計量社会学Ⅱ（2015年度秋学期 担当:保田（やすだ））
2015.9.24

第1回「なぜ推測統計が必要なのか」

問題
1．関大生全体（約2万8千人）の中で、アルバイトをしている学生が何％いるのかを知りたい。手抜きをして関大生100人だけを調査したならば、調査結果の誤差は最大何％くらい生じるだろう？
2．では、同じように日本全国の大学生（約255万人）のうち何％がアルバイトをしているか知りたいとして、大学生を100人だけ調査したとする。この場合は、誤差は何％くらい生じるだろう？
3．インターネット調査会社を使って、日本の大学生100人を調べた場合、誤差はどう考えるべきだろう？　同じく、街中で捕まえた100人の大学生の場合は？授業で100人に配布した場合は？

■全体的な目標

計量社会学（quantitative sociology）とは、社会を知るために積極的に数値（統計データ）を活用する社会学の一分野である。この講義では、ⅠとⅡを合わせて計量社会学の基本的な考え方を使いこなせるようになることをめざす。Iでは**記述統計**（descriptive statistics）の活用を、Ⅱは**推測統計**（inferential statistics）の活用（＋多変量解析の基礎）を学修する。合わせて修得することが望ましいが、一方だけでも理解できるように講義する。

記述統計……データがもつ情報を要約して記述する統計的方法
　例）関大生500人の調査を集計すると、1ヶ月の読書冊数は平均10.2冊だった
推測統計……一部のデータから調べてもいない全体を推し測る統計的方法
　例）関大生500人の調査から、大学全体でバイトをしているのは55～65％と予想される

逆に、この講義を終えても以下の点は限界として残ることを了承してほしい。あくまで「考え方」を身につけてもらう。

1）数学的な理解は最小限に留まる
2）逆に、実際的な統計分析ソフトの操作を練習するわけでもない
3）データの集め方（社会調査の方法）については解説しない

■社会学で推測統計を学ぶ理由

推測統計は次のような思考方法を持っている。この考え方は科学の中ではとても珍しく、画期的なものであった（ラオ 2010:第2章）。

$\boxed{\text{不確実な知識}}$ ＋ $\boxed{\text{不確実性の度合いについての知識}}$ ＝ $\boxed{\text{有益な知識}}$

例）天気予報

$\boxed{\text{明日は雨でしょう}}$ ＋ $\boxed{\text{降水確率は70％です}}$ ＝ $\boxed{\text{役に立つ}}$

同じデータ（天気図）でも、明日の天気は正確にはわからない。

過去のデータを集め、おそらく雨だろうという予想をする（不確実な知識)。

どのくらい不確実なのかを分析し、推し測る（不確実性の度合いについての知識)。

70％の確率で雨（有益な知識)。

人間の活動に関わる現象については、正確にはわからないことが非常に多い。もし、古典物理学のように正確な知識のみを積み重ねていこうとするならば、その研究はまったく進まなかっただろう。人間社会や人間心理にはわからない部分（不確実性）があることを受け入れ、どの程度不確実なのかを同時に考えることで、社会学をはじめとする社会諸科学は急速に進歩できるようになった。

それを可能にしたのは統計学であり、もはや統計学なしに社会科学を考えることは不可能である。社会学を志す我々は統計学（とくに推測統計）に慣れ親しまなければならない。

■手続きをたどるだけではダメなの？

社会学部の学生は、多くの場合、社会調査の手続きの一環として推測統計の考え方に触れるはずである。すでに学習をしている人は、母集団、標本、標本調査、推定、検定といった用語を見聞きしたことがあるかもしれない。たとえば、吹田市の高校生のレジャー活動について調査するときに、高校生全員調べるのではなく、その一部（たとえば200人だけ）を調査したりする。それでも、きちんとした手続きを踏んでいれば、その200人の調査結果から、調べてもいない全員の様子を正しく推し測ることができる。その「きちんとした手続き」をまとめた体系が、推測統計である。

しかしながら、実のところ、それほどしつこく推測統計を勉強しなくても、それを利用することは容易である。「統計分析ソフトのメニューでここを選んでこの数値を見てね」あるいは「この式にとにかくあてはめて計算してね」というだけでも、実用上は十分である。

例）2015年9月のNHK政治意識月例調査によれば、内閣支持率は43％だった。調査したのは有権者全員ではなく、1088人だけだが（計画調査数1640、回収率66.3％の電話調査)、下の式にあてはめれば、調べてもいない日本全体の内閣支持率が43±3％、つまり40～46％と推測できる。（http://www.nhk.or.jp/bunken/yoron/political/index.html　2015年9月23日取得）

$$p \pm 1.96 \times \sqrt{\frac{p(1-p)}{n}}$$

［※pは小数で表わした内閣支持率、nは調査人数（p=0.43, n=1088）］

$$\rightarrow \quad 0.43 \pm 1.96 \times \sqrt{\frac{0.43(1-0.43)}{1088}}$$

$$\rightarrow \quad 0.43 \pm 0.03$$

にもかかわらず、我々は推測統計の表面的な手続きだけでなく、その仕組みを理解しなければならない。なぜならば、推測統計の手続きは、「理想的な」データの状態を前提にしているだけだからである。たとえば、調査対象に選ばれた人々は全員調査に協力してくれるし（回収率100%）、けっして嘘をついたり、いい加減な答えをしたりはしないという前提である。当たり前であるが、現実の調査はそのようにはいかない。もっと言えば、卒業研究などの学生が行う調査で無作為抽出が可能なことはまれである。それでも、我々は推測統計の手続きを適用する。その際、どのような前提がどのくらい崩れていると、推測の結果にどういった歪みがありえるのかを想像できることが大切なのである。そのためには、推測統計の仕組みをある程度深く理解しなければならない。その重要度は、理系の場合よりもむしろ高い。実験環境における理系のデータ科学は、推測統計の前提を満たした環境を作りやすいため、前提が崩れることの影響をあまり考える必要がないからである。

推測統計の理解の程度は、社会調査のデータをどのくらい自信を持って扱えるかを決める重要なポイントになる。安らかな道のりではないものの、苦労するだけの価値はあるので、少しでも多くのことを身に付けてもらいたい。

授業予定表

	内容
第1回	なぜ推測統計が必要なのか
第2回	推測統計の基盤（1）　無作為の意義
第3回	〃　（2）　記述統計の復習
第4回	〃　（3）　正規分布の利用
第5回	推定と検定（1）　平均の推定
第6回	〃　（2）　平均と比率の関係
第7回	〃　（3）　平均の検定
第8回	〃　（4）　平均の差の検定
第9回	〃　（5）　独立性の検定
第10回	統計分析ソフトでの実際
第11回	なぜ多変量解析が必要なのか
第12回	多変量解析（1）　3変数のクロス表
第13回	〃　（2）　単回帰分析と相関係数
第14回	〃　（3）　重回帰分析と偏相関係数
第15回	まとめ：計量社会学がめざすもの

（事務連絡）

毎回、√の計算できる電卓を持参のこと（計量社会学Ⅰよりもかなり使う）。

学期末の試験のみで評価（持ち込み全て可）、出席による加点・減点なし

ただし、事前の4回の小テスト（持ち込みA4用紙1枚のみ）で60点以上を受験要件とする

質問は授業後か、研究室（C605）、メール（tyasuda@zf7.so-net.ne.jp）で。

〈文献〉

C.R.ラオ著、柳井晴夫ほか訳 2010 『統計学とは何か』 ちくま学芸文庫.

岩井紀子・保田時男 2007 『調査データ分析の基礎』 有斐閣.

片瀬一男・阿部晃士・高橋 征仁 2015 『社会統計学ベイシック』 ミネルヴァ書房.

計量社会学Ⅱ（2015年度秋学期 担当:保田）
2015.10.1

第2回「推測統計の基盤（1） 無作為の意義」

■母集団と標本

推測統計の目的は、「一部の人々しか調べていない調査データから、本来関心のある全体像を推し測ること」である。たとえば、日本の大学生の生活を知りたいときに、日本の大学生全員を調べるのではなく、1000人だけを調査したりする。元々関心のある集団全体のことを**母集団**（population）と呼び、全体の中から抽出した一部分のことを**標本［サンプル］**（sample）と呼ぶ。上の例では、日本の大学生全体が母集団、実際に実験に参加した一部の学生が標本である。標本、母集団という用語を使って推測統計の目的を言い直そう。

一部の標本から母集団全体を「きちんと」推し測ること

この目的を見失わないことは、極めて重要である（念仏のように唱えてほしいぐらい）。

そのため、推測統計においては、何らかの統計的な記述する際に、それが標本について述べているのか母集団について述べているのかを、はっきりと区別しなければならない。たとえば、同じように「平均得点が5.3点」と言ったときでも、標本の平均得点が5.3点と言っているのか、母集団の平均得点が5.3点と言っているのかを明確にしなければならない。下図に表される用語の使い分けに気を付けよう（図1）。

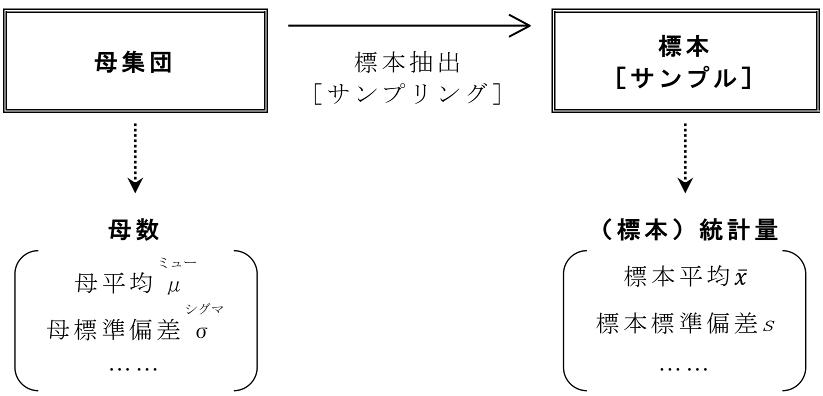


図1　母集団と標本の区別

母集団の分布の様子を指し示す**母数**（parameter）は、通常ギリシャ文字（μ、σ^2など）で表され、標本から算出される**標本統計量［あるいは単に、統計量］**（sample statistic）は、通常ラテン文字（ふつうのアルファベット）で表される。たとえば、以下のように同じ「平均」でも区別される。

> 日本の大学生から300人を抽出した調査で、月間読書冊数が平均4.5冊（$\bar{x} = 4.5$）であったとしても、この調査結果だけで、日本の大学生の月間読書冊数が平均4.5冊（$\mu = 4.5$）と結論づけるのは行き過ぎである。同じように300人を調査しても、平均値は4.2冊（$\bar{x} = 4.2$）になったり、4.7冊（$\bar{x} = 4.7$）になったりするであろうから、ある程度誤差があると考えて、平均は4.1～4.9冊（$4.1 < \mu < 4.9$）程度、といった見方をしなければならない。

■無作為抽出

推測統計のさまざまな手法は、標本が**無作為抽出［ランダム・サンプリング］**（random sampling）によって選ばれていることを大前提にしている。無作為抽出とは、標本の抽出に人間の作為が入る余地がまったくない、という意味である。つまり、母集団の中の1人1人がまったく等しい確率で標本として選ばれる。選ばれる確率を等しくする方法はさまざまにあるが、もっとも単純には、母集団全員の名簿をまず用意して1人1人に番号を振り、乱数表やサイコロを使って、調査対象者を1人1人選んでいき、標本集団を構成すればよい。

無作為抽出法がなぜ大切であるかという理由は、よく「母集団全体からまんべんなく標本をとるため」と誤解されている。無作為抽出された標本は、結果的に母集団をまんべんなくカバーしていることが多いことは事実である。しかしより重要な目的が別にある。

それは、その標本から母集団を推測する際に、確率にもとづいた計算を可能にすることである。例えば、何人かの人々に調査をしたところで、回答者に男性が多いことに気付き、女性を標本に追加したとしよう。その方が母集団をまんべんなく調べていることになるであろう。しかし、規則性を持たない方法で標本が選ばれると、確率論の知識が役に立たず、母集団の推測に統計学の手法を用いることが、もはやできなくなってしまう。

これに対して、無作為抽出によって選ばれた標本は、完全に確率的な規則性だけにもとづいて母集団からのずれが発生する。ということは、その標本調査で算出される平均値などの統計量も、何らかの規則性に従って、母集団の真の平均値から一定の確率でずれるということである。これらの規則性を逆算することによって、我々は標本調査の結果から（一定の誤差を覚悟しながら）母集団の平均値などを「きっちり」推測できるのである。

では、具体的には、無作為抽出によって標本集団にどのような規則性が生まれるのか。「標本調査で平均値を出す」という作業を何度も繰り返せば、その規則性を垣間見ることができる。

（実験）

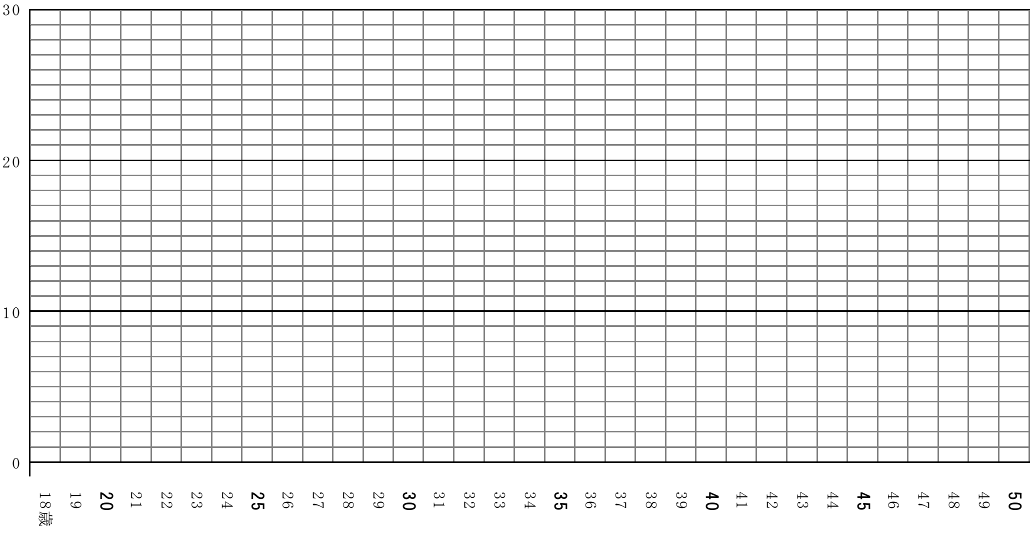
無作為抽出で生まれる規則性を、実験調査で確かめてみよう。

① 回答票を取る
② 理想の結婚年齢を回答
（結婚したくなくても回答。なるべく厳格に）
③ 保田が受講者全体で集計【＝全数調査】
④ 各自で無作為に5人を選ぶ【＝標本調査】
⑤ 5人の平均【＝標本平均】を算出
⑥ 皆の結果を持ち寄って、標本平均の分布を確かめる

	ID番号	理想の結婚年齢
1人目		
2人目		
3人目		
4人目		
5人目		

標本平均$\bar{x}$＝＿＿＿＿歳

（実験結果）標本調査を繰り返したときの「標本平均$\bar{X}$の分布」



■無作為抽出により生まれる規則性

実験の結果からも（たぶん）わかるように、無作為抽出で選ばれた標本の平均値は規則的できれいな形に分布する。このような規則性は、**中心極限定理**（central limit theorem）と呼ばれ、定式的には以下のように表現される。

> 平均がμ、標準偏差がσの分布に従う確率変数について、n個の標本の標本平均を出すとき、nの数が大きくなるにつれ、その標本平均の確率分布は、平均がμ、標準偏差が$\frac{\sigma}{\sqrt{n}}$の正規分布に近づく。

数学的な表現なのでわかりにくいが、要するに次のようなことを意味している。まず、標本平均の平均は母平均μと一致する。つまり、安心して標本の平均値を母集団の予測値として使用してよい。もちろん、ある程度のずれ（誤差）は生じるが、その誤差の程度（標本平均の標準偏差）は、$\frac{\sigma}{\sqrt{n}}$になる、という規則性をもつ。調査の人数nを増やせば、それだけ誤差は小さくなっていくことが期待でき、人数を増やせばどれだけ調査の精度が増すのかも正確にわかる。そして、最後にこれが重要であるが、驚くべきことに、元の母集団で人々がどのように分布していても、標本平均のずれ方は常に同じ規則性（正規分布）を示す（図2)。

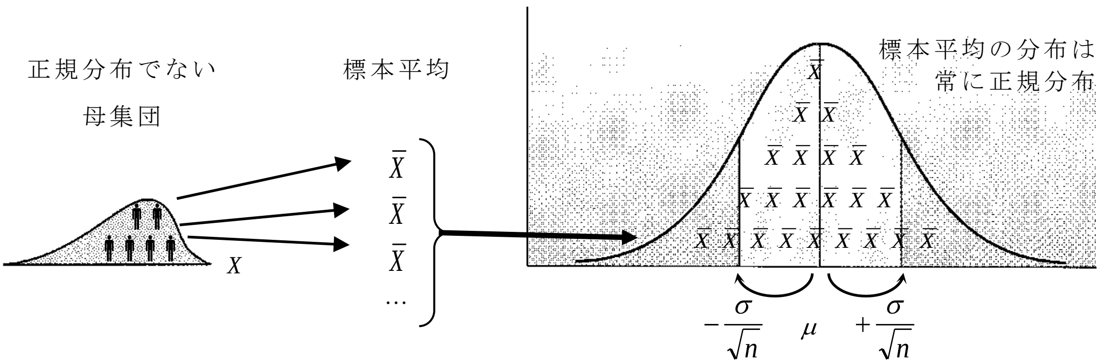


図2　中心極限定理

たとえば、平均μ＝25、標準偏差σ＝7の母集団から10人を無作為抽出して標本調査をおこなえば、その標本調査での平均値$\bar{x}$は、25（μと同じ）前後になる確率が高く、25から離れた平均値になる確率は、ベル型のカーブに沿って規則的に小さくなっていく。その標準偏差は、$\frac{\sigma}{\sqrt{n}}=\frac{7}{\sqrt{10}}=2.2$ということなので、その標本調査での平均値$\bar{x}$は、25±2.2くらいのずれが標準的（22.8歳～27.2歳くらいになることが普通）ということがわかる。

なぜ中心極限定理が成り立つのかという数理やその証明は数学的な話であるから深入りする必要はない。我々にとって大切なことは、この法則が正しそうなことを直感的に納得することと、この法則がなぜ標本調査や推測統計にとって重要なのかを理解することである。

標本調査の結果が規則的に分布するということは、「その規則性を逆算して」調査結果から母集団の様子（実際に知りたかったこと）をきっちり推測できる、ということを意味している（図3）。前回、試みに行なった内閣支持率の推測などは、このような逆算を定式化したものである。「無作為抽出をする」というただこれだけのことで、中心極限定理が生まれ、推測統計のすべての手続きが導かれる。それゆえに、社会調査は可能な限り無作為抽出に近づけることを目指すのである。

規則的な確率に従った標本抽出（**無作為抽出**） → 規則的な確率に従った標本の分布 → 規則的な確率に従った平均値の分布（**中心極限定理**）

↓

確率の逆算にもとづいて母集団が推測できる！

図3　無作為抽出と推測統計

※厳密には、中心極限定理が成り立つのは、ある程度標本サイズが大きい場合である。したがって、今回の実験のように5人程度の標本だと、実はよろしくないこともあるのだが、実際の調査ではそんな少人数なわけはないので、問題にはならない。

今日のポイント

①推測統計では、母集団と標本の区別が大切（例：$\bar{x}=4.2$なのか、$\mu=4.2$なのか）。

②無作為抽出によって、中心極限定理が生まれ、その規則性を逆算することで、標本から母集団を推測できる。だから無作為抽出は重要。

（連絡）

授業の配布資料は、授業後にWebで公開しているので、配布資料を紛失した場合などは、なるべく各自で補ってください。http://www2.itc.kansai-u.ac.jp/~tyasuda/

資料 1　「サイコロ 3 回→1～216 の乱数」換算表

1回目＝⚀

2回目＼3回目	⚀	⚁	⚂	⚃	⚄	⚅
⚀	1	2	3	4	5	6
⚁	7	8	9	10	11	12
⚂	13	14	15	16	17	18
⚃	19	20	21	22	23	24
⚄	25	26	27	28	29	30
⚅	31	32	33	34	35	36

1回目＝⚁

2回目＼3回目	⚀	⚁	⚂	⚃	⚄	⚅
⚀	37	38	39	40	41	42
⚁	43	44	45	46	47	48
⚂	49	50	51	52	53	54
⚃	55	56	57	58	59	60
⚄	61	62	63	64	65	66
⚅	67	68	69	70	71	72

1回目＝⚂

2回目＼3回目	⚀	⚁	⚂	⚃	⚄	⚅
⚀	73	74	75	76	77	78
⚁	79	80	81	82	83	84
⚂	85	86	87	88	89	90
⚃	91	92	93	94	95	96
⚄	97	98	99	100	101	102
⚅	103	104	105	106	107	108

1回目＝⚃

2回目＼3回目	⚀	⚁	⚂	⚃	⚄	⚅
⚀	109	110	111	112	113	114
⚁	115	116	117	118	119	120
⚂	121	122	123	124	125	126
⚃	127	128	129	130	131	132
⚄	133	134	135	136	137	138
⚅	139	140	141	142	143	144

