

正 誤 表

『数学嫌いのための社会統計学〔第3版〕』第1刷（2023年4月発行）において、下記の誤記がございました。謹んでお詫び申し上げますとともに、以下のとおり訂正いたします。

法律文化社

該当箇所	正	誤
97 頁 図表 7-15 中野区・下段	<u>63.1%</u> * 下記の表のとおり。	<u>64.9%</u>

108 頁⑤「発展」問題は図表 7-15 の数値データを利用しております。この問題も下記修正版：図表 7-15 の数値データを参照してください。

修正版：図表 7-15

千代田区	中央区	港区	新宿区	文京区	台東区	墨田区	江東区	品川区	目黒区	大田区	世田谷区
31.0%	26.3%	28.6%	24.9%	32.1%	20.5%	19.7%	20.9%	24.2%	27.5%	21.4%	26.8%
70.5%	70.4%	66.4%	66.2%	72.9%	63.2%	62.2%	65.3%	64.9%	67.5%	61.9%	67.3%
渋谷区	中野区	杉並区	豊島区	北区	荒川区	板橋区	練馬区	足立区	葛飾区	江戸川区	
29.9%	24.5%	26.7%	24.5%	19.7%	19.2%	20.1%	22.8%	15.0%	16.9%	16.5%	
64.4%	63.1%	64.9%	63.1%	63.1%	62.3%	60.5%	63.5%	62.4%	61.4%	59.4%	

図表 7-15 専門・管理職の割合（上段）と学力試験の平均正答率（下段）

（次頁より「第3版 練習問題の解答」）

第3版 練習問題の解答

第1章 社会を数字で捉える

①

(i) 1920 (ii) 10 (iii) 1 (iv) 0 (v) 5

②

(i) 正しい (ii) 間違っている。正しくは、全数調査 (iii) 正しい
(iv) 間違っている。正しくは、推定統計 (v) 間違っている 標本と母集団との間には、
つねにズレがある

③

(i) 順序づけ不可能な離散変数 (ii) 連続変数 (iii) 順序づけ可能な離散変数
(iv) 順序づけ可能な離散変数 (v) 順序づけ不可能な離散変数 (vi) 連続変数

④省略

⑤省略

第2章 可能性で考える

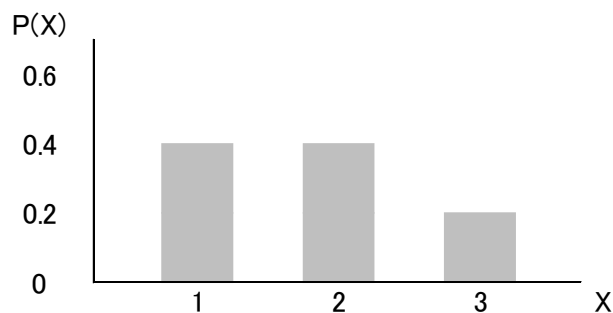
①

- (i) $X=4$ である確率は 0.24 である。
- (ii) 試験の点数が 30 点以下である確率は 0.017 である。
- (iii) I が a 以上 b 以下である確率は 0.67 である。
- (iv) 在職期間が 1 年以上 3 年未満の確率
- (v) $P(\text{在学年数} < 1 \text{ 年}) = 0.23$
- (vi) $1 - P$
- (vii) $P(\text{仮説が間違いである}) = 0.02$

②

- (i) 大吉 : 0.2、中吉 : 0.4、小吉 : 0.4
- (ii)

おみくじの幸福ポイントの確率分布



- (iii) 式 $0.2 \times 3 + 0.4 \times 2 + 0.4 \times 1 = 1.8$

解答 このおみくじを 1 回引くと、平均 1.8 ポイントが得られる。

③

離散変数の場合：縦軸は確率 横軸は確率変数の値。

連続変数の場合：縦軸は確率密度 横軸は確率変数の値。

④

- (i) ○「表」、○「表」
- (ii) 2 回連続で投げて 1 回だけ「表」が出るパターンは
 - 「表」 ●「裏」
 - 「裏」 ○「表」の 2 通りある。だから 1 回だけ「表」が出る確率は $0.50 \times 0.50 = 0.25$ と $0.50 \times 0.50 = 0.25$ を加えた 0.50

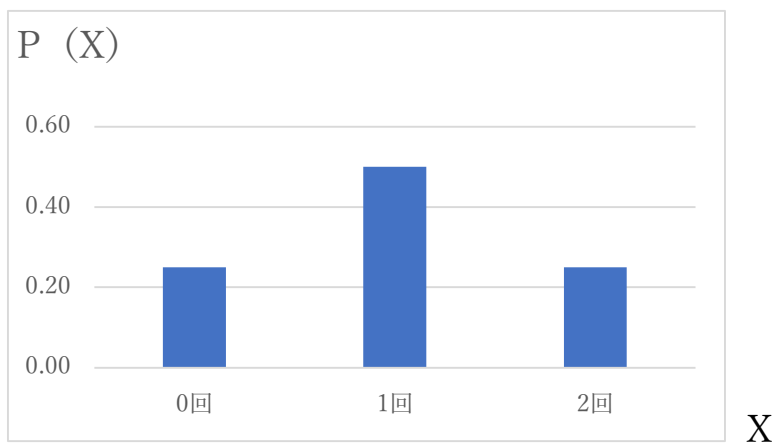
* $0.25 \times 2 \text{ 通り} = 0.50$ ということ。

(iii) 0 回「表」の確率は、 $0.25 \times 1 \text{ 通り} = 0.25$

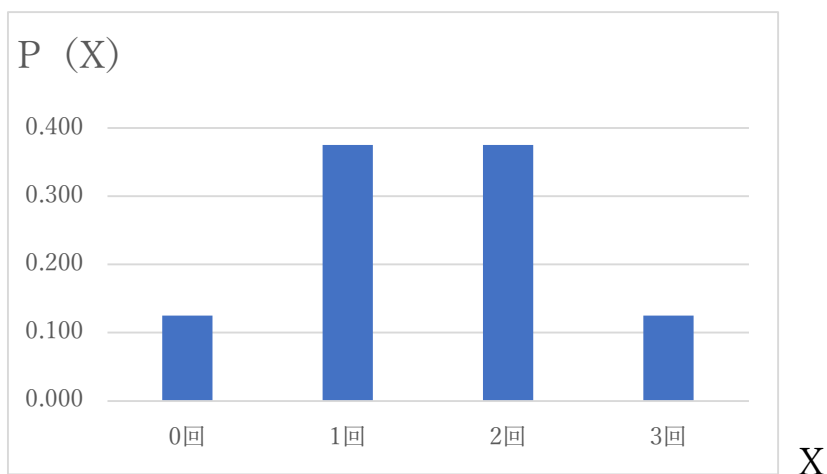
1 回「表」の確率は、 $0.25 \times 2 \text{ 通り} = 0.50$

2 回「表」の確率は、 $0.25 \times 1 \text{ 通り} = 0.25$

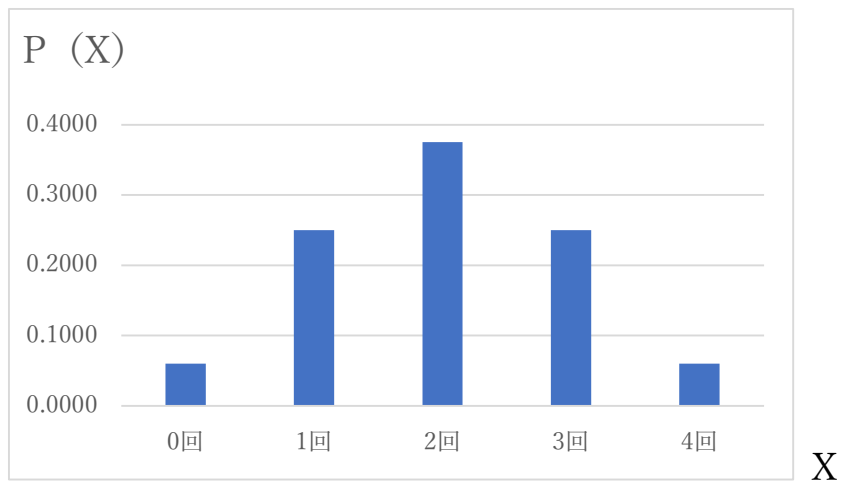
(iv)



(v)



3 回投げた場合



4 回投げた場合

⑤

四年制大学の大学生である確率：0.375

四年制大学や短期大学の学生ではない確率：0.584

第3章 ばらばらのデータを図表にまとめる

①

(i) 観測値 (ii) 度数分布 (iii) 度数分布表 (iv) 横軸 (v) 縦軸
(vi) ヒストグラム(度数分布図) (vii) 度数分布多角形

②

ヒストグラムには柱(棒)の間に隙間はないが、棒グラフには棒の間に隙間がある。ヒストグラムは連続変数または順序づけ可能な離散変数の度数分布表を図示するときを使う。例えば、図表3-3のような初任給(もとは連続変数)の度数分布表を図示するときである。一方、棒グラフは順序づけ不可能な離散変数の度数分布表を図示するときを使う。例えば、図表3-2のような就職先の業種(離散変数)の度数分布表を図示する場合には、棒グラフが使われる。

③

被疑者と被害者の関係	絶対度数	相対度数	百分率
家族	347	0.454	45.4
その他の親族	16	0.021	2.1
元配偶者	8	0.010	1.0
交際相手	83	0.108	10.8
知人・友人	106	0.139	13.9
職場関係	51	0.067	6.7
その他	59	0.077	7.7
面識なし	95	0.124	12.4
合計	765	1.000	100.0

