第 1 章

1 エ

解説 エは特定の個人の知能を測ることに関心があり，社会現象に関心があるわけではないので，明らかに社会調査ではない。イは質問紙法による典型的な社会調査である。対話による調査（ア）や観察による調査（ウ）も，質的調査として社会調査に含まれる。

2 解答例：実験も調査も科学的なデータ収集の方法であるが，実験は必要な状況を人工的につくったうえでデータを集めるのに対し，調査はいまあるものをいまあるところ（フィールド）で調べてデータを集める。

3 （1）名義尺度，質的変数　（2）比例尺度，量的変数　（3）順序尺度，質的変数
（4）間隔尺度，量的変数（厳密には，順序尺度，質的変数）　（5）間隔尺度，量的変数

解説 （1）数値の並び順に特別な意味はないので名義尺度。（2）2 は 1 の 2 倍，20 は 2 の 10 倍というように，数値の比が意味を持つので比例尺度。（3）数値の間隔が一定ではない（1 と 2 の間隔よりも 5 と 6 の間隔の方が広い）が，値が大きいほど期間が長いので順序尺度。（4）評定尺度であり，p.12 の例題と同様。（4）常識的に等間隔の数値なので間隔尺度。2 は 1 の 2 倍下層という意味はないので，比例尺度ではない。

4 （1）「Q55 飲酒頻度」は順序尺度による質的変数なので，平均を計算することはできない。
（2）「Q10 健康状態（本人）」は間隔尺度による量的変数なので，平均を計算することはできる。ただし，比例尺度ではないので，3.09 が 2.25 の 1.37 倍という計算はできない。

解説 （2）間隔尺度と比例尺度の変数はどちらも量的変数と考えられるので，違いが意識されることは少ないが，比が意味をもつかどうかが異なる。「Q10 健康状態（本人）」は間隔尺度なので，仮に 1〜5 の値の代わりに 0〜4 の値を用いてもよい。このとき，20 代の平均は 1.25，80 代の平均は 2.09 になるので，2.09÷1.25≒1.67 であり，値の定め方によって結果がまったく変わってしまう。つまり，このような計算はできない（意味がない）。

5 （1）集計データ　（2）素データ

解説 『労働力調査』は総務省統計局が行っている標本調査で，15 歳以上の人々の就業状態を毎月調べている。調査の単位は世帯なので，素データは 1 世帯を 1 ケースとしている。よって，都道府県ごとのデータは集計データである。『学校基本調査』は文部科学省が毎年行っている全数調査で，学校ごとの基本情報（在学者数，経費など）を調べている。調査票は都道府県，市町村を経て各学校に配布される。学校を単位とした調査であり，個人を単位とした調査ではないので，学校ごとのデータが素データである。これらの情報は，各調査の報告書や省庁のホームページを見ることで，簡単に知ることができる。

第 2 章

1 解答例 1：不安定な働き方をしている若者がどのくらいいるのか，高校の卒業年度別にその違いを明らかにする（非正規雇用の割合，短い期間で仕事を辞める人の割合などを調べ，比較する）。
解答例 2：どのような若者が不安定な働き方に陥っているのか，考えられる原因を検証する（在学時の成績，卒業後の進路，仕事への考え方の違いなどとの関連を調べる）。
解答例 3：若者が不安定な働き方に陥らないように用意されている学校や公的機関のサービスについて，その利用状況を調べ，雇用問題を解決する有効な方法を探る（各サービスの認知度や評価，利用されない理由などを調べる）。

解説 解答例 1 は問題の実態を，解答例 2 は問題の原因を，解答例 3 は問題の解決方法を探ることを目的としている。もちろん他にも多様な目的が考えられる。また，実際に調査をするには目的をよりはっきりとさせる必要があるだろう。

2 エ

3 イ

4 割り当て抽出は，条件に合う属性の人々の中から，調査をしやすい人が回答者に選ばれるのに対して，層化抽出は，条件に合う属性の人々の中から等確率に回答者を選び出す。つまり，割り当て抽出は無作為抽出ではないが，層化抽出は無作為抽出であるという点が大きく異なる。

5 ア

第 3 章

1 エ

解説 母集団は本来調べたい人々全体のことを指す。標本調査では，この調べたい人々全体（母集団）の

中から回答者が決まるので，この選択肢の中では解答はエである。

2 全数調査は，調べたい人々全体（母集団）を実際に調べる調査なので，全体の推測を行う必要がない。調査結果そのものが全体の様子を表していることになる。

3 （1）解答例 1：授業に欠席することが少ない学生ほど，現在の生活についての主観的な幸福感が高い。

解答例 2：ルールを守る意志が強い学生ほど，学内に友人が多い。

（2）解答例 1：

問 1　単位を取るつもりの授業を，あなたはどのくらい欠席していますか（公欠を除く）。

1	2	3	4	5	6
全く欠席しない	15 回に 1 回まで	10 回に 1 回まで	5 回に 1 回まで	3 回に 1 回まで	それより多い

問 2　あなたは，現在の生活がどのくらい幸せだと感じますか。0〜10 点の間で答えてください。

0　1　2　3　4　5　6　7　8　9　10

不幸せ ←→ 幸せ

解答例 2：

問 1　全体的に考えて，あなたは社会のルールを厳しく守る方だと思いますか。

1	2	3	4
強くそう思う	ある程度そう思う	あまりそう思わない	全くそう思わない

問 2　あなたは，この大学の中に悩み事を相談できる友人がどのくらいいますか。

1	2	3	4
1 人もいない	4〜6 人程度	7〜9 人程度	10 人以上

解 説　「まじめな学生」も「充実した学生生活」もあいまいな言葉なので，さまざまな作業仮説が考えられる。作業仮説によって分析の結論も大きく変わるので，作業仮説を明確にすることは，調査の意味を明確にすることにつながる。

4 解答例：「Q08c 生活満足度：家庭生活」は，満足度が高い方から順に「5，4，3，2，1」とコードを逆転させる。「Q19c 夕食の用意：頻度（本人）」については，「7，3，1，0.23，0.06，0.02，0」と 1 週間あたりのおよその回数に変換する。

解 説　数値が大きいほど満足度や頻度が高いことを表す方がわかりやすいので，コードを逆転させる方がよい（0 を最低として 4，3，2，1，0 や 6，5，……，1，0 としてもよい）。ただし，「Q19c 夕食の用意：頻度（本人）」は順序尺度の変数なので，コードを逆転させても数量として扱うことができず，このリコーディングの恩恵は少ない。より積極的に比例尺度の変数にリコーディングするならば，1 週間あたりのおよその回数に換算して，「7，3，1，0.23（7÷30），0.06（7÷365×3），0.02（7÷365），0」といったコードをあてることが考えられる。ただし，あまり確かな数値ではないので，分析の際には注意が必要になる。

第 4 章

1 それぞれ検索すると，以下のとおり調査データが見つかる（2007 年 1 月現在）。

（1）「0260　1995 年 SSM 調査」

（2）「0047　長寿時代の生活設計(生活者アンケート)」「0393　生活保障に関する調査」など

（3）「0296　介護サービス実態調査」「0045　高齢者の介護に関する調査」など

（4）「0395　老研−ミシガン大 全国高齢者パネル調査」「0247　日本人の民主主義観と社会資本に関する世論調査」など

解 説　「収録調査の検索」のページから，公開データを検索することができる。（3）のような場合には，「高齢者」「介護」などとキーワードを定めて検索する。検索の漏れを減らすためには，「老人」「ケア」など，幅広い関係キーワードで検索してみるとよい。

2 （1）留置調査票 Q46　（2）留置調査票 Q43e　（3）面接調査表 問 14

（4）それぞれ以下のとおり

（1）

Q46 安楽死の賛否

		度数	パーセント	有効パーセント	累積パーセント
有効	はい	1930	66.7	66.7	66.7
	いいえ	199	6.9	6.9	73.6
	わからない	754	26.1	26.1	99.7
	無回答	10	.3	.3	100.0
	合計	2893	100.0	100.0	

（2）

Q43e 意見についての賛否:性役割分担

		度数	パーセント	有効パーセント	累積パーセント
有効	賛成	393	13.6	13.6	13.6
	どちらかといえば賛成	1165	40.3	40.3	53.9
	どちらかといえば反対	882	30.5	30.5	84.3
	反対	409	14.1	14.1	98.5
	無回答	44	1.5	1.5	100.0
	合計	2893	100.0	100.0	

（3）

問14 支持政党		度数	パーセント	有効パーセント	累積パーセント
有効	自民党	592	20.5	20.5	20.5
	民主党	182	6.3	6.3	26.8
	公明党	96	3.3	3.3	30.1
	自由党	43	1.5	1.5	31.6
	共産党	47	1.6	1.6	33.2
	社民党	63	2.2	2.2	35.4
	保守党	2	.1	.1	35.4
	その他	6	.2	.2	35.6
	特に支持する政党はない	1767	61.1	61.1	96.7
	わからない	92	3.2	3.2	99.9
	無回答	3	.1	.1	100.0
	合計	2893	100.0	100.0	

