけることはあるかもしれません、私は科学的レポートの中で尖度を目にしたことはありません。

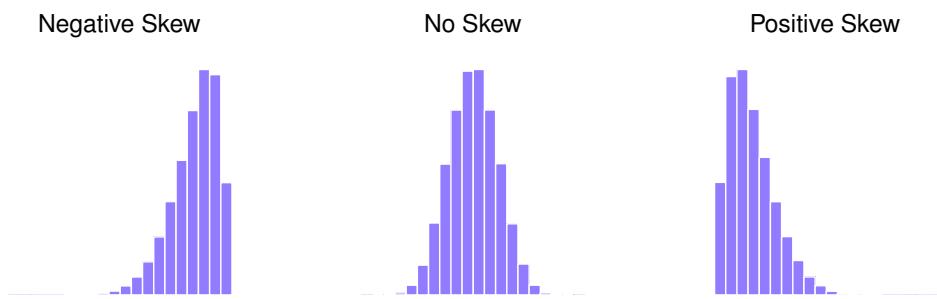


Figure 3.10 歪度のイメージ。左側は歪度が負 (歪度 = -.93), 真ん中は歪みなし (実際ほとんどありません。歪度 = -.006), そして右が正の歪度 (歪度 = -.006) をもつデータです。

.....

歪度の方が面白いので、こちらから話を始めましょう。歪度は基本的に非対称性の尺度で、図を書いてみれば理解は簡単です。図 3.10 にあるように、データに極端に小さな値（下の裾が上の裾よりも“長い”）を持っていて、極端に大きな値はそれほど持っていない（左図）場合、このデータは負の歪度をもつといいます。一方、極端に大きな値が小さい値より大きく多くあるようであれば（右図）、このデータは正の歪度をもつといいます。これが歪度の背後にある考え方です。平均よりも大きな値が相対的に多くあれば、分布は正、すなわち右に歪んでおり、裾も右に寄っています。負、すなわち左への歪みはその逆です。対称的な分布をしていれば、歪み度は 0 です。正に歪んだ分布の歪度は正の値であり、負の値は負の歪み分布だと言えます。

データセットの歪みについての定式化は次のとおりです。

$$\text{skewness}(X) = \frac{1}{N\hat{\sigma}^3} \sum_{i=1}^N (X_i - \bar{X})^3$$

ここで N は観測度数の数であり、 \bar{X} は標本平均、 $\hat{\sigma}$ は標準偏差（ただし “ $N - 1$ で割ったバージョン”）です。

ありがたいことに、JASP で歪度の計算することができます。'記述' - '記述統計量' の下にある '統計量' チェックボックスのオプションがそれです。変数 `afl.margins` について、その歪度を計算すると [0.780](#) です。この歪度の推定値を歪度の標準誤差で割れば、このデータがどれほど歪んでいるかの指標を得ることができます。経験的に行って、小さいサンプルでは ($N < 50$)、この値が 2 以下であればそれほど歪みは大きくなく、2 以上であればデータが統計的な分析をするに許される限界を超えて歪んでいる、と考えるのが目安です。これは経験則に過ぎず、この解釈にはっきりした共通見解があるわけではないことに注意してください。ということで、この分析をすると AFL の得点差データはちょっと歪んでいることになります ($0.780 / 0.183 = 4.262$ で、これは明らかに 2 より大きいです)。

時々つかわれる最後の尺度は、実際に使われることは非常に稀なのですが、データセットの尖度です。簡単にいえば、尖度は“尖っているかどうか”的尺度で、図 [3.11](#) にその状況を示しています。慣例によって、“正規分布”(黒い線) は尖度ゼロであり、データセットの尖り具合はこのカーブに比べて相対的に評価されます。

この図にあるように、左のデータはそれほど尖っておらず、尖度は負でこのデータは緩く尖った *platykurtic* データだと言われます右図はとても尖っており、尖度は正でこのデータは尖度の大きい *leptokurtic* データだと言われます。一方、真ん中のデータはちょうどいいぐらいの尖度で、これは中程度の尖度 *mesokurtic* と呼ばれ、尖度はゼロです。下の表にこれをまとめました。

| 一般的な言い方 | 専門的な言い方 | 尖度の値 |
|-------------|-------------|------|
| “かなりフラット” | platykurtic | 負 |
| “ちょうどいいぐらい” | mesokurtic | ゼロ |
| “とても尖っている” | leptokurtic | 正 |

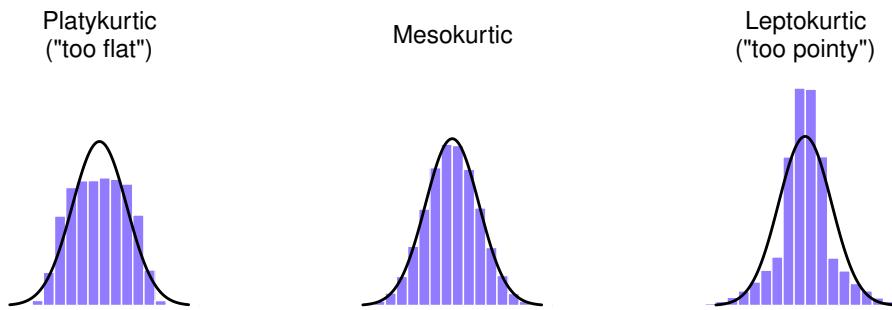


Figure 3.11 尖度の図。左側は“緩く尖った”データセット（尖度 = -.95）であり、これが意味するのはこのデータセットは“かなりフラット”だということです。真ん中の図は“中程度の尖り”をもったデータセット（尖度はほとんど 0）であり、これが意味するのはこのデータの尖度がちょうどいい感じであるということです。最後に、右側の図ですが、“尖度の大きい”データセット（尖度 = 2.12）であり、このデータセットは“とても尖っています”。尖度は正規分布（黒い線）と比べて評価されていることに注意してください。