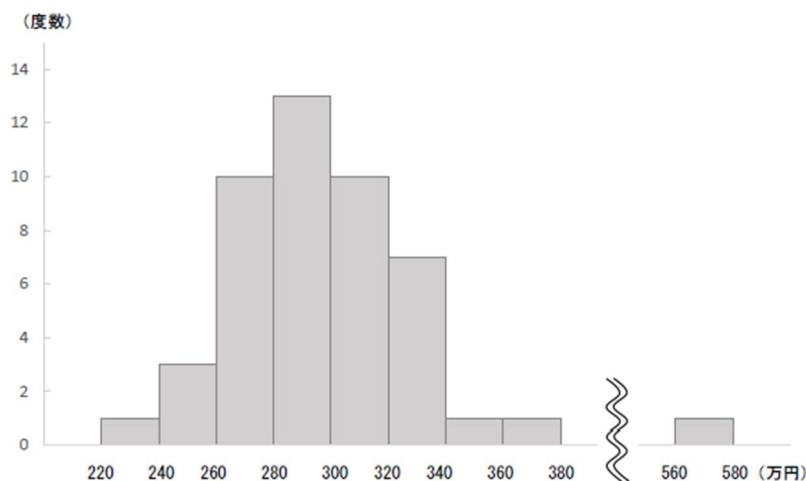
殺人の半数近くは、家族とその他の親族の間で起こっていることに注意したい。

④

(i) 度数分布表

階級番号	階級	階級値	絶対度数	相対度数	百分率
1	220 万円以上 240 万円未満	230 万円	1	0.021	2.1%
2	240 万円以上 260 万円未満	250 万円	3	0.064	6.4%
3	260 万円以上 280 万円未満	270 万円	10	0.213	21.3%
4	280 万円以上 300 万円未満	290 万円	13	0.277	27.7%
5	300 万円以上 320 万円未満	310 万円	10	0.213	21.3%
6	320 万円以上 340 万円未満	330 万円	7	0.149	14.9%
7	340 万円以上 360 万円未満	350 万円	1	0.021	2.1%
8	360 万円以上 380 万円未満	370 万円	1	0.021	2.1%
9	380 万円以上 400 万円未満	390 万円	0	0.000	0.0%
10	400 万円以上 420 万円未満	410 万円	0	0.000	0.0%
11	420 万円以上 440 万円未満	430 万円	0	0.000	0.0%
12	440 万円以上 460 万円未満	450 万円	0	0.000	0.0%
13	460 万円以上 480 万円未満	470 万円	0	0.000	0.0%
14	480 万円以上 500 万円未満	490 万円	0	0.000	0.0%
15	500 万円以上 520 万円未満	510 万円	0	0.000	0.0%
16	520 万円以上 540 万円未満	530 万円	0	0.000	0.0%
17	540 万円以上 560 万円未満	550 万円	0	0.000	0.0%
18	560 万円以上 580 万円未満	570 万円	1	0.021	2.1%
合計	—	—	47	1.000	100%

ヒストグラム



(ii)

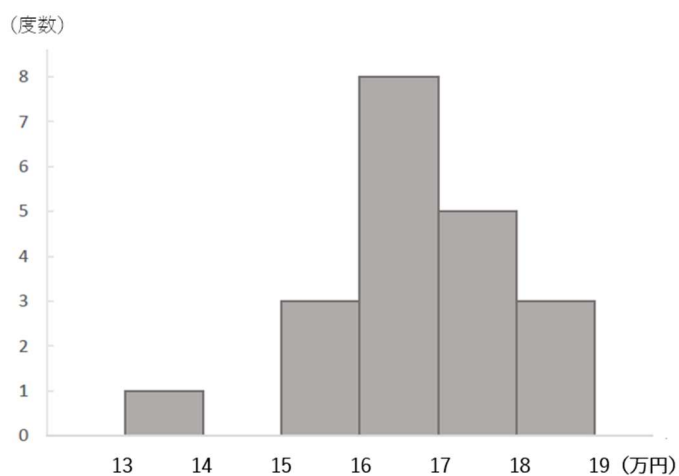
280 万円以上 300 万円未満の階級が一番多いことが見てとれる。次に多い階級は 280 万円以上 300 万円未満の階級の前後となっている。大半の都道府県の 1 人当たりの県民所得は、260 万円以上 340 万円未満の間にあることがわかる。それと同時に、東京都が極端に大きな値、他の 46 道府県から大きく外れた値をとっていることがわかる。

⑤

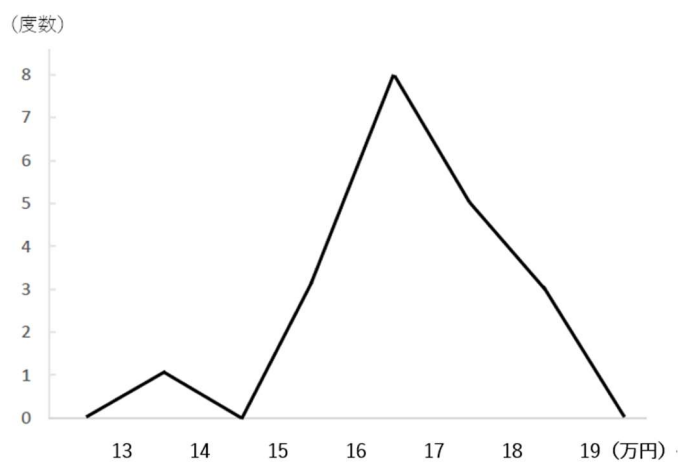
(i) 累積度数分布表

階級 番号	階級	階級値	絶対 度数	相対 度数	百分 率	累積 度数	累積 相対 度数	累積 百分 率
1	13 万円以上 14 万円未満	13 万 5 千円	1	0.05	5%	1	0.05	5%
2	14 万円以上 15 万円未満	14 万 5 千円	0	0.00	0%	1	0.05	15%
3	15 万円以上 16 万円未満	15 万 5 千円	3	0.15	15%	4	0.20	20%
4	16 万円以上 17 万円未満	16 万 5 千円	8	0.40	40%	12	0.60	60%
5	17 万円以上 18 万円未満	17 万 5 千円	5	0.25	25%	17	0.85	85%
6	18 万円以上 19 万円未満	18 万 5 千円	3	0.15	15%	20	1.00	100%
合計	—	—	20	1.00	100%	—	—	—

(ii) ヒストグラム

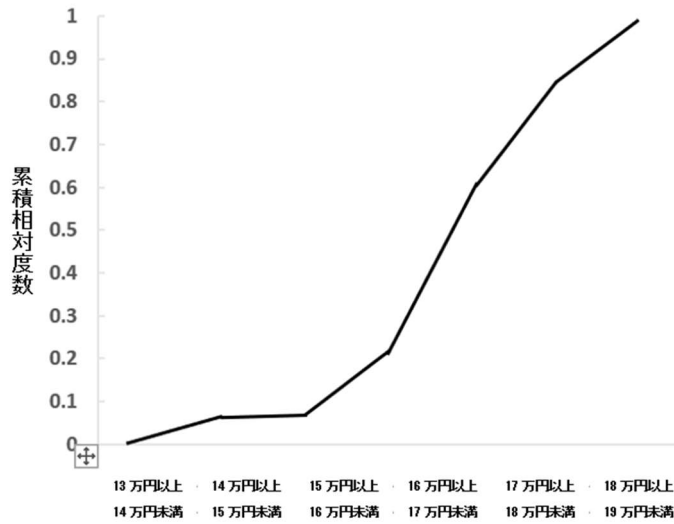


度数分布多角形



(iii)

累積相対度数分布図



(iv)

高卒新卒者（この練習問題のグラフ）は、大卒新卒者（本文中のグラフ）と比べて、初任給の低い人が多い。言い換えると、学歴が高いほど、初任給が高いということがわかる。

第4章 分布の特性を数字でつかむ

①

メディアンもモードも 290 万円で、平均値（302 万円）よりも低い。所得の平均値の分布はゆがんでいるため、平均以下の割合が半分よりも大きくなる。さらにこの場合、「はずれ値（東京の 575.7 万円）」によって平均値が引き上げられている。メディア等では「平均」だけしか報道しないことが多いため、実感とのズレを感じる人も少なくない。メディアンやモードを知れば、そのズレの意味を知ることができる。

②

メディアン 31 歳、モード 23 歳・28 歳・31 歳・39 歳、平均 32.3 歳

③

アメリカ：メディアン 2 人、モード 2 人、平均 2.2 人

中国：メディアン 2 人、モード 2 人、平均 1.8 人

④

	A 病院	B 病院
メディアン	5	16
モード	1	12
平均	6.6	17.1
レンジ	25.5	18
分散	69.0	34.1
標準偏差	8.3	5.8

A 病院の患者よりも B 病院の患者の方がメディアン・モード・平均罹病期間が長い、分散・標準偏差は小さい。つまり、A 病院の患者の罹病期間はばらついているが、B 病院の患者の多くは、病気が長期化していることが読みとれる。

⑤

・再婚者数は男女とも 1980 年から 2000 年までの 20 年間で 3 万人以上増加している。しかし、2000 年から 2020 年までの 20 年間では男女とも 2 万人弱減少している。

・メディアンは 40 年間で男性が 7 歳、女性が 6 歳上昇している。

・モードは 40 年間で男女ともに 5 歳上昇しているが、メディアンより 4～5 歳若い。

・平均は 40 年間で男性が 6 歳弱、女性が 7 歳弱上昇しており、メディアンやモードよりも高い。

・再婚年齢は 40 年間で散らばり（標準偏差）が男女ともに大きくなっている。40 年前よりも幅広い年齢層で再婚している、ということがわかる。

第5章 ふたつの離散変数を同時に扱う

①

度数クロス表

	賛成	反対	計
男	7	2	9
女	5	6	11
計	12	8	20

百分率クロス表

	賛成	反対	計
男	77.8%	22.2%	100.0%
女	45.5%	54.5%	100.0%
計	60.0%	40.4%	100.0%

②

	あてはまる	どちらかといえばあてはまる	どちらかといえばあてはまらない	あてはまらない	誤記入	無回答	計
小学校	13.9%	44.0%	32.5%	8.4%	0.1%	1.1%	100.0%
中学校	13.2%	43.2%	33.2%	9.2%	0.1%	1.1%	100.0%

③

全体で見ると「そう思う」が3割、「そう思わない」が4割だが、年齢別に見ると「そう思う」が若い世代（18～19歳、20歳代）で5割、4割を超えており、年齢が上がるにつれてその割合が低くなっている。世代間における意識の差が読み取れる。

	18・19歳	20歳代	30歳代	40歳代	50歳代	60歳代	70歳以上	計
そう思う	18	88	102	133	99	84	107	631
そう思わない	9	56	82	145	171	189	194	846
わからない	6	28	60	74	60	51	87	366
無回答	2	18	22	22	19	20	59	162
計	35	190	266	374	349	344	447	2005

*行和の度数（631, 846, 366, 162）は、この表の各行のセルの度数を合計したものである。図表5-14の総和（N=2005）と各行和の%（比率）を掛け合わせて求めた値ではない。後者の計算方法を使うと、丸め誤差のため上記の値にならない。

④

「ひきこもり」的生活開始時の属性

性別		中学生	高校生	高卒後	大学生	社会人	その他	計
	男	4 15.4%	8 30.8%	1 3.8%	9 34.6%	1 3.8%	3 11.5%	26 100.0%
	女	0 0.0%	2 16.7%	2 16.7%	2 16.7%	4 33.3%	2 16.7%	12 100.0%
	計	4 10.5%	10 26.3%	3 7.9%	11 28.9%	5 13.2%	5 13.2%	38 100.0%