解 説　本文の pp.46～47 のとおり，いろいろな方法で検索ができる。キーワードがぴたりと一致するとは限らないので，多様なキーワードで幅広く検索した方がよい。

3　（1）男性 1,318 人，女性 1,575 人　（2）2000 年 10～11 月　（3）中央調査社

解 説　いずれも，SSJ データアーカイブや JGSS のホームページで，JGSS-2000 の調査概要のページを見れば，簡単にわかる。データ分析の前に調査の概要を確認することは大切である。

4　JGSS-2000 で調査されているものに限ると，「Q03 テレビ視聴時間（HRTV）」「Q29h 組織への信頼：テレビ（TR3TVZ）」がある。他の調査年度も含めると，「政治に関わる情報収集：テレビ（JGSS-2003 FQ4PLTV）」「英語学習・経験：テレビやラジオの番組やニュース（JGSS-2002 XEMEDIA）」「娯楽の頻度：テレビゲーム（JGSS-2000~2003 FQ4GAME）」などもある。

解 説　大雑把な関心対象について，JGSS の調査項目を検索するには，JGSS ホームページの事項索引が向いている。調査年度をまたいで調査項目を探すことができる。

5　解答例：データ公開によって新たな調査を実施する必要が少なくなることは，よいことである。しかし，逆に考えれば，実施される調査の数が減り，調査データの蓄積が不十分になる危険も考えられる。また，研究者が調査の実施を経験する機会が少なくなり，その知識や技術の伝達が不十分になることも考えられる。

第 5 章

1

	度数	相対度数（％）	累積度数	累積相対度数（％）
賛成	609	21.1	609	21.1
どちらかといえば賛成	1,060	36.6	1,669	58.0
どちらともいえない	705	24.4	2,374	82.4
どちらかといえば反対	258	8.9	2,632	91.4
反対	248	8.6	2,880	100.0
無回答	13	0.4	—	—
計	2,893	100.0		

2　解答例：

	度数	相対度数（％）
11 時間以下	88	4.7
12～19 時間	105	5.6
20～27 時間	159	8.5
28～35 時間	259	13.8
36～43 時間	494	26.3
44～51 時間	427	22.7
52～59 時間	123	6.5
60～67 時間	147	7.8
68 時間以上	66	3.5
無回答	13	0.7
計	1,881	100.0

解説　8 の倍数に回答が集中する傾向があるので，8 の倍数が区切り目になることを避けつつ，階級の幅を 8 時間に設定した。最初の階級（11 時間以下）の幅が変則的であるが，度数が十分に少ないので特に問題はない。もちろん，他にも多様な階級の定め方が考えられる。

3　イ

解説　階級の幅が一貫していないので，全体の分布の形が想像しにくい。ウは階級の幅をよりばらばらにしてしまうので，不適切である。また，現状でも半数以上のケースが 1 つの階級に集まっているので，エのように区切りをおおまかにすることも不適切である。アのように比率だけを示すことは基本的に望ましくない。

4　Q15 少年法改正の賛否

解説　調査票で選択肢を確認すると，Q13d や Q17 は選択肢の順序に意味がある順序尺度の調査項目である。一方，Q15 は順序に特に意味がなく，名義尺度の調査項目と考えられる。名義尺度の調査項目の累積度数を求めることにはほとんど意味がない。

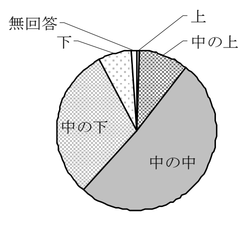
5　次の表のように，JGSS-2000 の回答は，国勢調査に比べ若干面積が広い方に偏っていることがわかる。ただし，回答者に未成年や 90 歳以上の高齢者を含んでいないことの影響は考えられる。

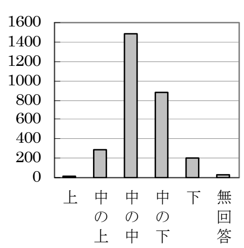
	平成 12 年国勢調査		JGSS-2000	
	度数	%	度数	%
0～19 ㎡	2,129,825	1.7	28	1.0
20～29 ㎡	3,975,118	3.2	55	1.9
30～39 ㎡	6,884,466	5.6	113	4.0
40～49 ㎡	8,654,208	7.0	148	5.2
50～59 ㎡	9,014,444	7.3	162	5.7
60～69 ㎡	10,720,661	8.7	163	5.7
70～79 ㎡	8,782,840	7.1	137	4.8
80～89 ㎡	7,516,834	6.1	164	5.8
90～99 ㎡	9,663,232	7.8	248	8.7
100～119 ㎡	13,529,750	11.0	334	11.7
120～149 ㎡	17,610,746	14.3	489	17.2
150～199 ㎡	15,862,599	12.8	502	17.7
200～249 ㎡	5,377,379	4.4	157	5.5
250 ㎡以上	3,768,291	3.1	144	5.1
計	123,490,393	100.0	2,844	100.0

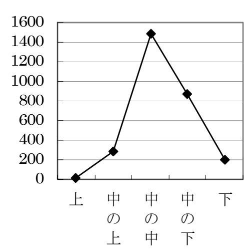
解説　必要な集計結果は，http://www.stat.go.jp/data/kokusei/2000/kihon1/00/hyodai.htm（2007 年 1 月現在）で表 19 として公表されている。国勢調査の集計表のタイトルは「○○別×× － △△」というパターンでつけらており，○○でグループ分けした××についての集計を，△△の地域区分（全国，都道府県など）で示すという意味である。複数の変数が併記されるのでわかりにくいが，○○に延べ面積，××に世帯人員数に近い用語が入っており，△△が全国である集計表を探せばよい。表 19 のタイトルは「延べ面積（14 区分），世帯人員（7 区分）,住宅の所有の関係（6 区分）別住宅に住む一般世帯数,一般世帯人員,1 世帯当たり人員,1 世帯当たり延べ面積及び 1 人当たり延べ面積 － 全国」であり，探している表に該当する。似たようなものとして表 20 に「……主世帯人員」という表があるが，主世帯は一般世帯から間借りの世帯を除いたものである。ここでは特に主世帯に限る必要はないので，一般世帯の集計である表 19 を用いた（国勢調査の用語については http://www.stat.go.jp/data/kokusei/2000/kihon1/00/yougo.htm などを参照）。国勢調査の集計表に合わせて，JGSS-2000 の度数分布表をまとめ直すと，上のように比較ができる。なお，100 ㎡以上で階級の幅がばらばらになるのは，あまり望ましいことではない。

第 6 章

1　（1）円グラフ　（2）棒グラフ　（3）折れ線グラフ





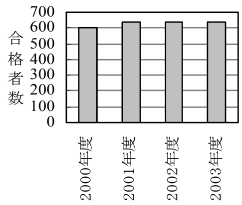


注：無回答の 32 ケースを除く。

解説　（1）割合を比較するには円グラフが最適である。細かい数値がある方が望ましければ，パーセントの実数も記す。また，帯グラフを用いてもよい。（2）絶対的な量を比較するには棒グラフがよい。（3）変化を表すには折れ線グラフがよい。

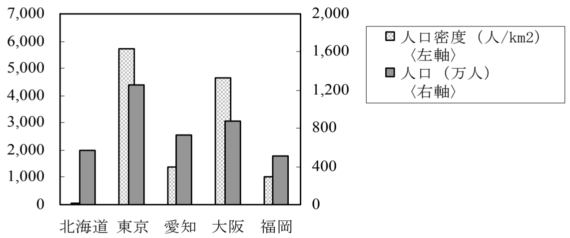
2　この棒グラフは，年度ごとの合格者数を直感的に比較するためのものである。しかし，示されているグラフでは，580 より下の棒が打ち切られており，棒の長さが合格者数に対応していない。また，棒を立体化し奥行きを出しているために，合格者数が同じでも手前（新しい年度）の棒ほど長くなってしまっている。これらのことから，この棒グラフでは直感的な比較ができない。

解説　比較したいもの（年度ごとの合格者数）を直感的に比較するための視覚情報（棒の長さ）に注目すれば，このグラフが不適切なことは，容易にわかる。ちなみに適切にグラフを描くと下のようになる。

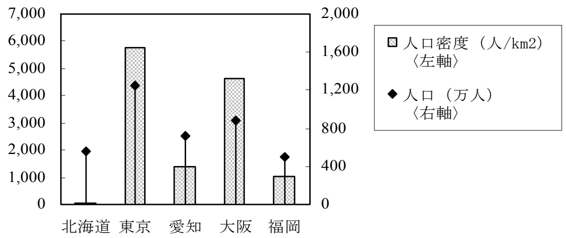


3　解答例

（a）

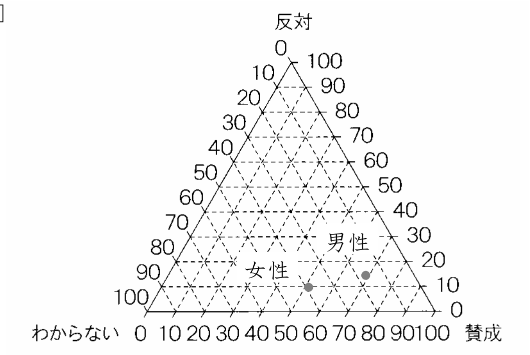


（b）



解説　2 つの棒の混同を避けるためには，（a）のように棒の色を変えたり，棒を少しずらしたりすることが多い。いずれも棒の長さという視覚情報を損なうことなく混同を避ける対策である。あまり見かけない形であるが，（b）のように一方の棒を極端に細くすることも有効だろう。