1回目＝⚄

2回目＼3回目	⚀	⚁	⚂	⚃	⚄	⚅
⚀	145	146	147	148	149	150
⚁	151	152	153	154	155	156
⚂	157	158	159	160	161	162
⚃	163	164	165	166	167	168
⚄	169	170	171	172	173	174
⚅	175	176	177	178	179	180

1回目＝⚅

2回目＼3回目	⚀	⚁	⚂	⚃	⚄	⚅
⚀	181	182	183	184	185	186
⚁	187	188	189	190	191	192
⚂	193	194	195	196	197	198
⚃	199	200	201	202	203	204
⚄	205	206	207	208	209	210
⚅	211	212	213	214	215	216

注：(1回目の出目－1)×6^2 ＋ (2回目の出目-1)×6 ＋ 3回目の出目　で算出できる。

第3回「推測統計の基盤（2） 正規分布の利用」

■実験調査を振り返る

前回の実験調査の目的は、無作為抽出によってどのような規則性が生まれるのかを確認することであった。受講生全体（81 人）を「母集団」と考えて、各自でその中から 5 人の「標本」を無作為抽出して、理想の結婚年齢の平均値$\bar{X}$を調べてもらった。

その結果、当然いろいろな 5 人が選ばれ、いろいろな平均値$\bar{X}$が算出される。どのような平均値になったか結果を集約すると、図 1 のようになった（整数に四捨五入）。これは標本平均$\bar{X}$としてどのような値が出やすく、どのような値が出にくいのかを表している。すぐに見てわかるように、$\bar{X} \fallingdotseq 27$という結果が一番多く、これを中心としてほぼ左右対称のきれいな形に分布している。

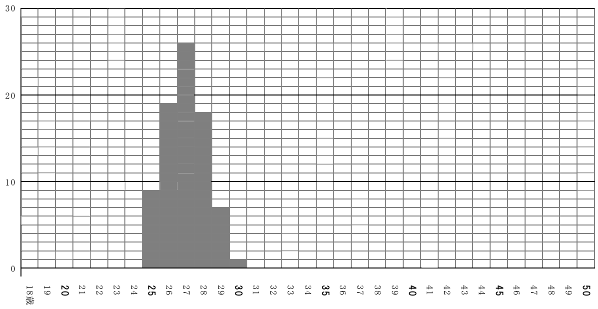


図 1 標本調査を何度も行った場合の標本平均$\bar{X}$の分布

一方で、もともとの母集団では理想の結婚年齢はどのように分布していたのだろうか（通常の母集団は人数が多すぎて調べられないが、今回は少ないので調べられてしまう）。全員の回答を集計してみると、図 2 のようになった。

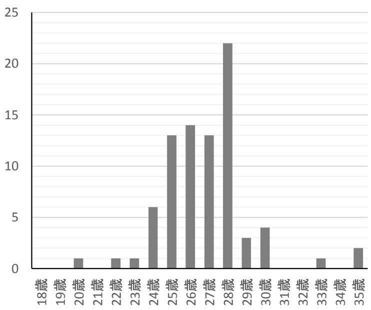


図2 母集団における理想の結婚年齢の分布

注：平均μ＝27.00歳、標準偏差σ＝2.69歳

ここで2つのことに驚いてほしい。1つは本当に全員調べた場合の平均値μ＝27.00歳を、たった5人しか調べていない標本調査の平均値$\bar{X}$が結構言い当てている、ということである。もう1つは、もともとの母集団の分布（図2）は28歳が突出しているなど、ややいびつな形であるにもかかわらず、標本平均の分布（図1）はきれいな形（正規分布）をしているとい

うことである。

これは、*きちんと無作為抽出さえしていれば*、中心極限定理という規則性が生まれるためである。今回の場合、母集団はμ＝27.00、σ＝2.69なので、無作為抽出したn＝5の標本の平均値$\bar{X}$は、平均27.00、標準偏差$\frac{2.69}{\sqrt{5}}=1.20$の正規分布を示すはずであり、実際にそのような結果が出ている。

一般的に表せば、平均値μ、標準偏差σの母集団から無作為抽出された標本の平均値$\bar{X}$は、必ず正規分布に沿った確率で出現し、その正規分布の平均値はμと一致し、標準偏差は$\frac{\sigma}{\sqrt{n}}$になる。とくに「標本平均$\bar{X}$の標準偏差」である$\frac{\sigma}{\sqrt{n}}$は**標準誤差**（standard error）と呼ばれる重要な数値である。標準誤差が小さいほど標本平均$\bar{X}$は母平均μに近い結果になる確率が高いということなので、標準誤差の小ささは調査の精度の高さを表す。

（問題）

1．日本全国のアルバイトをしている大学生を母集団とする。かりにその母集団で時給が平均μ＝970円、標準偏差σ＝130円で分布しているとする。次の問に答えなさい。

（1）無作為抽出で30人の標本調査をしたとき、その標本平均$\bar{X}$はどのように分布するか。

⇒　平均＝＿＿＿＿＿＿、標準偏差＝＿＿＿＿＿＿＿の正規分布になる

（2）この30人の標本調査で$\bar{X}=900$というような結果は起こりうると思うか。理由を含めて答えなさい。

（3）調査の人数を60人（30人の2倍）、120人（30人の4倍）と増やしたときに、調査の精度は何倍になるか（標準誤差は何分の1になるか）。

■正規分布

推測統計では、この中心極限定理を逆算して、標本調査の結果から母集団の様子を推し測る。具体的な手続きに話を進めるためには、**正規分布**（normal distribution）について理解を深める必要がある。ややまどろっこしいが、今回は正規分布の理解に集中しよう。

正規分布とは、図3のように、中心が盛り上がり、裾野が左右対称に広がるベル型のカーブを描く分布のことである。「無作為抽出の標本平均の分布」のように込み入った現象を考えなくとも、世の中の数量にはほぼ正規分布するものがたくさんある（たとえば、身長、体重、起床時間、スポーツテストの記録など）。

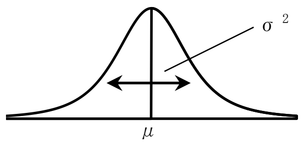


図3　平均μ、分散σ^2（標準偏差σ）の正規分布

正規分布の正体は、無数の誤差の積み重なり（和）であるということを、ドイツの数学者ガウスが1816年に示している。2、3個の原因が強い影響力をもって結果を規定するのではなく、小さな原因（誤差）がたくさん足し合わさることで結果が決まるような現象につ

ついて正規分布が現れる、と考えればよい。標本平均は、無数（n人）の個人の合計から成り立っているわけであるから、正規分布するのはごく自然なことである。

カール・フリードリヒ・ガウス
Karl Friedrich Gauss (1777-1855)

■**正規分布に従う変数の確率計算**

正規分布のグラフはヒストグラムと同じように面積が意味を持つ。その曲線の下に囲まれる面積が大きいほど、そのような値になる確率が高いことを意味する。つまり、ある数量が特定の範囲の値を取る確率（たとえば、テストで70〜80点を取る確率）は、下図のように、曲線の下で囲まれる面積全体のうち、この範囲が何％を占めているかを調べれば知ることができる。

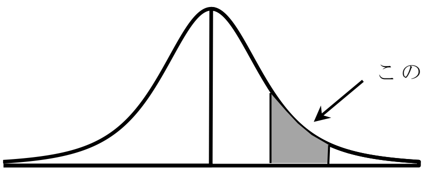


図4　正規分布と確率

正規分布の曲線は、下の式で表わされるので、これを積分して面積を算出すれば知りたい確率を知ることができる。しかし、この式を覚える必要や、積分の計算ができる必要は全くない。なぜならば、そのような面倒な手続きは、統計学者によってなされ、必要な情報は表にまとめられているからである。我々は、その表を利用する。

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

通常の統計学のテキストでは、正規分布の確率表が必ず掲載されている。これは、平均が0、標準偏差が1の正規分布についての情報を整理したものである。0.01刻みのzの値に対して、この正規分布に従う変数がz以上の値を取る確率（つまり曲線の下の面積）を記している。この表を見れば、例えば平均0、標準偏差1の正規分布に従う数量が1.23以上の値を取る確率は0.1093（＝10.93％）と瞬時に分かる。正規分布は左右対称であるから、この数量が-1.23以下の値を取る確率も0.1093（＝10.93％）である。

■**標準正規分布と標準化**

しかし、正規分布の曲線は平均値と標準偏差の違いによって無数に考えられる。無数の曲線について無数の表を提示することはできない。通常、提供されるのは平均値0、標準偏差1の正規分布についての表のみである。この正規分布のことを特に**標準正規分布**

（standard normal distribution）と呼ぶ。実は標準正規分布の表があれば、十分なのだ。なぜか。

例えば、平均値が0ではなく3、標準偏差が1ではなく2の正規分布に従う変数Xが6以上の値を取る確率を知りたいとしよう。変数Xの分布を表す曲線は、標準正規分布を正の方向(右)に平均値分の3だけずらして、左右に標準偏差分だけ2倍に引き伸ばしたものと言うことができる。すると、逆に変数Xの値から3を引き、2で割ってやれば、その値は標準正規分布の曲線を示すはずである。従って、変数Xが6以上の値を取る確率は、標準正規分布において$\frac{6-3}{2} = 1.5$以上の値を取る確率に等しい。先ほどの確率表から確認すると、その確率は0.0668（＝6.68％）である。

一般的には次のように言える。変数Xが平均μ、標準偏差σの正規分布に従う場合、下の式に従って、Xを**標準得点[標準化得点、z値、標準値]**(standard score; standardized score; z-score）に変換すれば、その値は標準正規分布（平均0、標準偏差1の正規分布）に従う。

$$z = \frac{x-\mu}{\sigma}$$　（ただし、μはXの平均、σは標準偏差）

この変換のことを**標準化［z-変換］**（standardization; z-transform）と呼ぶ。標準化を用いれば、どのような平均値・標準偏差の正規分布についても、1つの確率表であらゆる確率を知ることができる。

※なお、現代ではコンピュータが発達しているので、標準化の手続きを知らなくても直接知りたい確率を知ることができる。たとえばExcelでは「=norm.dist(70,60,10,true)」で平均60、標準偏差10の正規分布で70<u>以下</u>の値が出る確率が表示される（0.8413...＝84.13％）。しかし、推測統計の技法を理解するためには標準化の手続きを理解することが必須になるので、ここではあえて古典的な標準化の手続きを学習する。

（問題）

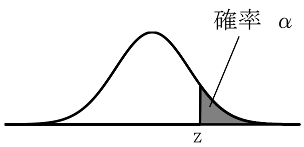
2. 知能指数（IQ）は、平均値100、標準偏差15の正規分布になるようにテストが作られている。ビートたけしのIQは132であるらしい。IQが132以上の人は、およそ何％の確率で現れるだろうか。

3. 日本全国の高校生を母集団とするとき、かりにその睡眠時間の平均値が7.40時間、標準偏差が1.60時間であるとする（正規分布）。このとき睡眠時間が6.00時間の高校生は下から何％程度の位置にあるといえるか。

4. 上の母集団から150人を無作為抽出すると、150人の平均睡眠時間$\bar{X}$は平均が7.40、標準偏差（標準誤差）が0.13の正規分布になるはずである。このとき、150人の平均睡眠時間$\bar{X}$が7.60時間以上になる確率は何％か。

今日のポイント
①数量が正規分布になると便利だ、ということを実感する。
②標準化と標準正規分布の確率表で、具体的な確率を出せるようになる。

【標準正規分布の確率表】



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.05050	.04947	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.00508	.00494	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002

※確率表はずっと使うので、毎回持っておくように。

第4回「推定と検定（1） 平均の推定」

■推測統計の2つの柱——推定と検定——

今回から、ようやく推測統計の具体的な手続きに入る。推測統計の手続きは、**統計的推定［あるいは単に推定］**（statistical estimation）と**統計的検定［あるいは単に検定］**（statistical test）という2本の柱からなっている。

推定とは、標本を分析することで、母集団の特徴を表す数量（母数）がどのような値を取るのかを探索する手続きである。たとえば次のような手続きが、典型的な統計的推定である。1）大阪府の中学生は月に何回くらい午前1時まで夜更かししているのか知りたい→2）母集団の中から200人の標本を抽出する→3）標本を調査して夜更かしの回数を分析する→4）大阪府の中学生全体の平均夜更かし回数が○回～○回くらいと推し測る。

一方、検定は、母集団についてのある仮説が正しいと言えるかどうかを、標本の分析から判定する手続きである。例えば、上の例で、「大阪府の中学生では、男子生徒よりも女子生徒の方が平均夜更かし回数が多い」という仮説を分析者が持っていたとする。この仮説が正しそうかどうかを、標本の得点の分布から判断する手続きが、統計的検定である。

今回は、推定に絞って解説をする。どのような数量（平均、標準偏差、比率など）を推定するのかによって、それぞれ推定の手続きがあるが、まずもっとも人々の関心を引きやすい「平均」の推定について解説しよう。他の場合も基本的な筋道は同じである。

■とりあえずやってみよう。

結論から述べてしまえば、ある程度人数の多い標本（100人程度が基準）を無作為抽出した場合には、以下の法則で、母集団の平均値を「推定」することができる。

95%の確からしさで、母集団の平均値 μ は、$\bar{X} \pm 1.96 \times \frac{s}{\sqrt{n}}$ の範囲に収まる。

一見すると難しそうだが、出てくる数値は標本の人数（n)、平均値（$\bar{X}$)、標準偏差（s）の3つだけである。たとえば、200人調査をして、夜更かし回数の平均値$\bar{X}$が7.5回、標準偏差sが2.1回という調査結果が得られたとき、母集団の平均値 μ は、$7.5 \pm 1.96 \times \frac{2.1}{\sqrt{200}}$、つまり7.5±0.29の範囲なので、7.21～7.79回であると、推定することができる。簡単である。

■中心極限定理と正規分布の利用から考える

なぜ、このような法則で推定が可能なのか。ここまでに学習してきた「中心極限定理」と「正規分布の利用」を組み合わせれば、その理屈が理解できる。理屈を理解していなくとも手続きをたどることは問題なくできるが、この講義のはじめの方で述べたとおり、原

理を理解していないと現実的な社会調査の場面（完全な無作為抽出でない場面など）で勘が働かないので、一度原理を理解することは重要である。

いま、母集団における夜更かし回数の平均値μと、標準偏差σがわかっているとする。このとき、無作為抽出したn人の標本で夜更かし回数の平均値$\bar{X}$を調べるならば、中心極限定理により、その標本平均は、平均μ、標準偏差$\frac{\sigma}{\sqrt{n}}$の正規分布に沿った確率で現れることが期待される。前回学習した正規分布の確率表を利用すれば、標本平均が「〇〇％の確率で××の範囲に収まるはずだ」という予測が正確にできるはずである。

科学的な予測はある程度高い確率で行うべきなので、通常は95％の確率で標本平均の「予測される範囲」を考える。この予測は上下合わせて2.5％ずつ計5％しか外れない（図1)。

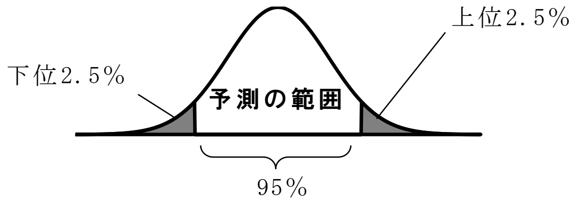


図1　正規分布における95％の確率範囲

標本平均を標準化すれば、そのz値は平均0、標準偏差1の正規分布に従うことになる（「標本平均の平均」と「標本平均の標準偏差」で標準化していることに注意)。

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

標準正規分布の確率表を確認すると、このz値は95％の確率で－1.96～＋1.96の範囲に収まることがわかる。なぜならば、確率表によると、z＞1.96の確率が.0250（2.5％）とあるので、左右対称であることから、z＜-1.96の確率も2.5％で、その間の－1.96～＋1.96の確率は95％になる。したがって、

$$-1.96 < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1.96$$

なので、標本平均$\bar{x}$は、95％の確率で次の範囲に収まることになる。

$$\mu - 1.96 \times \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

この不等式を変形した下の不等式も95％の確率で成り立つ。

$$\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

つまり、母集団の真の平均μの範囲は、95％の確からしさで、

$$\bar{X} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

である。

しかし、よく見ると、この式では母標準偏差σの値が必要である。母集団は調べてもい

ない全体なので標準偏差の値は通常わからない。そのため、通例として母標準偏差σは標本標準偏差sの値で代用する。標本の人数nが十分に大きければ（通常、100人程度）、σとsの値はほとんど一致するからである。ようやく、最初に示したのと同じ結論が導かれた。

95%の確からしさで、母集団の平均値μは、$\bar{X} \pm 1.96 \times \frac{s}{\sqrt{n}}$ の範囲に収まる。

■統計的推定に関する用語（概念）

推定に関して、知っておいてほしい用語（概念）は以下のとおりである。ここでおこなったように、1つの値に絞るのではなく、一定の範囲で母数を推定するやり方を**区間推定**（interval estimation）と呼ぶ。統計的推定は、通常、区間推定で行われる。区間推定が正しいと期待される確からしさ（ここでは95%）は、その推定の**信頼度**（confidence level）と呼ぶ。どの程度の信頼度で推定を行うかは、分析者が任意に決める。本来はどの程度確実な情報が必要かということを考慮すべきであるが、人文社会科学のトピックでは一般的に95%が用いられることが多い。より確実な情報が欲しい場合には99%の信頼度を設定し、ややあいまいでもよい場合には90%の信頼度を設定することが多い。信頼度のことを信頼係数と呼ぶこともある。逆に区間推定が外れる可能性のことを推定の危険度とか危険率と呼ぶことがある。信頼度が95%なら、危険度は5%になる。また、区間推定による具体的な一定幅の推定値のことを**信頼区間**（confidence interval）と呼ぶ。

先の例だと、「95%の信頼度で区間推定をすると、信頼区間は7.21～7.29であった」といった用い方をする。手続きが込み入ってくるので、決められた用語を正しく使うことは、コミュニケーションのために大切である。

（問題）

1．K市の高校1年生の家庭学習時間（/週）について、平均値を知りたい。無作為抽出で100人の生徒を選び調査をすると、平均$\bar{X}$＝4.50時間、標準偏差s＝2.06時間だった。K市の高校1年生全体の平均学習時間はどのあたりにあると推定できるか（信頼度95%で、母集団の平均家庭学習時間μを区間推定しなさい）。

2．信頼度が95%以外の場合について考えてみよう。

（1）信頼度を99%に上げた場合、推定の幅（信頼区間）は広くなるか、狭くなるか。計算をする前に、言葉の意味から考えてみよう。（「信頼度を上げる」ということは「もっと当たりやすい予想をする」ということである。その場合、予想の幅は広くなるのか、狭くなるのか？）

（2）信頼度99%の推定のためには、推定に用いる「1.96」を別の数値に置き換えなければならない。どのような数値にすべきか。（※信頼度99%ということは、正規分布の上下0.5%（=.0050）ずつを除く範囲ということである）

（3）信頼度を99%に変更して、問題1の推定をやり直しなさい。

（4）同様に、信頼度90%で問題1の推定をやり直しなさい。

3.「関大生を200人調査したところ、その通学時間は平均30.2分で、標準偏差が12.8分でした。この調査結果から、関大生全体の平均通学時間は28.4～32.0分程度であることが、95％の確からしさで推定できます」

この文章を、下の統計用語を使って書き直しなさい。

→関大生全体を＿＿＿＿＿＿＿＿として、200人の＿＿＿＿＿＿＿＿を調査したとろ、通学時間の＿＿＿＿＿＿＿＿は30.2分で、＿＿＿＿＿＿＿＿は12.8分であった。この調査結果から通学時間の＿＿＿＿＿＿＿＿を区間推定すると、＿＿＿＿＿＿＿＿が95％の場合の＿＿＿＿＿＿＿＿は$30.2 \pm 1.96 \times \frac{12.8}{\sqrt{200}}$、つまり28.4～32.0分と算出できる。

信頼度	母集団	信頼区間	標本平均	標本標準偏差	標本	母平均

4. 上の問題を参考に、「平均の推定」の自然な例文を自分で考えてみよう。

(1) まず文章の(　　　)内に自分で好きな語句・数値を自由に想像して書き入れなさい。

(2) 文章の..........に、適切な数値・式を書き入れなさい。

関大生から無作為抽出した(　　　　)人について、(　　　　　　　　　　　　)を尋ねた。すると、平均値$\bar{X}$が(　　　　)、標準偏差sが(　　　　)になった。この結果から母集団の平均値μを信頼度95％で区間推定すると、..........±..........という式でその範囲が算出できる。つまり、関大生全体の平均値μは、..........～..........の範囲にあると推定される。