「ひきこもり」的生活開始時の属性は、男性の場合、大学生が最も多く、高校生が次に多いものの、女性は、社会人が最も多いことがわかる。

⑤

図表 5-16 百分率クロス表 (%) (2008 年)

図表 5-17 百分率クロス表 (%) (2013 年)

	女性	男性	計
ドイツ	37.8	62.2	100.0
フランス	38.5	61.5	100.0
イタリア	33.2	66.8	100.0
スペイン	32.3	67.7	100.0
イギリス	34.6	65.4	100.0
スウェーデン	32.2	67.8	100.0
ハンガリー	36.3	63.7	100.0
アメリカ	42.7	57.3	100.0
カナダ	36.0	64.0	100.0
オーストラリア	36.7	63.3	100.0
タイ	23.7	76.3	100.0
韓国	9.6	90.4	100.0
日本	9.3	90.7	100.0

	女性	男性	計
ドイツ	28.8	71.2	100.0
フランス	36.2	63.8	100.0
イタリア	27.4	72.6	100.0
スペイン	30.9	69.1	100.0
イギリス	33.8	66.3	100.0
スウェーデン	37.0	63.0	100.0
ハンガリー	38.9	61.1	100.0
アメリカ	43.4	56.6	100.0
カナダ	36.4	63.6	100.0
オーストラリア	35.9	64.1	100.0
タイ	38.6	61.4	100.0
韓国	10.9	89.1	100.0
日本	11.2	88.8	100.0

5年間で女性の割合が、ドイツやフランスなどヨーロッパ諸国（スウェーデン、ハンガリーを除く）で減少しており、アメリカやカナダでは若干増加している。タイ、韓国、日本などアジア地域でも若干増加しているものの、韓国と日本は1割にとどまっている。

第6章 関連の強さをどう測る？

①

(i) 統計的独立 (ii) 相関 (iii) 因果 (iv) 独立 (v) 従属 (vi) 時間

②

(i)

	回復	非回復	計
A 群 (投与)	60.0%	40.0%	100.0%
B 群 (非投与)	45.0%	55.0%	100.0%
計	52.5%	47.5%	100.0%

(ii) $\phi = \frac{60 \times 55 - 40 \times 45}{\sqrt{100 \times 100 \times 105 \times 95}} \doteq 0.150$

(iii) $\chi^2 = 4.510$ (下表は計算過程)

セル	(1,1)	(1,2)	(2,1)	(2,2)	計
n_{ij}	60	40	45	55	200
e_{ij}	$e_{11} = \frac{100 \times 105}{200} = 52.5$	$e_{12} = \frac{100 \times 95}{200} = 47.5$	$e_{21} = \frac{100 \times 105}{200} = 52.5$	$e_{22} = \frac{100 \times 95}{200} = 47.5$	200
$n_{ij} - e_{ij}$	7.5	-7.5	-7.5	7.5	-----
$(n_{ij} - e_{ij})^2$	56.25	56.25	56.25	56.25	-----
$(n_{ij} - e_{ij})^2 / e_{ij}$	1.071	1.184	1.071	1.184	4.51

(iv) $v = \sqrt{\frac{4.510}{(2-1) \times 200}} \doteq 0.150$ (小数第4位を四捨五入)

(v) 薬を投与された人の回復率は60%で、投与されなかった人の45%を上回っており、 ϕ 係数が正の数だから、薬の効果はあったと言える。しかし0.150という係数の値を見る限り、強い効果があるとは言えない。

③

家族との食事	貧困層	非貧困層	計
月に1回超	75	771	846
月に1回以下あるいは全くない	113	556	669
計	188	1327	1515

$$0 = \frac{75 \times 556 - 771 \times 113}{\sqrt{(75 + 771)(113 + 556)(75 + 113)(771 + 556)}} = \frac{41700 - 87123}{\sqrt{846 \times 669 \times 188 \times 1327}} = \frac{-45423}{\sqrt{141196929624}}$$

$$= \frac{-45423}{375761.799048} = -0.12088243167$$

小数点以下第 3 位までとすると、-0.121

④

(i)

	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
20歳代	61.7%	25.5%	8.7%	4.1%	100.0%
30歳代	58.5%	27.9%	9.8%	3.8%	100.0%
40歳代	56.2%	25.4%	12.0%	6.3%	100.0%
50歳代	46.5%	25.0%	16.5%	11.9%	100.0%
60歳代	35.7%	22.9%	23.5%	17.9%	100.0%
70歳代以上	24.6%	19.4%	26.5%	29.5%	100.0%
計	45.9%	24.3%	16.9%	12.8%	100.0%

$$(ii) v = \sqrt{\frac{\chi^2}{(\min(k,l)-1)N}} = \sqrt{\frac{402.42}{(4-1)3394}} = \sqrt{\frac{402.42}{10182}} = \sqrt{0.03952 \dots \dots} = 0.1987 \dots \dots \div 0.199 \text{ (小数第 4 位}$$

を四捨五入)

(iii) クラメールの連関係数を見る限り、年齢と結婚に関する意見は関連していると言えるが、関連の度合いは強いとは言えない。

⑤

$$(i) v = \sqrt{\frac{\chi^2}{(\min(k,l)-1)N}} = \sqrt{\frac{90.79}{(4-1)1966}} = \sqrt{\frac{90.79}{5898}} = \sqrt{0.01539 \dots \dots} = 0.1240 \dots \dots \div 0.124 \text{ (小数第 4 位を}$$

四捨五入)

(ii) クラメールの連関係数は 0.124 なので、「年齢と自然の恩恵や食の生産活動への感謝の間には関係がある」と言える。ただし、その関係はあまり強くない。

第7章 連続変数同士の関連を分析する（その1）

①

（i）正の関連 （ii）正の関連

②

（i）解答例：負の関連が予想される。気温が低いと外出時間が減るため、その分、テレビ視聴時間は増えるかもしれない。

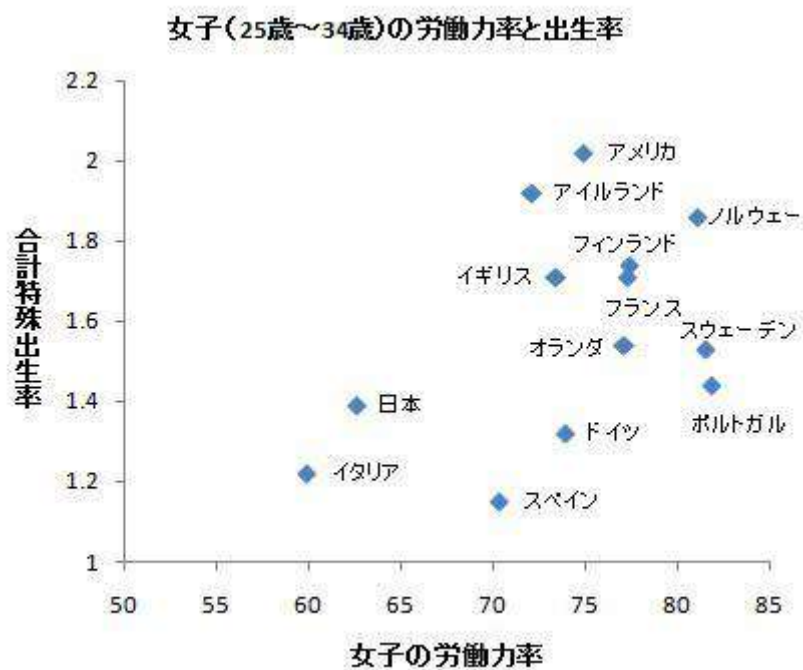
（ii）解答例：負の関連が予想される。いまのところ高齢者におけるインターネット普及率は低いから。

③

（i）アメリカ （ii）データの個数 （iii）日本 （iv）標準偏差

④

（i）



（ii）

$$(X_7 - \bar{X}) = 77.1 - 74.1 = 3 \quad (X_7 - \bar{X})^2 = 3^2 = 9 \quad (Y_7 - \bar{Y}) = 1.54 - 1.58 = -0.04$$

$$(Y_7 - \bar{Y}) = (-0.04)^2 = 0.0016 \quad (X_7 - \bar{X})(Y_7 - \bar{Y}) = 3 \times (-0.04) = -0.12$$

（iii）

$$s_{XY} = \frac{9.89}{13-1} = \frac{9.89}{12} \doteq 0.82$$

(iv)

$$s_X = \sqrt{s_X^2} = \sqrt{\frac{546.6}{13-1}} = \sqrt{\frac{546.6}{12}} = \sqrt{45.55} \doteq 6.7$$

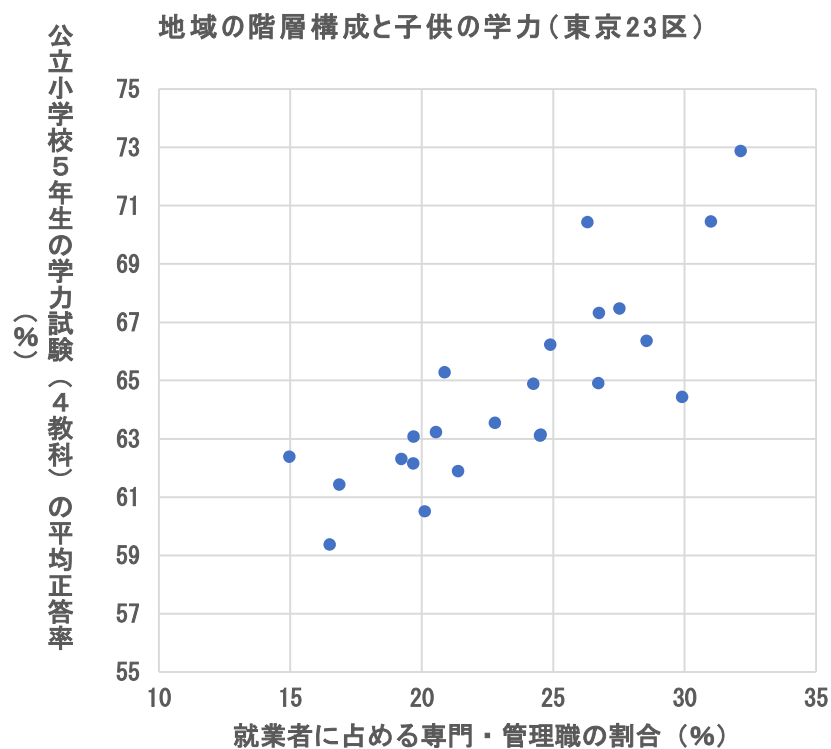
(v)

$$s_Y = \sqrt{s_Y^2} = \sqrt{\frac{0.890}{13-1}} = \sqrt{\frac{0.890}{12}} = \sqrt{0.074} \doteq 0.27$$