4

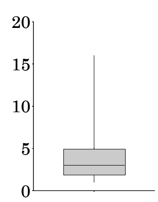


5　（略）

第 7 章

1　（1）　平均値 3.76，中央値 3，範囲 15，標準偏差 2.053，分散 4.2，四分領域 1.5

（2）



（3）正の側に裾野が長いので，歪度は 0 より大きい。

解 説 平均値，中央値，標準偏差は，リモート集計でそのまま算出できる。範囲は最大値－最小値なので，16－1＝15 である。分散は標準偏差の 2 乗である。四分領域は（第 3 四分位数－第 1 四分位数）÷2 なので，（5－2）÷2＝1.5 である。箱ヒゲ図は最小値・最大値と 3 つの四分位数から描くことができる。

2 男性の平均値は 3.27，女性の平均値は 3.59 なので，女性の方が平均値は大きい。

3 月給のばらつきの方が大きい。

解 説 月給の平均値は 304808.97，標準偏差は 179807.366 なので，変動係数は 179807.366÷304808.97 ≒0.59 である。一方，時間給の平均値は 848.25，標準偏差は 284.835 なので，変動係数は 284.835÷848.25 ≒0.34 である。

4 ウ

解 説 平均値は中央値よりもはずれ値の影響を受けて極端な値になりやすい。分布の裾野が正の側（右側）に広がっていれば，平均値は中央値よりも大きな値になり，逆に負の側（左側）に広がっていれば，平均値は中央値よりも小さな値になる。したがって，アやイのように断定するのは間違いである。

5 イ

6 イ

解 説 平均値を算出するためには，間隔尺度以上の水準で測定された変数でなければならない。アとウは比例尺度，エは間隔尺度とみなせるが，イは等間隔ではない順序尺度とみなせるので，そのままでは平均値を算出できない。

第 8 章

1

	度数	%
電子メール（仕事）	295	10.2
電子メール（私用）	462	16.0
パソコン（職場）	648	22.4
パソコン（自宅）	631	21.8
インターネットによる ショッピング・バンキング	139	4.8
インターネットによる株取引	24	0.8
携帯電話または PHS	1,361	47.0
ファックス	1,047	36.2
全ケース数（n）	2,893	100.0

2 円グラフは，その構成割合をパイの面積で比較するためのものである。ここで示されているグラフは，複数回答のため同じ回答者が重複しており，パイ全体が集計対象の回答者全体（100％）に対応していない。したがって，不適切である。

3 イ

解 説 教科の好き嫌いは，どこからが好きでどこからが嫌いと答えるべきなのか，はっきりとしない問題である。そのため，複数回答で尋ねると，回答者によって判断がばらばらで質の低い調査データになる恐れが強い。他の調査項目については，客観的に判断しやすいので，複数回答に向いている。

4 （略）

第 9 章

1 （1）列％。男の子を欲しい人の 81.9％（992÷1211×100）が既婚者で，女の子が欲しい人の 86.6％（1359÷1570×100）が既婚者なので，女の子が欲しい人の方が，既婚者の割合が高い。

（2）行％。既婚者は男の子が欲しい人が 42.2％（992÷2351×100）で女の子が欲しい人が 57.8％（1359÷2351×100）なのに対し，未婚者は男の子が欲しい人が 50.9％（219÷430×100）で女の子が欲しい人が 49.1％（211÷430×100）である。したがって，既婚者の方が女の子を欲しがる（未婚者の方が男の子を欲しがる）傾向がある。

（3）全体％。未婚で女の子を欲しい人の割合は 7.6％（211÷2781×100）である。

（4）行％。既婚者の中で女の子が欲しい人は 57.8％なのに対し，未婚者の中で女の子が欲しい人の割合は 49.1％である。したがって，既婚者の方が女の子を欲しがっている割合が高い。

2 無回答を除いて集計すると下のようなクロス表ができるので，男性の方がわずかにペットを飼っている割合が高いが，その差はほとんどない。

	飼っている	飼っていない	計
男性	508	809	1317
	38.6％	*61.4％*	*100％*
女性	597	977	1574
	37.9％	*62.1％*	*100％*

解説　男性と女性でペットを飼っている割合を比較するので，分布を知りたい中心的な変数（ペットの有無）を「列」に，グループ分けのための変数（性別）を「行」に指定して，クロス表を作る。読み取るべきパーセントは行％である。

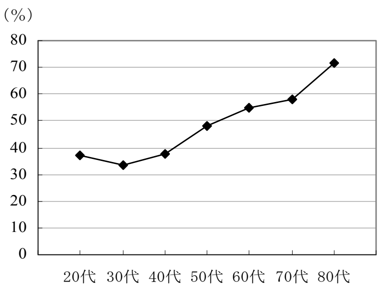
3 解答例：

	ほとんど毎日	週数回	それより少ない	計
下の方	87	9	32	128
	68.0％	*7.0％*	*25.0％*	*100％*
やや下の方	231	49	73	353
	65.4％	*13.9％*	*20.7％*	*100％*
真ん中のあたり	989	166	144	1299
	76.1％	*12.8％*	*11.1％*	*100％*
やや上の方	390	65	63	518
	75.3％	*12.5％*	*12.2％*	*100％*
上の方	339	32	37	408
	83.1％	*7.8％*	*9.1％*	*100％*

解説　もとのクロス表は下のようになり，やや煩雑である。「新聞を読む頻度」の回答は偏っており，たった1つの回答（ほぼ毎日）が4分の3程度を占める。そこで，少数回答の「週1回程度」〜「全く読まない」は1つのカテゴリー（それより少ない）に縮約した。一方で「中学3年生の頃の成績」は回答がある程度ばらついているので，そのまま5つのカテゴリーを残すことにした。こちらも縮約し3段階程度にしてもよいだろう。ただし，その場合，成績が「上の方」という回答者が新聞をかなりよく読むという傾向が読み取れなくなる。また，「新聞を読む頻度」をもっと縮約し，「ほとんど毎日」と「それ以外」にすることも考えられる。ただし，その場合，成績が「下の方」という回答者に新聞をほとんど読まない者が多い傾向が読み取りにくくなる。どの程度細かい情報を提示したいかによって必要な縮約の程度は変わってくる。

		Q01 新聞を読む頻度						
		ほぼ毎日	週数回	週1回程度	それ以下	全く読まない	無回答	合計
		度数	度数	度数	度数	度数	度数	度数
問30 中学3年生の頃の成績	下の方	87	9	4	11	17	1	129
	やや下の方	231	49	15	18	40	0	353
	真ん中のあたり	989	166	41	38	65	0	1299
	やや上の方	390	65	27	19	17	2	520
	上の方	339	32	15	12	10	1	409
	回答したくない	43	12	2	2	3	1	63
	わからない	77	16	5	4	14	0	116
	無回答	3	1	0	0	0	0	4

4 解答例：



注：「例外なく悪い」という回答の割合。
20代から順に，n＝393，416，495，634，535，332，88。

解 説　回答には 4 段階の選択肢があるが，半数程度の回答者が「例外なく悪い」と回答しているので，この割合のみを折れ線グラフで表せば，ほぼその変化を表現できるだろう。考え方は，クロス表の縮約と同じである。

第 10 章

1　（1）131÷230×100＝57.0％，958÷2479×100＝38.6％なので，外国人と一緒に勉強したことがある人の方が外国人増加に賛成であるという関連性が読み取れる（または，99÷230×100＝43.0％，1521÷2479×100＝61.4％なので，外国人と一緒に勉強したことがない人の方が外国人増加に反対であるという関連性が読み取れる）。

（2）ユールのQ＝$\dfrac{131\times1521-99\times958}{131\times1521+99\times958}\fallingdotseq 0.355$，$\phi$係数＝$\dfrac{131\times1521-99\times958}{\sqrt{230\times2479\times1089\times1620}}\fallingdotseq 0.104$，

オッズ比＝$\dfrac{131\times1521}{99\times958}\fallingdotseq 2.101$

2　ウ

3　$A=22\times(83+82+108+424)+33\times(82+424)+19\times(108+424)+83\times424=77332$

$B=19\times(33+8)+83\times8+17\times(33+8+83+82)+108\times(8+82)=14665$ なので，

グッドマンとクラスカルの $\gamma=\dfrac{77332-14665}{77332+14665}\fallingdotseq 0.681$

解 説　下の表が計算のもとになるクロス表である。あるセルの回答者からみるとそのセルの右下か左上にいる回答者とは大小関係が同方向のペアをつくることができる。例えば，真ん中のセルの 83 人の回答者は，右下（424 人）と左上（22 人）の回答者とは大小関係が同方向のペアとなる。しかし，これらを両方とも数えると同じペアを 2 度数えることになってしまうので，右下の回答者とのペアの個数だけを数えることにする。例えば，一番左上のセルの回答者（22 人）に関係して 22×（83＋82＋108＋424）の数だけ大小関係が同方向のペアができる。次に，その右のセルの回答者（33 人）に関係して 33×（82＋424），……以下同様にして大小関係が同方向のペアの数（A）をカウントする。大小関係が逆方向のペアの数（B）は，反対に考えて，それぞれのセルの回答者からみて右上（あるいは左下）にいる回答者とのペアの数を数えあげればよい。

