問題解答

ウォーミングアップ

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) 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
- ③ 省略
- ④ 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(在学年数<1年)=0.23
- (vi) 1-P
- (vii) P(仮説が間違いである)=0.02

2

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。

3

- (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」

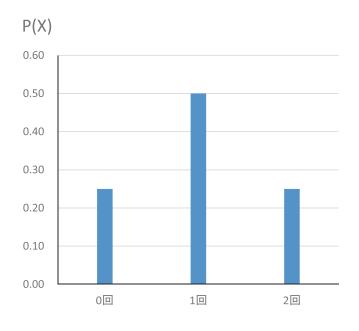
●「裏」○「表」

の二通りある。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 ※ $0.25 \times$ 二通り= 0.50 ということ。

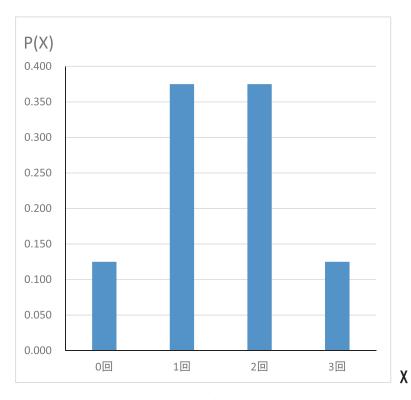
(iii) 0回「表」の確率は 0.25 × 一通り=0.25
1回「表」の確率は 0.25 × 二通り=0.50
2回「表」の確率は 0.25 × 一通り=0.25

 (i_{V})

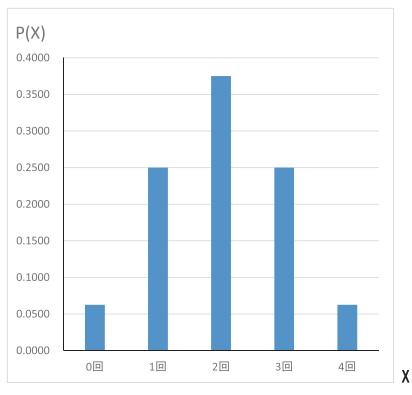


Χ

(v)



3回投げた場合



4回投げた場合

4

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 I 部 記述統計

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 度数分布表完成

| | 絶対度数 | 相対度数 | 百分率 |
|-----------|------|-------|--------|
| 家族 | 446 | 0.505 | 50.5% |
| その他の親族 | 27 | 0.031 | 3.1% |
| 知人·友人 | 178 | 0.201 | 20.1% |
| 職場関係 | 51 | 0.058 | 5.8% |
| その他 | 71 | 0.080 | 8.0% |
| 面識なし | 103 | 0.117 | 11.7% |
| 法人・団体・被害者 | 8 | 0.000 | 0.09/ |
| なし | | 0.009 | 0.9% |
| 合計 | 884 | 1.001 | 100.1% |

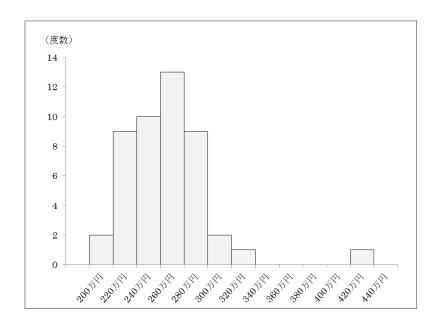
^{*}構成割合は四捨五入をしているため、その合計は100にならない。

3

(i) 度数分布表完成図

| 階級番号 | 階級 | 階級値 | 絶対度数 | 相対度数 | 百分率 |
|------|----------------|-------|------|-------|-------|
| 1 | 200万円以上220万円未満 | 210万円 | 2 | 0.043 | 4.3% |
| 2 | 220万円以上240万円未満 | 230万円 | 9 | 0.191 | 19.1% |
| 3 | 240万円以上260万円未満 | 250万円 | 10 | 0.213 | 21.3% |
| 4 | 260万円以上280万円未満 | 270万円 | 13 | 0.277 | 27.7% |
| 5 | 280万円以上300万円未満 | 290万円 | 9 | 0.191 | 19.1% |
| 6 | 300万円以上320万円未満 | 310万円 | 2 | 0.043 | 4.3% |
| 7 | 320万円以上340万円未満 | 330万円 | 1 | 0.021 | 2.1% |
| 8 | 340万円以上360万円未満 | 350万円 | 0 | 0 | 0.0% |
| 9 | 360万円以上380万円未満 | 370万円 | 0 | 0 | 0.0% |
| 10 | 380万円以上400万円未満 | 390万円 | 0 | 0 | 0.0% |
| 11 | 400万円以上420万円未満 | 410万円 | 0 | 0 | 0.0% |
| 12 | 420万円以上440万円未満 | 430万円 | 1 | 0.021 | 2.1% |
| 合計 | | _ | 47 | 1.000 | 100% |