(vi)

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} = \frac{0.82}{6.7 \times 0.27} = \frac{0.82}{1.809} \doteq 0.45$$

⑤ (i) 解答例



(ii) 正の関連

(iii) 解答例： 一般に、専門・管理職の就業者には高学歴の人が多く、また高学歴の親をもつ子どもほど学力が高くなりやすいことが知られている。よって、専門・管理職就業者の比率が高い地域ほど、子どもの学力の平均値も高くなるものと考えられる。

(ただし、この説明は一例にすぎず、他にも様々な説明がありうるだろう。)

第 8 章 連続変数同士の関連を分析する (その 2)

①

(i) 正 (ii) 回帰係数 (iii) 上がり (iv) 負 (v) 回帰係数 (vi) 下がり

②

親世代の大卒比率が 20% である都道府県の進学率 (%) の予測

$$\hat{Y} = 20.8 + 1.34 \times 20 = 20.8 + 26.8 = 47.6 \quad \therefore \text{進学率} = 47.6\%$$

親世代の大卒比率が 25% である都道府県の進学率 (%) の予測

$$\hat{Y} = 20.8 + 1.34 \times 25 = 20.8 + 33.5 = 54.3 \quad \therefore \text{進学率} = 54.3\%$$

③

(i) 回帰式を $\hat{Y} = a + bx$ とすると

$$b = \frac{s_{XY}}{s_X^2} = \frac{225}{187.5} = 1.2 \quad a = \bar{Y} - b\bar{X} = 55 - (1.2 \times 65) = -23$$

$$\therefore \hat{Y} = -23 + 1.2X$$

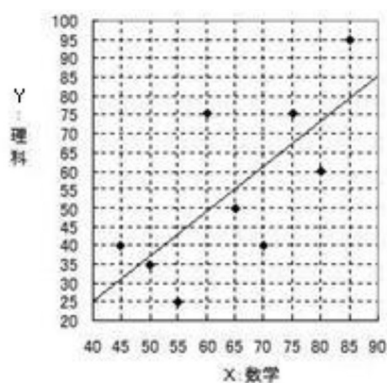
(ii) 決定係数は相関係数 (第 7 章参照) の 2 乗であり、分散は標準偏差の 2 乗 (第 4 章参照) であることから、以下のように計算できる。

$$R^2 = (r_{XY})^2 = \left(\frac{s_{XY}}{s_X s_Y} \right)^2 = \frac{s_{XY}^2}{s_X^2 s_Y^2} = \frac{225^2}{187.5 \times 525} = \frac{50625}{98437.5} = 0.5142 \dots \approx 0.514$$

(iii) 数学が 40 点の人の理科の点数は、 $\hat{Y} = -23 + 1.2 \times 40 = 25$ より、25 点と予測される。

数学が 90 点の人の理科の点数は、 $\hat{Y} = -23 + 1.2 \times 90 = 85$ より、85 点と予測される。

(iv)



④

(i) 回帰式を $\hat{Y} = a + bx$ とすると、

$$b = \frac{s_{XY}}{s_X^2} = \frac{81.25}{93.75} = \frac{8125}{9375} = \frac{13}{15}$$

$$a = \bar{Y} - b\bar{X} = 80 - \left(75 \times \frac{13}{15}\right) = 80 - 65 = 15$$

$$\therefore \hat{Y} = 15 + \frac{13}{15}X$$

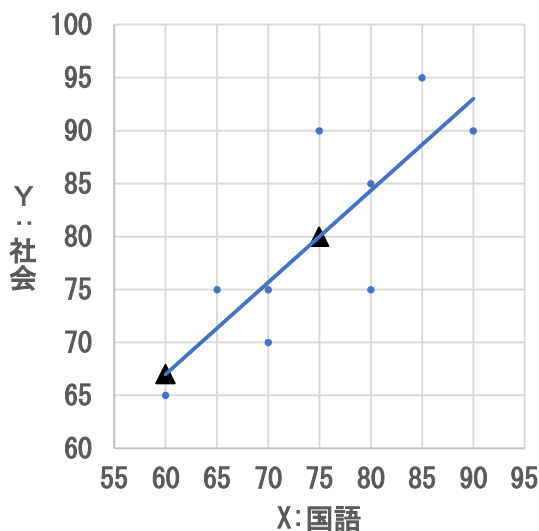
(ii) 決定係数は相関係数（第7章参照）の2乗であり、分散は標準偏差の2乗（第4章参照）であることから、以下のように計算できる。

$$R^2 = (r_{XY})^2 = \left(\frac{s_{XY}}{s_X s_Y}\right)^2 = \frac{s_{XY}^2}{s_X^2 s_Y^2} = \frac{81.25^2}{93.75 \times 106.25} = \frac{6601.5625}{9960.9375} = 0.6627 \dots \approx 0.663$$

(iii) 国語が60点の人の社会の点数は、 $\hat{Y} = 15 + \frac{13}{15} \times 60 = 15 + 42 = 67$ より、67点と予測される。

国語が75点の人の社会の点数は、 $\hat{Y} = 15 + \frac{13}{15} \times 75 = 15 + 65 = 80$ より、80点と予測される。

(iv)



⑤

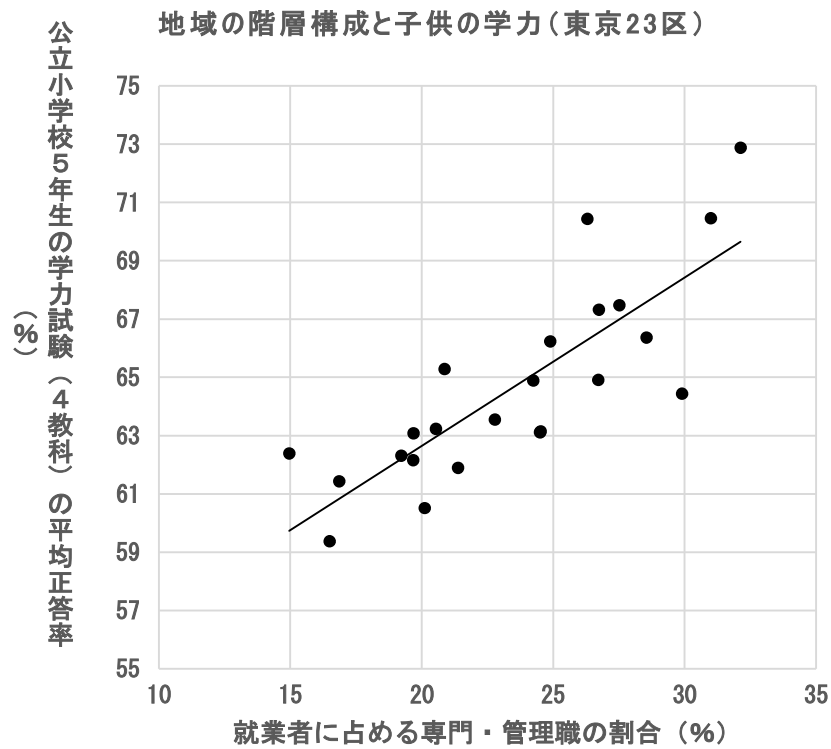
(i) 回帰式を $\hat{Y} = a + bx$ と表すと第8章の式①より、

$$b = \frac{s_{XY}}{s_X^2} = \frac{12.9}{22.3} = 0.5784 \dots \approx 0.578$$

$$\text{同じく式②より、} a = \bar{Y} - b\bar{X} = 64.6 - (0.578 \times 23.5) = 64.6 - 13.583 = 51.017 \approx 51.0$$

$$\therefore \hat{Y} = 51.0 + 0.578X$$

(ii)



(iii) 決定係数は相関係数（第7章参照）の2乗であり、分散は標準偏差の2乗（第4章参照）であることから、以下のように計算できる。

$$R^2 = (r_{XY})^2 = \left(\frac{s_{XY}}{s_X s_Y}\right)^2 = \frac{s_{XY}^2}{s_X^2 s_Y^2} = \frac{12.9^2}{22.3 \times 11.2} = \frac{166.41}{249.76} = 0.6662 \dots \approx 0.666$$

第8章2-3の目安によれば、あてはまりは「よい」。

第9章 見えない関係を探る

①

(i) ○ (ii) × (iii) × (iv) ○ (v) × (vi) ○

②

(i)

	優	良	可	計
法学部	20.0%	50.0%	30.0%	100.0%
経済学部	10.0%	30.0%	60.0%	100.0%
文学部	40.0%	40.0%	20.0%	100.0%
計	20.8%	39.2%	40.0%	100.0%

(ii) まず、期待度数を以下のように計算する。

	優	良	可	計
法学部	$e_{11} = \frac{200 \times 125}{600} = 41.667$	$e_{12} = \frac{200 \times 235}{600} = 78.333$	$e_{13} = \frac{200 \times 240}{600} = 80$	200
経済学部	$e_{21} = \frac{250 \times 125}{600} = 52.083$	$e_{22} = \frac{250 \times 235}{600} = 97.917$	$e_{23} = \frac{250 \times 240}{600} = 100$	250
文学部	$e_{31} = \frac{150 \times 125}{600} = 31.250$	$e_{32} = \frac{150 \times 235}{600} = 58.750$	$e_{33} = \frac{150 \times 240}{600} = 60$	150
計	125	235	240	600

つぎに、 χ^2 値を以下のように計算する。

	n_{11}	n_{12}	n_{13}	n_{21}	n_{22}	n_{23}	n_{31}	n_{32}	n_{33}	計
n_{ij}	40	100	60	25	75	150	60	60	30	600
e_{ij}	41.667	78.333	80	52.083	97.917	100	31.25	58.75	60	600
$n_{ij} - e_{ij}$	-1.667	21.667	-20	-27.083	-22.917	50	28.75	1.25	-30	-----
$(n_{ij} - e_{ij})^2$	2.779	469.459	400	733.489	525.189	2500	826.563	1.563	900	-----
$(n_{ij} - e_{ij})^2 / e_{ij}$	0.067	5.993	5	14.083	5.364	25	26.45	0.027	15	96.983

χ^2 値は 96.983 であったので、その数値を用いて、クラメールの連関係数を以下の式より求めると、0.284 となる。

$$V = \sqrt{\frac{96.983}{(3-1) \times 600}} = 0.2842 \dots \dots \approx 0.284$$

(iii) クラメールの連関係数が 0.284 であるので、学部と成績にはある程度の関連があると考えられる。すなわち、文学部の学生は、法学部や経済学部の学生より成績がよい傾向