	熱心である	まあまあ熱心である	そんなに熱心ではない	計
とても信頼している	22	33	8	63
少しは信頼している	19	83	82	184
ほとんど信頼していない	17	108	右下 424	549
計	58	224	514	796

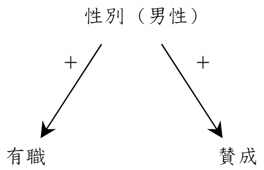
4　0.302

5　解答例：まず，因果の方向が逆である可能性がある。つまり，コーヒーを飲むことが原因で健康になるのではなく，健康であることが原因でコーヒーを飲む（飲める）という関係が正しいかもしれない。健康を害した人は，医者にコーヒーを止められることがあるからである。また，両者の関係が見せかけの関係である可能性もある。例えば，金銭的に余裕がある人ほどコーヒーをたくさん飲むことができ，また健康状態もよいという関係があるならば，コーヒーと健康の間に関連性は現れるが，それは因果関係とはいえない。

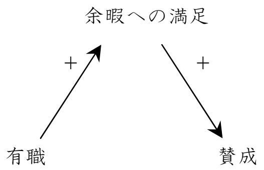
第 11 章

1　（1）行％から，有職の人の方が無職の人よりも賛成の割合が高いという関連性が読み取れる（または，列％から賛成の人の方が反対の人よりも有職の割合が高いという関連性が読み取れる）。

（2）男性の中だけで見ると，有職者も無職者も賛成の割合はで同じである（80％）。また，女性の中だけで見ても，有職者と無職者で賛成の割合は同じである（25％）。周辺度数に注目すると，男性の方が女性よりも有職者の割合が非常に高く，また賛成の割合も非常に高いことがわかる。このことから，職の有無と政策への賛否の間の関連性は，次の図のように性別を共通の原因（先行変数）とする見せかけの関係と考えることができる。



（3）第 3 の変数が余暇への満足の場合，2 つの変数よりも時間的に先行する変数と考えることはできない。下の図のように，職の有無と政策への賛否の間に入り，媒介関係をなしていると考えるのが自然である。つまり，有職者の方が余暇に満足しており，余暇に満足している人の方が政策に賛成する，といった関係である。職の有無と政策への賛否の間に因果関係が成立しているので，これは見せかけの関係ではない。



2 整理すると，次のような 3 重クロス表が作成できる。30 代の中だけで見ても 40 代の中だけで見ても，学歴が高い人の方がパソコンを使っている割合が高いので，学歴とパソコン利用の関係は，年齢層による見せかけの関係では説明できていない。ただし，周辺度数に注目すると，30 代の方が高学歴で，またパソコンを使っている割合が高いので，一部は見せかけの関係が発生しているはずであることがわかる。以上のことから，見せかけの関係は発生しているもののその影響はわずかであり，学歴とパソコン利用の関係は，大部分が年齢層では説明できない本質的な因果関係と考えるべきだろう。

		使っている		使っていない		計
30 代	短大以上	89	*53.0％*	79	*47.0％*	168
	高卒以下	59	*23.9％*	188	*76.1％*	247
	計	148		267		415
40 代	短大以上	68	*41.0％*	98	*59.0％*	166
	高卒以下	68	*21.0％*	256	*79.0％*	324
	計	136		354		490

30 代の方が短大以上が多い
30 代：168／415＝40.5％
40 代：166／490＝33.9％

30 代の方が使っている
30 代：148／415＝35.7％
40 代：136／490＝27.8％

見せかけの関係は
わずかしかない

解 説　「問 27 最終学校（本人）」を行に，「Q60_4 利用通信媒体：パソコン（自宅）」を列に，「年齢 10 歳刻み」を層に指定して 3 重クロス表を作成した上で，30 代〜40 代の部分を抜き出し，クロス表を縮約すると，上のような表ができる。実際の調査データの分析では，例題のようなわかりやすい結果が現れることはあまりなく，この問題のように入り組んでいることが多い。その中から正しい情報を読み解くには，基本的な読み取り方を熟知している必要がある。問題への解答からは脱線するが，このクロス表で注目すべきことは，40 代に比べて 30 代の方が学歴によるパソコン利用への影響が強まっていることだろう。

3 解答例：都市の方が賃貸住宅に住んでいる人は多く，賃貸住宅に住んでいる人は持ち家に住んでいる人に比べてペットを飼う（飼える）割合が低いように思える。つまり，市郡規模とペットの有無の関係は，住居の種類（一戸建てか集合住宅か）を媒介変数としている可能性が考えられる。この仮説を検証するために，次のような 3 重クロス表を作成すると，仮説が支持されることがわかる。ただし，市郡規模とペットの有無の関連性がすべて説明できたわけではなく，一戸建ての回答者の中だけで比べてもまだ都市に住んでいる人はペットを飼っている割合がやや低い。

		飼っている		飼っていない		計
一戸建て	13 大市	107	*36.9％*	183	*63.1％*	290
	その他の市	538	*41.3％*	765	*58.7％*	1303
	町村	320	*46.9％*	362	*53.1％*	682
	計	965		1310		2275
集合住宅	13 大市	50	*20.7％*	192	*79.3％*	242
	その他の市	77	*24.4％*	239	*75.6％*	316
	町村	8	*19.5％*	33	*80.5%*	41
	計	135		464		599

一戸建てには町村居住者が多い
一戸建て：
13 大市 290／2275＝12.7％
その他の市 1303／2275＝57.3％
町村 682／2275＝30.0％
集合住宅：
13 大市 242／599＝40.4％
その他の市 316／599＝52.8％
町村 41／599＝6.8％

一戸建ての方がペットを飼っている
一戸建て：965／2275＝42.4％
集合住宅：135／599＝22.5％

町村居住者は一戸建てであることが多いので，ペットを飼っている割合が高い。

解 説 「市郡規模」を行に，「Q62 ペットの有無」を列に，「問 16 住居形態」を層に指定して 3 重クロス表を作成する。住居形態を一戸建てか集合住宅かだけに注目して縮約すると，上のような表ができる（住居形態がその他や無回答のケースは除外している）。

第 12 章

1 信頼区間は $0.611 \pm 1.96 \times \sqrt{\dfrac{0.611\,(1-0.611)}{2890}}$ である。

つまり，無党派層の割合は 95％の信頼度で 59.3〜62.9％と推定できる。

解 説 無回答者 3 人を除くと，無党派（特に支持する政党はない）の割合は 1767÷2890＝0.611（61.1％）である。$p=0.611$，$n=2890$，$Z_{0.05/2}=1.96$ を 144 ページの式に当てはめればよい。

2 男性の喫煙者は，$0.500 \pm 2.58 \times \sqrt{\dfrac{0.500\,(1-0.500)}{1311}}$，つまり 46.4〜53.6％と推定できる。

女性の喫煙者は，$0.159 \pm 2.58 \times \sqrt{\dfrac{0.159\,(1-0.159)}{1565}}$，つまり 13.5〜18.3％と推定できる。

解 説 性別と喫煙習慣でクロス表を作成すると，無回答者を除いて，男性の喫煙者は 656÷1311＝0.500（50.0％）で，女性の喫煙者は 249÷1565＝0.159（15.9％）である。99％の信頼度なので，$Z_{0.01/2}=2.58$ として信頼区間を求めればよい。

3 ウ

解 説 推測統計法の目的は，一部の標本から母集団全体を推し測ることである。ここでは，S 大学の学生全体が母集団，その中から無作為抽出した 100 人が標本にあたるので，ウが正答である。

4 ウ

解 説 信頼度を下がれば，信頼区間は狭くなる（ある程度予想がはずれてもよければ，範囲を絞った予想が成り立つ）。その場合でも，信頼区間の中心（20％と 30％の間で 25％）は変わらないので，ウが正答である。

5 信頼度 90％の場合，信頼区間は 55.3〜76.7％
信頼度 95％の場合，信頼区間は 53.2〜78.8％
信頼度 99％の場合，信頼区間は 49.2〜82.8％

解 説 クロス表で確認すると，両生類／は虫類を飼っている人々は 53 人で，そのうち女性は 35 人（66.0％）である。信頼度 90％の場合は $Z_{0.10/2}=1.64$，信頼度 95％の場合は $Z_{0.05/2}=1.96$，信頼度 99％の場合は $Z_{0.01/2}=2.58$ である。ただし，標本の人数が少ないので推定の結果がやや不正確である可能性に注意が必要である。

6 世帯年収が 1,000 万円以上の人は，$0.195 \pm 1.96 \times \sqrt{\dfrac{0.195\,(1-0.195)}{1979}}$，つまり 17.8〜21.2％と推定できる。