ヒストグラム完成図



(ii)

260 万円以上 280 万円未満の階級が一番多いことが見て取れる。次に多い階級は 260 万円以上 280 万円未満の階級の前後となっている。この分布をみるとほとんどの都道 府県の1人当たり県民所得は 220 万円以上 300 万円未満の間にあることがわかる。

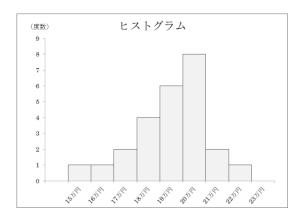
また、1つだけ 420 万円以上 440 万円未満の階級にあり、他の都道府県と比べてまったく異なる特徴をもっていることが推測される。

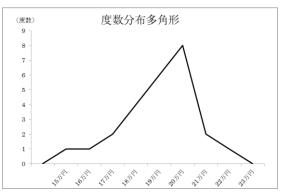
4

(i) 度数分布表完成図

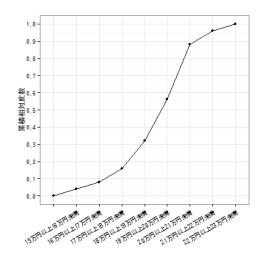
| 階級番号 | 階級 | 階級値 | 絶対度数 | 相対度数 | 百分率 | 累積度数 | 累積相対度数 | 累積百分率 |
|------|--------------|--------|------|-------|-------|------|--------|--------|
| 1 | 15万円以上16万円未満 | 15万5千円 | 1 | 0.040 | 4.0% | 1 | 0.04 | 4.0% |
| 2 | 16万円以上17万円未満 | 16万5千円 | 1 | 0.040 | 4.0% | 2 | 0.08 | 8.0% |
| 3 | 17万円以上18万円未満 | 17万5千円 | 2 | 0.080 | 8.0% | 4 | 0.16 | 16.0% |
| 4 | 18万円以上19万円未満 | 18万5千円 | 4 | 0.160 | 16.0% | 8 | 0.32 | 32.0% |
| 5 | 19万円以上20万円未満 | 19万5千円 | 6 | 0.240 | 24.0% | 14 | 0.56 | 56.0% |
| 6 | 20万円以上21万円未満 | 20万5千円 | 8 | 0.320 | 32.0% | 22 | 0.88 | 88.0% |
| 7 | 21万円以上22万円未満 | 21万5千円 | 2 | 0.080 | 8.0% | 24 | 0.96 | 96.0% |
| 8 | 22万円以上23万円未満 | 22万5千円 | 1 | 0.040 | 4.0% | 25 | 1 | 100.0% |
| 合計 | _ | _ | 25 | 1.000 | 100% | I | ı | _ |

(ii) ヒストグラム・度数分布多角形完成図





(iii) 類積相対度数分布表完成図



 (i_V)

この問題のグラフと本文中のグラフを比較すると、従業員 100 人~999 人規模の企業(この問題のグラフ)では、従業員 1000 人以上規模の企業(本文中のグラフ)に比べて、相対的に初任給の低い人が多い。つまり企業規模が大きいほど、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 270 万円で、平均値 (287.7 万円)よりも低い。「所得」の平均値は分布がゆがんでいるため、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 430.6 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

| | 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が、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。

| (4) | | 1961年 | 1986年 | 2011年 |
|-----|-------|----------|-------|-------|
| | メディアン | 30.5 | 35 | 37 |
| | モード | 28/29/34 | 34 | 36 |
| | 平均 | 31.2 | 35.5 | 38.6 |
| | レンジ | 8 | 29 | 31 |
| | 分散 | 8.2 | 80.1 | 104.3 |

2.9

標準偏差

初婚年齢と同様、この 50 年間に女性の再婚年齢の上昇と、再婚年齢のばらつきの増大が見られる。参考として、【図表 4-8】に実際の男女の再婚年齢の特性を示しておこう。