向がある。

- (iv) 文学部は法学部や経済学部より女子学生の比率が高い。女子学生は男子学生より一般的に大学の学業に真面目で、成績がよいかもしれない。したがって、学部間での成績の違いは疑似関係で、それは主に男子学生と女子学生の違いである可能性がある。

③

- (i) まず下のような百分率クロス表を作成する。

性別にみた性別役割分業意識と性道德への態度（百分率クロス表 その1）

	男性		計	女性		計
	よくない	かまわない		よくない	かまわない	
賛成	48.9%	51.1%	100.0%	60.5%	39.5%	100.0%
反対	65.9%	34.1%	100.0%	76.4%	23.6%	100.0%
計	52.3%	47.7%	100.0%	63.7%	36.3%	100.0%

性別にみた性別役割分業意識と性道德への態度（百分率クロス表 その2）

	男性		計	女性		計
	よくない	かまわない		よくない	かまわない	
賛成	74.8%	85.7%	80.0%	75.6%	86.7%	79.6%
反対	25.2%	14.3%	20.0%	24.4%	13.3%	20.4%
計	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

百分率クロス表（その1）で、女性の家庭外就労に「賛成」の人は「反対」の人に比べ、既婚女性の婚外性交渉を「かまわない」とする人の割合が高い。つまり性別役割分業意識の弱い人ほど性道德に対して柔軟な態度を示す。この傾向は男性にも女性にも見られる。また女性に比べ男性の方が既婚女性の婚外性交渉を「かまわない」とする人の割合が高い。つまり男性の方が性道德に対して柔軟な態度を示す。また百分率クロス表（その2）の「計」を見ると、女性の家庭外就労に賛成する人の割合は、男女間で大差が無い。つまり、性別と性別役割分業意識には関連性が見られない。

- (ii) 男性 $\phi = \frac{86 \times 15 - 90 \times 29}{\sqrt{176 \times 44 \times 115 \times 105}} \doteq -0.137$ (小数第4位を四捨五入)

$$\text{女性 } \phi = \frac{130 \times 13 - 85 \times 42}{\sqrt{215 \times 55 \times 172 \times 98}} \doteq 0.133 \quad (\text{小数第4位を四捨五入})$$

- (iii) 性別役割分業意識の弱い人ほど、性道德に対して柔軟な態度をとる傾向があり、この傾向は、男性にも女性にもほぼ同程度の強さで見られる。つまり性別による交互作用効果は見られない。

④

(i)

人口 20 万人未満

	満足+やや満足	やや不満+不満	計
18～29歳	76.4%	23.6%	100.0%
30～69歳	76.2%	23.8%	100.0%
70歳以上	86.5%	13.5%	100.0%
計	79.1%	20.9%	100.0%

人口 20 万人以上

	満足+やや満足	やや不満+不満	計
18～29歳	88.5%	11.5%	100.0%
30～69歳	86.4%	13.6%	100.0%
70歳以上	89.5%	10.5%	100.0%
計	87.3%	12.7%	100.0%

(ii)

まず人口 20 万人未満の都市のクロス表について下のように χ^2 値を求める。

	満足+やや満足	やや不満+不満	計
18～29 歳	$e_{11} = \frac{271 \times 2142}{2708} = 214.358$	$e_{12} = \frac{271 \times 566}{2708} = 56.642$	271
30～69 歳	$e_{21} = \frac{1669 \times 2142}{2708} = 1320.162$	$e_{22} = \frac{1669 \times 566}{2708} = 348.838$	1669
70 歳以上	$e_{31} = \frac{768 \times 2142}{2708} = 607.480$	$e_{32} = \frac{768 \times 566}{2708} = 160.520$	768
計	2142	566	2708

	n_{11}	n_{12}	n_{21}	n_{22}	n_{31}	n_{32}	計
n_{ij}	207	64	1271	398	664	104	2708
e_{ij}	214.358	56.642	1320.162	348.838	607.480	160.520	2708
$n_{ij} - e_{ij}$	-7.358	7.358	-49.162	49.162	56.520	-56.520	-----
$(n_{ij} - e_{ij})^2$	54.143	54.143	2416.877	2416.877	3194.504	3194.504	-----
$(n_{ij} - e_{ij})^2 / e_{ij}$	0.253	0.956	1.831	6.928	5.259	19.901	35.127

χ^2 値 = 35.127 から、クラメールの連関係数を以下の式より求めると、0.114 となる。

$$V = \sqrt{\frac{35.127}{(2-1) \times 2708}} = \sqrt{0.01297 \dots \dots} = 0.1138 \dots \dots \approx 0.114$$

次に人口 20 万人以上の都市のクロス表について下のように χ^2 値を求める。

	満足+やや満足	やや不満+不満	計
18～29 歳	$e_{11} = \frac{122 \times 994}{1138} = 106.562$	$e_{12} = \frac{122 \times 144}{1138} = 15.438$	122
30～69 歳	$e_{21} = \frac{741 \times 994}{1138} = 647.236$	$e_{22} = \frac{741 \times 144}{1138} = 93.764$	741
70 歳以上	$e_{31} = \frac{275 \times 994}{1138} = 240.202$	$e_{32} = \frac{275 \times 144}{1138} = 34.789$	275
計	994	144	1138

	n_{11}	n_{12}	n_{21}	n_{22}	n_{31}	n_{32}	計
n_{ij}	108	14	640	101	246	29	1138
e_{ij}	106.562	15.438	647.236	93.764	240.202	34.798	1138
$n_{ij}-e_{ij}$	1.438	-1.438	-7.236	7.236	5.798	-5.798	-----
$(n_{ij}-e_{ij})^2$	2.067	2.067	52.352	52.352	33.616	33.616	-----
$(n_{ij}-e_{ij})^2/e_{ij}$	0.019	0.134	0.081	0.558	0.140	0.966	1.898

χ^2 値=1.898 から、クラメールの連関係数を以下の式より求めると、0.041 となる。

$$V = \sqrt{\frac{1.898}{(2-1) \times 1138}} = \sqrt{0.0016 \dots \dots} = 0.0408 \dots \dots = 0.041$$

(iii)

人口 20 万人未満の地域では、年齢層と満足度との間に関連がみられる。すなわち 70 歳以上の満足度がその他の年齢層よりも高い。一方、人口 20 万人以上の地域では、年齢層と満足度との関連はほとんどみられない。このように図表 9-13 では、都市規模と年齢層の組み合わせが、地域での暮らしの満足度に影響を与えている。つまり、ここでは交互作用効果がみられる。

⑤

(i)

$$r_{XY.T} = \frac{r_{XY} - r_{TX}r_{TY}}{\sqrt{1-r_{TX}^2}\sqrt{1-r_{TY}^2}} = \frac{-0.340 - (-0.504 \times 0.681)}{\sqrt{1-(-0.504)^2}\sqrt{1-0.681^2}} = \frac{-0.340 + 0.343224}{0.8637036 \times 0.7322834} \cong 0.005$$

(ii)

虫歯経験率Xと視力1.0未満の児童比率Yは $r_{XY}=-0.340$ の弱い相関を示す。しかし、社会階層Tを第3変数としたときの、虫歯経験率と視力1.0未満の児童比率との偏相関係数

($r_{XY.T}$)は0.005とゼロに近い。よって、虫歯経験率と視力1.0未満の児童比率との関係は、社会階層という第3変数が作り出した「擬似関係」だと解釈しうる。つまり、社会階層の高い人々の住む地域ほど歯の悪い子が少なく($r_{XY}<0$)、目の悪いが多い($r_{XY}>0$)ため、結果的に歯の悪い子の多い地域ほど目の悪い子が少ない($r_{XY}=-0.340$)という関連性が現れていたと考えられる。

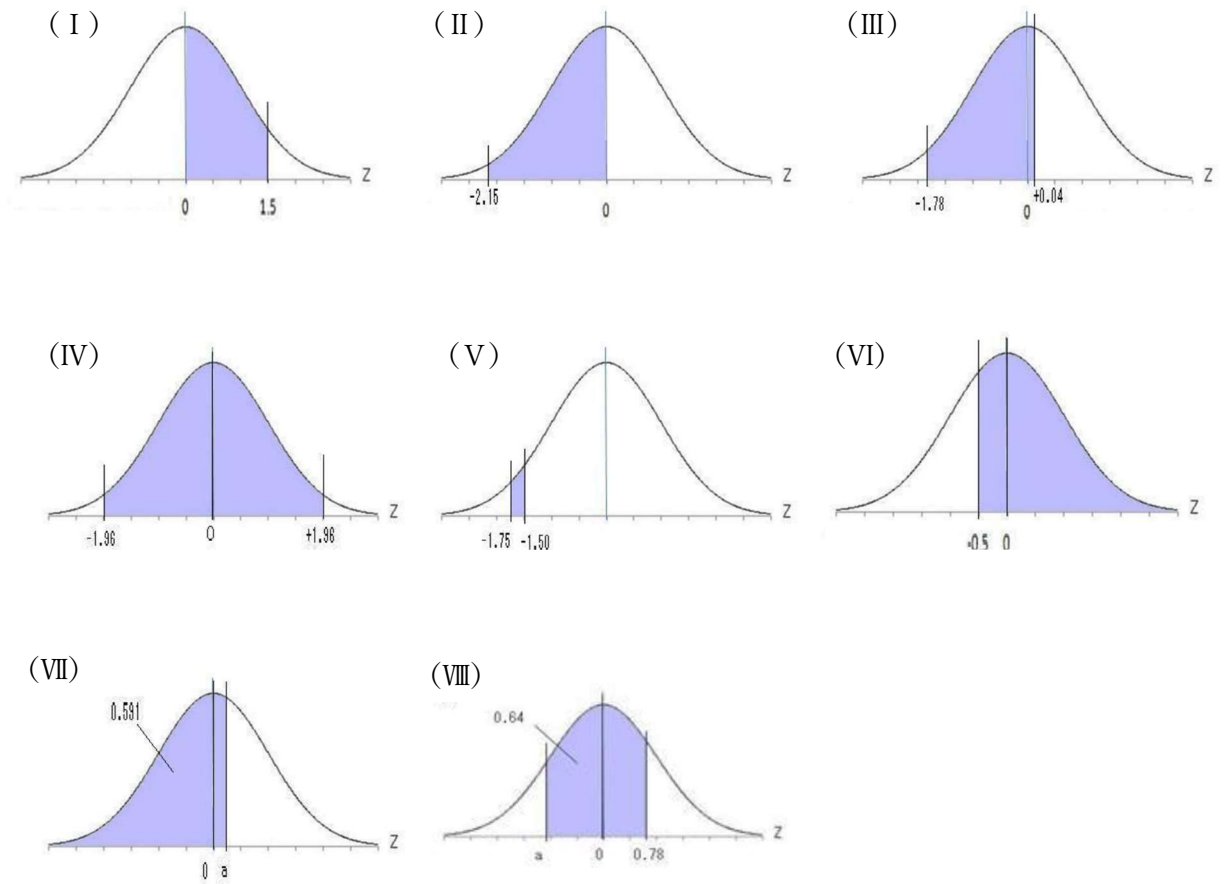
第10章 全体のなかでの位置を把握する

- ① 正解は (ii)、(iii)、(v)。ちなみに、(i) は高得点に、(iv) は低所得に人が集まる。したがって、(i) は左側に裾(すそ)が伸びた分布(左側にゆがんだ分布)、(iv) は右側に裾が伸びた分布(右側にゆがんだ分布)となる。