解 説 度数分布表で確認すると，世帯年収が 1,000 万円以上の人は 1,979 人中 386 人（19.5％）である。

第 13 章

1 調査結果からは，139 人中 79 人（56.8％）が男性である。
帰無仮説と対立仮説は以下のとおりである。
H_0：母集団では，男性は 50％である（$\pi = 0.50$）
H_1：母集団では，男性は 50％ではない（$\pi \neq 0.50$）
比率の検定（両側検定）を行うために，検定統計量を計算する。

$$Z = \frac{0.568 - 0.500}{\sqrt{\dfrac{0.568\,(1 - 0.568)}{139}}} \fallingdotseq 1.62$$

有意水準 5％における臨界値は ±1.96 である。
したがって，Z 値は帰無仮説の採択域に入り，帰無仮説が採択される。
つまり，「半数以上が男性である」と一般化することはできない。

2 調査結果からは，2,891 人中 1,105 人（38.2％）がペットを飼っている。
帰無仮説と対立仮説は以下のとおりである。
H_0：母集団では，ペットを飼っている人は 3 分の 1 である（$\pi = 0.333$）
H_1：母集団では，ペットを飼っている人は 3 分の 1 ではない（$\pi \neq 0.333$）
比率の検定（両側検定）を行うために，検定統計量を計算する。

$$Z = \frac{0.382 - 0.333}{\sqrt{\dfrac{0.382\,(1 - 0.382)}{2891}}} \fallingdotseq 5.42$$

有意水準 5％における臨界値は ±1.96 である。
したがって，Z 値は帰無仮説の棄却域に入り，対立仮説が採択される。
つまり，「ペットを飼っている人は 3 分の 1 を超えている」と一般化できる。

3 イ

解 説　統計的検定のもともとの考え方（153 ページ）に立ち返れば，正答はすぐにわかる。検定の具体的な手続きができることも大切であるが，その考え方を理解することも大切である。

4 解答例：世の中の人々の暮らし方や考え方を，アンケート用紙などを使って科学的に調べることを社会調査とよびます。よく新聞やテレビで「内閣の支持率，先月に比べ急落」といったニュースを聞くのは，社会調査の結果を知らせるものです。

社会調査は多くの人々を調べますが，それでも調べるべき人々を全員調べているわけではありません。例えば，内閣の支持率を正確に知るためには，日本の成人を全員調べなければなりませんが，そうするとたいへんな時間とお金がかかってしまうので，実際には数千人しか調べていないのです。そう聞くと，「そんな一部の人の意見で正しいことがわかるのか。」と疑問に思うかもしれません。その疑問は当然のものです。

そこで，社会調査の結果を人々に知らせる際には，「統計的検定」とよばれる判定テストを行います。統計的検定は，全員調べても同じような調査結果が出る可能性が高いかどうかを計算で判定します。例えば，かりに全員調べた場合でも「内閣の支持率が先月よりも下がった」という同じ調査結果が出るといってよいかどうかを判定します。調査をする少人数の人々を慎重に選んでさえいれば，このような判定が後からでも可能なことが，科学者たちの研究でわかっています。

統計的検定は，社会調査で間違ったことを人々に伝えないために欠かせない手続きなのです。

5 調査結果からは，20 代の女性 207 人中 96 人（46.4％）が賛成と回答している。
帰無仮説と対立仮説は以下のとおりである。
H_0：母集団では，賛成の人は半分である（$\pi = 0.50$）
H_1：母集団では，賛成の人は半分ではない（$\pi \neq 0.50$）
比率の検定（両側検定）を行うために，検定統計量を計算する。

$$Z = \frac{0.464 - 0.500}{\sqrt{\dfrac{0.464\,(1 - 0.464)}{207}}} \fallingdotseq -1.04$$

有意水準 5％における臨界値は ±1.96 である。
したがって，Z 値は帰無仮説の採択域に入り，帰無仮説が採択される。
つまり，「賛成が半数にも満たない」とはいえない。

解 説　3 変数のクロス表で確認すると，20 代の女性は 212 人で，「賛成」が 13 人，「どちらかといえば賛成」が 83 人いる。無回答者が 5 人いるので，207 人中 96 人（46.4％）が，賛成していることになる。この結果をもとに通常の比率の検定を行えばよい。

1 帰無仮説と対立仮説は以下のとおりである。

H_0：母集団では，予想と同じ分布である

H_1：母集団では，予想と同じ分布ではない

適合度検定を行うために，検定統計量を計算する。

$$\chi^2=\frac{(342-321.75)^2}{321.75}+\frac{(113-123.75)^2}{123.75}+\frac{(6-9.9)^2}{9.9}+\frac{(34-39.6)^2}{39.6}\fallingdotseq 4.54$$

自由度 4－1＝3 の χ^2 分布を参照すると，有意水準 5％における臨界値は 7.81 である。
したがって，χ^2 値は帰無仮説の採択域に入り，帰無仮説が採択される。
つまり，「予想される分布とかけはなれている」とはいえない。

解 説　ある理論に従った場合の期待度数は順に，495×0.65＝321.75，495×0.25＝123.75，495×0.02＝9.9，495×0.08＝39.6 と計算できる。これをもとに適合度検定を行う。手計算ではやや大変なので，期待度数を整数に四捨五入したとしても，検定の結果に大きな影響はない。

2 （1）行％を算出すると以下のようになるので，女性の方が反対する傾向が読み取れる。

	賛成	どちらかといえば賛成	どちらかといえば反対	反対	計
男性	9.0％	32.6％	42.7％	15.7％	100％
女性	4.4％	27.8％	47.3％	20.5％	100％
計	6.5％	30.0％	45.2％	18.3％	100％

（2）期待度数は以下のように算出できる（小数点以下第 2 位を四捨五入した場合）。

	賛成	どちらかといえば賛成	どちらかといえば反対	反対	計
男性	11.6	53.4	80.4	32.5	178
女性	13.4	61.6	92.6	37.5	205
計	25	115	173	70	383

解 説　例えば，「男性」で「賛成」の期待度数は，$\frac{178}{383}\times\frac{25}{383}\times 383\fallingdotseq 11.6$ であり，「男性」で「どちらかいえば賛成」の期待度数は，$\frac{178}{383}\times\frac{115}{383}\times 383\fallingdotseq 53.4$ である。
11.6＋53.4＋80.4＋32.5＝177.9 と合計が周辺度数と一致しないのは四捨五入による誤差である。また，簡便のために整数まで四捨五入しても，検定の結果に大きな影響はない。

（3）帰無仮説と対立仮説は以下のとおりである。

H_0：母集団では，性別と性役割分担への賛否は統計的に独立である（関連性がない）

H_1：母集団では，性別と性役割分担への賛否は統計的に独立ではない（関連性がある）

独立性の検定を行うために，検定統計量を計算する。

$$\chi^2=\frac{(16-11.6)^2}{11.6}+\frac{(58-53.4)^2}{53.4}+\frac{(76-80.4)^2}{80.4}+\frac{(28-32.5)^2}{32.5}$$
$$+\frac{(9-13.4)^2}{13.4}+\frac{(57-61.6)^2}{61.6}+\frac{(97-92.6)^2}{92.6}+\frac{(42-37.5)^2}{37.5}$$
$$\fallingdotseq 5.47$$

自由度（2－1）×（4－1）＝3 の χ^2 分布を参照すると，有意水準 5％における臨界値は 7.81 である。
したがって，χ^2 値は帰無仮説の採択域に入り，帰無仮説が採択される。
つまり，「2 つの変数の間に関連性がある」とはいえない。

（4）帰無仮説と対立仮説は以下のとおりである。

H_0：母集団では，性別と性役割分担への賛否は統計的に独立である（関連性がない）

H_1：母集団では，性別と性役割分担への賛否は統計的に独立ではない（関連性がある）

独立性の検定を行うために，検定統計量を計算する。

$$\chi^2=\frac{(16-9.1)^2}{9.1}+\frac{(86-69.1)^2}{69.1}+\frac{(51-63.9)^2}{63.9}+\frac{(24-34.8)^2}{34.8}$$
$$+\frac{(5-11.9)^2}{11.9}+\frac{(73-89.9)^2}{89.9}+\frac{(96-83.1)^2}{83.1}+\frac{(56-45.2)^2}{45.2}$$
$$\fallingdotseq 27.1$$

自由度（2－1）×（4－1）＝3 の χ^2 分布を参照すると，有意水準 5％における臨界値は 7.81 である。
したがって，χ^2 値は帰無仮説の棄却域に入り，対立仮説が採択される。
つまり，30 代では「2 つの変数の間に関連性がある」といえる（女性の方が反対する傾向がある）。

解説　リモート集計で 3 変数のクロス表をつくり確認すると，30 代の分布は下の表のようになる。

	賛成	どちらかといえば賛成	どちらかといえば反対	反対	計
男性	16	86	51	24	177
女性	5	73	96	56	230
計	21	159	147	80	407