10.2

8.9

| | | 図表4-8 | 図表4-8 再婚年齢の特性 | | | | | |
|----------|--------|--------|---------------|--|----------|--------|--------|--------|
| 〈男性〉 | | | | | 〈女性〉 | | | |
| | 1961年 | 1986年 | 2011年 | | | 1961年 | 1986年 | 2011年 |
| 総数(人) | 45,444 | 66,122 | 95,277 | | 総数(人) | 25,941 | 57,645 | 82,423 |
| メディアン(歳) | 35 | 37 | 39 | | メディアン(歳) | 31 | 34 | 37 |
| モード(歳) | 30 | 37 | 37 | | モード(歳) | 28 | 34 | 36 |
| 平均値(歳) | 38.1 | 39.7 | 42.1 | | 平均値(歳) | 33.4 | 36.0 | 38.9 |
| 標準偏差(歳) | 10.33 | 9.78 | 11.00 | | 標準偏差(歳) | 2.90 | 8.76 | 10.18 |

第5章 ふたつの離散変数を同時に扱う

1

| | 中学(高校) | 専門学校 | 短大•高専 | 大学(大学院) | 計 |
|----|--------|------|-------|---------|-------|
| 女子 | 15.7 | 18.5 | 9.5 | 56.3 | 100.0 |
| 男子 | 23.8 | 7.8 | 3.0 | 65.4 | 100.0 |
| 計 | 19.9 | 12.9 | 6.1 | 61.1 | 100.0 |

2

「生死は本人の判断」という考え方をする人の割合は、若い人が高く、高齢の人は低いことがわかる。

| 年齢・「生死は本人の判断に任せるべき」という考え別クロス表(人) | | | | | | | | | |
|----------------------------------|------------|------|------|------|------|-------|------|--|--|
| 〈生死は本人の | 〈生死は本人の判断〉 | | | | | | | | |
| | 20歳代 | 30歳代 | 40歳代 | 50歳代 | 60歳代 | 70歳以上 | 計 | | |
| そう思う | 102 | 115 | 123 | 123 | 110 | 84 | 658 | | |
| そう思わない | 68 | 123 | 186 | 173 | 202 | 95 | 847 | | |
| わからない | 23 | 41 | 53 | 44 | 76 | 37 | 274 | | |
| 無回答 | 11 | 13 | 19 | 30 | 64 | 102 | 238 | | |
| 計 | 204 | 292 | 381 | 370 | 452 | 318 | 2017 | | |

3

| | 「ひきこもり」的生活開始時の属性 | | | | | | | | |
|------|------------------|-------|-------|-------|-------|-------|-------|--------|--|
| | | 中学生 | 高校生 | 高卒後 | 大学生 | 社会人 | その他 | 計 | |
| | 男 | 4 | 8 | 1 | 9 | 1 | 3 | 26 | |
| J.d. | ח | 15.4% | 30.8% | 3.8% | 34.7% | 3.8% | 11.5% | 100.0% | |
| 性別 | 女 | 0 | 2 | 2 | 2 | 4 | 2 | 12 | |
| 733 | * | 0.0% | 16.7% | 16.7% | 16.7% | 33.2% | 16.7% | 100.0% | |
| | 計 | 4 | 10 | 3 | 11 | 5 | 5 | 38 | |
| | пІ | 10.5% | 26.3% | 7.9% | 28.9% | 13.2% | 13.2% | 100.0% | |

→1 行 4 列目(34.7%)、2 行 5 列目(33.2%)にまるめています。

「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。

4

| 主要国別·管理的職業従事者数(%) | | | | | | | |
|-------------------|------|------|-------|--|--|--|--|
| | 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 | | | | |
| 英国 | 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 | | | | |
| 米国 | 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 | | | | |

国際比較をしてみると、日本において管理的職業従事者数の女性割合がひじょうに少ない状況が確認できる。欧米諸国が3~4割、タイが2割強に対して、日本と韓国は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)
$$\phi = \frac{60 \times 55 - 40 \times 45}{\sqrt{100 \times 100 \times 105 \times 95}} = 0.150$$