②

(i) 標本統計 (ii) 標本統計 (iii) 母数 (iv) 標本統計 (v) 母数

③



- (i) $P(0 \leq z \leq +1.50) = 0.4332$
(ii) $P(-2.15 \leq z \leq 0) = P(0 \leq z \leq 2.15) = 0.4842$
(iii) $P(-1.78 \leq z \leq +0.04) = P(-1.78 \leq z \leq 0) + P(0 \leq z \leq 0.04)$
 $= P(0 \leq z \leq 1.78) + P(0 \leq z \leq 0.04) = 0.4625 + 0.0160 = 0.4785$
(iv) $P(-1.96 \leq z \leq 1.96) = P(0 \leq z \leq 1.96) \times 2 = 0.4750 \times 2 = 0.950$
(v) $P(-1.75 \leq z \leq -1.50) = P(+1.50 \leq z \leq +1.75)$

$$=P(0 \leq z \leq 1.75) - P(0 \leq z \leq 1.50) = 0.4599 - 0.4332 = 0.0267$$

$$\begin{aligned} \text{(vi)} \quad P(-0.5 \leq z) &= P(-0.5 \leq z \leq 0) + P(0 \leq z) = P(0 \leq z \leq 0.5) + P(0 \leq z) \\ &= 0.1915 + 0.5 = 0.6915 \end{aligned}$$

$$\begin{aligned} \text{(vii)} \quad P(z \leq a) &= P(z \leq 0) + P(0 \leq z \leq a) = 0.5 + P(0 \leq z \leq a) = 0.591 \\ \therefore P(0 \leq z \leq a) &= 0.591 - 0.5 = 0.091 \quad \therefore a = 0.23 \end{aligned}$$

$$\begin{aligned} \text{(viii)} \quad P(a \leq z \leq 0.78) &= P(a \leq z \leq 0) + P(0 \leq z \leq 0.78) = P(a \leq z \leq 0) + 0.2823 = 0.64 \\ \therefore P(a \leq z \leq 0) &= 0.64 - 0.2833 = 0.3577 \quad \therefore a = -1.07 \end{aligned}$$

④

(i)

$$\text{フツオ君の国語の標準得点} = (50 - 65.8) \div 7.5 = -15.8 \div 7.5 \div 2.11$$

$$\text{数学の標準得点} = (50 - 62.2) \div 14.3 = -12.2 \div 14.3 \div 0.85$$

(いずれも小数第3位を四捨五入)

国語の標準得点 < 数学の標準得点 $\therefore$ 数学の方が良かった。

(ii)

国語の上位14%が特進クラスということは、平均点（上位・下位50%の点）から分断点までの間には $(50\% - 14\%) = 36\%$ の人が入るとわかる。そこで正規分布表の中で0.36という割合を探すと、最も近いのは 0.3599であり、そのとき $z = 1.08$ だとわかる。すなわち分断点は国語の平均点 + 国語の標準偏差 $\times 1.08$ だから、 $65.8 + (7.5 \times 1.08) = 73.9$ となる。 $\therefore 73.9$ 点が分断点である。

⑤

(i)

ヒントより偏差値 $= 10 \times z + 50$ だから、偏差値 70 の場合、 $70 = 10 \times z + 50$ という式が成り立つ。よって $z = (70 - 50) \div 10 = 2$ となるから、偏差値 70 の人の標準得点 z は 2 だとわかる。そこで正規分布表より、 $P(0 \leq z \leq 2) = 0.4772$ であるから、偏差値が 70 よりも上の人の割合は、 $0.5 - 0.4772 = 0.0228$ となって、全体の 2.28 % だとわかる。 $\therefore$ 2.28

(ii)

ニガテくんは偏差値 42 だから、 $42 = 10 \times z + 50$ が成り立つ。よって $z = (42 - 50) \div 10 = -0.8$ となる。ニガテくんは平均点よりも標準偏差 $\times 0.8$ 分だけ下の点をとったことになる。よって $68.0 - 7.5 \times 0.8 = 62$ となる。 $\therefore$ ニガテくんの点は 62 点だった。

第11章 一部から全体を推し量る（その1）

①

- (i) 偽 (N=400 は標本の大きさである) (ii) 真
 (iii) 偽 (母数と標本統計の間には誤差〔ズレ〕がある) (iv) 真

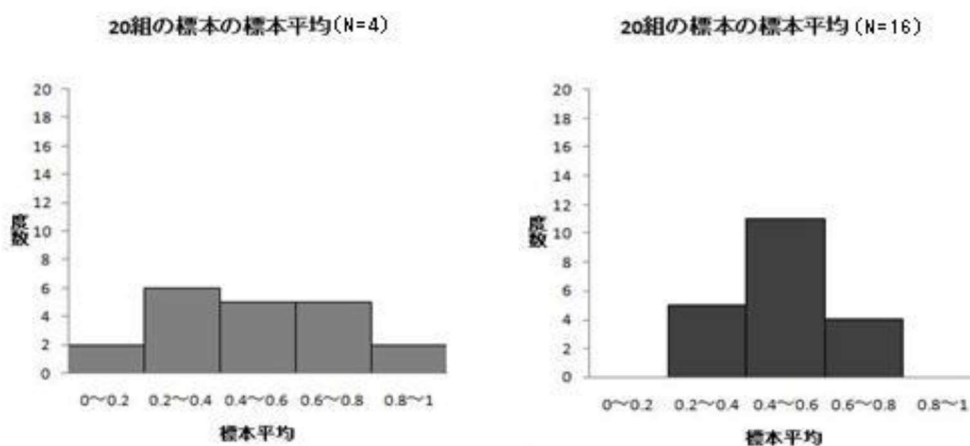
②

$\bar{\bar{X}}$ は標本平均、 $\bar{X}$ は標本平均 ($\bar{X}$) の平均である。

sは標本標準偏差、 $s_{\bar{X}}$ は標本平均 ($\bar{X}$) の標準偏差である。

③

- (i) (ii) (iii)



- (iv) 標本の大きさが4から16になると、標本平均のばらつきは小さくなり、理論的な平均値「0.5」を中心とした左右対称の山型の分布、すなわち正規分布に近づいた。

④

(i) N=4 のとき、標準誤差 $SE = \frac{\sigma}{\sqrt{N}} = \frac{15}{\sqrt{4}} = 7.5$

(ii) N=8 のとき、標準誤差 $SE = \frac{\sigma}{\sqrt{N}} = \frac{15}{\sqrt{8}} = 5.303 \dots = 5.30$

(iii) N=16 のとき、標準誤差 $SE = \frac{\sigma}{\sqrt{N}} = \frac{15}{\sqrt{16}} = 3.75$

⑤ 標本平均の平均 ($\bar{\bar{X}}$) は 170.0cm、標本平均の標準偏差は 1.9cm である。標本平均の標

標準偏差 $s_{\bar{x}}$ は $\frac{\sigma}{\sqrt{N}}$ に近づく。標本平均の標準偏差は標準誤差 SE と呼ばれ、

$$SE = \frac{\sigma}{\sqrt{N}} = \frac{5.7}{\sqrt{9}} = \frac{5.7}{3} = 1.9 \text{ と計算される。}$$

第12章 一部から全体を推し量る（その2）

①

ニガテくんは、区間推定の意味を誤解している。95%の女子高生の体重が52.79kgから53.01kgの間に収まるということではない。女子高生の真の平均体重（母平均 μ ）は95%の確からしさに52.79kg以上53.01kg以下に含まれる、ということを意味する。

②

母標準偏差 σ がわからないので、区間推定には次の公式を用いる。 $\bar{X} - Z \frac{s}{\sqrt{N}} \leq \mu \leq \bar{X} + Z \frac{s}{\sqrt{N}}$
信頼度は95%なので、Z 値は1.96 を用いる。公式に値を代入する。

<ふたり親>

$$626 - 1.96 \frac{335}{\sqrt{1162}} \leq \mu \leq 626 + 1.96 \frac{335}{\sqrt{1162}} \quad 626 - 19.261 \dots \leq \mu \leq 626 + 19.261 \dots$$
$$606.738 \dots \leq \mu \leq 645.261 \dots \quad \underline{606.7 \text{ 万円以上 } 645.3 \text{ 万円以下}}$$

<母子家庭>

$$294 - 1.96 \frac{224}{\sqrt{493}} \leq \mu \leq 294 + 1.96 \frac{224}{\sqrt{493}} \quad 294 - 19.773 \dots \leq \mu \leq 294 + 19.773 \dots$$
$$274.226 \dots \leq \mu \leq 313.773 \dots \quad \underline{274.2 \text{ 万円以上 } 313.8 \text{ 万円以下}}$$

<父子家庭>

$$550 - 1.96 \frac{261}{\sqrt{71}} \leq \mu \leq 550 + 1.96 \frac{261}{\sqrt{71}} \quad 550 - 60.711 \dots \leq \mu \leq 550 + 60.711 \dots$$
$$489.289 \dots \leq \mu \leq 610.711 \dots \quad \underline{489.3 \text{ 万円以上 } 610.7 \text{ 万円以下}}$$

③

男性の場合も女性の場合も、 $N=25 < 30$ なので、以下の小標本の場合の公式を用いる。

$$\bar{X} - t \frac{s}{\sqrt{N}} \leq \mu \leq \bar{X} + t \frac{s}{\sqrt{N}} \quad *s \text{ は標本標準偏差}$$