2 つの変数の関係が統計的に独立であった場合の期待度数は次のとおりである（小数点以下第 2 位を四捨五入した場合）。これらをもとに独立性の検定を行えばよい。

	賛成	どちらかといえば賛成	どちらかといえば反対	反対	計
男性	9.1	69.1	63.9	34.8	177
女性	11.9	89.9	83.1	45.2	230
計	21	159	147	80	407

3　帰無仮説と対立仮説は以下のとおりである。
　H_0：母集団では，三世代同居観と希望する子どもの性別は統計的に独立である（関連性がない）
　H_1：母集団では，三世代同居観と希望する子どもの性別は統計的に独立ではない（関連性がある）
独立性の検定を行うために，検定統計量を計算する。

$$\chi^2 = \frac{(842-786.8)^2}{786.8} + \frac{(962-1017.2)^2}{1017.2} + \frac{(353-408.2)^2}{408.2} + \frac{(583-527.8)^2}{527.8} \fallingdotseq 20.1$$

自由度（2－1）×（2－1）＝1 の χ^2 分布を参照すると，有意水準 5％における臨界値は 3.84 である。
したがって，χ^2 値は帰無仮説の棄却域に入り，対立仮説が採択される。
つまり，「2 つの変数の間に関連性がある」といえる（三世代同居に賛成の方が男の子を望む傾向がある）。

解説　リモート集計で集計すると左のようなクロス表ができる。統計的に独立の場合の期待度数は右のとおりである（小数点以下第 2 位を四捨五入した場合）。これらをもとに独立性の検定を行えばよい。

	男の子	女の子	計
望ましい	842	962	1804
望ましくない	353	583	936
計	1195	1545	2740

	男の子	女の子	計
望ましい	786.8	1017.2	1804
望ましくない	408.2	527.8	936
計	1195	1545	2740

第 15 章

1　（1）95％の信頼度における信頼区間は $2.85 \pm 1.96 \times \dfrac{1.745}{\sqrt{180}}$ である。

つまり，男性の平均テレビ視聴時間は 2.60～3.10 時間と推定できる。
（2）帰無仮説と対立仮説は以下のとおりである。
　H_0：母集団では，平均値は 3 時間である（$\mu = 3$）
　H_1：母集団では，平均値は 3 時間ではない（$\mu \neq 3$）
平均の検定（両側検定）を行うために，検定統計量を計算する。

$$t = \frac{2.85-3}{\dfrac{1.745}{\sqrt{180}}} \fallingdotseq -1.15$$

自由度 180－1 の t 分布は，標準正規分布とほぼ同じで，有意水準 5％における臨界値は ±1.96 である。
したがって，t 値は帰無仮説の採択域に入り，帰無仮説が採択される。
つまり，「平均テレビ視聴時間が 3 時間未満」とはいえない。
（3）帰無仮説と対立仮説は以下のとおりである。
　H_0：母集団では，平均テレビ視聴時間に差がない（$\mu_1 = \mu_2$）
　H_1：母集団では，平均テレビ視聴時間に差がある（$\mu_1 \neq \mu_2$）
平均の差の検定（両側検定）を行うために，検定統計量を計算する。

$$t = \frac{2.85 - 3.05}{0.193} \fallingdotseq -1.04 \quad \left[s_* = \sqrt{\frac{(180-1)1.745^2 + (211-1)2.033^2}{180+211-2} \times \left(\frac{1}{180} + \frac{1}{211} \right)} \fallingdotseq 0.193 \right]$$

自由度 180＋211－2 の t 分布は，標準正規分布とほぼ同じで，有意水準 5%における臨界値は±1.96 である。
したがって，t 値は帰無仮説の採択域に入り，帰無仮説が採択される。
つまり，「平均テレビ視聴時間に差がある」とはいえない。
（4）検定統計量が以下のようになる。

$$t = \frac{2.85 - 3.05}{0.061} \fallingdotseq -3.28 \quad \left[s_* = \sqrt{\frac{(1800-1)1.745^2 + (2110-1)2.033^2}{1800+2110-2} \times \left(\frac{1}{1800} + \frac{1}{2110} \right)} \fallingdotseq 0.061 \right]$$

有意水準 5%における臨界値である±1.96 を超えているので，帰無仮説の棄却域に入り，対立仮説が採択される。つまり，この場合は「平均テレビ視聴時間に差がある」といえる。

2 帰無仮説と対立仮説は以下のとおりである。

H_0：母集団では，平均満足度に差がない（$\mu_1 = \mu_2$）

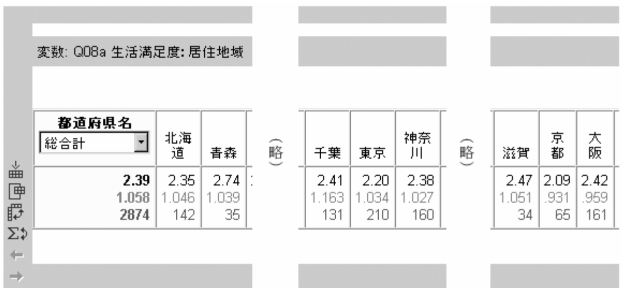
H_1：母集団では，平均満足度に差がある（$\mu_1 \neq \mu_2$）

平均の差の検定（両側検定）を行うために，検定統計量を計算する。

$$t = \frac{2.20 - 2.09}{0.143} \fallingdotseq 0.77 \quad \left[s_* = \sqrt{\frac{(210-1)1.034^2 + (65-1)0.931^2}{210+65-2} \times \left(\frac{1}{210} + \frac{1}{65} \right)} \fallingdotseq 0.143 \right]$$

自由度 210＋65－2 の t 分布は，標準正規分布とほぼ同じで，有意水準 5%における臨界値は±1.96 である。
したがって，t 値は帰無仮説の採択域に入り，帰無仮説が採択される。
つまり，「平均満足度に差がある」とはいえない。

解 説　リモート集計の「4．分類項目別度数分布と関連統計量の集計」を用いて，都道府県別の満足度を集計すると下のようになる（上の数値から順に，平均値，標準偏差，回答者数）。値が小さいほど満足度が高いことを表すので，京都の方がやや満足度が高い。この情報をもとに平均の差の検定を行えばよい。


変数: Q08a 生活満足度: 居住地域

都道府県名 総合計	北海道	青森	（略）	千葉	東京	神奈川	（略）	滋賀	京都	大阪
2.39	2.35	2.74		2.41	2.20	2.38		2.47	2.09	2.42
1.058	1.046	1.039		1.163	1.034	1.027		1.051	.931	.959
2874	142	35		131	210	160		34	65	161

3 エ

解 説　平均の差の検定で調べているのは，平均の差の有無だけである。帰無仮説は $\mu_1 = \mu_2$ であり，対立仮説は $\mu_1 \neq \mu_2$ なので，対立仮説が採択されたとしても，平均の差の量までは保証されない。

4 帰無仮説と対立仮説は以下のとおりである。

H_0：母集団では，2 つの満足度の差の平均値は 0 である（$\mu = 0$）

H_1：母集団では，2 つの満足度の差の平均値は 0 ではない（$\mu \neq 0$）

平均の検定（両側検定）を行うために，検定統計量を計算する。

$$t = \frac{0.29 - 0}{\frac{1.159}{\sqrt{212}}} \fallingdotseq 3.64$$

自由度 212－1 の t 分布は，標準正規分布とほぼ同じで，有意水準 5%における臨界値は±1.96 である。
したがって，t 値は帰無仮説の棄却域に入り，対立仮説が採択される。
つまり，「家庭生活より友人関係に満足している」といえる。

解 説　別々の調査項目間での平均の差の検定なので，通常の平均の差の検定を用いることはできない。「2 つの満足度の差」について，平均値が 0 でないといえるかどうか，平均の検定を行えばよい。

第 16 章

1 （1）全体的には，市郡規模が大きいほど家は狭い。

（2）全体的には，関東や九州で家が狭く，中部や中国・四国で家が広い。

（3）有意な交互作用効果は認められない。主効果についてはいずれも有意である。つまり，地域ブロックや市郡規模は家の広さに影響するが，それらが組み合わさることによる特別な影響は認められない。

解 説　（2）地域ブロックによる違いは，グラフからやや読み取りにくいが市郡規模別の回答者数（関東は町村が少ない，中部はその他の市が多いなど）も考慮してみると，およそ上記のように読み取れる。（3）まず交互作用効果の有無から読み取る。p 値が 0.156 なので 5％水準で有意ではない。そこで主効果に視点を移すと，いずれも 1％水準でも有意であることがわかる。

2 SPSS を用いて分散分析を行うと，下のような分散分析表が作成できる。家族の人数の影響は統計的に有意であることがわかる。分散分析表からは検定の結果しかわからないので，別途，グループごとにテレビ視聴時間の平均値を算出すると，1 人暮らしは 4.17 時間（n＝220），2 人暮らしは 3.77 時間（n＝707），3 人以上は 3.25 時間（n＝1,951）であった。一般的に，家族が多い方がテレビ視聴時間が短いといえる。