(iii) $\chi^2 = 4.510$ (下表は計算過程)

| セル | (1.1) | (1.2) | (2,1) | (2.2) | 計 |
|------------------------------------|----------|----------|----------|------------|------|
| nij | 60 | 40 | 45 | 55 | 200 |
| e ij | 100×105÷ | 100×95÷ | 100×105÷ | 100×95÷ | 200 |
| | 200=52.5 | 200=47.5 | 200=52.5 | 200 = 47.5 | |
| nij -eij | 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)
$$v = \sqrt{\frac{4.510}{(2-1)\times200}} \stackrel{..}{=} 0.150$$
 (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 |

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

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

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 | どちらかとい | 反対 | 計 |
|-------|------------|--------|--------|-------|--------|
| | | えば賛成 | えば反対 | | |
| 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 | 46.6%(まるめ) | 25.0% | 16.5% | 11.9% | 100.0% |

| 60 歳代 | 35.7% | 22.9% | 23.5% | 17.9% | 100.0% |
|--------|------------|-------|-------|-------|--------|
| 70 歳以上 | 24.6% | 19.4% | 26.5% | 29.5% | 100.0% |
| 計 | 46.0%(まるめ) | 24.3% | 16.9% | 12.8% | 100.0% |

(ii)
$$v = \sqrt{\frac{40242}{(4-1)\times3394}} = 0.199$$
 (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。

第7章 連続変数同士の関連を分析する(その1)

(1)

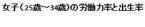
(i) 正の関連、(ii) 正の関連

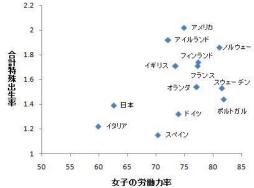
(2)

- (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)





(ii)

$$(X_7 - \overline{X}) = 77.1 - 74.1 = 3$$
 $(X_7 - \overline{X})^2 = 3^2 = 9$ $(Y_7 - \overline{Y}) = 1.54 - 1.58 = -0.04$
 $(Y_7 - \overline{Y})^2 = (-0.04)^2 = 0.0016$ $(X_7 - \overline{X})(Y_7 - \overline{Y}) = 3 \times (-0.04) = -0.12$

(iii)

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

 (i_V)

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

(v)

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

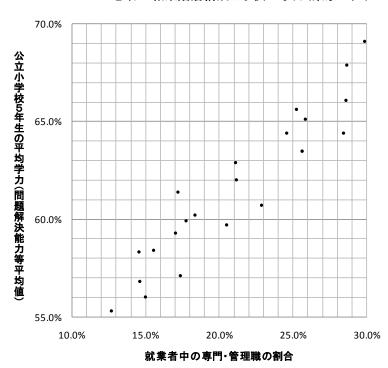
(vi)

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

4

(i)解答例

地域の職業階層構成と子供の学力(東京23区)



(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いこと、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

県民所得が200万円の都道府県の進学率(%)の予測

 $\hat{Y} = 9.98 + 0.138 \times 200 = 9.98 + 27.6 = 37.6$ \therefore # ? = 37.6 %

県民所得が300万円の都道府県の進学率(%)の予測

 $\hat{Y} = 9.98 + 0.138 \times 300 = 9.98 + 27.6 = 51.4$ ∴進学率 = 51.4%

3

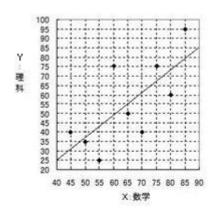
(i) 回帰式をŶ=a+bxとすると

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

- $\hat{Y} = -23 + 1.2X$
- (ii) 数学が40点の人の理科の点数は、 $\hat{\mathbf{Y}}=-23+1.2\times40=25$ 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。

(iii)



(iv) 相関係数は、

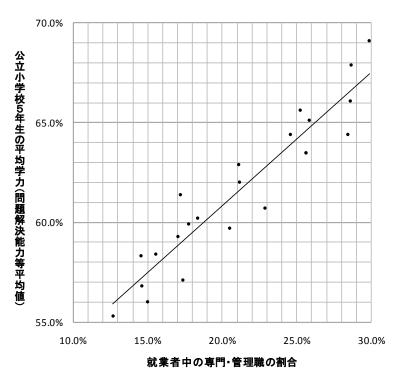
$$r_{XY} = \frac{s_{XY}}{s_X s_Y} = \frac{s_{XY}}{\sqrt{s_X^2} \sqrt{s_Y^2}} = \frac{225}{\sqrt{187.5} \sqrt{525}} = \frac{225}{13.693063 \times 22.912878} = \frac{225}{313.74748} = 0.7171372 \stackrel{...}{=} 0.72$$