（男性の場合） $\bar{X}=2.89$ 、 $N=25$ 、 $s=0.64$ 、 $t=2.064$ （t分布表の自由度24、信頼度95%の欄より）を公式に代入する

$$2.89 - 2.064 \frac{0.64}{\sqrt{25}} \leq \mu \leq 2.89 + 2.064 \frac{0.64}{\sqrt{25}}$$

$$2.63 \leq \mu \leq 3.15 \text{ (小数第3位を四捨五入)}$$

(女性の場合) $\bar{X}=3.29$ 、 $N=25$ 、 $s=0.39$ 、 $t=2.064$ (t分布表の自由度24、信頼度95%の欄より) を公式に代入する

$$3.29 - 2.064 \frac{0.39}{\sqrt{25}} \leq \mu \leq 3.29 + 2.064 \frac{0.39}{\sqrt{25}}$$

$$3.13 \leq \mu \leq 3.45 \text{ (小数第3位を四捨五入)}$$

④

信頼度95% ($\alpha=0.05$) の場合、 $z=1.96$ 、 $p=0.59$ 、 $N=1892$ を以下の公式に代入する。

$$p - z \sqrt{\frac{p(1-p)}{N}} \leq P \leq p + z \sqrt{\frac{p(1-p)}{N}}$$

$$0.59 - 1.96 \sqrt{\frac{0.59(1-0.59)}{1892}} \leq P \leq 0.59 + 1.96 \sqrt{\frac{0.59(1-0.59)}{1892}}$$

$$0.568 \leq P \leq 0.612 \text{ (小数第4位を四捨五入)}$$

信頼度99% ($\alpha=0.01$) の場合、 $z=2.58$ 、 $p=0.59$ 、 $N=1892$ を公式に代入する。

$$0.59 - 2.58 \sqrt{\frac{0.59(1-0.59)}{1892}} \leq P \leq 0.59 + 2.58 \sqrt{\frac{0.59(1-0.59)}{1892}}$$

$$0.561 \leq P \leq 0.619 \text{ (小数第4位を四捨五入)}$$

⑤

比率の区間推定の公式 $z \sqrt{\frac{p(1-p)}{N}}$ の部分が誤差を意味することを利用して、次の不等式をたて、これを解いてNを求める。pは0.808、信頼度は95%なので $z=1.96$ を代入する。

$$1.96 \sqrt{\frac{0.808(1-0.808)}{N}} < 0.01$$

$$\frac{1.96 \sqrt{0.155136}}{\sqrt{N}} < \frac{1}{100}$$

$$\frac{\sqrt{N}}{1.96 \sqrt{0.155136}} > 100$$

$$\sqrt{N} > 100 \times 1.96 \sqrt{0.155136}$$

$$N > 10000 \times 1.96^2 \times 0.155136 \div 5959.7$$

∴5960人以上の大きさの標本が必要である。

第13章 偶然と必然を見分ける

①

(i) 帰無仮説 (ii) 片側 (iii) 有意水準 (iv) 第1種の誤り (v) 棄却域

②

(i) マスクをしてもしなくても、インフルエンザの感染しやすさに違いはない。

(ii) 配偶者の有無と、寿命とは関連がない。

(iii) 読書をよくする人とそうでない人とで、英語の成績に差はない。

③

(i) 帰無仮説を棄却し、調査仮説を採択する。よって「統計王国の男女比はアンバランスである」と言える。

(ii) 第1種の誤り。

(iii) 帰無仮説を棄却せず、調査仮説を採択しない。よって「統計王国の男女比はアンバランスである」とは言えない。

(iv) 第2種の誤り。

④

(i) 1/12 (約8.3%) (ii) 1/12 (約8.3%) (iii) 18 (iv) 棄却し

(v) 採択する (vi) 言える

⑤ (解答例)

(i) 有意水準は調査仮説を採択するか否かを判定するために分析者が予め定めておく基準であるから、調査結果に基づいて計算されるものではない。むしろ調査前に定めておくべき値である。

(ii) 有意水準は、調査仮説を採択するか否かを調査結果から判定するための基準である。したがって調査結果を見て分析者に都合よく（調査仮説が採択されるように）変更することは望ましくない。これはスポーツの試合中に、試合経過に合わせてルール変更することが許されないのと同じである。

(iii) 有意水準を p 値が下回れば帰無仮説が棄却され調査仮説が採択されるのだから、問題文とは逆に「有意水準が高い値であるほど」調査仮説が採択されやすくなる。

(iv) 有意水準を p 値が下回れば帰無仮説が棄却され調査仮説が採択されるのだから、問題文とは逆に「 p 値が低い値であるほど」調査仮説が採択されやすくなる。

(v) 統計的検定は無作為抽出（ランダム・サンプリング）が前提となっている。無作為抽出では偶然のいたずらによって偏った標本が採られてしまう可能性が避けられな

い。したがって正しい検定方法を採用したとしても、本文中で述べた第1種、第2種の誤りが生じる可能性を完全には回避できない。ただし我々は調査する標本のサイズや有意水準を予め調整することによって、これらの誤りが生じる確率を一定の水準にコントロールすることはできる。

第 14 章 集団間で違いがあるか

①

(i) 平均 (ii) t (iii) 標準正規分布 (iv) 片 (v) 両

②

調査仮説は「大学新卒者のほうが高校新卒者より、平均初任給が高い」であり、帰無仮説は「大学新卒者と高校新卒者の間に平均初任給の差はない」である。また、この調査仮説は平均値の大小関係を明示しているので、片側検定を選択する。大学生を A 集団、高校生を B 集団とみなし、t 検定の公式に該当する数値を代入して計算する。

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{s_A^2}{N_A} + \frac{s_B^2}{N_B}}} = \frac{20.98 - 16.67}{\sqrt{\frac{1.60^2}{20} + \frac{1.20^2}{20}}} = \frac{4.31}{\sqrt{\frac{2.56}{20} + \frac{1.44}{20}}} = \frac{4.31}{\sqrt{0.2}} = 9.6374 \cdots \div 9.638$$

$$\text{自由度 df} = \frac{\left(\frac{s_A^2}{N_A} + \frac{s_B^2}{N_B}\right)^2}{\frac{\left(\frac{s_A^2}{N_A}\right)^2}{N_A - 1} + \frac{\left(\frac{s_B^2}{N_B}\right)^2}{N_B - 1}} = \frac{\left(\frac{1.60^2}{20} + \frac{1.20^2}{20}\right)^2}{\frac{\left(\frac{1.60^2}{20}\right)^2}{20 - 1} + \frac{\left(\frac{1.20^2}{20}\right)^2}{20 - 1}} = \frac{0.2^2}{\frac{0.128^2}{19} + \frac{0.072^2}{19}} = \frac{0.04}{0.00113} = 35.30 \cdots$$

t 分布表の「自由度 35」の行、「片側 0.05 (5%)」の列を見ると、限界値は 1.690 だとわかる。よって、 $9.638 > 1.690$ より、帰無仮説を棄却して、調査仮説は採択する。すなわち、「大学新卒者のほうが高校新卒者より、平均初任給が高い」と結論づける。

③

調査仮説は「男女間で、「自殺は仕方がない」と考える人の比率に差がある」、帰無仮説は「男女間で、「自殺は仕方がない」と考える人の比率に差はない」である。調査仮説の形式から両側検定を選択する。

「わからない」「無回答」を取り除いた度数・百分率クロス表を作成する必要がある。以下の表となる。

	そう思う	そう思わない	計
男	98 (13.1)	652 (86.9)	750 (100.0)
女	80 (9.5)	758 (90.5)	838 (100.0)
計	178 (11.2)	1410 (88.8)	1588 (100.0)

*括弧は百分率

男性を A 集団、女性を B 集団として、本文中の公式（比率の差の検定方法の step3 参照）にあてはめると、検定統計量 Z は以下のように計算できる。

$$Z = \frac{0.131 - 0.087}{\sqrt{0.112(1 - 0.112) \times (\frac{1}{750} + \frac{1}{838})}} = \frac{0.044}{\sqrt{0.112 \times 0.888 \times 0.00253}} = \frac{0.044}{0.01585} \doteq 2.776$$

このように、検定統計量 Z の値 2.776 は、有意水準両側 5 % の限界値 1.96 を上回っている。よって帰無仮説を棄却し、調査仮説を採択する。すなわち「男女間で、自殺は仕方がないと考える人の比率に差がある」といえる。

④

(i)

帰無仮説：階層帰属意識において、上層に帰属している者と下層に帰属している者とは、機会の平等が保たれれば、結果として貧富の格差がついてもよいという考え方に対して意識の差はない。

調査仮説：階層帰属意識において、上層に帰属している者と下層に帰属している者とは、機会の平等が保たれれば、結果として貧富の格差がついてもよいという考え方に対して意識の差がないとは言えない。

(ii)

t 値を計算すると、5.448 自由度を計算すると、714.314 となる。自由度が 714 となるが、巻末の t 分布表では 240 までしかない。この場合は ∞ の行を見る。そうすると、有意水準を 5 % (両側検定) に設定したときの t の限界値は 1.960 であることがわかる。計算した t 値と照らし合わせると $1.960 < 5.448$ だから、帰無仮説は棄却される。

(iii)

検定の結果から、母集団において二つの階層間で意識に差がないと結論づけるのは妥当でない。したがって、これを捨て、機会の平等が保たれれば、結果として貧富の格差がついてもよいという考え方に対する意識に二つの階層間で差がないとは言えないという判断になる。

⑤

(i)

北海道の青年の平均利用時間：3 時間 16 分 (196 分)

福岡県の青年の平均利用時間：2 時間 58 分 (178 分)

計算の仕方

階級に幅がある場合、計算には階級値を利用する。例えば、1 時間以上 2 時間未満は階級値の 90 分を計算に使う。北海道の場合は 90 分利用した人が 23 人いるとみなして計算する。そうすると北海道の青年の平均利用時間は、

$$\frac{30 \times 8 + 90 \times 23 + 150 \times 26 + 210 \times 24 + 270 \times 21 + 330 \times 23}{125} \approx 196$$

となり、あとは時間換算すればよい。福岡県についても同様に、

$$\frac{30 \times 10 + 90 \times 29 + 150 \times 30 + 210 \times 21 + 270 \times 8 + 330 \times 23}{121} \approx 178$$

となり、あとは時間換算すればよい。

(ii)

北海道の青年の利用時間の標準偏差：1 時間 13 分 (93 分)

福岡県の青年の利用時間の標準偏差：1 時間 15 分 (95 分)

計算の仕方例

北海道の場合

$$\sqrt{\frac{(30-19)^2 \times 8 + (90-19)^2 \times 23 + (150-196)^2 \times 26 + (210-196)^2 \times 24 + (270-196)^2 \times 21 + (330-196)^2 \times 23}{125-1}} \approx 93$$

福岡県の場合

$$\sqrt{\frac{(30-178)^2 \times 10 + (90-178)^2 \times 29 + (150-178)^2 \times 30 + (210-178)^2 \times 21 + (270-178)^2 \times 8 + (330-178)^2 \times 23}{121-1}} \approx 67$$