被験者間効果の検定

従属変数: hrtv Q03 テレビ視聴時間

ソース	タイプIII平方和	自由度	平均平方	F値	有意確率
修正モデル	266.575ª	2	133.287	32.152	.000
切片	19332.553	1	19332.553	4663.400	.000
szffherer	266.575	2	133.287	32.152	.000
誤差	11918.578	2875	4.146		
総和	46364.000	2878			
修正総和	12185.152	2877			

a. R2乗 = .022 (調整済みR2乗 = .021)

変動因	平方和	自由度	平均平方	F 値	p 値
級間（モデル）	266.575	2	133.287	32.152	0.000
級内（誤差）	11918.578	2875	4.146		
合計	12185.152	2877			

3 SPSS を用いて分散分析を行うと，下のような分散分析表が作成できる。交互作用の効果から注目すると，その効果は統計的に有意であることがわかる。つまり，結婚状況によって家族人数の影響の仕方は異なるということである（家族人数によって結婚状況の影響の仕方が異なるといってもよい）。交互作用の効果が有意であったので，主効果について読み取る意味はあまりない。

被験者間効果の検定

従属変数: hrtv Q03 テレビ視聴時間

ソース	タイプIII平方和	自由度	平均平方	F値	有意確率
修正モデル	570.226ª	8	71.278	17.606	.000
切片	7127.341	1	7127.341	1760.523	.000
domarry	209.916	2	104.958	25.926	.000
szffherer	31.790	2	15.895	3.926	.020
domarry * szffherer	79.929	4	19.982	4.936	.001
誤差	11614.927	2869	4.048		
総和	46364.000	2878			
修正総和	12185.152	2877			

a. R2乗 = .047 (調整済みR2乗 = .044)

変動因	平方和	自由度	平均平方	F 値	p 値
級内（モデル）	570.226	8	71.278	17.606	0.000
級間（誤差）	11614.929	2869	4.048		
合計	12185.152	2877			
結婚状況	209.916	2	104.958	25.926	0.000
家族人数	31.790	2	15.895	3.926	0.020
結婚状況×家族人数	79.929	4	19.982	4.936	0.001

分散分析表からは検定の結果しかわからないので，別途，グループごとにテレビ視聴時間の平均値を算出すると，下の表のようになる。未婚者は家族人数の影響をほとんど受けないのに対して，離死別者は 1 人暮らしになるとかなりテレビを長く見るようになるといった交互作用が読み取れる。

	1 人暮らし	2 人暮らし	3 人以上
有配偶	3.33 時間（n＝12）	3.82 時間（n＝571）	3.25 時間（n＝1,505）
離死別	5.08 時間（n＝117）	3.83 時間（n＝83）	3.78 時間（n＝153）
未婚	3.11 時間（n＝91）	3.13 時間（n＝53）	2.93 時間（n＝293）

4 （1）× （2）○ （3）○ （4）○

解 説 さまざまな表現を用いているが，いずれも分散分析の結果から，統計的に有意な影響が認められるかどうかを問題にしている。論文中では，F 値の横に有意確率に応じた＊印が記されているので，これを正しく読み取れば正答が得られるはずである。なお，論文中の表に注釈があるように，賛否と得点の大小の関係が変数によって異なっている。読み取りを誤らないように注意が必要である。

5 （1）○ （2）○ （3）× （4）○

解 説 （1）～（3）は分散分析の結果を問題にしている。（1）は親学歴と教育意識 A の交互作用効果，（2）は親学歴と教育意識 B の交互作用効果，（3）は教育意識 C の主効果の有無を読み取ればよい。いずれも F 値の横に有意確率に応じた＊印が記されている。（4）は母集団ではなく標本のことを問題にしているので，分散分析の結果とは関係なくグループ別の平均教育費負担を読み取ればよい。（分散分析の結果，交互作用効果は有意ではないが）高卒の親では教育意識 D に賛成の方が教育費負担が高いのに対して，大卒の親では教育意識 D に反対の方が教育費負担が高い。したがって，（4）は正しい。

第 17 章

1 SPSS で回帰分析を行うと，以下のような出力が得られる。わずかな誤差があるが，本文と同じ回帰式や決定係数，説明力の検定結果が得られていることがわかる。

モデル集計

モデル	R	R2 乗	調整済み R2 乗	推定値の標準誤差
1	.662a	.438	.394	6.97320

a. 予測値: (定数)、X。

分散分析b

モデル		平方和	自由度	平均平方	F 値	有意確率
1	回帰	491.868	1	491.868	10.115	.007a
	残差	632.132	13	48.626		
	全体	1124.000	14			

a. 予測値: (定数)、X。
b. 従属変数: Y

係数a

モデル		非標準化係数		標準化係数	t	有意確率
		B	標準誤差	ベータ		
1	(定数)	3.009	2.841		1.059	.309
	X	.095	.030	.662	3.180	.007

a. 従属変数: Y

2 $\beta = \dfrac{s_{XY}}{s_X^2} = \dfrac{-39.1}{105} \fallingdotseq -0.372$， $\alpha = \overline{Y} - \beta\overline{X} = 425 - (-0.372) \times 22 \fallingdotseq 433$ なので，最適な回帰線は

$$\hat{Y} = 433 - 0.372X$$

3 （1）$\hat{Y} = 18.533 + 24.202X$
（2）家族が 1 人増えると，家の広さは 24.202 ㎡大きくなる。
（3）18.533＋24.202×4≒115.341 なので，115.341 ㎡と予測される。
（4）調整済み決定係数が 0.171 なので 17.1％（決定係数をそのまま読むならば 17.3％）。
（5）F 値の横の有意確率が 0.000（0.001 未満）とあるので，5％水準で統計的に有意である。

4 それぞれについて，回帰分析の主な結果は次の表のようにまとめられる。30 代の女性や 40 代の男性については，年齢で月給の高さを有意に説明することはできない。調整済み決定係数の値が低く，説明力の検定結果も有意でないからである（調整済み決定係数がマイナスの値になることはほとんどないが，それだけ全く説明力がないということである）。20 代の女性については，5％水準で有意な説明力があり，年齢が 1 歳上がるごとに月給が約 5 千円上がると予測される。ただし，この回帰線で説明できるのは月給の 4.9％だけである。

	（1）30 代の女性	（2）20 代の女性	（3）40 代の男性
定数 α	173049.085	53745.517	420353.364
回帰係数 β	1214.642	5132.437	－298.236
調整済み決定係数	－0.17	0.049 ＊	－0.07
n	58	105	139

注：＊ $p<0.05$

18 章

1 （1）$\hat{Y} = 55.528 + 2.863X_1 - 0.017X_2 + 1.605X_3 + 0.329X_4 + 0.111X_5$

（2）調整済み決定係数が 0.174 なので 17.4％（決定係数をそのまま読むならば 17.7％）。

（3）t 値の横の有意確率が 0.050 未満の変数なので，男性ダミー，短大以上ダミーの 2 つ。

（4）1.605 歳高い。

（5）1.605－0.329＝1.276 なので，1.276 歳高い。

（6）解答例：女性，1960 年生まれ，高卒，中学 3 年の頃の成績はやや上の方の人の場合，55.528＋2.863×0－0.017×1960＋1.605×0＋0.329×1＋0.111×4＝22.981 なので，初婚年齢は 22.981 歳と予測される。実際の初婚年齢－予測値が残差である。

※ 第 1 版第 1 刷では、分析対象が誤って記述されていますが、問題の解答自体には影響はありません。

誤）1941～1965 年生まれ（調査時点でおよそ 35～59 歳）

正）1942～1980 年生まれ（調査時点でおよそ 20～58 歳）

2 無回答を除いた 494 人の 40 代の回答者について分析を行うと，下のような結果が出力される。最適な回帰式は $\hat{Y} = 2.238 + 0.41X_1 - 0.288X_2 - 1.082X_3 - 1.145X_4$ である。この回帰式から，テレビを見る時間は年齢が 1 歳上がるごとに 0.041 時間（約 2 分）長くなることや，女性よりも男性が 0.288 時間（約 17 分）短いことがわかるが，これらの影響は 5％水準で統計的には有意でない。統計的に有意なのは結婚状況の影響で，有配偶者や離死別者はいずれも未婚者より 1 時間ほどテレビを見る時間が短い。これらの影響は 20 代の場合とは大きく異なる。全体として 40 代の人々がテレビを見る時間を 2.3％だけ説明している（調整済み決定係数が 0.023）。説明力は低いが，統計的には有意な説明力が認められる（有意確率が 0.4％）。

ﾓﾃﾞﾙ集計

ﾓﾃﾞﾙ	R	R2 乗	調整済み R2 乗	推定値の標準誤差
1	.177[a]	.031	.023	1.815

a. 予測値: (定数)、domarry2 離死別ダミー, ageb 年齢, sex 男性ダミー, domarry1 有配偶ダミー。

分散分析[b]

ﾓﾃﾞﾙ		平方和	自由度	平均平方	F 値	有意確率
1	回帰	51.990	4	12.997	3.944	.004[a]
	残差	1611.535	489	3.296		
	全体	1663.524	493			

a. 予測値: (定数)、domarry2 離死別ダミー, ageb 年齢, sex 男性ダミー, domarry1 有配偶ダミー。

b. 従属変数: hrtv Q03 テレビ視聴時間

係数[a]