今日のポイント

①母集団の平均値μは、標本調査の結果から次の式を用いて区間推定できる（信頼度95％の場合）。

$$\bar{X} \pm 1.96 \times \frac{s}{\sqrt{n}}$$

※信頼度を95％から変更した場合、「1.96」が対応する値に置き換える。

信頼度90％→1.64　信頼度95％→1.96　信頼度99％→2.58　信頼度99.9％→3.29

②推定の仕組みは、中心極限定理と正規分布の利用から理解できる。

＊＊＊連絡＊＊＊

来週の授業始めに1回目の小テスト。

A4用紙1枚（裏表に何が書いていてもよい）は持ち込み可。確率表と電卓は常にOK。

※小テストは全部で4回。小テスト全体で60点以上取っていないと本試験を受験できない。

第5回「推定と検定（2）平均と比率の関係」

■これまでのポイント

ここまでに学んできたことのポイントを箇条書きすると次のようになる。

・計量社会学Ⅱのテーマは推測統計
・一部の標本から調べてもいない母集団全体を「きちんと」推し測る
・標本は$\bar{X}$, sなどで表記、母集団はμ, σなどで表記
・無作為抽出⇒中心極限定理（標本平均は標準誤差$\frac{\sigma}{\sqrt{n}}$で正規分布する）
・正規分布であれば、標準化によって確率計算ができる
・だから逆算すると、母集団の平均値が区間推定できる
　信頼度95%（危険度5%）ならば、その信頼区間は$\bar{X} \pm 1.96 \times \frac{s}{\sqrt{n}}$

■標本サイズが小さい場合の補正

前回示した区間推定の信頼区間$\bar{X} \pm 1.96 \times \frac{s}{\sqrt{n}}$は、本来は$\bar{X} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$になる。ところが、母集団の標準偏差σは当然わからないので、標本の標準偏差sで代用していた。標本のサイズnが十分に大きければ（通常100以上程度）σとsはほぼ一致するので問題ないが、標本サイズが小さいならば、この代用には無理があり、σとsは大きくずれている可能性がある。

このため、標本サイズが小さい場合には、その分のあいまいさを反映するように、標準正規分布を補正する。補正した分布のことを**スチューデントのt分布［あるいは単にt分布］**（t-distribution）と呼ぶ（図1）。t分布が正規分布と大きく異なる点は、**自由度**（degree of freedom; df）によって分布の形が変化することである。「自由度」という概念は非常に理解しにくいので、ここでは、単に「t分布の自由度はn－1」という定義で覚えておこう（後の方の回で、改めて説明する）。図1を見ればわかるように、t分布は自由度が小さい（標本サイズが小さい）ほど、両側の裾野が広がる、という特徴を持つ。つまり、標本サイズが小さい場合の、推定のあいまいさが反映される形になっている。自由度の値が大きくなると、t分布は正規分布と同じになる（自由度100程度でほぼ重なる）。

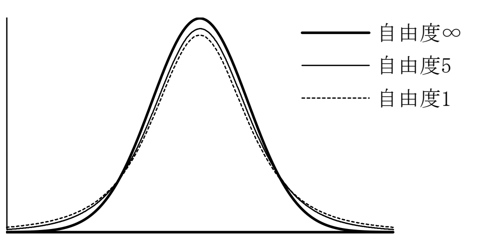


図1　自由度1、5、∞のt分布

t分布の大まかな様子は、統計学の教科書で表にまとめられている。ただし、無数にある自由度のt分布についてそれぞれの確率表を示すのは大変である。そこで、省略して、よく

用いる確率に対応する区切り値（**臨界値**（critical value）と呼ぶ）だけを示すのが通例である。標準正規分布表は区切り値に対応する確率を羅列しているのに対して、t分布表は確率に対応する区切り値の方を羅列している。意味が異なるので注意しよう。

t分布による補正の例を示そう。たとえば、5人しか調べていない調査で、平均値$\bar{X}$が7.5、標準偏差sが2.1だったとする。95%の信頼度で区間推定をするとき、標準正規分布を参照して1.96の係数を用いるのではなく、自由度4（df＝n-1＝5-1）のt分布を参照して、2.78の係数を用いる方が、補正が効いてより適切な区間推定ができる。

μの信頼区間は、$7.5 \pm 1.96 \times \frac{2.1}{\sqrt{5}}$　△

μの信頼区間は、$7.5 \pm 2.78 \times \frac{2.1}{\sqrt{5}}$　○

自由度が大きい（標本サイズが大きい）ときのt分布は標準正規分布と同じになるので、t分布表では自由度∞という欄で標準正規分布に対応する数値が示されている。だから、平均値の区間推定の作業としては、いつでもt分布の表を参照すると考えて差し支えない。人数が十分大きいときには、自由度∞の欄（すなわち標準正規分布）を参照すればよい。

また、ある程度自由度が大きくなると、自由度の値が多少違ってもほとんど確率は変わらないので、自由度が40の次は自由度が60というように途中を省いて表を作っている。若干計算結果は狂うが、一番近い自由度の欄を参照して差し支えない（たとえば、n＝45のときは自由度＝44だが、一番近い自由度40の欄を参照する）。

（問題）

1. ある大学の1年生が月曜1時限の授業に平均何回程度出席していたのかを調べたい。全員を調べるのはたいへんなので、無作為に抽出した10人に出席回数を聞いたところ、15回中の平均出席回数は10.25回、標準偏差は1.20であった。

（1）95%の信頼度で、母集団の平均出席回数を区間推定してみよう。

（2）かりに、上の調査結果が10人ではなく50人の標本を調べた結果だとすれば、区間推定の結果はどのように変わってくるか（信頼区間の幅は広くなるのか、狭くなるのか）。より大きな規模の調査結果であるということが、推定にどう影響するのかを考えて結果を予測しよう（※人数が多いことは、2つの理由で推定の結果に影響することに注意）。その上で、95%の信頼度で実際に計算をしてみよう。

■比率の推定

平均の推定には重要な応用がある。それは、同じ方法を比率（割合、%）の推定にそのまま適用できるとうことである。たとえば、「標本調査の結果、内閣支持率が75%だったとき、母集団での内閣支持率は何%程度か」というように母集団での比率を推定したい場面である。社会調査では、量的変数ではなく質的変数がよく扱われるので、平均よりもむしろ比率が問題になることが多い。

実は、比率は特殊な変数の平均値と考えることができるので、比率の推定は平均の推定とまったく同じように考えることができる。以下のように考えればよい。

①ある条件に該当する場合は値が1、該当しない場合は0になる変数があると考える。
例）内閣を支持する＝1、内閣を支持しない＝0

②このとき、その平均値は比率と一致する（比率とは1/0変数の平均値のことである)。
例）400人中300人が内閣を支持する場合

$$\frac{300}{400}\times 1+\frac{100}{400}\times 0=0.75=75\%$$

③だから、比率の推定は、1/0変数について平均の推定を行うのと同じである。

具体的には、比率pで1の値をとり、比率$(1-p)$で0の値をとる1/0変数について、

標本平均$\bar{X}$は $\frac{np\times 1+n(1-p)\times 0}{n}=p$

標本標準偏差sは $\sqrt{\frac{np(1-p)^2+n(1-p)(0-p)^2}{n-1}}=\sqrt{\frac{np(1-p)}{n-1}}\fallingdotseq\sqrt{p(1-p)}$ と整理できる。

④したがって、平均の推定の式にあてはめると、信頼度95％でμの信頼区間は、

$\bar{X}\pm 1.96\times\frac{s}{\sqrt{n}}$ が……

$p\pm 1.96\times\sqrt{\frac{p(1-p)}{n}}$ になる

※ただし、比率を0/1変数の平均値になぞらえるためには、ある程度標本サイズが大きくなければならない。100人程度の標本サイズは必要なので、t分布を利用する可能性はなく、必ず標準正規分布を利用することになる。

例）標本調査で400人中300人（75％）が内閣を支持するとき、母集団での内閣支持率は、信頼度95％で以下のように区間推定できる。

$$p\pm 1.96\times\sqrt{\frac{p(1-p)}{n}}$$

$$=0.75\pm 1.96\times\sqrt{\frac{0.75(1-0.75)}{400}}$$

$$=0.75\pm 0.042$$

$$=0.708\sim 0.792$$

つまり、母集団での内閣支持率は、0.708～0.792（70.8％～79.2％）

※本来は、0.75±0.042＝0.708～0.792といった数式の書き方はおかしいが便宜的にこうする。気になる人は、以下のように記す方が数学的には正しい。

$$p - 1.96 \times \sqrt{\frac{p(1-p)}{n}} < \mu < p + 1.96 \times \sqrt{\frac{p(1-p)}{n}}$$

$$0.75 - 0.042 < \mu < 0.75 + 0.042$$

$$0.708 < \mu < 0.792$$

（問題）

2．厚生労働省による「全国家庭児童調査」（2009年調査）によると、1098人の子ども（小5～18歳未満）のうち、「テレビをほとんど見ない」子どもは6.6%。5年前に比べて2.5倍に急増しているらしい。

（1）テレビをほとんど見ない＝1、見る＝0とすると、この変数の標本平均$\bar{X}$と標本標準偏差sはそれぞれいくらか。

（2）母集団で「テレビをほとんど見ない」子どもは何％程度いるだろう。信頼度95％で比率の区間推定をしなさい。

（3）同じ区間推定を、信頼度99％でおこないなさい。

3．関大生から無作為抽出した300人に調査をしたところ、朝食がごはん派の学生は34％だった（仮想データ）。関大生全体では、ごはん派は何％程度か。信頼度95％で区間推定しなさい。

今日のポイント

①母集団の平均値μは、標本調査の結果から次の式を用いて区間推定できる（信頼度95%の場合）。〈<u>標本サイズが小さい場合の注釈を追加</u>〉

$$\bar{X} \pm 1.96 \times \frac{s}{\sqrt{n}}$$

※ただし、標本サイズが小さい（およそ100以下の）場合は、自由度n-1のt分布を参照して、1.96の値を適切な臨界値に置き換える。

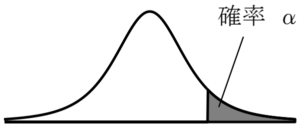
※また、信頼度を95%から変更した場合も、1.96を対応する値に置き換える。

②比率の推定は、0/1変数の平均値の推定で理解できる。具体的な式は下のとおり（信頼度95%の場合）。

$$p \pm 1.96 \times \sqrt{\frac{p(1-p)}{n}}$$

（pは標本での比率を表わす小数。例：40％はp＝0.40）

【t分布の確率表】



自由度	確率α（カッコ内は両側に同じ領域があった場合の確率）					
	0.1 (0.2)	0.05 (0.1)	0.025 (0.05)	0.01 (0.02)	0.005 (0.01)	0.0005 (0.001)
1	3.08	6.31	12.71	31.82	63.66	636.62
2	1.89	2.92	4.30	6.96	9.92	31.60
3	1.64	2.35	3.18	4.54	5.84	12.92
4	1.53	2.13	2.78	3.75	4.60	8.61
5	1.48	2.02	2.57	3.36	4.03	6.87
6	1.44	1.94	2.45	3.14	3.71	5.96
7	1.41	1.89	2.36	3.00	3.50	5.41
8	1.40	1.86	2.31	2.90	3.36	5.04
9	1.38	1.83	2.26	2.82	3.25	4.78
10	1.37	1.81	2.23	2.76	3.17	4.59
11	1.36	1.80	2.20	2.72	3.11	4.44
12	1.36	1.78	2.18	2.68	3.05	4.32
13	1.35	1.77	2.16	2.65	3.01	4.22
14	1.35	1.76	2.14	2.62	2.98	4.14
15	1.34	1.75	2.13	2.60	2.95	4.07
16	1.34	1.75	2.12	2.58	2.92	4.01
17	1.33	1.74	2.11	2.57	2.90	3.97
18	1.33	1.73	2.10	2.55	2.88	3.92
19	1.33	1.73	2.09	2.54	2.86	3.88
20	1.33	1.72	2.09	2.53	2.85	3.85
21	1.32	1.72	2.08	2.52	2.83	3.82
22	1.32	1.72	2.07	2.51	2.82	3.79
23	1.32	1.71	2.07	2.50	2.81	3.77
24	1.32	1.71	2.06	2.49	2.80	3.75
25	1.32	1.71	2.06	2.49	2.79	3.73
26	1.31	1.71	2.06	2.48	2.78	3.71
27	1.31	1.70	2.05	2.47	2.77	3.69
28	1.31	1.70	2.05	2.47	2.76	3.67
29	1.31	1.70	2.05	2.46	2.76	3.66
30	1.31	1.70	2.04	2.46	2.75	3.65
40	1.30	1.68	2.02	2.42	2.70	3.55
60	1.30	1.67	2.00	2.39	2.66	3.46
80	1.29	1.66	1.99	2.37	2.64	3.42
120	1.29	1.66	1.98	2.36	2.62	3.37
∞	1.28	1.64	1.96	2.33	2.58	3.29

注：それぞれの自由度の t 分布で、確率αに対応する臨界値だけを記している。

カッコ外のαは片側だけの確率。両側に同じ領域があれば確率はカッコ内の値になる。

・信頼度 95%（危険度 5%）の区間推定→αが「0.025（0.05)」の列を参照。

・有意水準 5%の両側検定→αが「0.025（0.05)」の列を参照。

・有意水準 5%の片側検定→αが「0.05（0.1)」の列を参照。

第6回「推定と検定（3） 平均の検定」

■検定は合否判定

ここまでに、推測統計の一方の柱である統計的推定について解説をした。今回は、もう一方の柱である統計的検定について解説をする。検定は、母集団について何らかの不確かな仮説があり、その仮説を支持するような調査結果が得られたとき、この証拠で十分に仮説が正しいと主張してよいかどうかを判定するための手続きである。検定に合格すれば、標本データが示す結果を信じて母集団でも仮説が支持されたといってよい。逆に、検定に合格できなければ、母集団で仮説が支持されたというには、まだ証拠が不十分とみなされる。

例）市民300人に対して意識調査を行うと、○○駅前の再開発に賛成の人は54％だった。
この結果から「母集団でも過半数が賛成している」という仮説は支持されるか？
検定の手続きを行った結果……
→ 合格！（300人調査で54％もあれば、少なくとも過半数と考えるには十分）
　 不合格（調査結果では過半数だが、300人程度では証拠不十分）

■帰無仮説と対立仮説

ただし、実際の手続きでは、合格・不合格といった言い方はせずに、**帰無仮説**（null hypothesis）と**対立仮説**（alternative hypothesis）と呼ばれる2つの仮説のいずれを採択すべきか、という言い方で判定をする。帰無仮説とは、標本データで得られた証拠が確率的な偶然にすぎず、「母集団についても同様だ」と主張できるだけの意味がない、とする仮説である。上の例の場合、標本調査で過半数（54％）が賛成票を投じたのは、「たまたま賛成派が多く抽出されてしまっただけだ」と考えるのが帰無仮説である。

対立仮説は、帰無仮説に対置するという意味である。つまり、調査データが示す証拠は偶然などではなく、「母集団の様子を反映した意味のある結果だ」と考える。調査実施者は基本的に調査結果が有効と信じて分析をおこなっているわけであるから、対立仮説が正しいことを示したいと思っている。しかし、全員を調べたわけでもない標本調査で簡単にそのような主張をしてもらっては困るので、一定の統計的な手続きを経ても対立仮説が支持される場合にだけ、データに意味があるという主張を許そう（合格と認めよう）、というのが統計的検定の基本的なシステムである。

帰無仮説と対立仮説は、それぞれ記号でH_0、H_1と表される。

帰無仮説H_0：標本調査の結果に意味がなく、無（偶然）に帰するという仮説。
対立仮説H_1：標本調査の結果が偶然ではなく、意味があるという仮説。

■平均の検定の考え方

さまざまな種類の仮説について、それぞれ検定の手続きがあるが、今回は平均の検定に話を絞る。まず手続きに慣れてもらいたい。

平均の検定の「考え方」は、次のようにまとめられる。

①母集団についての帰無仮説と対立仮説を定義する。

②帰無仮説が正しいと仮定してみる。

③このとき、実際に得られているような標本データが得られる確率が何%かを調べる。

④(a)確率が5%以上→十分にありうることと考え、帰無仮説を採択する。

④(b)確率が5%未満→ありえないことと考え、帰無仮説を棄却(対立仮説を採択)する。

たとえば、65歳以上の高齢者200人に対する標本調査で、1日にテレビを見る時間が平均2.2時間だったとしよう（n＝200、$\bar{X}$＝2.2）。この調査結果から、「高齢者のテレビ視聴時間の平均値は2時間を超えている」と結論付けてよいだろうか。調査結果は確かに「2時間を超えている」が、母集団（高齢者全体）ではもしかるすと違うかもしれない。母集団についてモノをいうには検定の手続きに合格しなければならない。

まず、検定の帰無仮説H_0と対立仮説H_1をはっきりと定義する（①）。違和感があると思われるが、この場合、帰無仮説と対立仮説は次のようになる。

帰無仮説H_0：高齢者のテレビ視聴時間の平均値は2時間である。($\mu=2.0$)

対立仮説H_1：高齢者のテレビ視聴時間の平均値は2時間を超えている。($\mu\neq2.0$)

帰無仮説とは、「標本調査の結果が偶然で意味がない」という状況であるから、母集団の平均値は2時間以下なのにたまたまテレビをよく見ている人を多く調べてしまったということである。その場合、母集団の平均値μは1.8かもしれないし、1.7かもしれないしわけだが、帰無仮説はその中でもっとも「平均が2時間を超える偶然」が起こりやすい1点で表現される（$\mu=2.0$）。対立仮説は帰無仮説の否定であるから$\mu\neq2.0$となる。

次に、唐突であるが、帰無仮説が正しいと仮定してみる（②）。そして、もし帰無仮説が正しいならば、手元にあるようなデータ（標本平均）が偶然に得られる確率は何%くらいあるのかを考える（③）。その確率が5%以上あるのならば、それは十分にありえる偶然と考える。根拠なく帰無仮説が正しいと仮定したが、そのまま帰無仮説を採択することにする（④a）。一方、帰無仮説が正しいと仮定すると手元にあるようなデータが偶然に得られる確率が5%未満しかないならば、ちょっとそういう偶然が起こったと考えるのは無理がある。つまり、帰無仮説を棄却して、対立仮説の方が正しいと考えるしかないだろうと結論付ける（④b）。これが検定に合格するということである。

■中心極限定理からの説明

では、その「確率」はどうやって知ることができるのだろうか。またしても、中心極限定理から考える。標本の平均値$\bar{X}$は95%の確率で以下の範囲に収まることがわかる。

$$\mu-1.96\times\frac{\sigma}{\sqrt{n}}<\bar{X}<\mu+1.96\times\frac{\sigma}{\sqrt{n}}$$

帰無仮説が正しいと仮定すると母平均$\mu = 2.0$である。母標準偏差σの値は不明だが、例によって標本標準偏差sで代用する。かりにs＝0.7だったとしよう。nは標本の人数200である。t分布を参照して「1.96」の数値を補正しなければならないが、今回は自由度が十分に大きいので（n-1＝200-1＝199)、自由度∞の欄で「1.96」のままでよい。以下のように具体的に計算できる。

$$2.0 - 1.96 \times \frac{0.7}{\sqrt{200}} < \bar{X} < 2.0 + 1.96 \times \frac{0.7}{\sqrt{200}}$$
$$2.0 - 0.097... < \bar{X} < 2.0 + 0.097...$$
$$1.90 < \bar{X} < 2.10$$

つまり、もし帰無仮説が正しいのならば（$\mu = 2.0$ならば)、標本平均$\bar{X}$は、95％の確率で1.9～2.1の範囲で現れるはずである。ところが、いま手元にあるデータでは$\bar{X} = 2.2$であり、このような極端な偶然が起こる確率は「非常に低い(5％未満)」ということがわかる(図1)。

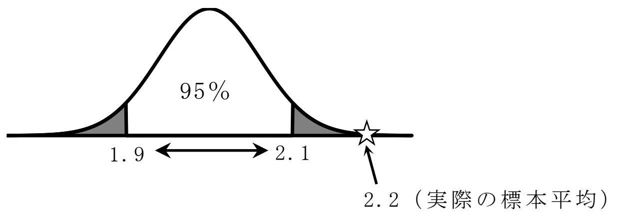


図1　H_0が正しい場合に予想される標本平均の分布

このような低確率の偶然が起こったと考えることは無理があるので､帰無仮説を棄却し、対立仮説を採択することになる。具体的には、この調査結果を元に「母集団でも、高齢者のテレビ視聴時間の平均値は2時間を超えている」ということができる（検定に合格)。

■平均の検定の実際

このように、中心極限定理から考えれば、平均値に関する検定の判断は適切に下すことができる。しかし、実際にはいちいち原理から考えるのは面倒なので、もっと定式的な手続きが定められている。平均の検定の「実際的な手続き」は、次のようにまとめられる。

①帰無仮説と対立仮説を正確に定義する。
②帰無仮説が正しい場合のt値を算出する。
③このt値がt分布（df=n-1）の臨界値を越えているかどうかを調べる。
④(a)t値が臨界値以下→帰無仮説を採択する。
④(b)t値が臨界値を越えている→帰無仮説を棄却（対立仮説を採択）する。

先ほどの例（n＝200、$\bar{X} = 2.2$、s＝0.7）で具体的に手続きをたどってみよう。帰無仮説と対立仮説の定義の仕方は全く同じである（①)。標本調査の結果から、母集団でも高齢者のテレビ視聴時間の平均値が2時間を超えているといってよいだろうか｡形式的な手続きで平均の検定をするときには、次の式にあてはめて帰無仮説（$\mu = 2.0$）が正しいときのt値

を算出する（②）。

$$t=\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}=\frac{2.2-2.0}{\frac{0.7}{\sqrt{200}}}=4.04$$

これは、図1を標準化して、同じことを標準正規分布（補正した場合はt分布）で考えているということである。t値は必ず平均値が0で標準偏差が1になり、自由度n-1のt分布に従う（図2）。t分布の確率表で分布の外側が5%未満になる区切り値を確認すると、自由度199は自由度∞とみなしてよいので、±1.96が区切り値であることがわかる（③）。