尖度の式は既に見た分散や歪度の式とかなり似ています。分散が偏差の二乗、歪度が偏差の三乗であったのに対し、尖度は四乗になっています。^a

$$\text{kurtosis}(X) = \frac{1}{N\hat{\sigma}^4} \sum_{i=1}^N (X_i - \bar{X})^4 - 3.$$

^aこの“-3”については正規分布の尖度がゼロになるように統計家が付け加えたものです。“-3”を式の最後に引っ付けておくのはちょっと馬鹿みたいですが、こうすることの数学的な理由があるのです。

大事なのは、JASP で尖度を計算するには歪度の下のチェックボックスをクリックするだけだということで、そうすると尖度の値 **0.101** がその標準誤差 **0.364** と共に表示されます。歪度をその標準誤差で割ったのと同じように計算すると、この値は 2 より小さい (**0.101 / 0.364 = 0.277**) ことがわかります。これは AFL の得点差データの尖度がちょうどいいぐらいだったことを意味しています。

3.4 _____

グループごとの記述統計

よくあることのひとつとして、記述統計量があるグループ変数ごとに分割してみたいと思うことがあります。JASP ではすごく簡単にできます。例えば、ある `clin.trial` データについて、`therapy` のタイプごとに記述統計量を見たいなと思ったとしましょう。これは今まで見せていない、新しいデータセットです。このデータセットは `clinicaltrial.csv` ファイルにあって、第 ?? 章でよく使うようになります（このデータの詳細についてはその時に説明します）。読み込んで、見てみましょう。

| | ID | drug | therapy | mood.gain | |
|----|----|----------|------------|-----------|--|
| 1 | 1 | placebo | no.therapy | 0.5 | |
| 2 | 2 | placebo | no.therapy | 0.3 | |
| 3 | 3 | placebo | no.therapy | 0.1 | |
| 4 | 4 | anxifree | no.therapy | 0.6 | |
| 5 | 5 | anxifree | no.therapy | 0.4 | |
| 6 | 6 | anxifree | no.therapy | 0.2 | |
| 7 | 7 | joyzepam | no.therapy | 1.4 | |
| 8 | 8 | joyzepam | no.therapy | 1.7 | |
| 9 | 9 | joyzepam | no.therapy | 1.3 | |
| 10 | 10 | placebo | CBT | 0.6 | |
| 11 | 11 | placebo | CBT | 0.9 | |
| 12 | 12 | placebo | CBT | 0.3 | |
| 13 | 13 | anxifree | CBT | 1.1 | |
| 14 | 14 | anxifree | CBT | 0.8 | |
| 15 | 15 | anxifree | CBT | 1.2 | |
| 16 | 16 | joyzepam | CBT | 1.8 | |
| 17 | 17 | joyzepam | CBT | 1.3 | |
| 18 | 18 | joyzepam | CBT | 1.4 | |

Figure3.12 `clinicaltrial.csv` ファイルにある変数を写した JASP スクリーンショット

三つのドラッグがあるのがわかりますね。プラセボと、“anxifree”と“joyzepam”と呼ばれるものです。そしてそれぞれに 6 人割り当てられています。そして 9 人が認知行動療法 (CBT) を受けています。

て、9人が心理療法は何も受けていない状態です。そして `mood.gain` 変数の‘記述’をみてみると、ほとんどの人が気分の向上(平均 = 0.88)を示していますが、この尺度が何なのかわからないまでは、それ以上のことは言えません。でも、それはそれでわるくないのです。全体的には何か勉強になった気になります。

さて、さらに他の記述統計量を見て行きましょう。こんどはセラピーのタイプごとに分けて。JASPで‘統計量’オプションから標準偏差、歪度、尖度にチェックを入れます。同時に、`therapy` 変数を‘分割’ボックスに入れます。すると図 3.13 のような結果が得られます。