よって決定係数は、 $R^2 = r_{XY}^2 = 0.717137 \stackrel{?}{2} = 0.5142857 \stackrel{...}{\Rightarrow} 0.51$

(4)

(i) 回帰式を \hat{Y} =a+bX と表すと第8章の式①より $\mathbf{b} = \frac{\mathbf{S}_{XY}}{\mathbf{S}_{X}^{2}} = \frac{\mathbf{19.1}}{\mathbf{28.5}} \leftrightarrows \mathbf{0.670}$ 、同じく式②より $\mathbf{a} = \overline{Y} - \mathbf{b}\overline{X} = \mathbf{61.6} - \mathbf{0.670} \times \mathbf{21.2} \leftrightarrows \mathbf{47.4}$ となる。 \therefore \hat{Y} =47.4+0.670X (ii)

地域の職業階層構成と子供の学力(東京23区)



(iii) 第7章の STEP3 より、相関係数 $=\frac{s_{XY}}{s_X s_Y} = \frac{19.1}{\sqrt{28.5} \times \sqrt{14.5}} = 0.940$ 。決定係数は相関係数の2乗だから決定係数 $=0.940^2 = 0.884$ 。第8章 2-3 の目安によれば、あてはまりは「よい」。

第9章 みえない関係を探る

(1)

$$(i) \bigcirc, (ii) \times, (iii) \times, (iv) \bigcirc, (v) \times, (vi) \bigcirc$$

2

共通の原因として「年齢」が考えられる。一般に年齢が若い人は高い人よりも教育年数が長く(昔と比べて高等教育機関への進学率が上昇しているため)、しかも携帯電話を所持している傾向にあると考えられる。したがって、教育年数と携帯電話の所持率とのクロス表を作成すると、教育年数の長い人ほど、携帯電話を所持しているかのような擬似関係が生じるものと考えられる。

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徳への態度(百分率クロス表 その1)

| | 男性 | | | 女性 | | |
|----|-------|-------|--------|-------|-------|--------|
| | よくない | かまわない | 計 | よくない | かまわない | 計 |
| 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徳への態度(百分率クロス表 その2)

| | 男性 | | | 女性 | | |
|----|--------|--------|--------|--------|--------|--------|
| | よくない | かまわない | 計 | よくない | かまわない | 計 |
| 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徳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徳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
$$\phi = \frac{86 \times 15 - 90 \times 29}{\sqrt{176 \times 44 \times 115 \times 105}} = -0.137$$
 (小数第4位を四捨五入)

女性
$$\phi = \frac{130 \times 13 - 85 \times 42}{\sqrt{215 \times 55 \times 172 \times 98}} = -0.133$$
 (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、性道徳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)

$$\mathbf{r}_{\mathbf{XY\cdot T}} = \frac{\mathbf{r}_{\mathbf{XY}} - \mathbf{r}_{\mathbf{TX}} \mathbf{r}_{\mathbf{TY}}}{\sqrt{1 - \mathbf{r}_{\mathbf{TX}}^2} \sqrt{1 - \mathbf{r}_{\mathbf{TY}}^2}} = \frac{-0.537 - (-0.749 \times 0.628)}{\sqrt{1 - (-0.749)^2} \sqrt{1 - 0.628^2}} = \frac{-0.537 + 0.470372}{0.6625699 \times 0.7782133} = -0.129$$
 (小数第 4位を四

捨五入)

(ii)

虫歯経験率Xと眼鏡等着用率Yは r_{XY} =-0.537と中程度の相関を示すが、社会階層Tを第3変数とした虫歯経験率と眼鏡着用率の偏相関係数($r_{XY:T}$)は-0.129とゼロに近い。よって虫歯経験率と眼鏡等着用率との関係は、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虫歯の子が少なく(r_{TX} <0)、眼鏡の子が多い(r_{TY} >0)ため、結果的に虫歯の子の少ない地域ほど眼鏡の子が多い(r_{XY} =-0.537)という関連性が現れていたと考えられる。