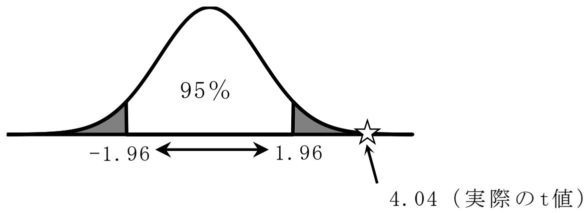


図2　H_0が正しい場合に予想されるt値（df＝∞）の分布

もし本当に帰無仮説が正しいのならば、このt値は0付近の値をとるはずである。多少ずれるにしても、95%の確率で±1.96の範囲に収まるはずである。ところが、現実のデータでのt値は4.04であり、この臨界値を大きくオーバーしている。したがって、偶然によってこのようなデータが得られたとは考えにくく、根拠なく前提とした帰無仮説を棄却して、対立仮説の方を採択する（④b）。つまり、「母集団でも、高齢者のテレビ視聴時間の平均値は2時間を超えている」ということができる。もし仮に、t値が±1.96の範囲に収まったならば、帰無仮説をそのまま採択する（④a）。「母集団でも、高齢者のテレビ視聴時間の平均値は2時間を超えている」とはいえない（このデータでは証拠不十分）と結論づける。

■用語の整理

検定の手続きで使う独特の用語を整理しておこう（図3）。検定のために定式的に算出される統計量のことを**検定統計量**（test statistic）と呼ぶ。平均の検定の場合には、t値が検定統計量である。検定統計量が区切り値を超えているかどうかで、知りたい確率が5%未満かどうかを判断するが、この区切り値のことを**臨界値**（critical value）と呼ぶ。臨界値の内側の範囲を帰無仮説の**採択域**（acceptance region）と呼び、外側の範囲を帰無仮説の**棄却域**（rejection region）と呼ぶ（図3）。

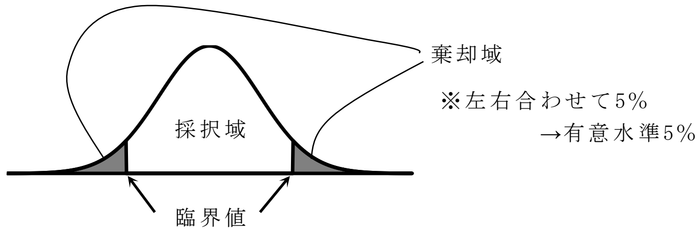


図3　検定に関する用語

また、この確率5％という基準は分析者が変更してもよい（慣習として5％がもっともよく用いられるだけ）。検定で基準とする確率のことを**有意水準**（signigicance level）と呼ぶ。有意水準は「このような偶然は起こらないものとみなす」基準なので、実際にそのような偶然が起こって「判断を誤る危険率」を指し示している。だから、もっと厳しい判定が必要ならば有意水準（危険率）を1％や0.1％に小さくすればよい。逆にもっと大らかな判定でよければ、有意水準を10％などに大きく設定する。

（問題）

1. ある会社の社員から無作為に抽出した150人の標本について、通勤時間を尋ねる調査をおこなった。平均通勤時間が31.9分、標準偏差が13.2分であったとき、母集団の平均通勤時間は30分を超えていると判断してよいかどうか。5％の有意水準で検定しなさい。
2. ある大学の全学生の中から無作為に選んだ20人に対して、自分の容貌を1点（非常に悪い）から5点（非常によい）の5段階で評価してもらったところ、平均3.5点、標準偏差0.8という結果になった。この大学の学生の平均評価はふつう（3点）とは違うと言えるか。5％の有意水準で統計的に検定しなさい。
3. 問題2について、10％の有意水準と1％の有意水準でそれぞれ検定しなさい。
4. S市に住む20代の市民から無作為抽出した30人に対して、親しい友人の人数を尋ねる調査をおこなった。その結果、平均が8.9で、標準偏差が3.8であった。この調査結果から、S市の20代市民は親しい友人の人数が平均10人未満である、と結論づけてよいか。5％の有意水準で検定しなさい。

今日のポイント

①検定は、「標本調査の結果を証拠にして、母集団についても同様とみなせる」といってよいかどうかを判断するための合否判定。

②有意水準5％で平均値の検定をするためには、t値を計算して、臨界値（±1.96）を超えているかどうかを確かめる。超えていればH_0を棄却してH_1を採択。

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

※ただし、標本サイズが小さい（およそ100以下）の場合は、自由度n-1のt分布を参照して、臨界値（±1.96）を置き換える。有意水準（5％）を変更した場合も、対応する値に置き換える。

（連絡）

再来週小テスト②（比率の推定、平均の検定）

第7回「推定と検定（4） 平均の差の検定」

■2つの母集団を比較する

前回までに学習した平均の推定および検定は、「1つの」母集団の特徴を推し測るためのものであった。しかし、我々が社会調査で知りたいことは、単に1つの母集団の特徴ではなく、2つの集団を比較した結果であることが多い。たとえば、調査結果を男女で比較したり、年齢層で比較したりすることがよくある（属性比較）。あるいは、去年の結果と今年の結果を比較することや（時間的比較）、日米での比較（文化・空間的比較）など、さまざまな側面から調査結果の比較はなされる。

こうした比較は、典型的にはそれぞれの集団で何らかの平均値を算出して値の大小を比べることでさなれる。たとえば、「男子学生の平均通学時間が15分なのに対して、女子学生の平均通学時間は20分と、5分長い」といった比較である。しかし、例によってここでわかったことは、男子学生と女子学生の標本の間での差異である。母集団でも同じように男女に差異があるかはわからない。この標本データから、母集団でも差異があるとみなしてよいのだろうか（検定に合格するのか）。統計的な判断のためには、「**平均の差の検定**」と呼ばれる手続きが必要となる。

■平均の差の検定

たとえば、次のように男女差を問題にする場面を想定する。

2009年全国家族調査※では、結婚生活の満足度を1～4の4点満点で尋ねている。結婚3年目の人々（男性32名、女性28名）の調査結果によると、以下のとおり男性の方が満足度が高かった。この調査結果から、母集団（全国の結婚3年目の人々）でも、「男性の方が結婚満足度が高い」といってよいだろうか？

男性： $n_1 = 32$，平均 $\bar{x}_1 = 3.3$，標準偏差 $s_1 = 0.82$

女性： $n_2 = 28$，平均 $\bar{x}_2 = 3.0$，標準偏差 $s_2 = 0.79$

（※全国家族調査（NFRJ）は日本家族社会学会が1999、2004、2009年に実施した全国無作為標本調査で、家族関係や家族意識を幅広く調べている）

この場合、帰無仮説と対立仮説は次のようになる。

帰無仮説H_0：男女で結婚満足度に差はない（$\mu_1 = \mu_2$）

対立仮説H_1：男女で結婚満足度に差がある（$\mu_1 \neq \mu_2$）

平均の検定と同様に、まず帰無仮説が正しいと仮定する。つまり男女で結婚満足度の平均が同じでと仮定する。しかし、実際のデータでは0.3ポイントの差がある。この0.3ポイントの差が偶然の範囲内と言えるのかどうかを判断しなければならない。

通常、有意水準は5%と定める。帰無仮説が正しいときでも0.3ポイントの平均差が生じるという確率が5%以上あれば、この0.3ポイントの差は偶然に起こりうる範囲内とみなさ

れる（帰無仮説を採択）。逆に、その確率が5%未満であれば、この差は偶然には表れえないもので、母集団の平均に実際に差があったのだとみなす（対立仮説を採択）。

地道に考えれば､その正確な確率を計算することは可能であるが､それは頭が混乱する。そのため、やはり「検定統計量が一定の値（臨界値）を越えているかどうか」で、どちらの仮説を採択するかを決定する。統計的検定の常套手段である。

平均の差の検定では、検定統計量のt値は次の計算式で算出される。

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}}$$

前回の平均の検定の場合とは式が異なるのに「t値」という同じ名前なのは、臨界値を調べるときに参照するのが同じt分布だからである。このt値は、帰無仮説が正しい場合、自由度df＝$(n_1-1)+(n_2-1)=n_1+n_2-2$ のt分布を示すことが知られている。t値がこの分布から大きくかけ離れていれば、帰無仮説が棄却されることになる。「平均の検定」でもt分布を利用していたが､t分布を利用する検定でもっとも多用されるのは｢平均の差の検定｣なので、**t検定**といえばふつう平均の差の検定を指す。

この式の $S_{\bar{X}_1 - \bar{X}_2}$ は、「標本平均の差」の標準偏差で、次のように算出される。$S_{\bar{X}_1 - \bar{X}_2}$ は、2つの標本集団のばらつき具合を標本サイズの違いを加味しながら足し合わせることで算出されている。

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

2つのグループの標本数が等しくnの場合、簡単になる。 $\sqrt{\frac{S_1^2 + S_2^2}{n}}$

※ n_1=1つ目の標本グループの標本サイズ　n_2 2つ目の標本グループの標本サイズ
$\bar{X}_1$ 1つ目の標本グループの平均　$\bar{X}_2$ 2つ目の標本グループの平均
S_1 1つ目の標本グループの標準偏差　S_2 2つ目の標本グループの標準偏差

結婚婚満足度の男女差での、実際に検定の例

帰無仮説H_0：男女で結婚満足度に差はない（$\mu_1 = \mu_2$）
対立仮説H_1：男女で結婚満足度に差がある（$\mu_1 \neq \mu_2$）

有意水準5%で平均の差の検定を行う。

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}} = \frac{3.3 - 3.0}{0.21} = 1.43$$

（※ $S_{\bar{X}_1 - \bar{X}_2}$ は右のとおり計算）

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$
$$= \sqrt{\frac{(32-1)0.82^2 + (28-1)0.79^2}{32+28-2} \times \left(\frac{1}{32} + \frac{1}{28}\right)} = 0.21$$

自由度＝n_1+n_2-2＝32＋28-2＝58のt分布（自由度60で代用）において臨界値は±2.00。

t＜2.00なので、帰無仮説の採択域に入る。

「男女で結婚満足度に差がある」とはいえない。

■平均の差の区間推定

平均の差については、検定ではなく推定も当然できる。つまり、母集団でグループ間の平均の差がどの程度になるのかを、「95%の確信をもって、平均の差は○○～××の範囲にある」という形で区間推定することができる。

考え方自体は、平均の推定の場合とまったく同じなので省略する。結論としては、母平均の差$\mu_1-\mu_2$は、危険度5%（信頼度95%）で次の信頼区間に収まると推定できる。

$$\left(\bar{X}_1-\bar{X}_2\right)\pm 1.96\times S_{\bar{X}_1-\bar{X}_2}$$

（$S_{\bar{X}_1-\bar{X}_2}$の算出は検定の場合と同様）

※ただし、標本の人数が少ない場合（合計しておよそ100以下の場合）、自由度n_1+n_2-2のt分布に照らし合わせて、±1.96の値を補正する。信頼度を95%から変更する場合も、同様に確率表で対応する値を参照して置き換える。

先ほどの結婚満足度の例で、実際に平均の差の区間推定を行うと次のようになる（信頼度95%）。つまり、「差異を大きく見積もれば、男性の方が平均0.72大きいと予想されるが、少なく見積もれば-0.12（男性の方が平均値が小さい）という予想も成り立つ」ということである。

$$\begin{aligned}\mu_1-\mu_2&=(3.3-3.0)\pm 2.00\times 0.21\\&=0.3\pm 0.42\\&=-0.12\sim 0.72\end{aligned}$$

ただし、実際にはこの種の「平均の差の推定」はあまり行われない。この場合、男女それぞれのグループについて「平均の推定」を行い比較した方が、よほどわかりやすいからである。この場合、男性の平均値は$3.3\pm 2.04\times\frac{0.82}{\sqrt{32}}=3.0\sim 3.6$と推定され、女性の平均値は$3.0\pm 2.05\times\frac{0.79}{\sqrt{28}}=2.7\sim 3.3$と推定できる。「平均の差の推定」は参考程度に留めておこう。

今日のポイント
①平均の差の検定（t検定）は、社会調査のデータ分析で頻繁に用いられる。
②有意水準5%で「平均の差の検定」をするためには、下式でt値を計算して、臨界値（±1.96）を超えているかどうかを確かめる。超えていればH_0を棄却してH_1を採択。

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}} \qquad \left[S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{(n_1 - 1)S_1{}^2 + (n_2 - 1)S_2{}^2}{n_1 + n_2 - 2} \times \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right]$$

※ただし、標本サイズが小さい（およそ100以下）の場合は、自由度n_1+n_2-2のt分布を参照して、臨界値（±1.96）を置き換える。有意水準（5%）を変更した場合も、対応する値に置き換える。

（問題）

1. ある会社で社員の中から無作為に選ばれた正規雇用の20人、非正規の12人に、会社への満足度を尋ねて得点化した。その結果、正規雇用の満足度は平均50点、標準偏差10であり、非正規の満足度は平均43点、標準偏差8であった。母集団において正規と非正規の間に満足度に差があるといえるか。有意水準5%で検定せよ。

2. 2009年全国家族調査の結果から、30代前半（30〜34歳）で働く男女の通勤時間を整理すると次のようになった。男性の方が8.1分通勤時間が長いことがわかる。

 男性：228名、平均34.0分、標準偏差23.3分
 女性：173名、平均25.9分、標準偏差20.7分

（1）この調査結果ら、母集団でも「男性の方が平均通勤時間が長い」と結論づけてもよいだろうか。有意水準5%で平均の差を検定しなさい。

（2）現代日本社会において、このような通勤時間の差があるのはなぜだろうか。考察しなさい。

3. 大阪府の成人から300人を無作為に選び、現在の内閣を支持するかどうかを尋ねた。結果は55%が支持しているというものであった（300人全員が回答したものとする）。

（1）母集団において内閣の支持率は50%より高いといえるか。有意水準5%で検定せよ。
（2）先月、やはり300人に調査をしたときには、支持率が58%であったという。母集団において、先月と今月の間に差があったといえるか。有意水準5%で検定せよ。

（連絡）

来週小テスト②（比率の推定、平均の検定＋平均の差の検定）

第8回「小休止（復習）」

1．一般に、「検定に合格」とはどういう状態を指すか。以下の（1）～（8）のそれぞれについて、正しい方に〇を付けなさい。

（1）　A　調査でわかったことが母集団についても成り立つ
　　　B　調査でわかったことが母集団については成り立たない

（2）　A　調査結果が偶然の産物にすぎない可能性が結構ある
　　　B　調査結果が偶然の産物である可能性はほとんどない

（3）　A　帰無仮説を棄却する
　　　B　帰無仮説を採択する

（4）　A　対立仮説を棄却する
　　　B　対立仮説を採択する

（5）　A　手元のような調査結果が偶然に得られる確率が5%以上ある
　　　B　手元のような調査結果が偶然に得られる確率は5%未満しかない

（6）　A　検定統計量が臨界値を超えている（臨界値の外側）
　　　B　検定統計量が臨界値を超えていない（臨界値の内側）

（7）　A　調査結果が統計的に有意である
　　　B　調査結果が統計的に有意ではない

（8）　A　平均の差の検定の場合、「平均の差がある」といえる
　　　B　　　〃　　　　　、「平均の差がある」といえない

2．次のような場合に、検定の結果はどのように判断されるのか。（1）～（4）のそれぞれについて、正しい方に〇を付け、結論を具体的な文で記しなさい。

（1）顧客満足度調査では、男性の平均が5.5点、女性の平均が7.2点だった。平均の差を検定すると、t＝-3.27で、臨界値が±2.02。
⇒　検定統計量が臨界値を［超えている・超えていない］
　　検定に［合格・不合格］
　　H_0を［棄却・採択］
　　結論（…といえる/いえないという形で）:

　　母集団で____

（2）去年の4年生は選挙に行ったことのある学生が45%しかいなかったが、今年の4年生は52%いた。今年の方が選挙に行った比率が高いといってよいか、0/1のダミー変数と考えて、比率の差を検定すると、t＝1.44で、臨界値が±1.96。
⇒　検定統計量が臨界値を［超えている・超えていない］
　　検定に［合格・不合格］
　　H_0を［棄却・採択］
　　結論（…といえる/いえないという形で）:

　　母集団で____

(3) 小学生のテレビ視聴時間は1日1時間程度と予想していたが、調査では平均1.3時間という結果が出た。平均は1時間でないと結論づけてよいか、平均の検定をすると、t＝1.69で、臨界値が±1.96。

⇒ 検定統計量が臨界値を［超えている・超えていない］
検定に［合格・不合格］
H_0を［棄却・採択］
結論（…といえる/いえないという形で）：

母集団で______________________________

(4) －5点～＋5点の範囲で、自分の性格を自己評価してもらったところ、平均値は－0.8点になった。性格の自己評価は標準（±0）よりも低いといってよいか。平均の検定の結果は、t＝-4.29で、臨界値が±2.04。

⇒ 検定統計量が臨界値を［超えている・超えていない］
検定に［合格・不合格］
H_0を［棄却・採択］
結論（…といえる/いえないという形で）：

母集団で______________________________

3.「幼児期に母親が外で働くことは子どもの成長に悪い」という意見に賛成か反対かを＋2点（とても賛成）～－2点（とても反対）の5段階で答えてもらう調査を行った。996人から回答を得て、平均値が-0.52、標準偏差が1.25であった。

(1) 95%の信頼度で平均得点の信頼区間を区間推定せよ。

(2) 母集団でも、「平均値が0より小さい（全体的には、母親の仕事は有害と考えられない傾向がある）」と判断してよいであろうか。有意水準5%で検定せよ。（両側検定で）

4. ある2つの地域で5年生を20人ずつ無作為に選んで、算数の学力試験を行った。その結果、地域Aでは、平均得点が73点、標準偏差が10.0であった。一方、地域Bでは、平均得点が68点、標準偏差が8.0であった。
2つの地域で算数の学力に差があると言ってよいであろうか（平均得点に差があるといってよいであろうか）。有意水準5%で検定を行え。（両側検定で）

5. ある県の35歳の男性700人を調査して、父親が存命の場合と亡くなっている場合に分けて分析をした。その結果、父親が存命の男性は576人いて、そのうち360人（62.5%）が結婚していた。一方、父親が亡くなっている男性は124人いて、結婚しているのは62人（50.0%）だった。

（1）この結果から、「父親が亡くなっている男性の方が、35歳で結婚している割合が低くなる」と一般化してよいだろうか。平均の差の検定を応用して、比率の差を有意水準5%で検定しなさい。

（2）父親が亡くなっている男性と亡くなっていない男性のそれぞれについて、母集団での既婚率を95%の信頼度で区間推定しなさい。

（補足）両側検定と片側検定

統計的検定において、帰無仮説の棄却域は正の側にも負の側にも両方設けられている。しかし、これを片側だけにすべきだ、という考え方もある。標本データで得られる情報には、ふつう方向性があるからである。

たとえば、標本データの平均通学時間が男子学生で15分、女子学生で20分だったとき、得られた情報は「男女で平均通学時間が異なる」というよりも、「女子学生の方が平均通学時間が長い」と考えるのがふつうである。この場合、対立仮説が支持される（標本データと同じ傾向が母集団についてもあてはまるとみなす）t値の領域は、片側だけに設けられると考える方が自然であろう。

両側に帰無仮説の棄却域を設ける通常の検定を**両側検定**（two-tailed test）と呼び、情報の方向性を考えて棄却域を片側だけに限定する検定を**片側検定**（one-tailed test）と呼ぶ。棄却域のある方向によって右側検定（あるいは上側検定）、左側検定（あるいは下側検定）と呼ぶこともある。この授業では、検定は両側検定でおこなうことを前提にしているが、概念としては知っておこう。

片側検定をおこなう場合、有意水準が同じ5%でも臨界値が変わってくることに注意が必要である。なぜならば、両側検定では両方を足して5%であったのが、片側検定では一方だけで5%の面積を占めるのであるから、その臨界値はより内側になる。具体的には、両側検定における有意水準10%の臨界値が、片側検定における有意水準5%の臨界値に相当する。したがって、片側検定は、結局両側検定よりも検定の基準を甘くしていることになる。これは片側検定の使用が実際上は避けられる理由の1つである。

第9回「推定と検定（5） 独立性の検定」

■クロス表についての検定

以前学習した「平均の差の検定」は、非常によく用いられる検定手続きであるが、これが使えるのは、注目している変数が、平均を算出できる変数、すなわち量的変数の場合である。注目している変数が質的変数の場合、同様の手続きはクロス表について検定をおこなうことに置き換えられる。

仮に次のような話を考えよう。K市に住んでいる30代で独身の男女の中から無作為に標本を抽出し、アンケート調査を行った。男女別にその結婚観（今後、結婚をするつもりかどうか）の回答をクロス集計表に整理したところ、下のような表になった。ざっと見たところ、女性の方が男性よりも「結婚をするつもり」という傾向が読める。

表1 性別と結婚観のクロス表（観察度数）

	するつもり	どちらともいえない	しないつもり	計
男性	14 *35.0%*	12 *30.0%*	14 *35.0%*	40 *100%*
女性	28 *63.6%*	9 *20.5%*	7 *15.9%*	44 *100%*
計	42 *50.0%*	21 *25.0%*	21 *25.0%*	84 *100%*

つまり、性別と結婚観の間には何らかの関連性があるように見える。統計学では、2つの変数の間に関連性がないことを、「2つの変数は統計的に**独立**（independent）である」と言う。ここの例では、性別と結婚観は独立ではない、と言い換えられる。