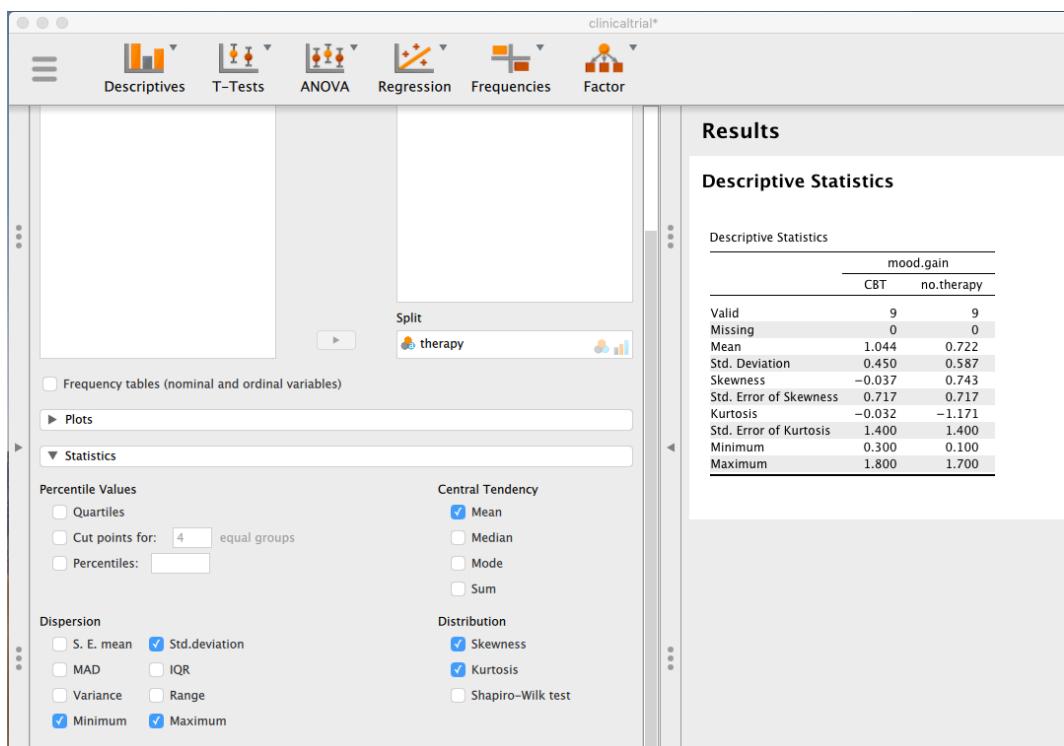


Figure 3.13 セラピータイプごとに分割した記述統計量を示した JASP のスクリーンショット

3.5

標準得点

私の友人が“不機嫌さ”を測定するための新しい質問紙を作ろうとしているとしましょう。この調査票は 50 の質問からなり、不機嫌かどうかについて答えるものとします。大きなサンプルをとって

(仮に百万人ぐらいとったとしましょう!), このデータが正規分布しており, 50 問中 17 点が平均不機嫌スコアで, 標準偏差が 5 だとしましょう。これに比べて, 私の得点は 50 問中 35 点だったとします。私はどれぐらい不機嫌なんでしょうね? これについて考える一つの方法は, 私は 35/50 が不機嫌なのだから, 70% ぐらい不機嫌だと考えることです。しかしちょっと考えてみれば, おかしい気もしますよね。もし私の友人が, その質問紙を少し違った捉え方で答えていたとしたら, その問い合わせ本当に問うていることに比べて, 全体的な分布が簡単に上がったり下がったりしてしまいます。ですから, 私が 70% 不機嫌だというのは, 調査票の質問セットに応じて変わることになります。とても良い質問項目であったとしても, これではあまり意味のある表現にはなりません。

これについての良いやり方の一つは, 私の不機嫌の程度を周りの人と比べることです。驚くべきことに, 私の友人は 1,000,000 人のサンプルを持っていて, その中でたった 159 人だけが私と同じ程度の不機嫌さ (本当ははそなことありませんよ) であれば, 私はトップ 0.016% の不機嫌度ということになります。このほうが, ロウデータを解釈しようとする時にはより意味があるのではないでしょうか。この考え方は, 私の不機嫌さの程度を人の全体的な不機嫌分布にあわせて記述しようとするものであり, 標準化がしようとしているのはまさにこれなのです。これを正しくやる方法の一つは, さっきやって見せたように, パーセンタイルで表現することです。しかし問題があるのは, この方法だと “トップが寂しい” ということです。私の友人が集めたサンプルが 1000 人に過ぎなかったとしましょう (これでもまだ新しい質問紙を検証するためには大きいサンプルですが)。そして今回, 平均が 50 問中 16 点で標準偏差が 5 だったとします。問題は, このサンプルでは私と同じぐらいの不機嫌度を持っている人が一人もいないということです。

しかし, 全てが失われたわけではありません。もう一つのアプローチとして, 私の不機嫌スコアを **標準スコア** に変換するのです。これは z -スコアとも言われています。標準スコアは私の不機嫌スコアが平均から標準偏差いくつ分上にあるかを表すのです。これを “数学っぽく” いうと, 標準偏差は次のように計算できます。：

$$\text{標準スコア} = \frac{\text{ロースコア} - \text{平均}}{\text{標準偏差}}$$

実際数学的には, z スコアについての式は次のようになります。

$$z_i = \frac{X_i - \bar{X}}{\hat{\sigma}}$$

不機嫌さデータに戻っていようと, ダニーの生の不機嫌さデータを標準化された不機嫌スコアに変換することができます。

$$z = \frac{35 - 17}{5} = 3.6$$

この値を解釈するときに, セクション 3.2.5 で触れた, 平均から 3 標準偏差範囲にだいたい 99.7%

が入るという、概算を思い出してください。ですから、私の不機嫌さスコアを z スコアにして 3.6 になったということは、実際私はかなり不機嫌状態にあるということです。実際この推論からいくと、私は全体の 99.98% の人よりも不愉快なのです。そうですよね。