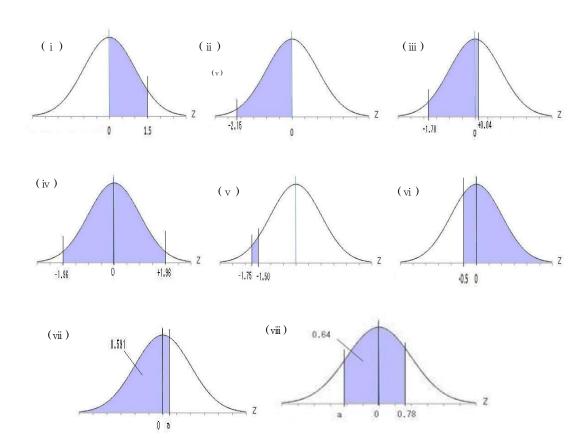
第Ⅱ部 推定統計

第10章 全体のなかでの位置を把握する

1

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

2



- (i) $P(0 \le z \le +1.50) = 0.4332$
- (ii) $P(-2.15 \le z \le 0) = P(0 \le z \le 2.15) = 0.4842$
- (iii) $P(-1.78 \le z \le +0.04) = P(-1.78 \le z \le 0) + P(0 \le z \le 0.04)$ = $P(0 \le z \le 1.78) + P(0 \le z \le 0.04) = 0.4625 + 0.0160 = 0.4785$
- (iv) $P(-1.96 \le z \le 1.96) = P(0 \le z \le 1.96) \times 2 = 0.4750 \times 2 = 0.950$
- (v) $P(-1.75 \le z \le -1.50) = P(+1.50 \le z \le +1.75)$ = $P(0 \le z \le 1.75) - P(0 \le z \le 1.50) = 0.4599 - 0.4332 = 0.0267$
- (vi) $P(-0.5 \le z) = P(-0.5 \le z \le 0) + P(0 \le z) = P(0 \le z \le 0.5) + P(0 \le z)$ =0.1915+0.5=0.6915

- (vii) $P(z \le a) = P(z \le 0) + P(0 \le z \le a) = 0.5 + P(0 \le z \le a) = 0.591$ $\therefore P(0 \le z \le a) = 0.591 - 0.5 = 0.091 \quad \therefore a = 0.23$
- (viii) $P(a \le z \le 0.78) = P(a \le z \le 0) + P(0 \le z \le 0.78) = P(a \le z \le 0) + 0.2823 = 0.64$ $\therefore P(a \le z \le 0) = 0.64 - 0.2833 = 0.3577 \quad \therefore a = -1.07$

(3)

(i) フツオ君の国語の標準得点=(50-65.8)÷7.5=-15.8÷7.5≒-2.11
数学の標準得点=(50-62.2)÷14.3=-12.2÷14.3≒-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 : 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×1.08だから、65.8+(7.5×1.08)=73.9となる。 ∴ 73.9点が分断点である。

4

- (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、 $P(0 \le z \le 2) = 0.4772$ 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×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。∴ 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)

(典型的な実験結果と、その際の解答例)

- 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

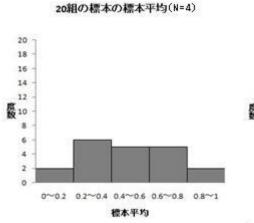
2

- Xは標本平均、Xは標本平均(X)の平均である。

sは標本標準偏差、sxは標本平均(X)の標準偏差である。

3

(i)(ii)



20組の標本の標本平均 (N=16)

20 18 16 14 12 10 8 6 4 2 0 0~0.2 0.2~0.4 0.4~0.6 0.6~0.8 0.8~1

(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= $\frac{S}{\sqrt{N}}$ = $\frac{15}{\sqrt{4}}$ =7.5
- (ii) N=8 のとき 標準誤差 SE= $\frac{S}{\sqrt{N}}$ = $\frac{15}{\sqrt{8}}$ =5.303 \cdots =5.30
- (iii) N=16 のとき 標準誤差 SE= $\frac{S}{\sqrt{N}}$ = $\frac{15}{\sqrt{16}}$ =3.75

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%の確からしさで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。X-Z $\frac{s}{\sqrt{N}} \le \mu \le X$

$$+Z\frac{S}{\sqrt{N}}$$

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>

$$626-1.96 \frac{336}{\sqrt{1356}} \le \mu \le 626+1.96 \frac{336}{\sqrt{1356}}$$