しかし、例によって、これは標本の集計に過ぎない。母集団では独立であるにもかかわらず、偶然に偏った標本が選ばれたのかもしれない。母集団でも関係が独立でないことを確認するためには、やはり検定が必要である。帰無仮説と対立仮説は次のようになる。

H_0: 性別と結婚観は、独立である（関連性がない）。

H_1: 性別と結婚観は、独立ではない（関連性がある）。

このような検定を**独立性の検定**（test of independence）と呼ぶ。クロス表の検定とか、カイ二乗検定と呼ぶこともある。とくに社会調査データでは質的変数が多いので、独立性の検定は、平均の差の検定と並んでもっとも頻繁に利用される。

■期待度数の算出

検定の常套手段として、形式的な検定統計量が臨界値を超えるかどうかで帰無仮説を採択するか、それとも対立仮説を採択するかを決定する。独立性の検定で用いる検定統計量は以下のような**カイ二乗値［χ^2値］**（chi-square）である。

$$\chi^2 = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j}$$ （jは何個目のセルかを表わし、k個のセルがあると定めている）

カイ二乗値を算出するためには、**期待度数**（expected frequency）という考え方を理解する必要がある。検定の一般的な考え方は、「帰無仮説が正しいと仮定したときに、手元にあるデータが、実際に偶然によって得られる確率が何％程度あるかを調べ、判断の根拠とする」というものであった。独立性の検定でも、帰無仮説が正しいと仮定してみるところから話が始まる。つまり、クロス表の2つの変数の間にまったく関連性がない（独立）ならば、どういったクロス表になるべきと期待されるのか、を考える。

もし帰無仮説が正しく、性別と結婚観が統計的に独立であるならば、たとえば、ある標本ケースが「男性」で「結婚をするつもり」である確率は、単純に「男性である確率」と「結婚をするつもりである確率」のかけ算で求まるはずである。それぞれの本当の確率は不明だが、データから一番もっともらしい数値を決めるならば、周辺度数（合計値）の割合を根拠にすればよい。すなわち、あるケースが「男性である確率」は$\frac{40}{84}$である。同じように、「結婚をするつもりである確率」は$\frac{42}{84}$である。したがって、あるケースが「男性」かつ「結婚をするつもり」である確率は、$\frac{40}{84} \times \frac{42}{84}$である。

そして標本ケースは全部で84人いるわけだから、「男性」で「結婚をするつもり」のセル度数はこの確率が84回繰り返されて、$\frac{40}{84} \times \frac{42}{84} \times 84 = 20$と、20人になることが予測される。このように帰無仮説を前提として算出されるセル度数を「期待度数」と呼ぶ。

他のセルの期待度数も同様のかけ算で求まる（表2）。たとえば、「女性」で「結婚しないつもり」のセルの期待度数は、$\frac{44}{84} \times \frac{21}{84} \times 84 = 11$人である。

表2　性別と結婚観のクロス表（期待度数）

	するつもり	どちらともいえない	しないつもり	計
男性	20	10	10	40
女性	22	11	11	44
計	42	21	21	84

■期待度数を用いたカイ二乗値の算出

もともと標本データで実際に観察された度数（**観察度数［観測度数］**（observed frequency）と呼ぶ）と、理論的に計算された期待度数を比較すると、表3のように整理される。

表3　観察度数と期待度数の比較

	観察度数O_j	期待度数E_j
男性・するつもり	14	20
男性・どちらともいえない	12	10
男性・しないつもり	14	10
女性・するつもり	28	22
女性・どちらともいえない	9	11
女性・しないつもり	7	11
計	84	84

観察度数と期待度数の間にはいくらかのずれがある。独立性の検定で問題にしているのは、この程度のずれが偶然でも起こりうるのか、それとも偶然には起こりえないのか（確率が5%未満なのか）、ということである。そのため、検定統計量のカイ二乗値では、すべてのセルについて、標準化されたずれの程度を算出し、それを足し合わせることで全体としてのずれの大きさを評価している。今回のデータの場合、以下のようになる。

$$\chi^2 = \frac{(14-20)^2}{20} + \frac{(12-10)^2}{10} + \frac{(14-10)^2}{10} + \frac{(28-22)^2}{22} + \frac{(9-11)^2}{11} + \frac{(7-11)^2}{11}$$
$$= 1.8 + 0.4 + 1.6 + 1.64 + 0.36 + 1.45 \quad = 7.25$$

帰無仮説が本当に正しければ、ずれの大きさ（カイ二乗値）は図1のようなカイ二乗分布に従い、平均的には0付近のそれなりに小さい値をとるはずである。ただし、偶然によっていくらかは大きな値をとることもあるので、実際のずれ（カイ二乗値）が偶然ではありえない程度の大きさとみなすべきかどうかは、臨界値を超えるかどうかで判別する。

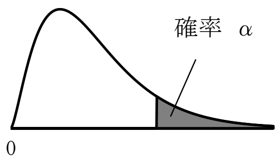


図1　χ^2分布

臨界値はカイ二乗分布の確率表を参照するが、カイ二乗分布にも、t分布と同じように自由度が存在する。独立性の検定における自由度は、(クロス表の行数-1) ×（クロス表の列数-1) である。ここでは、df=(2-1)×(3-1)=2となる。自由度の意味は、後の回で解説する。

自由度2のカイ二乗分布の表を確認すると、有意水準5%での臨界点は5.99である。したがって、7.25というカイ二乗値は、偶然には起こりえない。帰無仮説は棄却され、対立仮説の方が採択される。つまり、母集団でも「性別と結婚観の関係が独立ではない」ということができる。要するに、母集団でも「男女で結婚観が異なる」という結論は信用できる。

■度数サイズの必要性

独立性の検定について、一点だけ大切な注意をしておこう。カイ二乗値を用いた検定は、各セルの期待度数がある程度大きくないと、不適切な確率を導いてしまうことがある。そのため、いずれかのセルの期待度数が5未満になることは避ける、という経験則がある。経験則なので厳密な根拠はないが、実際に期待度数が小さすぎると、確率計算がおかしくなることは多くなるので、注意しなければならない。

社会調査で期待度数が小さくなるのは、ほとんどの場合、質問項目の選択肢が多すぎて、細かいクロス表になっているときである。このような場合には、可能な範囲で複数の選択肢をまとめてクロス表を縮約するべきである（たとえば、「とても賛成」と「ある程度賛成」を1つの「賛成」にまとめるなど)。別の手段として、カイ二乗値を使わずに確率を計算する「正確検定」という技法もある。しかしながら、社会調査のデータは正確検定に耐えうるほど厳格に収集されていないことが多い。やはり、データの身の丈に合うようにクロス表に縮約して、できる範囲の分析をおこなう方がよいだろう。

今日のポイント
①クロス表に対する独立性の検定は、社会調査のデータ分析で頻繁に用いられる。
②有意水準5%で「独立性の検定」をするためには、観察度数O_jと期待度数E_j※1を元に、下式でカイ二乗値を計算して、臨界値※2を超えているかどうかを確かめる。超えていればH_0を棄却してH_1を採択。

$$\chi^2 = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j}$$

※1　たとえば、下表のセルaの期待度数は$\frac{72}{127} \times \frac{88}{127} \times 127 = 49.9$と算出できる。

	Y=1	Y=2	計
X=1	a	b	72
X=2	c	d	45
計	88	39	127

※2　自由度＝(クロス表の行数-1)×(クロス表の列数-1)のカイ二乗分布で確認する。

（問題）

2008年実施の全国調査JGSS-2008では、安楽死を法律で認めることへの意識を尋ねている。30代の回答者と70代の回答者で結果を比較するために、下の表を作成した。

	安楽死に賛成か			
	はい	いいえ	わからない	計
30代の回答者	116 (　　　)%	8 (　　　)%	40 (　　　)%	164
70代の回答者	178 (　　　)%	20 (　　　)%	93 (　　　)%	291
計	294	28	133	455

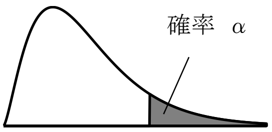
注：質問文は「不治の病におかされた患者が、痛みを伴わない安楽死を望んでいるとします。その家族も同意している場合に、医者が安楽死を行なえる法律をつくるべきだと思いますか」

(1) 比較に適するように行%か列%を記入して、30代と70代の標本の間にどのような違いがあるか記述しなさい。

(2) この表に見られるような世代と意識の関連性が、母集団についてもあてはまると考えてよいか。有意水準5%で独立性の検定をおこなって判定しなさい。なお、検定のためにはまず期待度数を算出する必要がある。(割り切れない期待度数は、小数点以下第一位くらいまで求めれば、ほとんど誤差なく計算できる)

（期待度数）	安楽死に賛成か			
	はい	いいえ	わからない	計
30代の回答者				164
70代の回答者				291
計	294	28	133	455

【カイ二乗分布（χ^2分布）の確率表】



自由度	α 0.1	0.05	0.01	0.001
1	2.71	3.84	6.63	10.83
2	4.61	5.99	9.21	13.82
3	6.25	7.81	11.34	16.27
4	7.78	9.49	13.28	18.47
5	9.24	11.07	15.09	20.52
6	10.64	12.59	16.81	22.46
7	12.02	14.07	18.48	24.32
8	13.36	15.51	20.09	26.12
9	14.68	16.92	21.67	27.88
10	15.99	18.31	23.21	29.59
11	17.28	19.68	24.72	31.26
12	18.55	21.03	26.22	32.91
13	19.81	22.36	27.69	34.53
14	21.06	23.68	29.14	36.12
15	22.31	25.00	30.58	37.70
16	23.54	26.30	32.00	39.25
17	24.77	27.59	33.41	40.79
18	25.99	28.87	34.81	42.31
19	27.20	30.14	36.19	43.82
20	28.41	31.41	37.57	45.31
21	29.62	32.67	38.93	46.80
22	30.81	33.92	40.29	48.27
23	32.01	35.17	41.64	49.73
24	33.20	36.42	42.98	51.18
25	34.38	37.65	44.31	52.62
26	35.56	38.89	45.64	54.05
27	36.74	40.11	46.96	55.48
28	37.92	41.34	48.28	56.89
29	39.09	42.56	49.59	58.30
30	40.26	43.77	50.89	59.70

注：それぞれの自由度のカイ二乗分布で、確率αに対応する臨界値だけを記している。

※一応もう一度お知らせしておきますが、配布プリントは授業後にWebにアップしています。とくに、確率表はずっと使うので、なくしたら各自で取得しておいてください。

http://www2.itc.kansai-u.ac.jp/~tyasuda/

第10回「統計分析ソフトでの実際」

■補足：比率の差を検定する2つの方法

以前、「比率とは 0 か 1 しかとらない変数（ダミー変数と呼ぶ）の平均値にあたる」ことを解説した。これを応用すると、比率の推定や検定に加えて、2 つのグループの間の比率の差の検定もできる。一方で、比率の差の検定には、前回学習した独立性の検定を用いることもできる。どちらの方法であってもよいので、実践できるようにしておこう。

※標本の人数が少ないとき（およそ 100 人未満のとき）にはダミー変数で平均値を考えることに無理が生じる。その場合は独立性の検定を利用する方がよい。

たとえば、ある大学で 1 年生と 2 年生を 200 人ずつ調べたとき、アルバイトをしている割合が 52%と 58%という違いがあったとする。このとき、母集団でも学年によってアルバイト率が違うといってよいだろうか。検定は以下のように 2 通りの方法でできる。

ダミー変数で平均の差の検定をした場合

1 年生　$n_1 = 200,\quad \bar{X}_1 = 0.52 \quad s_1 = \sqrt{0.52 \times (1 - 0.52)} = 0.50$

2 年生　$n_2 = 200,\quad \bar{X}_2 = 0.58 \quad s_2 = \sqrt{0.58 \times (1 - 0.58)} = 0.49$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}} = \frac{0.52 - 0.58}{0.05} = -1.2$$

$$\left[S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{(n_1 - 1)S_1{}^2 + (n_2 - 1)S_2{}^2}{n_1 + n_2 - 2} \times \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \sqrt{\frac{199 \times 0.50^2 + 199 \times 0.49^2}{398} \times \left(\frac{1}{200} + \frac{1}{200} \right)} = 0.05 \right]$$

自由度∞のt分布（標準正規分布）で、有意水準5%の臨界値は±1.96なので、H_0を採択。
「アルバイト率に差がある」とはいえない。

クロス表で独立性の検定をした場合

	バイトしている	していない	計
1 年生	104（52%）	96（48%）	200
2 年生	116（58%）	84（42%）	200
計	220	180	400

↓

（期待度数）

	バイトしている	していない	計
1 年生	110	90	200
2 年生	110	90	200
計	220	180	400

$$\chi^2 = \frac{(104-110)^2}{110} + \frac{(96-90)^2}{90} + \frac{(116-110)^2}{110} + \frac{(84-90)^2}{90} = 1.45$$

自由度(2-1)×(2-1)＝1 のカイ二乗分布で、有意水準 5%の臨界値は 3.84 なので、H_0を採択。
「アルバイト率に差がある」とはいえない。

■あとは同様

ここまでに、以下のような推定・検定の手続きを学習してきた。

・平均の推定
・平均の検定　（比率の推定・検定を含む）
・平均の差の検定
・独立性の検定
・比率の差の検定　（平均の差の検定の応用でも、独立性の検定でもOK）

社会調査の分析で頻繁に用いられる推測統計の手続きは、以上である。他の種類の推定や検定を行うこともあるが、推定・検定の基本的な考え方さえわかっていれば、検定統計量などの計算式が変わるだけで、まったく同じように扱えばよい。

たとえば、相関係数（2つの量的変数の関係性を－1～＋1の範囲で表す）が統計的に有意（母集団でも相関がある）かどうかを検定したければ、その検定統計量は下のとおりである。この数値が自由度n-2におけるt分布の臨界値を超えていれば、検定に合格ということになる。

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

■統計分析ソフトでの実際の推定・検定

ここまでに用いてきた手続きでは、決められた検定統計量について、有意水準に相当する値（臨界値）を調べ、それを超えるかどうかで帰無仮説を採択するか、対立仮説を採択するかを決めてきた。算出された検定統計量の値ごとに対応する確率を載せていると、膨大な表になってしまうため、区切りになる重要な値の対応だけを掲載していたわけである

例：自由度１のカイ二乗分布では……

χ^2値		知りたい確率
2.71	⇔	10%
3.84	⇔	5%
6.64	⇔	1%
10.83	⇔	0.1%

しかしながら、コンピュータ（統計分析ソフト）を使って、検定を実行させた場合には、このような配慮をする必要がなくなる。算出された検定統計量がどのような値でも、対応した確率を計算してくれるからである。そのような確率のことを**有意確率**（significance probability）と呼ぶ。有意確率は記号「p」あるいは「p値（p-value)」と表され、「p=.043」といった表現で示される（この場合、有意確率が4.3%、つまり標本で得られたようなデータが偶然に得られる確率は4.3%という意味)。

そのため、統計分析ソフトを使う場合は、もっと直接的に検定の手続きを進められる。有意確率が有意水準を下回っているかどうか（通常は5%未満、つまり「p＜.05」かどうか）

を確認するだけで、検定に合格か不合格かを判断でき、非常に簡単である。

> 有意確率が 5%未満（p＜.05）
> 　⇒検定に合格したので、調査結果は統計的に有意（信頼して母集団についてものをいってよい）
> 有意確率が 5%以上（p≧.05）
> 　⇒検定に不合格なので、調査結果は統計的に有意でない（データを鵜呑みにするのは危険）

図1、2に、**統計分析ソフトSPSS**を用いて平均の差の検定と独立性の検定をおこなった場合のアウトプットを示した。この中から、自由度（df）、検定統計量（tなど）、有意確率（p）という3つの数値を抜き書きすればよい。調査データ分析の実践では、こういったアウトプットの読み取りに慣れることが必要になる。なお、有意確率だけを見ればよいので、他の数値（自由度と検定統計量）は要らないのではないかと思えるかもしれないが、有意確率は自由度と検定統計量を根拠として概算されているので、レポート等ではこれらの情報も明記することがふつうである。

（レポート等での書き方の例）
「恋人に友人が多いとうれしいか」を4点満点で尋ねると、男性は平均2.32に対して、女性は平均2.68で、女性の方が喜ぶことがわかった。有意水準5%で平均の差の検定をすると統計的に有意だったので、母集団についても男女差があるといってよい（df＝183、t＝-2.814、p＝.005）。

グループ統計量

	Q1 性別	度数	平均値	標準偏差	平均値の標準誤差
Q12C 恋人に友人が多い	1 男性	94	2.32	.858	.088
	2 女性	91	2.68	.893	.094

独立サンプルの検定

		等分散性のための Levene の検定		2 つの母平均の差の検定						
									差の 95% 信頼区間	
		F	有意確率	t	df	有意確率 (両側)	平均値の差	差の標準誤差	下限	上限
Q12C 恋人に友人が多い	等分散が仮定されている	.111	.739	-2.814	183	.005	-.362	.129	-.616	-.108
	等分散が仮定されていない			-2.812	182.037	.005	-.362	.129	-.616	-.108

図1　SPSSの出力例（平均の差の検定）

Q5 高校は男女共学でしたか と Q6 恋人がいますか のクロス表

			Q6 恋人がいますか		
			1 いる	2 いない	合計
Q5 高校は男女共学でしたか	1 男女共学である	度数	35	129	164
		Q5 高校は男女共学でしたか の %	21.3%	78.7%	100.0%
	2 男女共学でない	度数	3	19	22
		Q5 高校は男女共学でしたか の %	13.6%	86.4%	100.0%
合計		度数	38	148	186
		Q5 高校は男女共学でしたか の %	20.4%	79.6%	100.0%

カイ 2 乗検定

	値	df	漸近有意確率 (両側)	正確有意確率 (両側)	正確有意確率 (片側)
Pearson のカイ 2 乗	.708[a]	1	.400		
連続修正[b]	.314	1	.575		
尤度比	.771	1	.380		
Fisher の直接法				.575	.299
線型と線型による連関	.705	1	.401		
有効なケースの数	186				

a. 1 セル (25.0%) は期待度数が 5 未満です。最小期待度数は 4.49 です。
b. 2x2 表に対してのみ計算

（レポート等での書き方の例）
高校が男女共学だったかどうかによって恋人がいる割合は異なり、男女共学の方が恋人がいる割合が7.7ポイント高かった（21.3%と13.6%）。ただし、独立性の検定をすると有意水準5%で統計的に有意ではなかった。母集団でこの2つの変数に関連があるとはいいきれず、偶然の結果という可能性がある（df＝1、χ^2＝.708、p＝.400）。

図2　SPSSの出力例（独立性の検定）

■補足：自由度の意味

これまで「自由度」という言葉を何の断りもなしに使用してきたが、簡単に説明しておこう。抽象的に表現するならば、自由度とは「ある確率分布について自由に（確率的に）値が決まる数値の個数」である。

たとえば、独立性の検定では、自由度が（クロス表の行数-1）×（クロス表の列数-1）であった。この検定では、帰無仮説が正しければ、各セルの度数は期待度数のようになることが予想される。しかし、それは「必ず期待度数どおりでなければならない」ということではない。期待度数からいくらかずれた「自由な（確率的な)」値をとることができる。ただ、全てのセル度数が完全に自由なわけではない。なぜならば、期待度数を求める際に、周辺度数（周りの合計値）を利用して確率を定めているので、自由な中でも周辺度数は固定されていなければならないからである。例えば、図3の3×4のクロス表では、太枠の中の度数を自由に決めると、網がけの部分の度数は自動的に決まることになる。このため、3×4のクロス表では、独立性の検定の自由度は、(3-1)×(4-1)=6となる。

	B_1	B_2	B_3	B_4	計
A_1					20
A_2					30
A_3					50
計	20	20	30	30	100

自由でない

図3　自由度の考え方

自由度の大きさは、確率にしたがって自由に動ける値の個数を表しているので、自由度が大きいとそれだけχ^2値が大きくなる可能性が高い。そのため、同じχ^2＝6.55であっても、自由度の大きさによって評価が異なる。例えば、自由度が2の場合、「自由に動ける値が2個しかないにもかかわらず、6.55ものχ^2値（期待度数との乖離の度合いを表す）を示している」とみなされ、帰無仮説は棄却される。一方、自由度が10の場合、「10個もの自由に動ける値があるにもかかわらず、χ^2値(期待度数との乖離)は6.55にしかなっていない」と評価され、帰無仮説が採択されることになる。