ロースコアをより大きな母集団に広げて解釈することを許すとするなら（そしてそれによって任意の尺度の変数を意味のあるものにするなら）、標準スコアは第二の便利な機能を持っていると言えます。標準スコアはロースコアができないような状況でも互いに比較することができます。たとえば、私の友人が 24 項目からなる外向性を測る別の質問紙を持っていたとしましょう。この尺度が全体的に、平均が 13 で標準偏差 4 であり、私のスコアが 2 だとたつとしましょう。想像の通り、私の外向性のロースコア 2 を、不機嫌さ質問紙のロースコア 35 と比較するのは意味がありません。この二つの変数のロースコアは基本的に違うもので、いわばりんごとオレンジを比較するようなものです。

標準スコアではどうでしょう？ これはちょっと事情が違います。標準スコアを計算すると、不機嫌さは $z = (35 - 17)/5 = 3.6$ 、外向性は $z = (2 - 13)/4 = -2.75$ となります。この二つの数字は相互に比較することができます^{*5}私はほとんどの人の中では外向性が低く ($z = -2.75$)、不機嫌さが高い ($z = 3.6$) のです。しかし私のハズレ具合は外向性よりも不機嫌さの方が大きいといえます。3.6 が 2.75 よりも大きな数字だからです。それぞれの標準化スコアはその観測値がその母集団においてどのあたりに落ちるのかを示すので、全く異なる変数についても標準スコア同士を比較することができるのです。

3.6 _____

要約

基礎統計量を計算することは、あなたが実際にデータを取ったとき真っ先にすべきことの一つであり、記述統計量は推測統計よりも単純で理解しやすいので、他の統計の教科書と同じように私も記述統計から説明しました。この章では、以下のトピックスについて議論しました。

- 中心化傾向の指標 一般的に、中心化傾向はデータがどのあたりにあるのか教えてくれます。典型的に報告される指標は次の三つでしょう；平均値、中央値、最頻値です（セクション Section 3.1）。
- 変動の指標 それに対して、変動の指標はデータがどのように“散らばっているか”を教えてくれます。鍵になる指標としては、次のものがあるでしょう；範囲、標準偏差、四分位範囲です（セクション Section 3.2）。

^{*5}いくつかの注意は必要です。変数 A についての 1 標準偏差が、変数 B の 1 標準偏差と“ある意味”対応しているとは言えないからです。二つの変数に関する z スコアが意味のある比較ができるかどうかを決めるには、常識をはたらかせねばなりません。

- 歪みと尖りの指標 変数の分布が非対称さの指標 (歪度) と、尖り具合 (尖度) もみてきました (セクション 3.3)。
- JASP で群ごとに変数の要約をする この本では JASP でデータ分析をすることに焦点化していますから、異なるサブグループそれぞれについて記述統計量を計算するにはどうするかについても触れました (セクション 3.4)。
- 標準化スコア z -スコアはちょっと変わった野獣です。これは記述統計量とはちょっと違いますし、推測統計の話でもありません。これについてはセクション 3.5 で触れました。この章も理解してもらえたと思います。また後で触れることになります。

次の章では、どうやって絵を描くのかについての話題に移りたいと思います! 誰だって可愛い絵が好きですもんね? しかしその前に、重要な点を抑えておきたいと思います。統計の伝統的な入門コースは、記述統計について小さな配分しかせず、1,2 回授業で触れる程度です。授業時間のほとんどの時間は、推測統計学に使われます。というのも、そこが本当に大変なところだからです。それはそれで意味があるのですが、良い記述統計量を選択するという、日々の重要な実践を覆い隠してしまいます。このことを覚えておいて欲しいのです…

3.6.1 エピローグ: 良い記述統計量とは記述的である!

一人の死は悲劇である。

数百万の死は統計である。

– Josef Stalin, Potsdam 1945

$950,000 - 1,200,000$

– ソ連における弾圧の死者数,
1937-1938 (**Ellman2002**)

スターリンの悪名高き、数百万人の死に関する統計の特性についての引用は、少し考えてみる必要があります。彼の主張意図は明らかに、個々人の死は我々の心に触れ、無視することはできないけれども、非常に多くの死については理解できないし、結果的に単なる統計であって、無視してしまうことも簡単である、というところにあります。スターリンは、半分は正しいと思います。統計というのは抽象化であり、個々人の経験を超えた出来事の記述であり、可視化されにくいものです。百万人の死が“本当に”どういうことなのかを想像できる人はほとんどいませんが、一人の死は簡単に想像できますし、孤独な死は悲劇の感情を呼び起こし、Ellman の冷たい統計的記述の感覚が失われたように感じます。

これはそんなに簡単な話ではないのです。数字がなければ、数えなければ、何が起こったのかの記述がなければ、われわれは本当に何が起こったのかを理解する機会すらもてず、この失われた感覚を

呼び起こす機会さえ持つことができません。そして実際には、私はこれを気持ちの良い土曜日の朝に腰掛けながら書いており、世界の半分そしてこれまでの人生でずっと、ソ連の強制収容所から離れたところにいるのですが、Ellman の推定値とスターリンの引用を書く時には鈍い恐怖がズッシリ胃にきて、寒気を覚えます。スターリン主義の弾圧は私の経験を超えたところにありますが、統計データと結びつき、そこに記録された個人史を思うと、私の理解を完全に超えているとはいえないません。なぜなら、Ellman の数字は私たちに教えてくれるからです。2年以上のスターリンの弾圧によって、私の住んでいる街に今生きている全ての男性、女性、子供たちと同じ数の人が消え去ってしまったのだ、ということを。この死の一つ一つに、独自の物語があって、それぞれの悲劇があって、その幾らかは私たちにも知られています。ですから、注意深く選ばれた統計量を見ながら、残虐行為のスケールに焦点化していきましょう。