 $626-17.884\dots \le \mu \le 626+17.884\dots$

$$608.116 \dots \le \mu \le 643.884 \dots$$

608.1 万円以上 643.9 万円以下

<母子家庭>

$$294-1.96 \frac{223}{\sqrt{699}} \le \mu \le 294+1.96 \frac{223}{\sqrt{699}}$$

 $294-16.531\dots \le \mu \le 294+16.531\dots$

 $277.469 \dots \leq \mu \leq 310.531 \dots$

277.5 万円以上 310.5 万円以下

<父子家庭>

$$550-1.96 \frac{260}{\sqrt{84}} \le \mu \le 550+1.96 \frac{260}{\sqrt{84}}$$

 $550\text{-}55.601\cdots \leqq \mu \leqq 550\text{+}55.601\cdots$

 $494.399 \dots \leq \mu \leq 605.601 \dots$

494.4 万円以上 605.6 万円以下

(3)

信頼度95%(α=0.05)の場合、z=1.96、p=0.66、N=2084を以下の公式に代入する。

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

$$0.66 - 1.96\sqrt{\frac{0.66(1 - 0.66)}{2084}} \le P \le 0.66 + 1.96\sqrt{\frac{0.66(1 - 0.66)}{2084}}$$

$0.640 \le P \le 0.680$ (小数第4位四捨五入)

信頼度99%(α=0.01)の場合、z=2.58、p=0.66、N=2084を公式に代入する。

$$0.66 - 2.58\sqrt{\frac{0.66(1 - 0.66)}{2084}} \le P \le 0.66 + 2.58\sqrt{\frac{0.66(1 - 0.66)}{2084}}$$

 $0.633 \le P \le 0.687$ (小数第4位四捨五入)

(4)

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

$$1.96\sqrt{\frac{0.804(1-0.804)}{N}} < 0.01$$

$$\frac{1.96\sqrt{0.157584}}{\sqrt{N}} < \frac{1}{100}$$

$$\frac{\sqrt{N}}{1.96\sqrt{0.157584}} > 100$$

$$\sqrt{N} > 100 \times 1.96 \sqrt{0.157584}$$

N > $10000 \times 1.96^2 \times 0.157584 \rightleftharpoons 60537$ (小数第2位四捨五入)

∴ 6054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) 第一種の誤り、(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) 25、(ii) 25、(iii) 21、(iv) 棄却し、(v) 採択する、(vi) 言える

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「規模の大きい企業のほうが規模の小さい企業より、新卒者の平均初任給が高い」であり、帰無仮説は「規模の大きい企業と規模の小さい企業とを比べると、新卒者の平均初任給に差はない」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。

規模の大きい企業をA集団、小さい企業を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。

$$t = \frac{\mathbf{A}$$
集団の標本平均 $-\mathbf{B}$ 集団の標本平均
$$\frac{\mathbf{X}_A - \overline{\mathbf{X}_B}}{\sqrt{\frac{\mathbf{A}$$
集団の標本分散}{\mathbf{A}集団の標本の大きさ $+\frac{\mathbf{B}$ 集団の標本の大きさ
$$\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

$$t = \frac{19.88 - 19.58}{\sqrt{\frac{1.50^2}{20} + \frac{1.58^2}{25}}} = \frac{0.3}{\sqrt{\frac{2.25}{20} + \frac{2.4964}{25}}} = \frac{0.3}{\sqrt{0.212356}} = 0.6510..... = 0.651$$

自由度
$$df = \dfrac{\displaystyle \left(\dfrac{A$$
集団の標本分散}{A集団の標本分散} + \dfrac{B集団の標本分散}{B集団の標本の大きさ}^2 - \dfrac{\displaystyle \left(\dfrac{s_A^2 + s_B^2}{n_A}\right)^2}{\displaystyle \left(\dfrac{A集団の標本分散}{A集団の標本の大きさ}\right)^2 + \displaystyle \left(\dfrac{B集団の標本の大きさ d B集団の標本の大きさ d B集団の解本の大きさ d B集団の解本の大き

$$df = \frac{\left(\frac{1.50^2}{20} + \frac{1.58^2}{25}\right)^2}{\left(\frac{1.50^2}{20}\right)^2 + \left(\frac{1.58^2}{25}\right)^2} = \frac{0.212356^2}{\frac{0.1125^2}{19} + \frac{0.099856^2}{24}} = \frac{0.04509507}{0.00108159} = 41.6934.... = 41.693$$

t分布表の「自由度 41」の行、「片側 0.05(5%)」の列を見ると、限界値は 1.683 だとわかる。よって、0.651<1.683 より、帰無仮説は棄却されず、調査仮説は採択されないことがわかる。すなわち、「規模の大きい企業のほうが規模の小さい企業より、新卒者の平均初任給が高い」とは言えない。