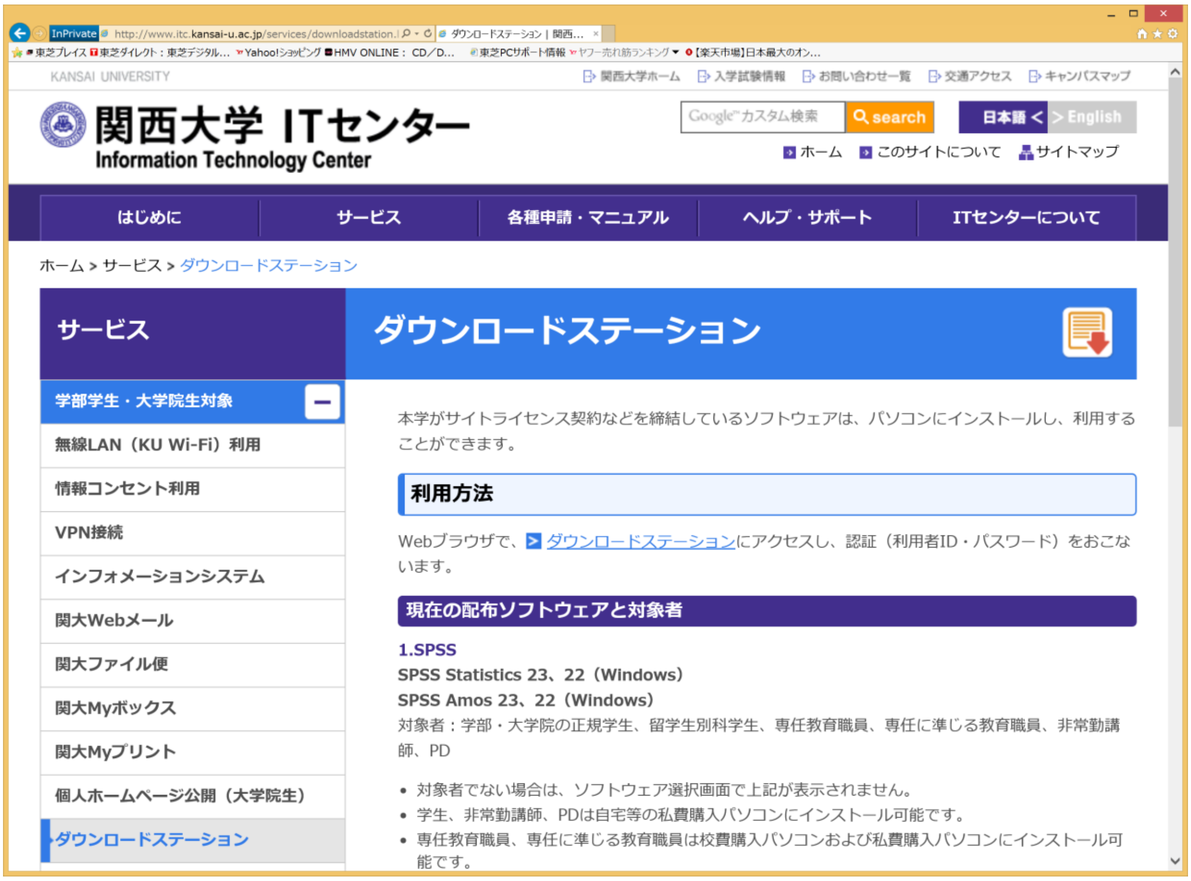
同じように平均の検定等で用いたt分布の自由度も考えることができる。平均の検定の自由度は「n-1」であった。これは、平均の検定では、標本の標準偏差sを計算に利用しているためこの値は一定に保たなければならないためである。そのため、最後の1人の値だけは自由に決めることができない。平均の差の検定では、2つのグループの標準偏差をそれぞれ一定にしなければならないので、n_1+n_2-2が自由度になる。

今日のポイント
①比率の差の検定はできるようになっておこう
（※平均の差の検定の応用でも、独立性の検定の応用でも、どちらでもよい）
②統計分析ソフトでの検定は、直接、有意確率（p）が5%未満かを見ればよい

（連絡）

は統計分析ソフトSPSSはITセンターで利用できる。

また、インストール手続きはややめんどうだが、自宅のPCにもインストールして自由に在学中は利用できる。基礎研究、基礎演習、ゼミ活動等での調査データの分析に活用しよう。



（問題1）

ある市で街の再開発への賛否を尋ねる標本調査をおこなったところ、北部の住人は300人中66%が賛成で、南部の住人は300人中59%が賛成、という結果が出た。

(1)「北部の方が賛成率が高い」といってよいかどうか。ダミー変数（0か1しかない変数）の平均の差の検定とみなして、比率の差を有意水準5%で検定してみよう。
(2) 同じデータにつて、独立性の検定の応用で、比率の差を検定してみよう。

（問題2）

「学生の恋愛観に関する調査」について、知りたい集計をいくつか提案してみよう。実際に、統計分析ソフト SPSS で結果を出力し、推定や検定を含めて結果を読み取ってみよう。

問1.　別をお答えください。
　1．男性
　2．女性

問2.　学年をお答えください。
　（　　　）年

問3.　一人暮らしですか。
　1．一人暮らしである
　2．一人暮らしでない

問4.　兄弟構成について、当てはまるもの全てに○をつけてください。
　1．兄がいる
　2．姉がいる
　3．弟がいる
　4．妹がいる
　5．どれもいない

問5.　通っていた高校は男女共学でしたか。
　1．男女共学である
　2．男女共学でない
　3．高校には行っていない

問6.　現在、恋人がいますか。
　1．いる
　2．いない

問7.　今までに恋人は何人いましたか（現在を含む)。
　（　　　）人

問7で0人と答えた方は問10へお進みください。

問8.　今までで付き合った恋人の中で最も長く続いたのは何か月でしたか。
　（　　　）か月

問9.　今までで一番長く付き合った恋人の兄弟構成について、当てはまるものす全てに○をつけてください。
　1．兄がいる
　2．姉がいる
　3．弟がいる
　4．妹がいる
　5．どれもいない
　6．知らない

問10.　理想の恋人を思い浮かべたときに過去に付き合った人数を気にしますか。
　1．とても気にする
　2．気にする
　3．少し気にする
　4．気にしない

問11.　以下の項目について、どれくらい自信がありますか。数字に○をつけてください。

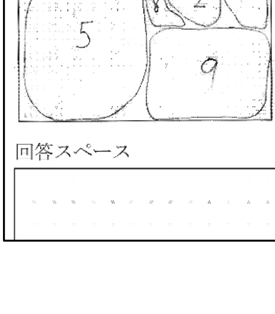
	まったく自信がない	自信がない	自信がある	とても自信がある
A)　顔の良さ	1	2	3	4
B)　スタイル	1	2	3	4
C)　ファッション	1	2	3	4
D)　頭の良さ	1	2	3	4
E)　性格の良さ	1	2	3	4
F)　ユーモア	1	2	3	4
G)　将来性	1	2	3	4

問12.　以下の項目について、どれくらいうれしく感じますか。数字に○をつけてください。

	うれしくない	少しうれしい	うれしい	とてもうれしい
A)　異性にナンパされた場合	1	2	3	4
B)　同性に告白された場合	1	2	3	4
C)　付き合っている恋人に友人が多い場合	1	2	3	4
D)　恋人がデートの計画を進んで立ててくれる場合	1	2	3	4
E)　恋人が処女または童貞だった場合	1	2	3	4
F)　人前で恋人に甘えられた場合	1	2	3	4

問13.　理想の恋人を思い浮かべたときに、以下の項目をどれくらい重視しますか。例のように、重視する度合いを面積の大きさで表現してください。すべての項目を記入する必要はありません。

①．顔の良さ　　②．スタイル　　③．ファッション
④．頭の良さ　　⑤．性格の良さ　　⑥．ユーモア
⑦．将来性　　⑧．趣味が合うか　　⑨．価値観が合うか
⑩．その他（具体的に：　　　　　　　　　　　）



(例)

この例の場合、面積の大きな1，5，9は重視する度合いが高く、2，4，6，8は、そこまで重視しないということになる。また、描かれていない3，7，10に関しては全く重視しないということになる。

回答スペース

第11回「統計分析ソフトでの実際（追補）」

■補足：検定結果の正確な意味

検定の結果から下される判断は、誤っている可能性が非常に低い（普通は5%未満）からこそ、信頼される。ただ、「誤った判断」には2種類あることに注意が必要である（表1）。

表1　検定の2種類の誤り

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">真実</td></tr>
<tr><td>対立仮説が正しい</td><td>帰無仮説が正しい</td></tr>
<tr><td rowspan="2">検定の判断</td><td>対立仮説を採択</td><td>正解</td><td>第1種の誤り
（危険度は有意水準未満）</td></tr>
<tr><td>帰無仮説を採択</td><td>第2種の誤り
（危険度は？？）</td><td>正解</td></tr>
</table>

本当は帰無仮説が正しいのに誤って対立仮説が正しいと判断してしまうことを、**第1種の誤り［第1種の過誤］**（Type I error）と呼ぶ。手続きに沿って検定を行えば、我々が第1種の誤りを犯す確率は、有意水準未満（つまり、通常は5%未満）になる。第1種の誤りを犯す確率は、十分に低くなるようにコントロールされているわけである。このように、第1種の誤りを犯す危険性をなるべく小さくしようとするのは、我々が「母集団について○○とわかった」というときには、十分に慎重でなければならないからである。

一方で、本当は対立仮説が正しいのに誤って帰無仮説が正しいと判断してしまうことを、**第2種の誤り［第2種の過誤］**（type II error）と呼ぶ。検定の手続きは、第1種の誤りを犯すことはほとんど防いでくれるが、第2種の誤りは基本的に容認する。これは誤って「○○とわかった」ということに比べれば、「○○とはわからなかった」という誤りの方が罪が小さいからである。「帰無仮説が正しい」という判断は、正確には「どちらの仮説が正しいのかわからないので、対立仮説が正しいと主張するのは差し控えておき、とりあえず偶然の結果（帰無仮説）と考えておくことにしよう」といった程度のものなのである。

■SPSSに触れよう

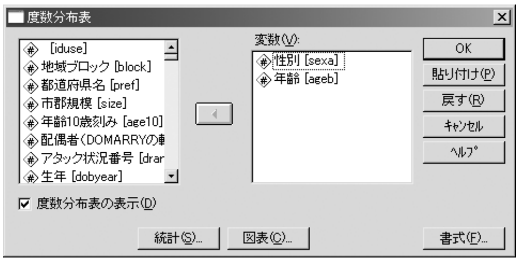
ごく一部のSPSSの操作だけを記す。SPSSのしっかりした入門書としては、小田（2007）や秋川（2007）を推薦する。SPSSの本格的な利用は、社会学研究法aや社会調査演習・実習などで学習できる。とにかく恐れずに触れてみよう。

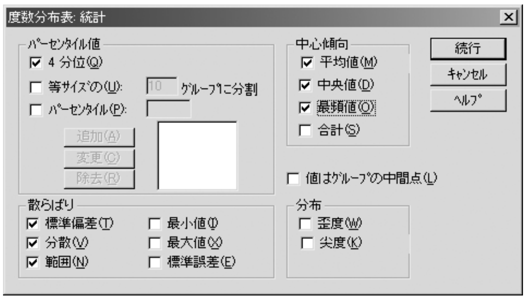
小田利勝 2007 『ウルトラ・ビギナーのためのSPSSによる統計解析入門』プレアデス出版.
秋川卓也 2007 『文系のためのSPSS超入門 新装版』 プレアデス出版.

なお、SPSSのメニューやボタンで使われている用語には、あまり一般的でない用語が多いので、意味がよくわからなくてもあまり気にしない方がよい（勉強不足のせいではない）。

■度数分布表・基本統計量（平均など）

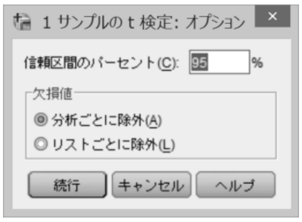
1. 分析→記述統計→度数分布表
2. 集計する項目を「変数」へ
3. 「統計」ボタンの中の必要な要約統計量にチェック





■平均の推定・検定

1. 分析→平均の比較→1サンプルのt検定
2. 平均を調べる項目を「検定変数」へ
3. 推定の場合は「検定値」を「0」にして、
   「オプション」の「信頼区間の％」に信頼度を入力
4. 検定の場合は「検定値」に帰無仮説で仮定する母平均を入力して、
   「オプション」の「信頼区間の％」に「100-有意水準」を入力（有意水準5％なら「95」）



1サンプルの検定

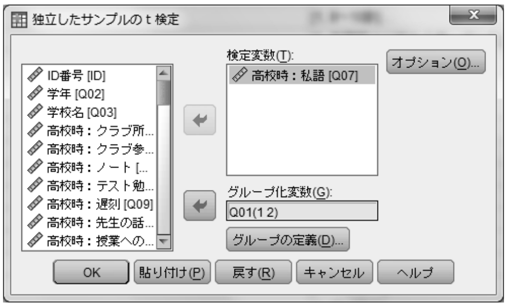
	検定値 = 0					
	t	df	有意確率 (両側)	平均値の差	差の 95% 信頼区間 下限	差の 95% 信頼区間 上限
Q7 今までに恋人は何人いましたか	11.975	183	.000	1.511	1.26	1.76

<u>平均の検定</u>
（この場合、df＝183、t＝11.975、p＝.000［0.1％未満ということ］で、母集団の平均値は0ではないと言える）

<u>平均の推定</u>
（この場合、信頼度95％で母集団の平均値は1.26～1.76の間にある）

■グループ別の平均値（グループが2つだけの場合）

1. 分析→平均の比較→独立したサンプルのT検定
2. 平均を調べる項目を「検定変数」へ
3. グループ分けのための変数を「グループ化変数」へ
4.「グループの定義」ボタンで2つのグループの値を指定



t検定（平均の差の検定）

・報告する値は、2グループの「平均値」と、検定の「自由度」「t値」「有意確率」

グループ統計量

	性別	N	平均値	標準偏差	平均値の標準誤差
満足度	男性	40	2.88	1.09	.17
	女性	40	3.45	1.15	.18

独立サンプルの検定

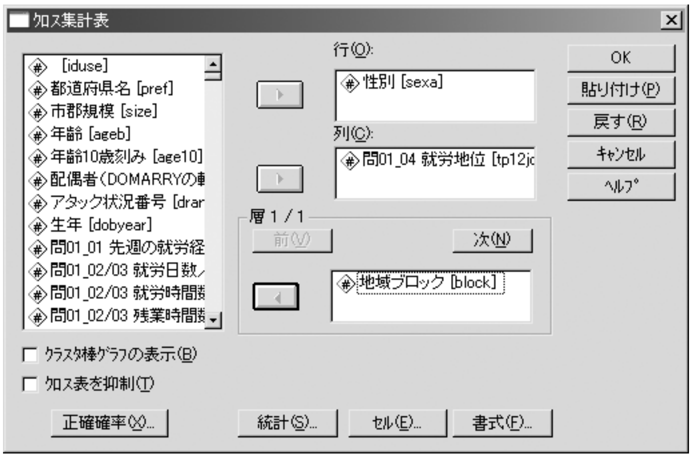
		等分散性のための Levene の検定		2 つの母平均の差の検定				
		F 値	有意確率	t 値	自由度	有意確率（両側）	平均値の差	差の標準誤
満足度	等分散を仮定する。	.746	.390	−2.291	78	.025	−.58	
	等分散を仮定しない。			−2.291	77.754	.025	−.58	

通常は上段を読む。左の等分散性の検定が有意（有意確率が.05未満）の場合だけ下段を読む

※検定が必要なければ、「分析→平均の比較→グループの平均」で何グループでも比較可能

■クロス表

1. 分析→記述統計→クロス集計表
2. グループ分けの変数を「行」へ、関心の中心の変数を「列」へ
3.「セル」ボタンの中の「パーセンテージ（行)」にチェック
4.「統計」ボタンの中の「カイ2乗」にチェック



χ^2検定（独立性の検定）

・検定には、各セルの期待度数が5程度は必要。度数の小さすぎるセルがある場合には、リコーディングした変数で縮約したクロス表を作成して検定する

・報告する値は、「自由度」「χ^2値（値)」「有意確率（漸近有意確率)」

カイ2乗検定

	値	自由度	漸近有意確率(両側)	正確有意確率(両側)	正確有意確率(片側)
Pearson のカイ2乗	5.208[b]	1	.022		
連続修正[a]	4.219	1	.040		
尤度比	5.277	1	.022		
Fisher の直接法				.039	.020
線型と線型による連関	5.143	1	.023		
有効なケースの数	80				

a. 2x2 表に対してのみ計算

b. 0 セル (.0%) は期待度数が 5 未満です。最小期待度数は 16.00 です。

今日のポイント
①とにかく統計分析ソフトに触れてみよう。
②手計算の結果と照らし合わせて、自信を持って出力を読み取れるようになろう。

（連絡）

来週小テスト③（独立性の検定、統計分析ソフトの結果の読み取り、各種の補足の理解［片側検定と両側検定、自由度の意味、第１種・第２種の誤り］）

※自由課題

IT センターのパソコン等で SPSS を起動して、「大学生の恋愛観に関する調査」のデータに触れてみる。以下の課題を提出した人は、最大 10 点最終評価に加点する。

・データは http://www2.itc.kansai-u.ac.jp/~tyasuda/に置いている。
・提出の際は必ずホッチキス留めして、最初に学籍番号と氏名を書くこと。
・15 回目の授業（1/21）まで提出を受け付ける。

練習1　Q1とQ2について、度数分布表を出力しなさい。（出力を貼り付け）

練習2　Q7について、平均値と標準偏差を出力しなさい。（出力を貼り付け）

練習3　Q8の平均値を95％の信頼度で区間推定しなさい。（出力を貼り付け）

練習4　Q13_2の「相手のスタイルを重視する度合い」の平均点を男女で比較しなさい。また、男女差があるという結果を母集団に一般化してもよいか検定をしなさい。
（出力を貼り付けた上で、以下の文章を完成）
※この結果から、男性の方がスタイルを重視する度合いの平均点が【　　　　】点大きい。平均の差の検定を行うと、有意確率が5%未満【である・ではない［一方を選択]】ので、この結果は一般化【できる・できない［一方を選択]】

練習5　「Q3 一人暮らしですか」と「Q6 恋人がいますか」でクロス表を出力し、一人暮らしの方が恋人がいる割合が高いのかどうか、確かめなさい。また、独立性の検定の結果も出力して、読み取りなさい。（出力を貼り付けた上で、以下の文章を完成）
※この結果から、一人暮らしの方が【　　　　】％だけ恋人がいる割合が高い。独立性の検定の結果は、有意確率が5%未満【である・ではない［一方を選択]】ので、この結果は一般化【できる・できない［一方を選択]】

練習6　何か自分で1つ分析を考えて、結果を出力し、簡単に結果を読み取りなさい。

計量社会学Ⅱ（2015年度秋学期 担当:保田（やすだ））
2015.12.10

第12回「多変量解析（1） 回帰分析」

■〇〇分析とは

今回と次回は、いわゆる「〇〇分析」というより高度な統計分析技法について論じる。計量社会学では「〇〇分析」という名称が付いた技法が無数にある。1つ1つの技法をこの講義で深く扱うことはできないので、別の機会に必要に応じて学習を進めなければならない。しかしながら、どのような分析技法であっても、その基盤となる知識を皆さんはすでに身につけているということを、ここでは理解してもらいたい。

一般に、〇〇分析は、3つ以上の変数を同時に扱うもので、まとめて**多変量解析**（multivariate analyses）と呼ばれる。これまでに我々が学習してきたことは、1つの変数の分布を記述すること、2つの変数の関係性を記述すること（計量社会学1)、および、それぞれの推定・検定（計量社会学2）である。すでにややこしいのに、3つ以上の変数を同時に扱って……というと難しそうで手を出してはいけない世界のように感じるかもしれない。しかし、そのようなことはない。多変量解析は、これまでに学習した基本的な統計技法のいくつかを組み合わせてパッケージ化したものにすぎない。

つまり、〇〇分析のそれぞれのパーツは、すでに学習していることでだいたい理解できるものばかりなのである。大切なことは、それらのパーツを組み合わせて結局何をしようとしているのかという「その技法全体の目的」と、その目的を果たすために「各パーツが果たしている役割」を理解することである。その点にだけ注意すれば、〇〇分析を恐れる必要はない。

■回帰分析とは

具体的に、**回帰分析**（regression analysis）を例にして考えよう。回帰分析は多変量解析（〇〇分析）の中でもっとも有名で、よく使用され、他の〇〇分析の基礎にもなっているものなので、多変量解析に入門するためには、これをまず理解しなければならない。

回帰分析は、ある1つの変数の値を他の複数の変数で説明しようとする。表1は典型的な回帰分析の結果である。学術論文などで、実際にこのような形で出版される。分析対象は子育て中の妻であり、夫婦関係の満足度が高い妻や低い妻がいる理由を分析している。一つの常識的な見方として、「夫の家事参加が多い方が妻の満足度は高い」という予想が成り立つ。しかし、この分析者は、役割期待という社会学的概念に照らし合わせると、この見方だけでは不十分と考えた。つまり、夫も家事をするという役割を妻が期待している場合に限って、夫の家事参加が満足感を与えるのであって、夫の家事役割を妻が期待していなければ、いくら夫が家事をしても妻は満足しない、と予想される。下の回帰分析では、この仮説を調査データから検証している。

表1　回帰分析の結果の例（妻から見た夫婦関係の満足度が従属変数）

	偏回帰係数							
	自営・自由業 (n=104)		常勤 (n=107)		パート・アルバイト (n=275)		専業主婦 (n=400)	
	モデル 1	モデル 2	モデル 1	モデル 2	モデル 1	モデル 2	モデル 1	モデル 2
定数	2.305	3.318	2.156	2.633	1.143	2.124	3.036	3.734
結婚年数	−.031	−.003	−.041	−.043	.004	−.017	−.026	−.018
結婚年数2乗	.001	.000	.001	.001	.000	.001	.000	.000
親・義理の親同居ダミー	.185	.025	−.101	−.166	.041	.032	−.089	−.135
妻の最終学歴ダミー	−.254	−.200	.037	.085	.112	.129	.088	.088
世帯年収	.000	.000	.000	.000	.000	.000	.000	.000
末子0〜6歳ダミー	−.120	−.431	−.534	−.380	.085	−.055	−.884**	−.985**
末子7〜12歳ダミー	.279	.241	−.432	−.265	−.041	.011	−.752*	−.822**
末子13〜18歳ダミー	.258	−.002	−.116	−.068	.006	−.029	−.656**	−.739**
末子19歳以上ダミー (RG)								
妻の性別役割意識	.076	.060	−.026	−.056	.047	−.014	.198***	.122*
夫の情緒的サポート	.209**	.194**	.314***	.319***	.271***	.250***	.155***	.142***
夫の家事参加度	.082*	.030	.092**	.049	.090***	.047*	.062***	.038*
期待充足度		.103***		.083**		.080***		.079***
R^2	.203	.332	.454	.512	.286	.341	.201	.257
Adjusted R^2	.107	.243	.390	.449	.256	.311	.179	.233