統計家と科学者の最初の仕事である、データを集めて要約し、何が起ったのかを聴衆に知らせる数字を見つけてくるというのは、簡単なことではないのです。これは記述統計の仕事ですが、数字だけを使って何が言えるかはその仕事ではありません。あなたはデータアナリストであり、統計ソフトパッケージではないのです。あなたがすべきはこれらの統計量を取り出して、記述に持っていくことです。あなたがデータを分析するとき、数字のコレクションをリストアップするだけでは十分ではありません。忘れてはいけないのは、あなたは人間の、聴衆を相手にコミュニケーションしようとしているということです。数字は重要ですが、あなたの聴衆が理解できるような意味のあるストーリーと一緒にでなければなりません。あなたはフレーミングについて考える必要がある、ということです。文脈について考えなければなりません。あなたの統計量が要約した、一つ一つの出来事について考えなければなりません。

4. グラフを描く

何よりもまず、データを見せろ

—Edward Tufte^{*1}

データを可視化することは、データを分析しようとするものにとって最も重要な課題です。これが重要なのは、二つの異なる、しかし相互に関係し合う理由によります。まず、“提示するグラフ”を描くこととは、あなたのデータをスッキリと提示し、読者にとってあなたが言いたいことを簡単に理解させるために視覚的に訴えかけるようにすることです。同じぐらい、あるいはもっと重要なことは、グラフを描くことであなた自身がデータを理解できるようになることです。そのために、“探索的なグラフ”を描くことは、あなたがいざ分析しようとしているデータについて理解するのを助けることになるのが重要なのです。このことは当たり前のようでもありますが、私はこれを人に何回言ったかわからないほどです。

この章の重要さを示すために、優れたグラフというものがいかに有用なのかを示す典型例から始めたいと思います。そのために、図 4.1 に最も有名なデータの可視化の例の一つを示しています。これは 1854 年、John Snow によるコロナの死亡者数の地図です。この図はその単純さにおいて、非常にエレガントだといえます。背景として、われわれは見る人の方向性を示すストリートマップを持っている、というのがあります。地図上には多数の小さな点があり、それぞれがコロナの発祥地点を表しています。大きな文字は水のポンプの位置を示していて、その名前ラベルがついています。この図をちょっと見ただけでも、アウトブレイクの源は Broad Street ポンプを中心にしていることが明らかです。このグラフを見て、Dr.Snow はポンプからハンドルを取り除き、500 人以上を殺したアウトブレイクを終わらせたのです。これが、良いデータの可視化の力です。

この章の目標は二つあります。まず、データを分析したり表示したりするとき、私たちがよく使うグラフについて説明し、続いてこれらのグラフを JASP で作成するにはどうすれば良いかを示します。このグラフそのものは、直接的なものなので、この章のある側面は非常にシンプルだと言えるでしょう。人がよく困惑するのは、グラフをどうやって作るかを学ぶとき、特に良いグラフをどうやっ

^{*1} この言葉の原典は、Tufte の本『量的情報を可視化する』です。

て作れば良いかを学ぶときです。幸い、JASPでのグラフの書き方は、あなたがグラフの見え方にそれほどこだわらなければ、かなりシンプルなものです。私がこれをいうことの意味は、JASPのデフォルトのグラフがかなり良いものだということで、ほとんどの場合すっきりとクオリティの高いグラフィックを提供できるということです。しかし、標準的でない図を描きたいとあなたが思ったとき、あるいは図にかなり特殊な変更を加える必要があるとき、JASPのグラフィック関数は発展的な仕事や詳細な編集にはまだ向いていないということはあります。

Snow's cholera map of London

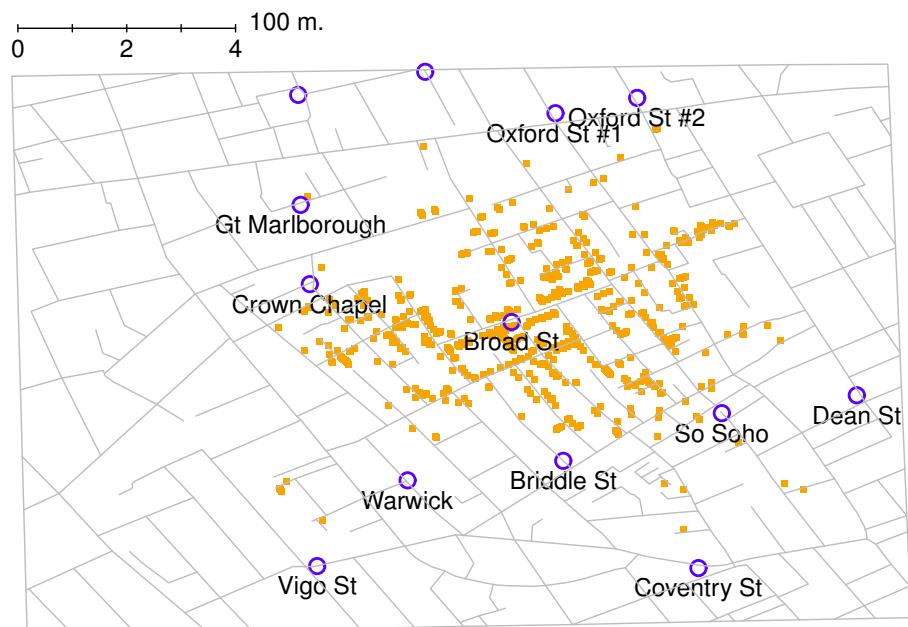


Figure4.1 John Snow のスタイリッシュなコロナマップのオリジナル。小さな各点はコロナ発生点で、大きな円は井戸の位置を示しています。このプロットが明らかにしたように、コロナのアウトブレイクは Broad St のポンプを中心にしていることがわかります。

