# 1. なぜ統計学を学ぶのでしょう?

「汝，質問紙調査に応えるべからず
世界情勢に関するクイズにも答えてはならぬ，　コンプライアンス上，いかなるテストも受けてはならない。統計学者のよこに座ることも社会科学に関わることも許されない。　"

– W.H. Auden[*1]

## 1.1 心理学における統計

ほとんどの学生が驚くのですが，統計学は心理学教育のとても重要な要素です。誰も驚かないことですが，統計学は心理学教育の中でほとんど好まれることのない要素です。つまり，もしあなたが本当に統計学を愛するのであれば，心理学のクラスではなく統計のクラスに今すぐ入るべきだということです。ですから，それほど驚くべきことでもないですが，心理学がその営みの中でかなり統計学を使っているという事実を，快く思っていないほうが学生の多数派です。ですから，統計について人か持っている一般的な疑問に答えるところから始めるのがいいだろうと思います。

この問題の大半は，統計学についての考え方に大きく関係します。それは何なの? 何のためにあるの? そしてなぜ科学者がそれに血眼になるの? 考えてみれば，いずれも良い質問です。最後の一つから始めましょう。全体として，科学者は何にでも統計的検定をかけることにこだわっているようです。事実，統計を使うときに，なぜわたしたちがそうするのかを人に説明するのを忘れてしまいがちです。まるで科学者間，特に社会科学者の間では，統計にかけるまでは自身の発見も信用できないと

[*1] この詩は Auden の 1946 年，*Under Which Lyre: A Reactionary Tract for the Times* が出典で，ハーバード大学の卒業式での演説の一部から持ってきました。この種のポエムの歴史は面白いですね。http://harvardmagazine.com/2007/11/a-poets-warning.html

いう信仰じみたものがあるようです。学生諸君は，誰も次の単純な質問に答えてくれないので，わたしたちが完全にトチ狂っているのではないかと思わずにいられません。

なぜ統計をするのですか？ なぜ科学者は 常識的に考えないんですか？

これはある意味素朴な質問ですが，良い質問でもあります。いい答え方もいくつかありますが[*2]，私に言わせると，ベストな答えは単純な次のものです: 私たちは私たち自身のことをそれほど信用してないのです。私たちは人間なので，あらゆるバイアス，誘惑，弱さから影響を受けてしまうことを心配しているのです。統計のほとんどは基本的に安全装置なのです。“常識” を使って証拠を表 k するというのは，直感を得ることを信用することを意味しますが，それは言語的な表現に依存しますし，正しい答えに辿り着くための人の理由づけに対する力そのものを利用するということです。ほとんどの科学者は，そのアプローチがうまく行くとは思っていません。

現に，そう考えることは心理学的な質問を投げかけているように思えます。そして私は心理学の大学で働いていますから，この少し根深い問題を掘り下げてみるのもいいかもしれないと思うのです。この “常識的に” 考えるというアプローチが信用に足る，という考え方は本当に妥当でしょうか? 言葉による表現は言語で構成されていて，全ての言語はバイアスを含みます–あるものは他のものよりも言いにくいですよね。それが間違っているわけでもないのに (例えば，量子電気力学は良い理論ですが，言葉で説明するのはとても難しい)。わたしたちの “肚に落ちる” という直感は，科学的な問題を解決するのには向いておらず，日々の推論のためにデザインされたものです–そしてこの生理学的な評価は文化的な変化よりも遅く，それらはわたしたちが今生きているのとは異なる世界における，日々の問題を解決するためにデザインされたもの，というべきでしょう。最も基本的なことですが，理由づけは人が “帰納的に” 考えるためのもので，うまく推論し，感覚的な証拠を超えて世界を一般化するためのものなのです。もしあなたが，自分は世界のさまざまな障壁から影響されることなく考えられるんだ，と思うのであれば，そうだな，ロンドンにかかっている橋を一つあなたに売ってあげますよ。次のセクションで説明するように，私たちは既に存在するバイアスからの影響を離れて，“演繹的な” 問題 (推論を必要としないもの) 解決をすることはできないのです。

### 1.1.1 信念バイアスの呪い

人は大抵の場合，とても賢いものです。私たちはこの星に共生する他のどの種よりも賢いでしょう (ほとんどの人は同意しないかもしれませんが)。私たちの心は不思議なもので，思考と理性について信じられないような特徴を発揮できるように思えます。しかしそれは，私たちが完璧に思考できるということではありません。そして心理学者が何年にもわたって示してみせたことのほとんどは，私たちは自然な状態でいること，エビデンスを公平に評価することはとても難しく，既に存在するバイ

[*2] 科学者には常識が欠如している，というのも含みます。

アスによってそれらが揺らいでしまうということです。これの良い例が論理的思考における**信念バイアスの影響**です。人に特定の表現が論理的に妥当かどうか (例えば，前提が正しければ結論が正しいといえるかどうか) 判断させるとき，そうすべきではないとわかって吐いても，その結論が信じられる程度に影響される傾向があるのです。例えば，ここに結論が信じやすい妥当な議論があります。

> 全てのタバコは高価である (前提 1)
> 中毒性のあるものは安価である (前提 2)
> ゆえに，中毒性のあるものの中には，タバコでないものがある (結論)

さて，結論が信じにくい，妥当な議論もあります。

> 中毒性のあるものは全て高価である (前提 1)
> タバコの中には安価なものがある (前提 2)
> ゆえに，タバコの中には中毒性がないものがある (結論)

議論#2 の論理的な構造は，議論#1 のそれと同じで，どちらも妥当なものです。しかし，第二の議論については，前提 1 が正しくないと思われる十分な理由がありますから，結果的に結論もまた正しくないと思われます。しかしトピックがどうであるかは全体には無関係です。結論が前提条件から論理的に導かれる以上，この議論は疑うべくもなく妥当なのです。つまり，妥当な議論というのは含まれる命題が真である必要はないのです。