(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。

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。

$$Z = \frac{0.137 - 0.078}{\sqrt{0.107(1 - 0.107) \times (\frac{1}{731} + \frac{1}{721})}} = \frac{0.059}{\sqrt{0.09555100 \times 0.00275495}} = \frac{0.059}{0.01622462} \stackrel{\rightleftharpoons}{=} 3.636$$

このように、検定統計量 Z の値、3.636 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では、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では、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。

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、乗却域は 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 |
|-----------------------|------------|------------|------------|-------------------------|------|
| nij | 512 | 472 | 476 | 540 | 2000 |
| eij | 984×988÷ | 984×1012÷ | 1016×988÷ | $1016 \times 1012 \div$ | 2000 |
| | 2000=486.1 | 2000=497.9 | 2000=501.9 | 2000=514.1 | |
| nij -eij | 25.9 | -25.9 | -25.9 | 25.9 | |
| $(n_{ij} - e_{ij})^2$ | 670.81 | 670.81 | 670.81 | 670.81 | |
| (n ij -eij)²/ eij | 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

次にあげる表は、「臓器移植への関心の有無」を居住する都市規模別に見たものである。 都市規模によって「臓器移植への関心」に違いがあるかどうか知りたいとする。以下の問に答えなさい。

| | 関心がある | 関心がない | 計 |
|-----|-------|-------|------|
| 大都市 | 262 | 181 | 443 |
| 中都市 | 477 | 313 | 790 |
| 小都市 | 230 | 200 | 430 |
| 町村 | 106 | 86 | 192 |
| 計 | 1075 | 780 | 1855 |

(出典:内閣府大臣官房広報室、平成25年実施、「臓器移植に関する世論調査」)

- (i) 調査仮説:居住する都市規模によって臓器移植への関心に違いがある
- (ii) 対立仮説(帰無仮説):居住する都市規模によって臓器移植への関心に違いがない

(iii) 期待度数を計算しなさい

| | Oi | | | | e_i |
|-----|------|-----|------|------|----------|
| n11 | 262 | 443 | 1075 | 1855 | 256.725 |
| n12 | 181 | 443 | 780 | 1855 | 186.275 |
| n21 | 477 | 790 | 1075 | 1855 | 457.817 |
| n22 | 313 | 790 | 780 | 1855 | 332.183 |
| n31 | 230 | 430 | 1075 | 1855 | 249.191 |
| n32 | 200 | 430 | 780 | 1855 | 180.809 |
| n41 | 106 | 192 | 1075 | 1855 | 111.267 |
| n42 | 86 | 192 | 780 | 1855 | 80.733 |
| 計 | 1855 | | | | 1855.000 |

(iv) χ^2 値を求めなさい。

| o _i e _i o _i -e _i (o _i - | $(o_i - e_i)^2$ $(o_i - e_i)^2 / e_i$ |
|--|---------------------------------------|
|--|---------------------------------------|

| n11 | 262 | 256.7251 | 5.2749 | 27.8249 | 0.1084 |
|------------------|-----|----------|----------|----------|--------|
| n12 | 181 | 186.2749 | -5.2749 | 27.8249 | 0.1494 |
| n21 | 477 | 457.8167 | 19.1833 | 367.9986 | 0.8038 |
| n22 | 313 | 332.1833 | -19.1833 | 367.9986 | 1.1078 |
| n31 | 230 | 249.1914 | -19.1914 | 368.3089 | 1.4780 |
| n32 | 200 | 180.8086 | 19.1914 | 368.3089 | 2.0370 |
| n41 | 106 | 111.2668 | -5.2668 | 27.7397 | 0.2493 |
| n42 | 86 | 80.7332 | 5.2668 | 27.7397 | 0.3436 |
| χ ² 値 | | | | | 6.2773 |

(v) このクロス表の自由度を述べなさい。

$$df = (4-1) \times (2-1) = 3$$

(vi) χ^2 値6.2773 < 限界値7.8147である。したがって、帰無仮説を棄却することはできない。居住する都市規模によって、臓器移植への関心に違いがあるとは言えない。

4

- (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。