ヒストグラム

普通のヒストグラムの話から始めましょう。ヒストグラムは最もシンプルで最も一般的な、データ可視化手法の一つです。あなたが間隔尺度水準、あるいは比率尺度水準のデータ（例えば、第3章の `afl.margins` データなど）を持っていて、その辺図宇野全体的な印象を掴みたいと思った時に、ヒストグラムは有効です。ヒストグラムがどんなものかは、ほとんどの人が知っていると思います。広く使われていますからね。でも完璧を期すために、しっかり説明しておきます。あなたがすべきことは、あり得る値をビン幅に分割し、各区間にに入る観測度数の数を数え上げることだけです。この数のことを頻度とかビンの密度といい、それが垂直に伸びるバーとして表示されます。AFLの勝利数データでは、得点が10点未満だったゲームが33ゲームあり、これが以前示した第3の図3.2中、左端のバーの高さとして表されています。以前のグラフはJASPの能力を超えた、Rの発展的プロットパッケージの力を使って描かれていました。しかしJASPもそれに近いことをしてくれます。JASPでのヒストグラムの描画はとても簡単です。「記述」-「記述統計」メニューの下にある「プロット」をひらき、「分布のプロット」チェックボックスをクリックしたのが、図4.2に示されています。JASPのデフォルトでは、y軸が「度数」とラベルされていて、x軸が変数名になっています。ビンは自動的に選択されます。度数が表示されますが、実際の数字はそれほど問題にならないことに注意してください。むしろ、われわれが本当に興味を持っているのは、分布の形状からくる印象なのです。それが正規分布しているのか、それが尖っていたり歪んでいたりしないか？私たちの第一印象は、ヒストグラムから作られるのです。

JASPの特徴を一つ付け加えるなら、「密度」曲線をこのヒストグラムの上に書き加えられるというところです。これをするには「プロット」の下にある「密度を表示」のチェックボックスをクリックしてください。これが図4.3に示されているプロットです。密度プロットは連続した区間や時系列全体をカバーする分布を可視化します。この図は、プロットされた値にカーネルスムージングを使ったヒストグラムの一種で、ノイズを除去した平滑化によって分布をよりスムーズにしたものです。密度プロットのピークは、区間中の値がどこに集中しているかを示してくれています。ヒストグラムの上に密度プロットを描くことの利点は、分布の形をわかりやすくすることにあります。なぜならこれはビン（ヒストグラムで使われている各バー）の数に影響されないからです。たった4つのビンしかないヒストグラムは、20のビンをもつヒストグラムに比べて分布の形をうまく表現できません。でも密度プロットでは、そういう問題が生じません。

この画像はプレゼンテーション用のグラフィック（例えばレポートに入れるもの）にするには、かなり修正する必要がありますが、データを描画する分にはかなりいい仕事をしてくれます。実際、ヒストグラムや密度プロットの強みは（適切に使えば）、データの全体的な広がりを表示し、それがどんな形をしているのかについてかなり良い直感を与えてくれることです。ヒストグラムの欠点は、コンパクトさに欠けるところです。他のプロットと違って、20から30ものヒストグラムを一つの図に詰め込んで人に説明するのはとても難しいのです。そしてもちろん、データが名義尺度水準であれば