一方で，妥当でないのに結論が信じやすい議論の例もあります。

> 中毒性があるものは全て高価である (前提 1)
> タバコの中には安価なものがある (前提 2)
> ゆえに，中毒性のないタバコもある (結論)

最後に，信じられないような結論を導く妥当でない議論の例も挙げておきましょう。

> 全てのタバコは高価である (前提 1)
> 中毒性のあるもののなかには，安価なものもある (前提 2)
> ゆえに，中毒性のないタバコもある (結論)

さて，人に何が正しくて何が正しくないかに関する事前にあるバイアスを完璧に避け，論理的な美しさだけを純粋に評価できるものとしましょう。100% のひとが，妥当な議論は妥当であり，妥当でない議論を妥当だという人は 0% だと期待しますよね。さてこの実験をやってみると，あなたの取るデータは次のような感じになります。

| | 結論が正しそう | 結論が間違っていそう |
|---|---|---|
| 議論は妥当 | 100% が「妥当である」という | 100% が「妥当である」という |
| 議論は妥当でない | 0% が「妥当である」という | 0% が「妥当である」という |

心理学のデータがこのよなものであれば (あるいは，これによく似た感じになっていれば)，私たちは安心して肚の奥底で感じた直感を信じてしまうかもしれません。つまり，科学者の常識に基づいてデータを評価させてこそ，完璧に OK な状態になって曖昧な統計情報に惑わされることはなくなるのです。しかし，みなさんは心理学の授業を受けてきているので，これがどうなるかはわかるでしょう。

昔の研究では，Evans1983 がほぼこれと同じような実験を実施しました。彼らが見つけたのは，既存のバイアス (例えば信念) がデータの構造と一致していれば，全てその人の希望する通りに進んでいくことがわかりました。:

| | 結論が正しそう | 結論が間違っていそう |
|---|---|---|
| 結論が妥当 | 92% が「妥当である」という | |
| 結論が妥当でない | | 8% が「妥当である」という |

完璧ではないですが，結構いいでしょう。しかし結論の真偽についての直感が，議論の論理的構造に反するときはどうなるでしょうか。

| | 結論が正しそう | 結論が間違っていそう |
|---|---|---|
| 結論が妥当 | 92% が「妥当である」という | **46% が「妥当である」という** |
| 結論が妥当でない | **92%「妥当である」という** | 8% が「妥当である」という |

やれやれ，これではお話にならないですね。どうやら，人が私たちが事前に持っている信念に矛盾する正しい議論を表示されたとき，それが正しい議論になっているということを受け止めるのはかなり難しいようです (たった 46% がそうしただけです)。さらに悪いことに，わたしたちが事前に持っているバイアスに合致する，間違った議論を提示されたとき，ほぼ誰もその議論が間違っていると認識できないのです (間違った方を取るのが 92% もいます!)[*3]

If you think about it, it's not as if these data are horribly damning. Overall, people did do better than chance at compensating for their prior biases, since about 60% of people's judgements were

[*3]皮肉なことに，この事実から私がインターネットで読んだものの 95% を説明できてしまうような気がします。

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):

[*4] Earlier 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.

| | Males | | Females | |
|---|---|---|---|---|
| Department | 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 **??** 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

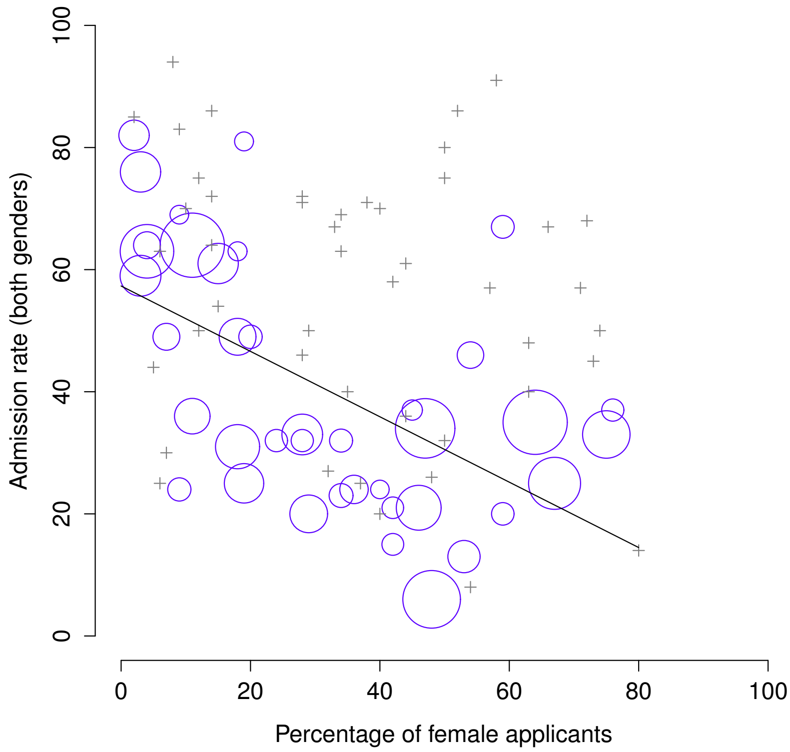


Figure1.1 The Berkeley 1973 college admissions data. This figure plots the admission rate for the 85 departments that had at least one female applicant, as a function of the percentage of applicants that were female. The plot is a redrawing of Figure 1 from **Bickel1975**. Circles plot departments with 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 **??**. In later versions of this book I'll try to include more anecdotes along those lines.

## 1.5 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.