(iii)

有意水準 5 % で t 検定したところ、北海道の青年が趣味・娯楽にインターネットを利用する時間の平均と福岡県の青年が趣味・娯楽にインターネットを利用する時間の平均には、統計的に差があるとは言えない。

計算の仕方

検定に利用する帰無仮説は、「北海道の青年の利用時間の平均と福岡県の青年の利用時間の平均には差がない」である。

調査仮説は、「北海道の青年の利用時間の平均と福岡県の青年の利用時間の平均には差がある」

である。

北海道を A、福岡県を B として、平均の差の検定をする。t 値と自由度 df を計算する

$$t = \frac{196-178}{\sqrt{\frac{93^2}{125} + \frac{95^2}{121}}} = 1.501 \quad df = \frac{\left(\frac{93^2}{125} + \frac{95^2}{121}\right)^2}{\frac{\left(\frac{93^2}{125}\right)^2}{125-1} + \frac{\left(\frac{95^2}{121}\right)^2}{121-1}} \doteq 243.29$$

となる。

有意水準 5 %、両側検定の際の自由度 243.29 の限界値は 1.970 となる。限界値と算出した t 値を比べると $1.501 < 1.970$ で算出した t 値は棄却域にないことがわかる。その結果、帰無仮説を棄却できない。「北海道の青年の平均利用時間と福岡県の青年の平均利用時間には差がない」という結論になる。

第 15 章 関連の真偽を判断する

①

- (i) クロス表におけるふたつの離散変数が関連しているか否かを調べるため。
- (ii) $(k-1) \times (l-1)$
- (iii) χ^2 検定は必ず両側検定であり、 χ^2 値が限界値よりも大きいかどうかを確かめればよいので、棄却域は χ^2 分布の右側のみとなる。
- (iv) 一つは標本が小さい時には、 χ^2 値が χ^2 分布に従わなくなるので、利用することができないこと。もう一つは、検定の結果が標本の大きさによって左右され、標本が大きいほど、調査仮説が採択されやすくなること。
- (v) 標本の相関係数

②

- (i) 調査仮説：男性と女性で、消費税に関する意見に違いがある。
- (ii) 帰無仮説：男性と女性で、消費税に関する意見に違いはない。
- (iii) (iv)

$\chi^2 = 5.37$ (小数第 3 位を四捨五入：下表は計算過程)

セル	(1,1)	(1,2)	(2,1)	(2,2)	計
n_{ij}	512	472	476	540	2000
e_{ij}	$e_{11} = \frac{984 \times 988}{2000} = 486.1$	$e_{12} = \frac{984 \times 1012}{2000} = 497.9$	$e_{21} = \frac{1016 \times 988}{2000} = 501.9$	$e_{22} = \frac{1016 \times 1012}{2000} = 514.1$	2000
$n_{ij} - e_{ij}$	25.9	-25.9	-25.9	25.9	-----
$(n_{ij} - e_{ij})^2$	670.81	670.81	670.81	670.81	-----
$(n_{ij} - e_{ij})^2 / e_{ij}$	1.38	1.35	1.34	1.3	5.37

- (v) 行数 = 2、列数 = 2 より、自由度 = $(2-1) \times (2-1) = 1$
 自由度 1、有意水準 5% (0.05) の χ^2 検定の限界値は、巻末の χ^2 分布表より、3.8415 とわかる。5.37 > 3.8415 だから、帰無仮説は棄却され、調査仮説を採択する。
- (vi) 男性と女性で消費税に関する意見に違いがあると言える。

③

- (i) 居住する都市規模によって、科学技術の発展に関する意見に違いがある。
- (ii) 居住する都市規模によって、科学技術の発展に関する意見に違いはない。

(iii)

	プラス面が多い	両方同じくらいである	マイナス面が多い	計
大都市	$e_{11} = \frac{441 \times 926}{1678} = 243.365$	$e_{12} = \frac{441 \times 630}{1678} = 165.572$	$e_{13} = \frac{441 \times 122}{1678} = 32.063$	441
中都市	$e_{21} = \frac{675 \times 926}{1678} = 372.497$	$e_{22} = \frac{675 \times 630}{1678} = 253.427$	$e_{23} = \frac{675 \times 122}{1678} = 49.076$	675
小都市	$e_{31} = \frac{389 \times 926}{1678} = 214.669$	$e_{32} = \frac{389 \times 630}{1678} = 146.049$	$e_{33} = \frac{389 \times 122}{1678} = 28.282$	389
町村	$e_{41} = \frac{173 \times 926}{1678} = 95.470$	$e_{42} = \frac{173 \times 630}{1678} = 64.952$	$e_{43} = \frac{173 \times 122}{1678} = 12.578$	173
計	926	630	122	1678

(iv)

*セルの数が多いことから、下の計算過程の表は、これまでの表の行と列を入れ替えて表示している。

	n_{ij}	e_{ij}	$n_{ij} - e_{ij}$	$(n_{ij} - e_{ij})^2$	$(n_{ij} - e_{ij})^2 / e_{ij}$
n_{11}	249	243.3647	5.6353	31.7564	0.1305
n_{12}	165	165.5721	-0.5721	0.3273	0.0020
n_{13}	27	32.0632	-5.0632	25.6357	0.7995
n_{21}	361	372.4970	-11.4970	132.1815	0.3549
n_{22}	257	253.4267	3.5733	12.7685	0.0504
n_{23}	57	49.0763	7.9237	62.7853	1.2793
n_{31}	223	214.6687	8.3313	69.4113	0.3233
n_{32}	139	146.0489	-7.0489	49.6865	0.3402
n_{33}	27	28.2825	-1.2825	1.6448	0.0582
n_{41}	93	95.4696	-2.4696	6.0990	0.0639
n_{42}	69	64.9523	4.0477	16.3837	0.2522
n_{43}	11	12.5781	-1.5781	2.4903	0.1980
計	1678	1678	-----	-----	3.8524

χ^2 値は 3.852 となる。

(v) このクロス表の自由度は次のように計算される。

$$df = (4 - 1) \times (3 - 1) = 6$$

自由度が 6 で有意水準が 5% だから、巻末の χ^2 分布表より限界値 = 12.592 だとわかる。よって χ^2 値 3.852 < 限界値 12.592 となる。したがって、帰無仮説を棄却するこ

とはできない。

(vi)「居住する都市規模によって、科学技術の発展に関する意見に違いがある」とはいえない。

④

まず、つぎのように調査仮説と帰無仮説を立てる。

調査仮説：年齢によって、「選択的夫婦別姓制度」に関する意見に違いがある。

帰無仮説：年齢によって、「選択的夫婦別姓制度」に関する意見に違いはない。

つぎに、期待度数を以下のように計算する。

	現在の制度である夫婦同姓制度を維持した方がよい	現在の制度である夫婦別姓制度を維持した上で、旧姓の通称使用についての法制度を設けた方がよい	選択制夫婦別姓制度を維持した方がよい	計
18～29 歳	$e_{11} = \frac{322 \times 778}{2828} = 88.584$	$e_{12} = \frac{322 \times 1217}{2828} = 138.569$	$e_{13} = \frac{322 \times 833}{2828} = 94.847$	322
30～39 歳	$e_{21} = \frac{321 \times 778}{2828} = 88.309$	$e_{22} = \frac{321 \times 1217}{2828} = 138.139$	$e_{23} = \frac{321 \times 833}{2828} = 94.552$	321
40～49 歳	$e_{31} = \frac{472 \times 778}{2828} = 129.850$	$e_{32} = \frac{472 \times 1217}{2828} = 203.120$	$e_{33} = \frac{472 \times 833}{2828} = 139.030$	472
50～59 歳	$e_{41} = \frac{451 \times 778}{2828} = 124.073$	$e_{42} = \frac{451 \times 1217}{2828} = 194.083$	$e_{43} = \frac{451 \times 833}{2828} = 132.844$	451
60～69 歳	$e_{51} = \frac{490 \times 778}{2828} = 134.802$	$e_{52} = \frac{490 \times 1217}{2828} = 210.866$	$e_{53} = \frac{490 \times 833}{2828} = 144.332$	490
70 歳以上	$e_{61} = \frac{772 \times 778}{2828} = 212.382$	$e_{62} = \frac{772 \times 1217}{2828} = 332.222$	$e_{63} = \frac{772 \times 833}{2828} = 227.396$	772
計	778	1217	833	2828

つぎに、 χ^2 値を計算すると、以下のように 311.424 となる。

	n_{ij}	e_{ij}	$n_{ij} - e_{ij}$	$(n_{ij} - e_{ij})^2$	$(n_{ij} - e_{ij})^2 / e_{ij}$
n_{11}	52	88.584	-36.584	1338.401	15.109
n_{12}	141	138.569	2.431	5.908	0.043
n_{13}	129	94.847	34.153	1166.459	12.298
n_{21}	41	88.309	-47.309	2238.146	25.344
n_{22}	155	138.139	16.861	284.294	2.058
n_{23}	125	94.552	30.448	927.082	9.805
n_{31}	80	129.850	-49.850	2485.030	19.138
n_{32}	205	203.120	1.880	3.534	0.017
n_{33}	187	139.030	47.970	2301.149	16.551
n_{41}	95	124.073	-29.073	845.230	6.812
n_{42}	214	194.083	19.917	396.683	2.044
n_{43}	142	132.844	9.156	83.831	0.631
n_{51}	128	134.802	-6.802	46.267	0.343
n_{52}	233	210.866	22.134	489.899	2.323
n_{53}	129	144.332	-15.332	235.061	1.629
n_{61}	382	212.382	169.618	28770.301	135.465
n_{62}	269	332.222	-63.222	3997.030	12.031
n_{63}	121	227.396	-106.396	11320.117	49.782
計	2828	2828	-----	-----	311.424

このクロス表の自由度は、以下のように 10 となる。

$$df = (6 - 1) \times (3 - 1) = 10$$

自由度 10、有意水準 5%の限界値は 18.3070 である。 χ^2 値は 311.424 であった。したがって、 χ^2 値 311.424 > 限界値 18.3070 となり、帰無仮説は棄却される。「年齢によって、選択的夫婦別姓制度に関する意見に違いがある」といえる。

⑤

(i) 帰無仮説：英語の得点と数学の得点の相関はゼロである。

(ii) 相関係数の検定表の標本数「10」の行、有意水準「片側 5 %」の列を見ると、限界値は 0.549 とわかる。0.549 < 0.79 なので帰無仮説を棄却し、調査仮説を採択する。
英語の得点と数学の得点は正の相関を示すと言える。