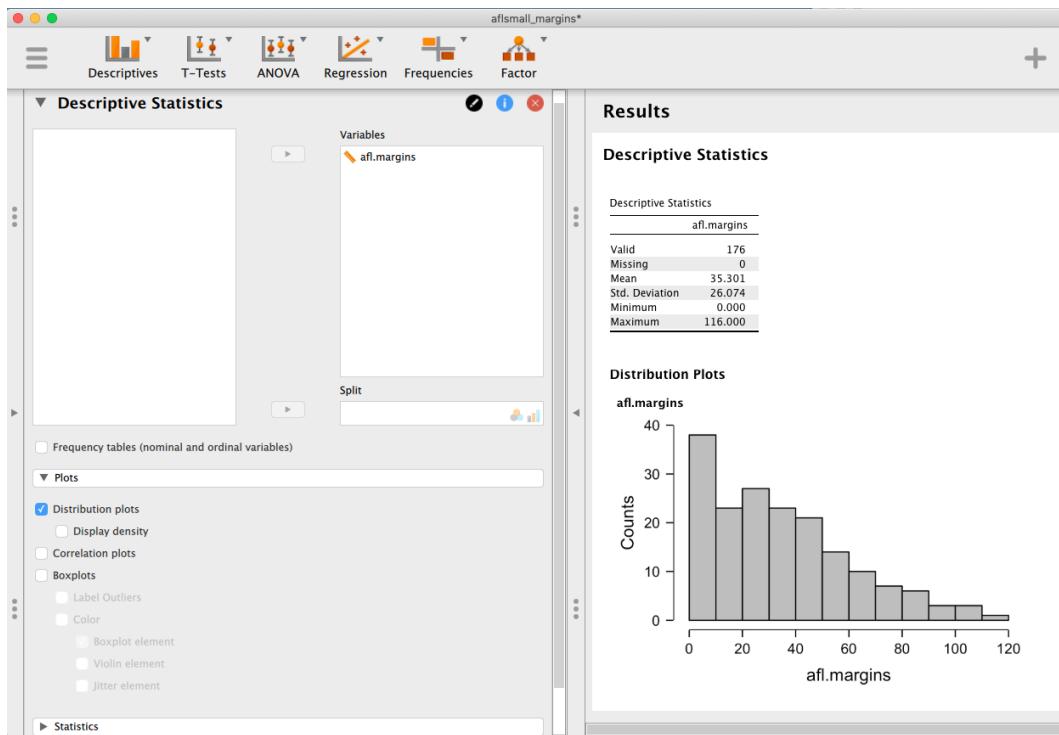


Figure4.2 ‘分布のプロット’ オプションによって作られたヒストグラムを描いた JASP のスクリーンショット

ヒストグラムは適用できません。

4.2

ボックスプロット

ヒストグラムの代わりになるのは、**ボックスプロット**、別名“箱ヒゲ図”と呼ばれるものです。ヒストグラムのように、間隔あるいは比率尺度水準のデータに適しています。ボックスプロットの背後にある考え方とは、中央値、四分位範囲、データの幅を単純に示して見せようというものです。ボックスプロットによる表現は非常にコンパクトで、特にデータ分析の探索的な段階でデータがどんなものかを理解しようとする時の手法としてとてもポピュラーなものになっています。ではそれがどういうものか、`afl.margins` のデータを例にしてみていきましょう。

ボックスプロットがどんなものかを見るために、まず描いてみるのがいいでしょう。‘ボックスプロット’をクリックすれば、右下に図 4.4 のようなものが示されると思います。デフォルトでは、JASP は最も基本的なボックスプロットを示します。このプロットを見れば、そこから何がわかるか

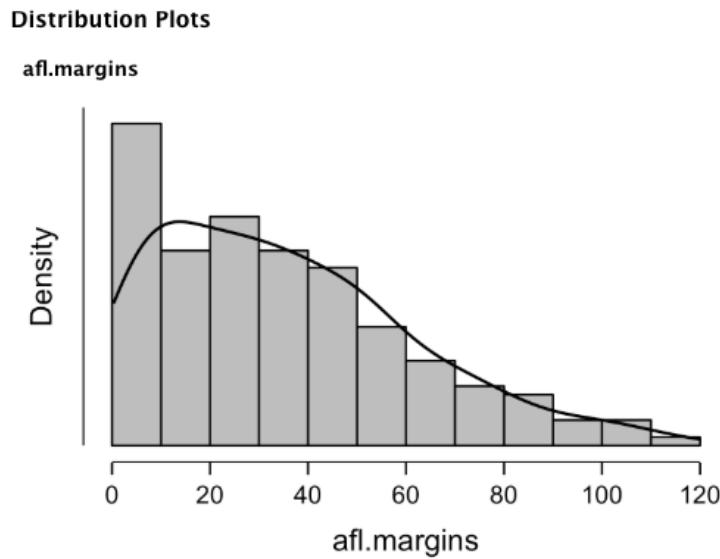


Figure4.3 afl.margins 変数の JASP による密度プロット

一目瞭然です。箱の中心にある太い線が中央値です。箱の幅は 25 パーセンタイルと 75 パーセンタイルの幅になっています。そして “ひげ” の部分はある限界値を超えない最も極端なデータポイントです。デフォルトでは、この限界値は四分位範囲 (IQR) の 1.5 倍で、下限は [25 パーセンタイル点 - \(1.5*IQR\)](#)、上限は [75 パーセンタイル点 + \(1.5*IQR\)](#) になっています。この範囲の外に入る点は、髭でカバーできないので円あるいは点で示され、これは一般的に外れ値とよばれます。私たちの AFL 勝率データでは、二つの観測点がこの範囲の外に落ちており、この観測データは点で表されています（上限は 107 で、スプレッドシートのデータをみると 2 件これより大きいもの、108 と 116 があり、それぞれの点が打たれています）。

4.2.1 Violin plots

伝統的なボックスプロットのバリエーションとして、バイオリンプロットというのがあります。バイオリンプロットはボックスプロットに似ていますが、異なる値におけるデータのカーネル確率密度も表示してくれます。典型的には、バイオリンプロットはデータの中央値と、標準的なボックスプロットと同じような四分位範囲を示すボックスも同時に示します。JASP では、この種の機能は ‘バイオリンの要素’ と ‘ボックスプロット要素’ のチェックボックスをチェックすることができます。図 4.5 では、データ点もプロットしました（これは ‘Jitter 要素’ のチェックボックスを選択することで、