		非標準化係数		標準化係数		
ﾓﾃﾞﾙ		B	標準誤差	ﾍﾞｰﾀ	t	有意確率
1	(定数)	2.238	1.332		1.680	.094
	ageb 年齢	.041	.029	.062	1.402	.162
	sex 男性ダミー	-.288	.168	-.078	-1.722	.086
	domarry1 有配偶ダミー	-1.082	.300	-.202	-3.610	.000
	domarry2 離死別ダミー	-1.145	.465	-.137	-2.465	.014

a. 従属変数: hrtv Q03 テレビ視聴時間

3 （略）

4 （1）○　（2）○　（3）×　（4）○　（5）×

解 説　（1）就労時間数の偏回帰係数から正しい。（2）標準化回帰係数のサイズ（絶対値）を見ると，年収の標準化回帰係数（0.191）が最大なので正しい。（3）決定係数 R^2 からわかるが，0.043＝4.3％なので誤りである。修正済み決定係数（この論文では記されていない）を採用するとしても，決定係数よりもやや小さくなるので，43％になることはない。（4）働きかた変数の基準カテゴリー（参照カテゴリー）は常用雇用とあるので，この読み取りで正しい。（5）男性ダミーの偏回帰係数（−0.206）を参照し，0.206 低くなると読まなければならない。−0.109 は標準化回帰係数である。

5 （1）○　（2）×　（3）×　（4）×

解 説　（1）2.28－0.47×0＋0.04×1－0.02×25＋0.08×12＋0.19×0＝2.78 と予測値が計算できる。（2）性別の変数は男性を 1，女性を 0 とするダミー変数なので，偏回帰係数が負の値であることは，男性が女性よりも相談相手が少ないことを意味する。（3）標準化回帰係数のサイズ（絶対値）が最大なのは年齢なので，間違いである。（4）それぞれの独立変数の影響について有意確率（p 値）と＊印の両方が示されている。配偶者の有無による影響は有意確率が 57.2％（当然＊印はつかない）なので，有意ではない。

第 19 章

1 （1）$\log\left(\dfrac{\pi}{1-\pi}\right) = -4.379 + 1.611X_1 + 0.064X_2 + 1.268X_3 + 1.627X_4 + 0.090X_5$

（2）回帰係数（0.090）が正の値なので，上げる。

（3）exp（β）＝5.006 とあるので 5.006 倍（β＝1.611 から改めて計算すると誤差で 5.008 となる）。

（4）37.8％　（5）約 10.5 年　（6）性別（男性ダミー），学歴（高校ダミー，短大以上ダミー），就労年数

（7）0.96〜2.29（より厳密には 0.97〜2.28）　（8）有意　（9）大きくなる　（10）適合している

解 説　（4）$-4.379+1.611\times1+0.064\times3+1.268\times0+1.627\times1+0.090\times5=-0.499$ なので，

$$\pi=\frac{\exp(\alpha+\beta_1X_1+\cdots+\beta_5X_5)}{1+\exp(\alpha+\beta_1X_1+\cdots+\beta_5X_5)}=\frac{2.71828^{-0.499}}{1+2.71828^{-0.499}}\fallingdotseq0.378$$
（37.8％）と算出できる。

（5）$-4.379+1.611\times1+0.064\times3+1.268\times0+1.627\times1=-0.949$ なので，この属性をもつ人々の回帰式は

$\log\left(\dfrac{\pi}{1-\pi}\right)=-0.949+0.090X_5$ と考えることができる。役職者率が 50％になるのは $X_5=-\dfrac{\alpha}{\beta}$ のときなので，

$-\dfrac{\alpha}{\beta}=-\dfrac{-0.949}{0.090}\fallingdotseq10.5$ と計算できる。

（6）Wald 統計量の横の有意確率を参照すればよい。なお，高校ダミーと短大以上ダミーの上に xxlstschr という欄があり，Wald 統計量が 23.951，自由度が 2 で有意確率が 0.000 と記されているのは，高校ダミーと短大以上ダミーを合わせて学歴全体の影響を検定した結果を示している。（7）信頼区間は β を中心として±標準誤差の 2 倍（厳密には 1.96 倍）である。したがって，$1.627\pm0.333\times2$（厳密には 1.96），つまり 0.96〜2.29（厳密には 0.97〜2.28）が信頼区間である。（8）モデル全体の説明力を検定するための尤度比統計量（263.732）の有意確率は 0.000（0.1％未満）とあるので，5％水準で有意である。（9）-2 対数尤度はモデルで説明できない程度を表す。独立変数を減らした場合，当然説明できない部分が大きくなるので，-2 対数尤度は大きくなる。（10）Hosmer & Lemeshow の適合度検定は χ^2 値が 8.682 で有意確率が 0.370 なので，観察度数と期待度数は有意にはかけはなれていない。

2　無回答者を除いて 2,828 人の回答者について分析を行うと，下のような結果が出力される。男性で，年齢が高く，フルタイムやパートタイムで働いている人ほど，毎日お酒を飲む確率が高い（家族人数は有意な影響がない）という結果である。モデル全体の説明力は有意であり，擬似決定係数によると Cox & Snell の R^2 では 16.0％，Nagelkerke の R^2 では 24.2％の説明力が認められる。ただし，Hosmer & Lemeshow の適合度検定では有意に適合度が悪いことが示されており（有意確率 0.2％），実際と合致しない部分も大きいことがわかる。

モデル係数のオムニバス検定

		カイ2乗	自由度	有意確率
ステップ1	ステップ	494.645	5	.000
	ブロック	494.645	5	.000
	モデル	494.645	5	.000

モデルの要約

ステップ	-2 対数尤度	Cox & Snell R 2乗	Nagelkerke R 2乗
1	2585.542[a]	.160	.242

a. パラメータ推定値の変化が .001 未満なので, 反復回数 5 で推定が打ち切られました。

Hosmer と Lemeshow の検定

ステップ	カイ2乗	自由度	有意確率
1	24.648	8	.002

Hosmer と Lemeshow の検定の分割表

		do7drnkr 飲酒ダミー = .00 飲まない		do7drnkr 飲酒ダミー = 1.00 飲む		
		観測値	期待値	観測値	期待値	合計
ステップ1	1	256	269.698	28	14.302	284
	2	267	264.481	17	19.519	284
	3	261	260.961	22	22.039	283
	4	260	257.931	23	25.069	283
	5	257	253.424	26	29.576	283
	6	234	220.456	50	63.544	284
	7	189	178.011	92	102.989	281
	8	169	167.696	115	116.304	284
	9	139	155.951	147	130.049	286
	10	133	136.392	143	139.608	276

方程式中の変数

		B	標準誤差	Wald	自由度	有意確率	Exp (B)
ステップ1[a]	sexa(1)	1.947	.115	288.258	1	.000	7.010
	ageb	.017	.003	24.810	1	.000	1.017
	szffhere	.016	.031	.262	1	.609	1.016
	job			25.645	2	.000	
	job(1)	.663	.131	25.636	1	.000	1.941
	job(2)	.435	.186	5.449	1	.020	1.545
	定数	−3.699	.272	185.440	1	.000	.025

a. ステップ1: 投入された変数 sexa, ageb, szffhere, job

3　（略）

4 （1）× （2）○ （3）○ （4）○

解 説 （1）有意確率が 24.3％なので間違いである。（2）回帰係数が正の値なので正しい。（3）ロジスティック回帰曲線の接線の傾きを求めると，π＝0.40 のとき，と，βπ（1－π）＝0.023×0.40（1－0.40）≒0.006（0.6％）なので正しい。（4）この属性を持つ者が外国人増加に反対する確率πを求めればよい。
－0.846－0.124×1＋0.023×30－0.623×1＋0.742×0.05－0.351×0－0.837×1－0.772×0－1.283×0＝－1.7029 なので，

$$\frac{\exp(\alpha+\beta_1 X_1+\cdots+\beta_8 X_8)}{1+\exp(\alpha+\beta_1 X_1+\cdots+\beta_8 X_8)}=\frac{2.71828^{-1.7029}}{1+2.71828^{-1.7029}}\fallingdotseq 0.154$$ （15.4％）と計算できることから正しい。

5 （1）× （2）○ （3）○ （4）○

解 説 （1）Wald 統計量の横に＊印で記されている有意確率を参照すればよい。＊印がないので，5％水準でも統計的には有意でないことがわかる。（2）本人就労地位が常雇であることの回帰係数（－1.057）から $\exp(\beta)=2.71828^{-1.057}\fallingdotseq 0.35$ と求まるので，正しい。（3）「モデルカイ 2 乗」と記されているのが全体の説明力を検定するための尤度比統計量であり，横に 0.1％水準で有意であることを示す＊印がついているので，5％水準でも有意である。（4）信頼区間はβを中心として±標準誤差の 2 倍（厳密には 1.96 倍）である。S.E.が標準誤差（Standard Error）なので，0.037±0.010×2（より厳密には 1.96），つまり 0.017〜0.057 が信頼区間である。