

1. なぜ統計学を学ぶのでしょうか？

「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。
統計学者のよかに座することも
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

^{*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>

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。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、わたし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

^{*2}科学者には常識が欠如している、というのを含みます。

ちは自然な状態にいること、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、前提が正しい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粹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のでしょうか。

	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

そう考えたとしても, これらのデータがそれほど恐ろしいものであるわけではありません。全体的に, 偶然よりも高い確率で人は事前のバイアスを補正できているのです。なぜなら 60% のひとが正

*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。

しく判断しているのですから (偶然は 50% ですからね)。そうであったとしても、もしあなたが “” 証拠に基づいて判断する” ことのプロであり、誰かが正しい判断をする確率を 60% から 95% にまで上げてくれる魔法の道具を提供してくれることになったら、あなたはそれに飛びつくでしょうね? もちろんね。有難いことに、そういうことができるツールがあるのです。マジックじゃありません。スタティスティクス (統計) です。これが #1 なぜ科学者が統計を愛するのか、の理由です。私たちは “” 意図も容易く” 信じたいものを信じる” のです。言い換えれば, “” データを信じる” ようになるためには、個人的なバイアスを制御するためのちょっとした手助けが必要です。これこそ統計のすることであり、統計は私たちを正直者にしてくれるのです。

1.2

シンプソンのパラドックスへの警告

以下の話は実話 (だと思う!) です。1973 年、カリフォルニア大学バークレー校では、大学院への入学者数についてある悩みが生じていました。特に問題になったのは、入学者の性別の内訳が以下のようになっていたことです。:

	志願者	合格率
男性	8442	44%
女性	4321	35%

これを見て、大学関係者は訴えられるのではないかと思ったのです*4 約 13,000 人の志願者がいて、男女の合格率に 9% の差が出たのは、偶然にしては大きすぎるからです。かなり説得力のあるデータだと思いませんか? そしてもし私が、女性に対して 実際には弱いバイアスしかないとデータが示している、といったらあなたは私の頭がおかしいか、性差別主義者だと考えるでしょうね。

ところがどっこい、これは正しいのです。この志願者のデータをもっと注意深く見てみると、別の話が見えてきます (Bickel1975)。特に、学部ごとのバイアスを見てみると、ほとんどの学部において女性志願者の方が男性志願者よりも少し 高い合格率になっていることがわかります。以下に示す表は、6 大学部における志願者の概観です (プライバシーの観点から学部名は削除しています):

*4以前の版では実際に訴えられた、と書いていましたが、事実ではありませんでした。これについてはここに素敵な解説があります:<https://www.refsmmat.com/posts/2016-05-08-simpsons-paradox-berkeley.html> これを教えてくれたウィルフライド・ヴァン・ハートンに感謝します。

学部	男性		女性	
	志願者	合格率	志願者	合格率
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%

驚くべきことに、ほとんどの学部では女性の方が男性よりも 高い合格率を示しています！大学全体での合格率を見ると、女性が男性よりも 低いのに。なんでこんなことに？なぜこれらの説が同時に成立するのでしょうか。

ここで何が起こったか見ていきましょう。まず、学部ごとに合格率が等しく ないことに注意が必要です。すなわち、ある学部 (例えば A,B) では有資格者を高い割合で合格させていますが、他では (例えば F) 質の高い志願者であってもほとんどの候補者を不合格にするのです。ですから、ここで示した 6 学部の間で、A 学部は最も寛大で、B,C,D,E,F の順で変わっていきます。次に、男女の傾向も学部ごとで違うことがわかります。男性志願者数で学部を順序づければ、**A>B>D>C>F>E** の順になりますね (より “” 簡単” な学部を太字にしました)。全体として、男性は合格率の高い学部挑戦する傾向があるようです。では女性志願者の分布はどうなっているか、比べてみましょう。女性志願者の総数が多い順に学部を並べると、少し違った並びで次のようになります。**C>E>D>F>A>B**。言い換えると、このデータは女性がより “” 困難な” 学部” に挑戦する傾向があることを示しています。そして事実、図 ?? を見るとわかるように、この傾向が全体的で顕著であることがわかります。この効果は**シンプソンのパラドクス**として知られています。これは一般的ではないですが、実際の生活でも起こりうることで、ほとんどの人は最初驚き、現実であることさえ信じようとしません。現実なんです。そして、そこには非常に微妙な統計学的教訓が含まれているのですが、そこから私はより重要な点を指摘したいと思います。すなわち、研究することは難しく、そこには 多くの微妙な問題があり、直感に反する罠があちこちにある、ということです。これがなぜ科学者が統計を愛するか、なぜ研究法を教えるのかについての第二の理由なのです。科学的営みは難しく、真実は時に複雑なデータの隅々に、うまく隠されているのです。