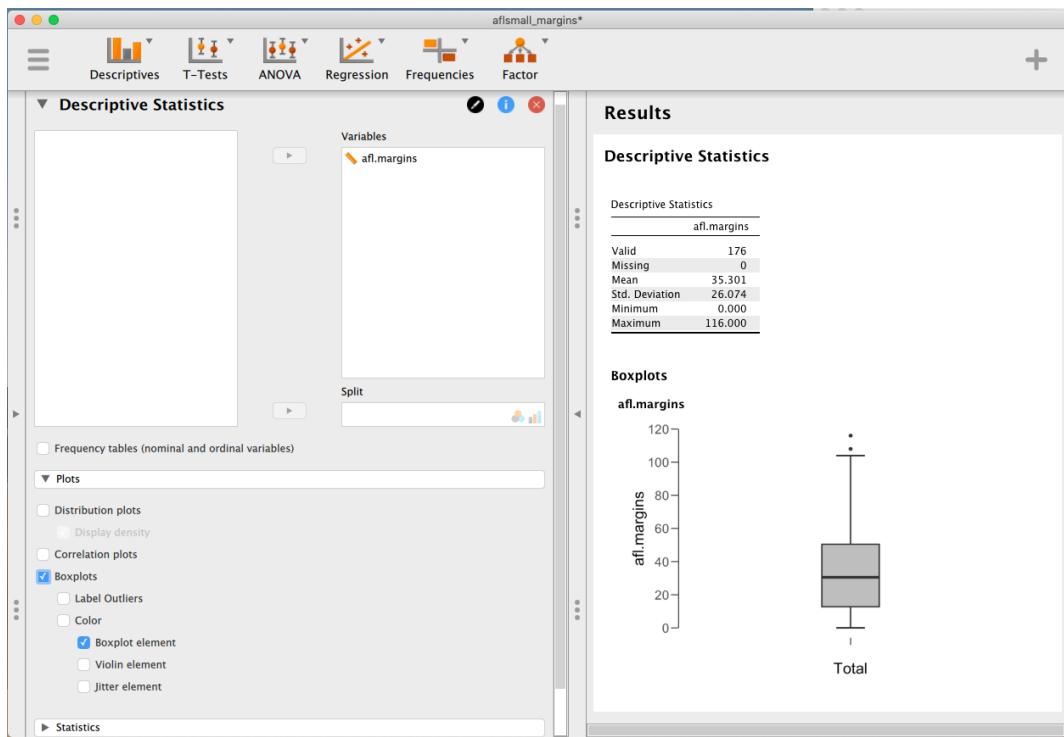


Figure4.4 JASPによるafl.margins変数のボックスプロット

プロットに実際のデータ点を追加します)。

4.2.2 複数のボックスプロットを描画する

最後にもう一つだけ。複数のボックスプロットを一度に書くにはどうしたらいいでしょう？ 例えば、2010年のAFL勝率データだけでなく、1987年から2010年までの各年度のボックスプロットを個別に描きたいと思ったとしましょう。これをするためには、まずデータを見つけなければなりません。このデータはaflsmall12.csvファイルにあります。ではJASPに読み込んで、みてみましょう。これはちょっと大きなデータセットであることがわかると思います。ここには4296ゲームとその変数が含まれています。JASPで**勝率**変数についてのボックスプロットを描く時に、**年度**ごとに分けたいですね。それをするためには、**年度**変数を名義尺度水準の変数に変換し、**年度**にわたってボックスの‘分割’をします。

その結果が図??です。このバージョンのボックスプロットは、年度ごとに分割されており、ヒストグラムよりボックスプロットを使った方がいいこともあるのはなぜか、ということがすぐにわかりますね。これを見ると、データの詳細に入り込まなくとも年度ごとにどうなっているか、わかりやすくなっています。もしこのスペースに24個のヒストグラムを詰め込もうとしたら、何が起こるか考

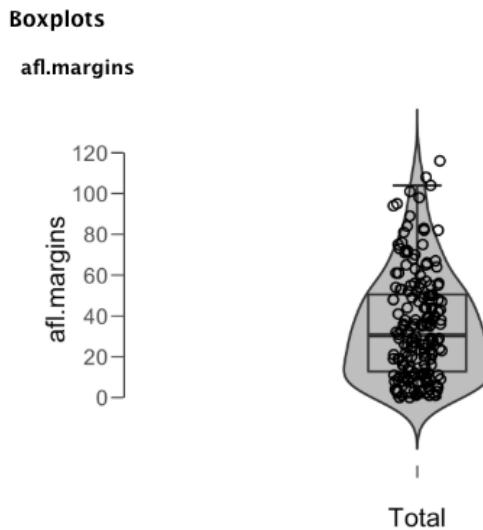


Figure4.5 JASPにおける[afl.margins](#)変数のバイオリンプロットにボックスプロットとデータ点も重ねてみました

.....

えてみてください。そんなことをしても、読者が何かを学べるとは思いませんけどね。

4.3 _____

JASPで画像を保存する方法

ちょっと待って、と思ってるかもしれませんね。JASPでいい図が欠けてもそれを保存したり友達に送り、私のデータがいかに素晴らしいかを語れないようでは意味がありません。図を保存するにはどうしたらいいでしょう？簡単です。プロットの上部、横についている三角形をクリックして、「名前をつけて画像を保存’を選ぶだけです。いくつかのフォーマットを選んで保存することができ、選択できる形式は‘png’, ‘pdf’, ‘eps’, ‘tif’があります。これらのフォーマットで友達に画像を送ったり、(もしかするとさらに重要なことには)それらを課題や論文に含めることができます。

4.4 _____