※$*p<.05$, $**p<.01$, $***p<.001$

出典:李基平 2008「夫の家事参加と妻の夫婦関係満足度--妻の夫への家事参加期待とその充足度に注目して」『家族社会学研究』 vol.20, no.1, pp.70-80.
注：データは「夫婦の生活意識に関する調査」。生命保険文化センターが2004に20〜49歳の既婚男女3000人を対象に実施（2355人回収）。分析は、子どもがいる有配偶女性に絞り、データに欠損がない886人を使用。
　従属変数の「夫婦関係の満足度」は1〜5点の5点満点の尺度。独立変数の「夫の家事参加度」は、5つの家事項目について、5点満点で測定したものを合計した5〜25点の25点満点の尺度。同じように、夫の期待する夫の家事参加の水準を25点満点で尋ね、その差を取ったのが独立変数の「期待充足度」である。

回帰分析の目的は、ある1つの変数の値がなぜ人によって異なるのかを、別の複数の変数を原因として説明しようとする。説明を与えたい1つの変数を**従属変数**（dependent variable）と呼び、その原因と考える変数のことを**独立変数**（independent variable）と呼ぶ。結果変数と原因変数とか、目的変数と説明変数と呼ばれることもある。ここでは、妻の満足度が従属変数であり、夫の家事参加の絶対量や、その家事参加が妻の期待をどの程度満たしているか（期待充足度）が独立変数である。他にも原因と考えられる事柄が独立変数になる。

つまり、分析者は、「妻の満足度という従属変数にとって、夫の家事参加の絶対量よりも、その期待充足度の方が重要な独立変数だ」ということを回帰分析で示そうとしている。

■回帰分析でパッケージ化されているもの

表1ではたくさんの数値が並んでいるが、この回帰分析では、次の3つの手続をパッケージ化している。

①それぞれの独立変数が従属変数に対してどのくらい強く関係しているか
　→回帰係数
②それぞれの関係性は統計的に有意か
　→t値による回帰係数の検定
③全体として複数の独立変数で従属変数の値がどのくらい説明できているか
　→決定係数

具体的に読んでみよう。

①回帰係数は、各独立変数の横に並んでいる数値である。表1の「自営・自由業」の「モデル1」では結婚年数の横に「-.031」という数値があるのは、「結婚年数が1年長くなれば、妻の満足度は0.031点だけ低くなる」という関係性を示している（-0.31と読み間違えないように注意)。ちなみに、満足度は1～5の5点満点なので、ほとんど関係ないことがわかる。

②そして、それぞれの関係性（回帰係数）が偶然でないかどうかを検定する。検定の結果は、検定統計量t値をもとに有意確率（p値）をコンピュータに算出させているが、これらは表1では有意確率の具体的な値を省略し、それがどのくらい小さかったか（どのくらい統計的に有意だったか）ということだけを＊の記号で表現している。一般的に、5％を切っていれば＊1つ、さらに1％を切っていれば＊2つ、0.1％も切っていれば＊3つで表す。検定の数が多いときには、このような簡便法が常套手段である。自営・自由業についてのモデル1の場合、夫の情緒的サポートや夫の家事参加度だけが、それぞれ1％水準、5％水準で統計的に有意な原因として認められている。

③決定係数はR^2という記号で表わされている。これは、複数の独立変数の情報から得られる従属変数の予測値と、実際の従属変数のデータの間の相関係数の2乗で算出できる（だから、相関係数の記号Rの2乗で表現されている)。決定係数の意味は、この回帰分析によって従属変数の値が人々によってばらばらであることの何％くらいが説明できたのか、ということである。たとえば、自営・自由業についてのモデル1ではR^2が「.203」とあるが、これはこれらの独立変数によって自営・自由業の妻の満足度の違いが20.3％説明できることを意味する。ただし、ふつうは人数の少なさや独立変数の数を考慮に入れて、この説明力をやや割り引いて解釈する。すぐ下の調整済み決定係数（adjusted R^2）がそれでありふつうはこちらを用いる。この場合は10.7％しか説明できていないと考える。

■パッケージの中身は知っているものか？

さて、さっぱり知らない用語ばかりが出てきて、「話が違う！」と思うかもしれない。しかし、ここで行なわれていることは、すべてすでに学習してきた基本的な手続きの範囲内の姿勢である。すなわち、①は、2つの変数の関係を1つの数値で表わそうとする姿勢であり、相関係数やユールのQなどでこのようなことを学習してきた（計量社会学1の内容)。ここではそれが**回帰係数**（regression coefficient）という初めて見る概念になっただけである。そして、むしろその意味は相関係数よりもわかりやすい（独立変数が1ポイント増えれば、従属変数が何ポイント増えるのかという具体的な数値)。

同じように、②で行なわれている検定は、独立性の検定や平均の差の検定でおこなってきたことと同じで、2つの変数の関係性を検定している。その検定統計量であるt値の計算式は知らずとも、結果として出てきた検定結果の意味は理解できるはずである。

そして、③も2つの変数の関係を1つの数値で表わす姿勢で、①と同じである。ただ、その2つの変数の一方が「予測値」という変わったものなだけに過ぎない。また、その**決定係数**（coefficient of determination）の読み取り方も「何％説明できているか」という非常に具体的な意味である。

やや乱暴かもしれないが、このようにこれまでに学習してきた1つ1つの手続きとその姿

勢が理解できていれば、未知の○○分析についても恐れる必要はない。もちろん、もっと深く「回帰係数はどうやって算出されるんだろう？」「なぜその算出でよいのだろう？」といったことを知ろうと思えば、一定の集中的な学習が必要である。ただ、そのような深い学習をしていなければ回帰分析の結果を読めない、あるいは自分で使えない、ということはまったくない。実際の数値の計算は統計分析ソフトがおこなってくれる。これまで学習してきた基本的な姿勢さえしっかり身についていれば、十分に読みこなせるはずである。また、少し勉強すれば十分に使いこなせるので、この先も、ぜひ自信をもって取り組んでほしい。（※社会学研究法aでは、実際に統計分析ソフトを使って○○分析に取り組む）

（問題）

1. 下の表は、月給をもらっている30代の女性について、月給の額（円）を従属変数とした回帰分析をおこなった結果である。この結果を読み取ってみよう。

	回帰係数
定数	-18553.81
年齢	1620.41
勤続年数	6772.44 ***
中3時の成績［1-5］	33703.64 **
n＝56	
調整済みR^2＝.283	

＊ p<.05, ＊＊ p<.01, ＊＊＊ p<.001

30代女性の月給についてその個人差を説明するために回帰分析を行った。月給の高さを左右する要因として独立変数を全部で（　　　　）つ設定したが、このうち、（　　　　）の影響力は統計的に有意な効果が確認されなかった。他のものは有意水準5%で統計的に有意であり、とくに（　　　　）は有意水準を0.1%にしても有意な効果が認められる。回帰係数の値を見ると、たとえば勤続年数が1年長くなると、月給が約（　　　　）円高くなることがわかる。総合すると、これらの独立変数で30代女性の中での月給の個人差が（　　　　）%説明できることになる。

2. 表1で示した回帰分析からは、結局、分析者が思ったとおりの結果が得られているのか。夫の家事参加の度合いよりも、期待充足度の方が重要という予想が支持されているかどうか、数値を読み取りなさい。

今日のポイント

①多変量解析はすでに知っている技法を組み合わせたパッケージであることを理解する。

②回帰分析について、どのようなパーツが組み合わさっているのかを理解し、数値の意味を読み取れるようになる。

【回帰係数、回帰係数の検定、決定係数】

第13回「多変量解析（2） さまざまな多変量解析」

■多変量解析のバリエーション

前回、もっとも一般的な多変量解析（3つ以上の変数の関係を一度に分析する技法）として、回帰分析を紹介した。多変量解析には無数のバリエーションがある。今回は、いくつかの技法を簡単に紹介しておこう。どういうときにどの技法を使うべきなのか、技術的にはいくつかのポイントがある。次の3点はよく注目されるポイントであろう。

1）従属変数がはっきりしている技法か、そうでないか

多くの技法は1つの従属変数について､いくつかの独立変数で説明を加えようとするものである。前回扱った回帰分析（もう少し正確に言えば、線形回帰分析、あるいは一般線形モデル）は、その中でもっとも単純なものである。一方で、1つの従属変数をはっきり決めるのではなく、複数の変数を同等に扱って全体的な関係構造を明らかにしようとするような分析技法もある。後者の場合は読み取り方が特殊なので事前の学習が欠かせない。一方で、前者の場合は多少の違いはあるものの基本的には回帰分析と同じような読み取り方ができると思ってよい。

2）量的変数を扱っているのか、質的変数を扱っているのか

多くの多変量解析は、扱う変数が量的変数（平均などが計算できる変数）であることを前提としている（回帰分析もそうである)。一方、質的変数は平均などの計算ができないので､分析技法も全く違ってくる｡社会調査のデータには質的変数が多く含まれているので、計量社会学では質的変数を扱える特殊な技法を用いることが比較的多い。社会調査データの分析を念頭に置いたテキストでないと解説していない技法がけっこう多いので注意しよう。

3）実際に調べた顕在変数だけを扱っているのか、調べてもいない潜在変数を扱うのか

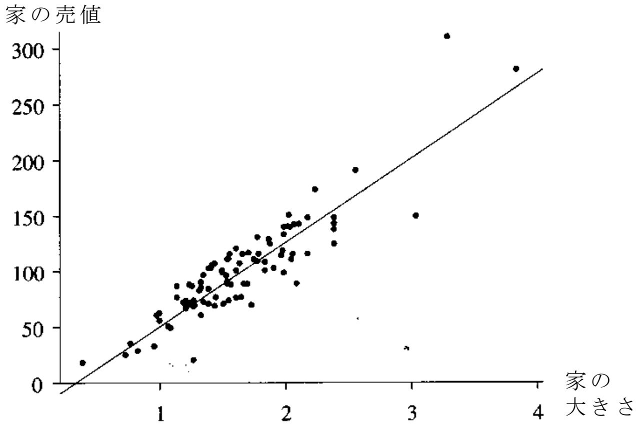
当たり前のことであるが、多くの分析技法では実際に調べた変数群の中で変数同士の関係性を明らかにしようとする。ところが、一部の技法は調べてもいない変数（潜在変数）の存在を仮定して、潜在変数と顕在変数の関係、あるいは潜在変数同士の関係を調べたりしようとする。ややトリッキーな技法であるが、社会学の中では直接調べることが難しい概念が多いので、このような技法は重宝される。潜在変数を扱う技法は1つ慣れておけば、他のものもだいたい読み取り方がわかるようになるので、一度は腰を据えて勉強してみてほしい。

いろいろなことを書いたが、現時点ではたぶんピンとは来ないことと思う。多変量解析を理解する一番の近道は、実際にその技法が使われている研究に触れることである。恐れずに「○○分析」が使われている論文を読んでみてほしい。

■回帰分析（一般線型モデルや分散分析とも関連）

1つの量的変数の値がケースによって違うのがなぜなのかを、別の（複数の）変数で説明しようとする分析技法。たとえば、人によってIT技術に関心が高かったり低かったりするのはなぜなのかを、その人の年齢や仕事、趣味などによって説明しようとする。具体的に独立変数の値が1段階違うことで、従属変数の値がいくら上下するのかを、統計的に推定する。

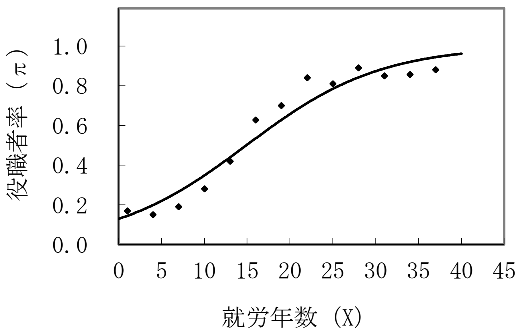
関心の中心になる変数が決まっていて、とにかくその値を左右する原因が知りたい、という考え方は、ごく自然なものなので、回帰分析はもっとも頻繁に活用される技法である。



■ロジスティック回帰分析（ロジット分析、ロジットモデル）

ロジスティック回帰分析は、基本的に回帰分析の仲間である。従属変数（説明したい変数）が連続的な量的変数ではなく、カテゴリーを表わす質的変数の場合に用いられる。たとえば、タバコを吸う原因は何なのか、田舎に住む原因は何なのか、といったようなことを問題にする。

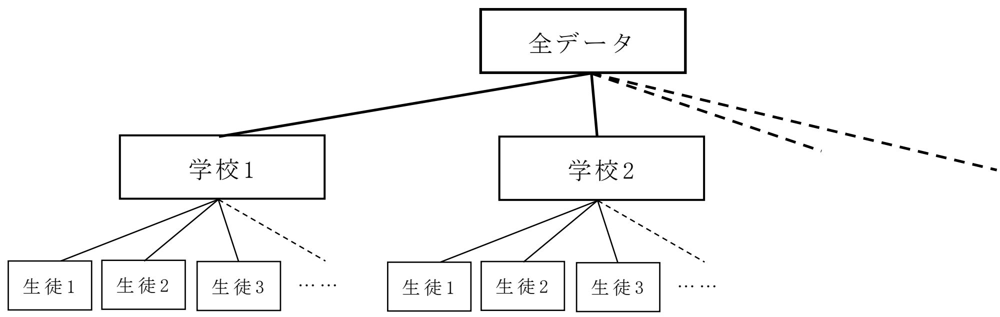
社会調査のデータには、連続的な量的変数では測りにくいものが多いので、ロジスティック回帰分析は、実際的に重宝する。



■マルチレベル分析（階層線型モデル、HLM）

マルチレベル分析は、データの構造が複層的になっている際に、それぞれの層で適切に回帰分析をおこなうための分析技法である。たとえば、ある県の中からまずいくつかの中学校を抽出し、さらにその学校の中で何人かの生徒を抽出したようなデータを考える。このデータで、生徒の生活態度がよい/悪いことの原因を回帰分析で調べる場合、生徒個人の要因もあるが、学校単位で生活態度の改善に取り組んでいるといった、学校の要因もある。このような複層的なデータをマルチレベル分析は、適切に扱うことができる。

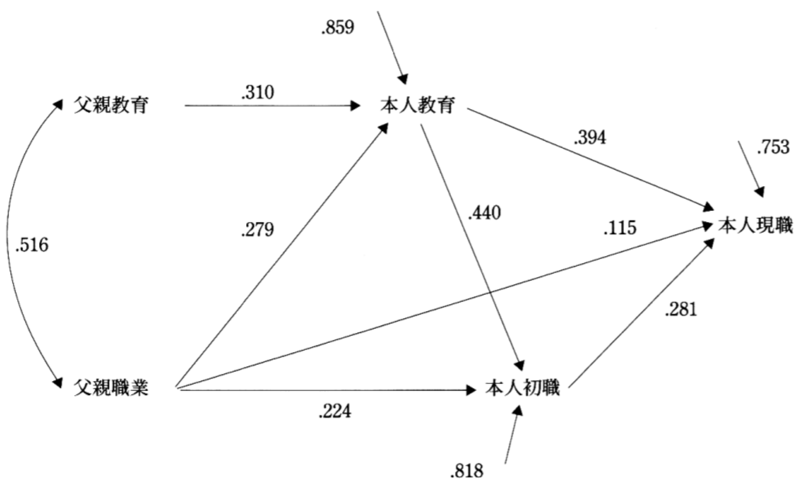
マルチレベル分析は、統計的に適切な分析技法で望ましい、という面もあるが、それ以上に、ふつうの回帰分析では考えられない複数の層の原因についていろいろなヒントが得られる魅力がある。



■構造方程式モデル（SEM、共分散構造分析、パス解析）

構造方程式モデルは、関心のある複数の変数間の関係性を1つの図式に整理する分析技法である。関係の有無と方向性を矢印の記号で表わし、その関係の強さを0～1の間の数値で表わす。

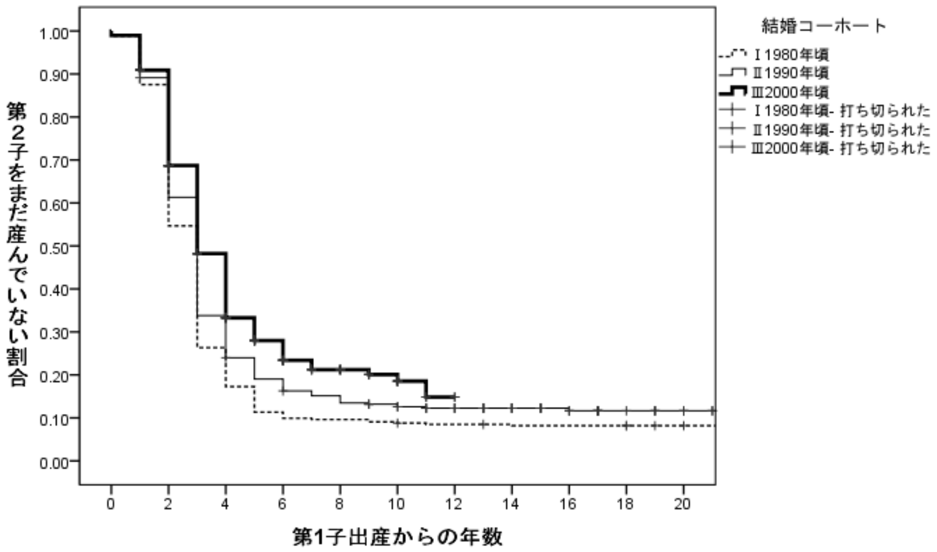
直感的に意味がわかりやすい構造方程式モデルは、とくに同じ視点からいろいろな集団を比較するとき（たとえば、年齢層による違いを比較したり、国・文化による違いを比較したりするとき）に、有効な技法である。



■イベントヒストリー分析（生存分析）

イベントヒストリー分析は、時間が経つにつれて「ある状態」に移行する人が徐々に増えていくようなことが想定される現象について、どういう人が早く移行して、どういう人が遅く移行するのかを調べる技法である。たとえば、「就職先が決まる」という状態に移行する時期の早い／遅いについて考えたりすることができる。

就職や結婚、出産など、社会学でこの種のデータを問題にすることは意外と多い。ロジスティック回帰分析でも似たような分析はできるが、イベントヒストリー分析を用いることの1つの大きな利点は、打ち切りデータ（調査時点ではまだ就職していないので、いつまで就職しない状態なのかが不明なデータ）を含めた分析ができることである。



■ログリニア・モデル（対数線型モデル）

ログリニア・モデルは、クロス表で表現される変数間の関係性を整理するための分析技法である。2変数のクロス表であれば、よく見ればその内容を読み間違えることはないが、3変数以上のクロス表では、どの変数とどの変数がどう関係しているのか、読み取ることは難しい。ログリニア・モデルでは、たとえば、「性別」と「仕事の有無」と「結婚」の3変数のクロス表から、何と何が関係しているのかを整理してくれる。

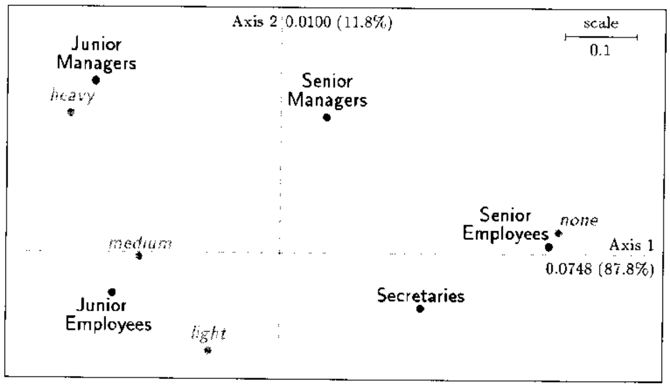
繰り返すが、社会調査のデータでは質的変数が多いので、クロス表で結果を整理する場面が多く見られる。ログリニア・モデルは、クロス表を作る延長線上で多変量解析を進めようとするので、ある意味で自然な分析技法である。

母と娘の類似性
（2×3×3のクロス表）

■対応分析（コレスポンデンス分析、相対尺度法、数量化Ⅲ類）

対応分析は､質的変数同士の対応関係を1つの図式に整理することができる分析技法である。扱っている変数が連続的な量的変数であれば、散布図を描くなどの方法で、その関係性を直感的に表現できる。これに対して、対応分析では「仕事の種類」と「趣味の種類」のように、数量にできないようなカテゴリー同士の関係（たとえば、管理職の人はゴルフが趣味の人が多い、など）を図式にする。