この話から離れる前に、研究法のクラスでは見過ごされがちな、重要な点を指摘しておきたいと思います。統計は問題の 一部を解くだけです。思い出して欲しいのですが、我々はパークレー校の入試が女性志願者に対して不当なバイアスをかけているのではないかと、という懸念から始めたのでした。“” 総合的な” データを見た時、大学は女性に対して差別的であるように見えてましたが、“” バラバラにして” 個々人の行動を全ての学部についてみると、実際学部レベルでは、微妙ではありましたが、やや女性に有利なように動いていたのです。入試全体の女性に対するバイアスは、女性がより難しい学部挑戦しようと自ら選択する傾向が引き起こしたものでした。法律的な観点からは、

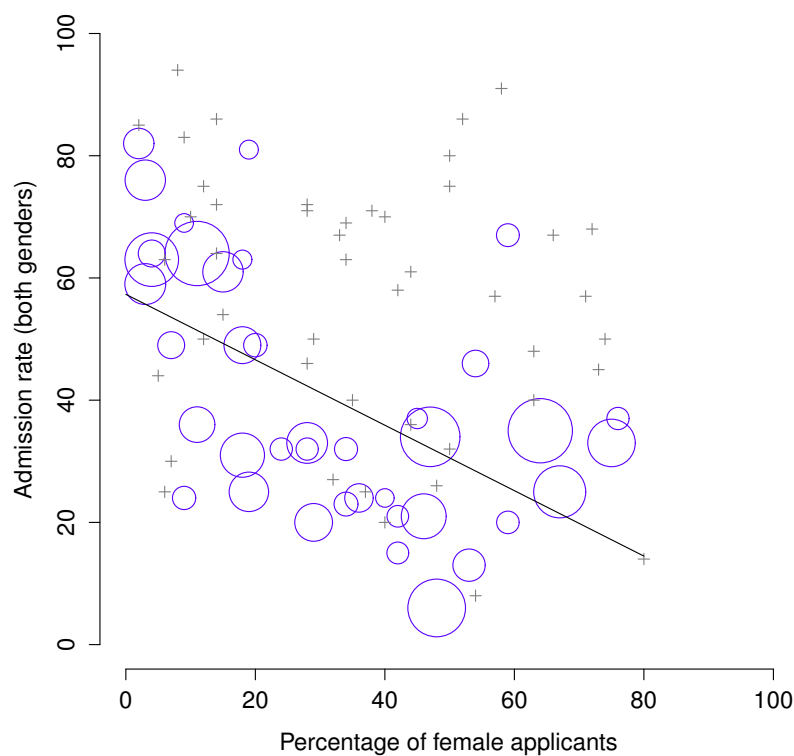


Figure1.1 1973 年、バークレー校の入試データ。この図は少なくとも一人の女性志願者がいた 85 の学部と、女性の志願者割合の関数をプロットしたもの。このプロットは **Bickel1975** の図 1 を描き直したもの。円は 40 名以上の志願者がいた学部を表している。円の面積は志願者の総数の割合。十字は応募者が 40 人未満の学部を表している。

.....

大学に非はないとするでしょう。大学院の入試は個々の学部のレベルに応じて決められますが、そうするには真っ当な理由があるのです。個々の学部のレベルにおいて、判断はほぼ偏りがありません(この微かな女性最賃のバイアスは小さなものですし、学部間で一貫したものではありません)。大学は受験生にどの学部が良いかをあてがうことはできませんし、学部レベルで判断がなされますから、その選択が生み出すバイアスに対して責任を取ることはできないのです。

とまあ私のいつもの軽口が出ましたが、それで全部というわけではないですね。もちろん、社会学や心理学的な観点から考えると、「実際には、なぜこんなに大きなジェンダー差があるのか」、ということを知りたいくなります。なぜ男性が女性よりも工学部を志望するのか、そしてなぜ英文学ではそれが逆転するのでしょうか。さらに、なぜ女性が出願しがちなバイアスがある学部は全体的に低い入学率で、男性が出願しがちな学部がそうならないのでしょうか。それぞれの学部ではバイアスがな

いのに、ジェンダーバイアスが反映されていないのでしょうか。そうかもしれません。仮に、男性は“”“ハードサイエンス”を、女性は“” 人文科学”を好む傾向があると仮説を立ててみましょう。そしてさらに、なぜ人文学部が低い合格率なのかという理由について、政府は人文学に投資していないからだと考えてみましょう（例えば、Ph.D は政府の資金提供を受けたプロジェクトと繋がっていますから）。これがジェンダーバイアスを生み出しているのでしょうか？あるいは人文科学の価値についての、無教養なものの見方が表れているだけでしょうか？もし政府の高官が人文科学に対して、“役に立たない学問だ”ということで資金援助をカットしたらどうでしょう。これは あからさまなジェンダーバイアスに見えます。これらはいずれも統計学の範疇から外れますが、研究プロジェクトとしては重要な点です。もしあなたがジェンダーバイアスの全体的な構造的影響に興味があるのなら、おそらく集計されたデータも集計されていないデータも いずれもみたいと思うでしょう。もしバークレー校の意思決定プロセスに興味があるのなら、あなたは集計されていないデータだけ注目するでしょう。

要するに、あなたが統計学をつかって答えられない決定的な質問というのは多く存在しますが、それらの質問に対する答えはあなたがどのようにデータを分析し、解釈するかに大きな影響を与えるのです。そしてこれこそ、あなたがいつもあなたのデータを学ぶ手がかりになる 道具として統計学を考えるべきだという、理由なのです。それ以上でもそれ以下でもありません。強力なツールではありますが、慎重に考えることの代わりになるものではないのです。

1.3

心理学における統計

ここまでの説明で、なぜ化学全体において統計学に注目するのか、伝わっていれば幸いです。しかし心理学で統計がどういう役割を果たしているのか、特になぜ心理学のクラスで統計に関しての講義が多いのかについては、まだ多くの疑問を持たれているのではないかと思います。ここではそれに少し答えてみようと思います。

■ なぜ心理学はそんなに多くの統計をやるの？

もっとも正直に答えるなら、これにはいくつかの異なる答えがあり、あるものは他の理由より良いものでしょう。もっとも重要な理由は、心理学は統計科学だということです。ここでいいたいのは、私たちが学ぶ“” こと”は、人間であるということです。リアルで、複雑で、素晴らしいぐらい煩雑で、腹が立つほどネジクリ曲がった人間なのです。物理学が学ぶ“” こと”は、電子のような物質であり、物理の中では複雑な振る舞いを見せますが、電子自体は心を持っていません。意見も持っていないし、それぞれが互いに異なる奇妙で気まぐれな動きをするわけでもないし、実験の途中で飽きたりすることもないし、実験者に腹を立ててわざとデータを妨

害しようなんて考えません (私は決してそんなことしませんけど!)。基本的なレベルで、心理学は物理学よりも厳しいのです。^{*5}

基本的に皆さんには心理学者として統計を学んでもらいますが、それは皆さんが物理学よりも統計に強くなってもらう必要があるからです。実際に物理ではよく言われることなのですが、“もし統計が必要になったのなら、もう少し良い実験をするべきだ”ということがあります。このように贅沢なことを言われるのは、彼らの研究の対象が、社会学者が立ち向かっている膨大な混乱に比べると、悲しいほど単純だからです。心理学だけではありません。社会科学のほとんどが、統計学にどっぷり依存しているのです。私たちが実験者として未熟なのではなく、私たちが解くべき問題がより難しいからなのです。私たちが皆さんに統計を教えるのは、みなさんが本当に、本当にそれを必要としているからです。

■ 誰か他の人に統計をやってもらえないの？

ある程度はできるかもしれませんが、完全には無理です。皆さんが心理学をするにあたって、完璧に訓練された統計学者になる必要はありませんが、あるレベルの統計的能力に届いている必要はあります。私の考えでは、全ての心理学研究者が基本的な統計力を持つべき3つの理由があります。

- － 第一に、基本的な理由から。：統計は研究デザインに深く関わっています。もしあなたが心理学的研究をうまくデザインしようと思うなら、基本的な統計について最低限の理解をしていなければなりません。
- － 第二に、研究の心理学的な側面について強くなりたいのなら、心理学的な文法を理解できるようになる必要がありますよね？心理学の文法に沿った論文のほとんどは、統計的な分析の結果を報告しています。ですから、もしあなたが本当に心理学を理解しようとするなら、他の人がそのデータをどう扱ったのかを理解する必要があります。そしてそのためには、ある程度の統計を理解していなければなりません。
- － 第三に、あなたの統計に関する部分を他の人に完全に依存することは、現実的な問題があります。：統計分析は 高価なんです。もしあなたが嫌になって、でもオーストラリア政府が大学の学費にどれぐらい補助を出しているのか知りたいと思ったら、面白いことに気づくでしょう。：統計学は “” 国家的優先事項 ” のカテゴリに配置されており、他の研究領域よりもずっとずっと、やすい学費で学べるのです。これは統計学者が圧倒的に不足しているからです。ですから、心理学研究者の観点からみれば、需要と供給のルールはあなたに味方してくれません！結果として、心理学研究がしたいというあなたの実生活上の環境において、統計屋さんを雇う十分な資金がないという決定的な事実がつきつけられます。経済的な事情から、あなたは自分でやらなければならないのです。

これらの理由は研究者一般に適用できます。あなたが心理学の実践家として、その領域のトップを目指すなら、科学的な文法を読めるようになるべきですし、それらはかなり統計に依存し

^{*5}これが、物理学が私たちよりも進んでいる理由かもしれません。

ていることなのです。

- **仕事や研究，臨床的な仕事に興味はないよ。そrでも統計が必要？**

オーケイ，冗談はよしこさん。それでも統計はあなたにとって重要なことだと思いますよ。統計は 誰にとっても重要なものであるのと同じく，あなたにとっても重要なのです。私たちは 21 世紀に生きていて，データは そこらじゅうにあります。率直に言って，我々が生きているこの世界では，統計の基礎知識はサバイバルツールに近いものです！これについては次のセクションで論じましょう。

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.