社会学では、やはりこのように1つの数直線上に表現できないデータ（質的変数）を扱うことが多いので、対応分析は意外と便利な場面が多い。

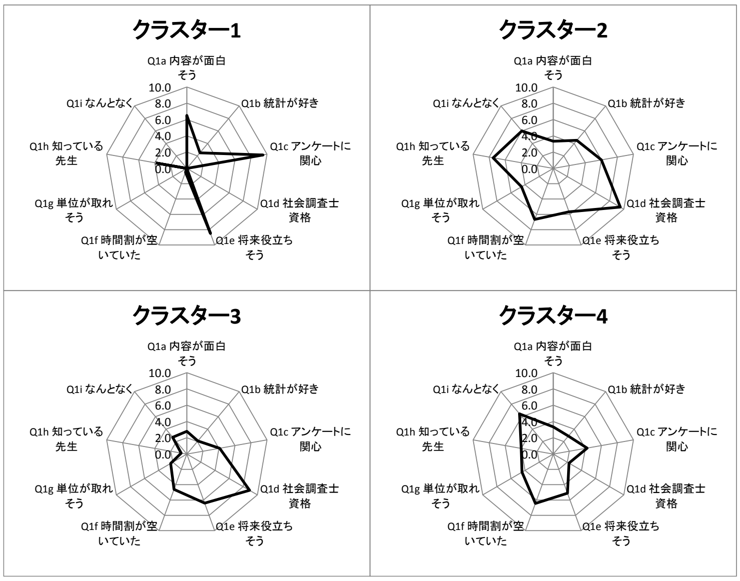


喫煙程度と役職の対応

■クラスター分析

クラスター分析は、複数の量的変数から、よく似た回答傾向の人々をグループ（クラスター）にして、全体としてどのような種類の人々に分類できるのかを探索する分析技法である。たとえば、各科目への好き嫌いのデータから、中学生を分類して、どのような好き嫌いパターンが存在するのかを整理することができる。

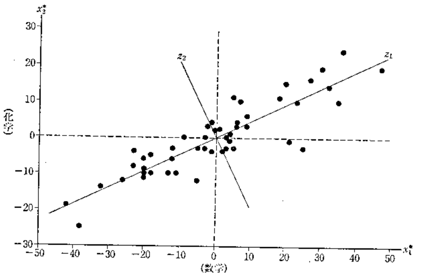
人々を分類したいという気持ちはごく自然なもので、特別な分析技法を使わなくても直感的に分類作業を進めることもできる。しかし、クラスター分析を用いれば、分類過程を客観化できるだけでなく､自分では思いも付かなかった分類方法が示される可能性がある。



■主成分分析

主成分分析は、複数の変数から1つの合成得点を算出するための分析技法である。たとえば、5教科の好き嫌いをそれぞれ5点満点で尋ねて、合計得点（25点満点）を算出すれば、総合的な「勉強好き」得点を算出できるかもしれない。しかし、単純な合計では、科目によって好きな生徒が多い/少ないといったことが考慮されていないし、そもそも本当に1つの得点に合成できるくらい一貫性のあるものなのかも検討されていない。

複数の変数から総合的な得点を出したい場面は非常に多い。主成分分析を使えば、単純に合計するよりも適切な方法で総合得点を合成できることが、魅力である。



■因子分析

因子分析は、多くの量的変数（意識項目であることが多い）が少数の見えない要素に規定されていると想定して、その見えない要素（潜在因子）を導出する分析技法である。たとえば、結婚の価値観について、50個くらいの質問をしたデータから、背後でその価値観を決めている要素は「愛情を重視するか」「経済状態」「子どもをほしいか」の3つであることをはじき出したりできる。

とくに人々の意識については、直接的に1つの質問で測りにくい事柄が多いので、たくさんの質問の結果から、根本的に大事な要素を導き出そうとする因子分析は重宝される。

ﾊﾟﾀｰﾝ行列a

	因子	
	1	2
q4mnmga f 男性の幸せは結婚だ	.993	-.265
q4wnmga b 女性の幸せは結婚だ	.901	-.117
q4mghapp k 結婚する方が幸せ	.670	-.029
q4noccmg h 子どもをもつ必要はない	-.463	.051
q4wwhphh j 妻は夫の仕事を助けるべき	.371	.366
q4wwhhx e 夫は仕事・妻は家庭	.334	.569
q4wwjbia a 妻は仕事をすべきでない	.190	.562
q4jbmmfm c 働く母親もよい親子関係を築ける	.176	-.437
q4jbmmcc g 幼児の母は働くべきでない	.207	.384
q4mncook d 男性も家事をすべき	-.005	-.371
q4wnjb2l i 女性の自立には仕事が必要	.216	-.351

因子抽出法: 最尤法
回転法: Kaiser の正規化を伴うﾌﾟﾛﾏｯｸｽ法
a. 3 回の反復で回転が収束しました。

〈多変量解析をするために参考になる文献〉

筒井淳也ほか編 2015 『計量社会学入門』 世界思想社.
与謝野有紀ほか編 2006 『社会の見方、測り方』 勁草書房.
村瀬洋一・高田洋・廣瀬毅士編 2007 『SPSSによる多変量解析』オーム社.
三輪哲・林雄亮編 2014 『SPSSによる応用多変量解析』 オーム社.
小塩真司 2011 『SPSSとAmosによる心理・調査データ解析（第2版）』 東京図書.
平井明代編 2012 『教育・心理系研究のための データ分析入門』 東京図書.

※次回の終わりに、小テスト④（回帰分析の理解、40ページの内容、14回目の話の内容）

第14回「まとめ：計量社会学がめざすもの」

■学習してきたこと

この講義では、計量社会学（社会調査データの分析）で用いる推測統計の技法について、1つずつ学習してきた。実際に取り組んだのは、以下の推定・検定の手続きであるが、他の統計量についても、基本となる考え方は同じである。講義では電卓での手計算を中心にしたが、後半では実際の活用場面での統計分析ソフト（SPSS）の利用についても触れた。

・平均の推定
・平均の検定　（比率の推定・検定にも応用できる）
・平均の差の検定
・独立性の検定　（比率の差の検定にも応用できる）

また、最後には、回帰分析を例にして、より高度な分析技法もすでに学習している考え方のパターンをパッケージ化したものにすぎないことを、体験した。さらに、講義の初めの方では、推測統計の基盤になる無作為抽出と中心極限定理について、理解を深めた。

これらの技術は、社会調査のデータを用いて計量社会学を実践する上で、直接的に役に立つ。あえてもう一度強調しておくが、大切なことは正確に計算をこなすことではない。推測統計の考え方を理解することである。つまり、何を目的としてそんな手続きをしているのかを忘れないことや、帰無仮説、有意水準といった概念（用語）を正しく理解していることが重要である。これまで、実際に数式を追ったり電卓を叩いて計算をしたりしてきたのは、すべてその考え方を理解し、身体に染みつかせるためである。

この講義で「太字」で取り上げた用語を羅列してみた。これらの用語がどのような場面で、何のために使われているか、すべて説明できるならば、推測統計の考え方は身についているだろう。自信を持って活用してほしい。

計量社会学
記述統計
推測統計
母集団
標本［サンプル］
母数
標本統計量
［あるいは単に統計量］
無作為抽出
［ランダム・サンプリング］
中心極限定理
正規分布
標準誤差
標準正規分布
標準得点
［標準化得点、z値、標準値］
標準化［z-変換］
統計的推定［あるいは単に推定］
統計的検定［あるいは単に検定］
区間推定
信頼度
信頼区間
スチューデントのt分布
［あるいは単にt分布］
自由度
帰無仮説
対立仮説
検定統計量
臨界値
採択域
棄却域
有意水準
平均の差の検定（t検定）
両側検定
片側検定
独立
独立性の検定
カイ二乗値［χ^2値］
期待度数
観察度数［観測度数］
有意確率（p、p値）
統計分析ソフトSPSS
第1種の誤り［第1種の過誤］
第2種の誤り［第2種の過誤］
多変量解析
回帰分析
従属変数
独立変数
回帰係数
決定係数

■最後のメッセージ①：無作為な標本との乖離

推測統計の学修を終え、社会学に活用しようとする皆さんに対して、最後に、3つのメッセージを発しておきたい。1つ目は、「実際の社会調査の標本は無作為ではない」ということである。このことは、この授業の最初の辺りで強調してきた。たとえば、皆さんが卒業研究のために、大教室の授業を借りて学生から調査データを集めたとする。これは当然、無作為ではない。また、本格的な社会調査では無作為抽出が普通だが、求めに応じて協力してくれる人は半分程度であることが多い。このデータも、純粋に無作為な標本にはならない（協力的な人に偏っている）。

にもかかわらず、社会調査の分析には、無作為が重要な前提のはずの推測統計を適用する。なぜか？　社会学にたずさわる者は、皆、この疑問に対して何らかの答えをもっていなければならない。1つの回答は、理想状態を想定した推定や検定の結果によって、データが持つ価値の上限を知ることができる、ということであろう。つまり、「手元の標本データが理想的な状態（無作為）で収集されたとするならば、母集団は○○％の確率で××だ」と推測しているわけであるから、実際の推測の精度は「必ず」それよりも劣る。データが持つ情報量は、推測統計の適用によって得られる結果が最大限だということになる。

そして、その最大限からどの程度劣っているのか、どういう方向に歪みがあるのかを考えるためには、理想状態（無作為な標本）から現実の標本がどのように乖離しているのかを検討することによって、慎重に推理するしかない。これは統計的な話ではなく、もっと広い話（場合によっては、経験則に基づく感覚的な話）である。

推理を適切におこなうには、標本と母集団のそれぞれについて、できるだけ豊かな知識を持つことが望まれる。大阪の大学生が母集団と考えて、関西大学の授業の受講生に調査をおこなったのであれば、まず、大阪の大学生（母集団）について得られる知識を可能な限り収集しなければならない。また、調査に協力した受講生（標本）の特徴も可能な限り知ろうとしなければならない。これには、統計的情報も質的情報も含まれ、観察やインタビューなどの質的研究との協力も不可欠になる。母集団の範囲を想定せずに推測統計の技法を用いることは、もちろん論外である。無作為から乖離しているからこそ、母集団と標本を意識することがことさら重要になる。

■最後のメッセージ②：反証できることにこそ価値がある

次に、仮説が正しいということを確かめる「検証」だけが計量社会学の目的ではない、ということを強調しておきたい。統計的検定の手続きを身に付けると、検定に合格することがデータ分析の目的であるかのように勘違いしてしまう人がいる。

検証は大切であるが、より大切なことは予想外の「反証」である。つまり、「当然データはこうなっているだろうと予想していたのに、そうではなかった」という事実の方が重要なのである。反証は、新しい発見の出発点となる。だから、調査結果が予想とまったく違った、あるいは検定に合格できなかった、というときに、がっかりする必要はない。むしろ、自分の頭の中にはなかった意外な事実から、新しい考え方が発現できる可能性を喜ばなければならない。

もちろん、反証を基に、よりよい考え（理論、仮説）をまとめあげることができるかは不透明で、いつもうまくいくわけではない。というか、うまくいかないことの方が多い。しかし、少なくとも、反証から目をそらし、「仮説どおり」という検証にばかり目を向けるのは、計量社会学の態度ではない。それではデータを分析している意味がない。一般に、人間は「こうじゃないかな」という仮説を確かめようとするとき、自分の予測と合致する

事例ばかりに目を向けてしまう傾向がある（確証バイアス）。データは、そんな我々の傾向とは関係なく、ただ冷徹に反証を示してくれるからこそ、価値があるのである。

■最後のメッセージ③：不確実性への意識

初回に、推測統計の考え方について説明する中で次のことを強調した。学修を終えた今であれば、ここで強調したことがある程度具体的に感じられるのではないだろうか。

> 推測統計は次のような思考方法を持っている。この考え方は科学の中ではとても珍しく、画期的なものであった（ラオ 1993:第2章）。
>
> 不確実な知識 ＋ 不確実性の度合いについての知識 ＝ 有益な知識
>
> 例）天気予報
>
> 明日は雨でしょう ＋ 降水確率は70％です ＝ 役に立つ
>
> 同じデータ（天気図）でも、明日の天気は正確にはわからない。
> 過去のデータを集め、おそらく雨だろうという予想をする（不確実な知識）。
> どのくらい不確実なのかを分析し、推し測る（不確実性の度合いについての知識）。
> 70％の確率で雨（有益な知識）。

推測統計の手続きは、その答えを確実・正確に出すことを放棄し、一定の不確実性を許容する。「5％の確率で間違えることはありえるのだけれども」という言い訳をした上で、結論を出す。

ただし、ここでポイントになるのは、不確実なことを許容するのだけれどもどの程度不確実なのかは精密に考える、ということである。つまり、「間違える確率は 5％」と数値で明確にする。統計は不確実なので間違いを犯すが、不確実性の度合いを精密に考えているおかげで、どこで間違えた可能性が高いのかを振り返り、反省、吟味することができる。そして、数値化されている不確実性は、皆で共有できる。それぞれの研究の反省、批判、改善について、他人が関わることが容易になる。

この発想こそが社会学に取り組む皆さんに必ず身に付けてほしい考え方である。人間の行動・心理・社会現象については、正確には分からないことが多すぎる。物理学のような厳密さを求めると、その研究は一歩も進まない。しかし、同時に、あいまいさを無限に許容するわけではない。それぞれの研究がどの程度のあいまいさをもって進められているのかという情報は、皆で共有しなければならない。

数人の友人の事例をもとに「大学生とはこういうものだ」と想像するのはよくあることである。あるいは、自分の住む地域での経験だけをもとに「日本の地域コミュニティでは〇〇な変化が進んでいる」と考えることもあるだろう。それぞれ、そういったことがあっても構わない（思考の出発点になる）。しかし、そのとき同時に、自分の考えの根拠がどのくらい弱いものなのか気になってしまう、その不確かさを他人とも共有したうえで議論をしたい、そんなことを自然に考える感性を身に付けてほしい。そして、より強い根拠を探索して不確実性を下げるすべを考えてほしい。これは、数字を使うかどうかとは関係のない話である。情報の不確実性への意識こそが、社会学部の学生であれば誰もが留意すべき、極めて重要な考え方であり、「常識を疑う」ための強力な基盤になる。

推定・検定の復習問題

1.「幼児期に母親が外で働くことは子どもの成長に悪い」という意見に賛成か反対かを＋2点（とても賛成）～－2点（とても反対）のいずれかで答えてもらう調査を行った。996人から回答を得て、平均値が-0.52、標準偏差が1.25であった。

（1）95%の信頼度で平均得点の信頼区間を区間推定せよ。

（2）母集団でも、「平均値が0とは異なる」と判断してよいであろうか。有意水準5%で検定せよ。

2. ある2つの地域で5年生を20人ずつ無作為に選んで、算数の学力試験を行った。その結果、地域Aでは、平均得点が73点、標準偏差が10.0であった。一方、地域Bでは、平均得点が68点、標準偏差が8.0であった。2つの地域で算数の学力に差があると言ってよいであろうか（平均得点に差があるといってよいであろうか)。有意水準5%で検定を行え。

3. 次の調査結果について以下の問いに答えよ。1991年に、18歳以上のアメリカ在住の人々全体を母集団として行われた社会調査では、死後の世界を信じるかどうかについて、男女別に次のような結果が得られた（実データ）。

	信じる	信じない	分からない	合計
男性	435	89	58	582
女性	275	84	50	409
合計	710	173	108	991

（1）男性、女性それぞれについて、死後の世界を信じている者の比率を算出せよ。

（2）母集団において「性別」と「死後の世界を信じるかどうか」の間に関連性があると言えるかどうか。有意水準5%で独立性の検定を行え。ただし、期待度数は整数に四捨五入してよい。

（3）検定の結果を踏まえて、上のクロス表から分かる内容を簡単に記せ。

4. ある中学校で無作為抽出した150人の生徒に対して、「大山のぶ代を知っているか？」と尋ねたところ、64.0%の生徒が知っていた。この調査結果をもとに以下の問いに答えなさい。

（1）この学校全体で大山のぶ代を知っている生徒は何%いると予想されるか。95%の信頼度で区間推定しなさい。

（2）この中学校で3年前に同様の調査をおこなったときには、150名の生徒のうち76%の生徒が大山のぶ代を知っていた。3年前と現在で大山のぶ代の知名度に違いがあるといってよいか、有意水準5%で検定しなさい（平均の差の検定の応用でも、独立性の検定の応用でもどちらでもよい)。

■「推定・検定の復習問題」の解答

1．（1）

自由度996－1＝995のt分布（標準正規分布ともみなせる）で臨界値は±1.96
だから、母平均μの信頼区間は、

$$-0.52 \pm 1.96 \times \frac{1.25}{\sqrt{996}}$$

つまり、-0.60～-0.44

（2）

H_0：μ＝0
H_1：μ≠0

平均の検定を行う。

$$t = \frac{-0.52 - 0}{1.25 / \sqrt{996}} = -13.13$$

自由度996－1＝995のt分布（標準正規分布ともみなせる）で臨界値は±1.96。
t＜－1.96なので、帰無仮説を棄却し、対立仮説を採択する。
平均値は0とは異なる、と言える。

2．

H_0：地域で平均に差がない（$\mu_1 = \mu_2$）
H_1：地域で平均に差がある（$\mu_1 \neq \mu_2$）

平均の差の両側検定を行う。

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(20-1)\times 10.0^2 + (20-1)\times 8.0^2}{20+20-2} \times \left(\frac{1}{20} + \frac{1}{20}\right)} \fallingdotseq 2.86$$ なので、

$$t = \frac{73-68}{2.86} \fallingdotseq 1.75$$

自由度20＋20－2＝38のt分布において臨界値は約±2.02である。
（t分布の表に自由度38の欄がないので、40の欄で代用する）
－2.02＜t＜2.02なので、帰無仮説を採択する。
2つの地域で学力に差がある、とは言えない。

3．（1）

男性は、435÷582×100＝74.74（%）
女性は、275÷409×100＝67.24（%）

（2）

期待度数を算出すると以下のとおり。

$\frac{582}{991} \times \frac{710}{991} \times 991 \fallingdotseq 417$などと計算。

	信じる	信じない	分からない
男性	417	102	63
女性	293	71	45

これをもとに検定をする。

H_0：関連性がない（独立である）
H_1：関連性がある（独立でない）

χ^2値で独立性の検定を行う。

$$\chi 2 = \frac{(435-417)^2}{417} + \frac{(89-102)^2}{102} + \frac{(58-63)^2}{63} + \frac{(275-293)^2}{293} + \frac{(84-71)^2}{71} + \frac{(50-45)^2}{45}$$

≒0.78+1．+606．+410．+121．+308．5 6

≒6．8 9

自由度(2-1)×(3－1)＝2のχ^2分布において臨界値は5.99である。
χ^2＞5.99なので、帰無仮説を棄却し、対立仮説を採択する。
性別と信仰の間に関連性がある、と言える。

（3）

解答例）標本の集計から、男性の方が女性よりも死後の世界を信じる割合が高いことがわかる。また、検定の結果から、この傾向は母集団（1991年に18歳以上のアメリカ在住の人々全体）についても当てはまると考えられる。

4．（1）

知っている比率がpということは、知っている場合を1点、知らなかった場合を0点としたときに、平均値がpで、標準偏差が$\sqrt{p(1-p)}$と同じことになる。
臨界値は±1.96なので、母集団での比率の信頼区間は、

$$0.64 \pm 1.96 \times \sqrt{\frac{0.64(1-0.64)}{150}} = 0.64 \pm 0.0768$$

つまり、56.32%～71.68%の生徒が大山のぶ代を知っていると推定できる

（2）

（平均の差の検定の応用の場合）

$$n_1 = 150,\ \overline{X}_1 = 0.73,\ s_1 = \sqrt{0.76(1-0.76)} = 0.427$$

$$n_2 = 150,\ \overline{X}_1 = 0.64,\ s_1 = \sqrt{0.64(1-0.64)} = 0.480$$

H_0：3年前と現在で比率に差がない（$\mu_1 = \mu_2$）
H_1：3年前と現在で比率に差がある（$\mu_1 \neq \mu_2$）

平均の差の両側検定を行う。

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{(150-1)\times 0.427^2+(150-1)\times .480^2}{150+150-2}\times\left(\frac{1}{150}+\frac{1}{150}\right)} \fallingdotseq 0.052$$ なので、

$$t = \frac{0.76-0.64}{0.052455} \fallingdotseq 2.29$$

自由度150＋150－2＝298のt分布（標準正規分布ともみなせる）において、
臨界値は±1.96である。
t＞1.96なので、帰無仮説を棄却して対立仮説を採択する。
3年前と現在で知名度に違いがある、と言える。

（独立性の検定の応用の場合）
下のようなクロス表が得られたということになる（観察度数）。

	知っている	知らない	計
3年前	114	36	150
現在	96	54	150
計	210	90	300

期待度数を算出すると以下のとおり。

$\frac{150}{300}\times\frac{210}{300}\times 300 = 105$などと計算。

	知っている	知らない	計
3年前	105	45	150
現在	105	45	150
計	210	90	300

これをもとに検定をする。

H_0：関連性がない（独立である）
H_1：関連性がある（独立でない）

χ^2値で独立性の検定を行う。

$$\chi 2 = \frac{(114-105)^2}{105}+\frac{(36-45)^2}{45}+\frac{(96-105)^2}{105}+\frac{(54-45)^2}{45} = 5.14$$

自由度(2-1)×(2－1)＝1のχ^2分布において臨界値は3.84である。
χ^2＞3.84なので、帰無仮説を棄却し、対立仮説を採択する。
3年前と現在で知名度に違いがある、と言える。