統計学入門 練習問題解答集

この解答集は 1995 年度ゼミ生
椎野英樹（4 回生)、奥井亮（3 回生)、北川宣治（3 回生）
による学習の成果の一部です. ワープロ入力はもちろん井戸温子さんのおかげです. 利用される方々のご意見を待ちます. （1996 年 3 月 6 日）
趙君が 7 章 8 章の解答を書き上げました. （1996 年 7 月）
線型回帰に関する性質の追加. （1996 年 8 月）

ホーム頁に入れるため、1999 年 7 月に再度編集しました.

改訂にあたり、
久保拓也（D3)、鍵原理人（D2)、奥井亮（D1)、三好祐輔（D1)、
金谷太郎（M1）
の諸氏にお世話になりました. （2000 年 5 月）

森棟公夫

606-8501 京都市左京区吉田本町京都大学経済研究所
電話 075-753-7112
e-mail morimune@kier.kyoto-u.ac.jp

第 1 章追加説明　Tschebychv (1821-1894)の不等式 [離散ケース]

命題：1 よりも大きな k について、観測値の少なくとも $(1-(1/k^2))$ の割合は (平均値－k標本標準偏差) から (平均値＋k 標本標準偏差) の区間に含まれる.例えば 2 シグマ区間の場合は $(1-(1/2^2))=\frac{3}{4}=75\%$ 以上. 3 シグマ区間の場合は $(1-(1/3^2))=\frac{8}{9}$ 以上. 4 シグマ区間の場合は $(1-(1/4^2))=\frac{15}{16}\approx 93.75\%$ 以上.

証明：観測個数を n、変数を x、平均値を $\dot{x}$、標本分散を $\hat{\sigma}^2$ とおくと、定義より

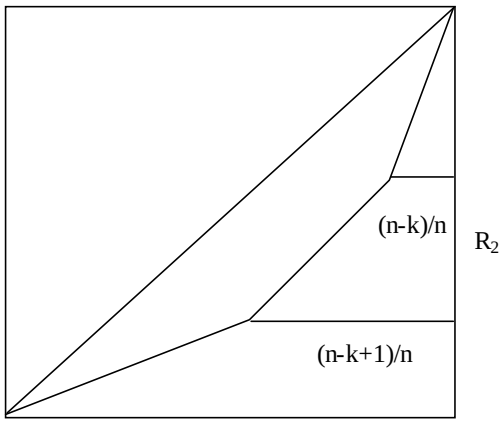
$$n\hat{\sigma}^2=\sum_{i=1}^{n}(x_i-\bar{x})^2 \qquad \cdots (1)$$

ここで $k>1$ の条件の下で $|x_i-\bar{x}|\le k\hat{\sigma}$ となる x を $x_{(1)},\cdots,x_{(a)}$,$|x_i-\bar{x}|\ge k\hat{\sigma}$ となる x を $x_{(a+1)},\cdots,x_{(n)}$ とおく. この分割から、(1)の右辺は

$$n\hat{\sigma}^2\ge\sum_{i=1}^{n-a}(x_{(a+i)}-\bar{x})^2\ge(n-a)(k\hat{\sigma})^2 \qquad \cdots (2)$$

となる. だから、$n-a<\frac{1}{k^2}\cdot n$. あるいは $a>(1-\frac{1}{k^2})n$ となる.

ジニ係数の計算



$$\text{ジニ係数}=1-\frac{\text{ﾛｰﾚﾝﾂ曲線下の面積}}{\text{三角形の面積}}$$

ローレンツ曲線下の図形を右のように台形に分割する.両端は三角形となる.**原データが利用可能である**として、各人の相対所得をR_1からR_nまでとしよう.この場合、下かから k 段目の台形は下底が $(n-k+1)/n$、上底が $(n-k)/n$ である.（相対順位の差は$1/n$だから、この差だけ上底が短い.）台形の高さはR_kだから、台形の面積は $R_k(2n-2k+1)/(2n)$ となる.（$k=n$ では台形は三角形になっているが、式は成立する.）台形と三角形の面積を足し合わせると、ローレンツ曲線下の面積 $=\sum_{k=1}^{n} R_k(2n-2k+1)/(2n)$ となる.したがってこの面積と三角形の面積の比は、$\sum_{k=1}^{n} R_k(2n-2k+1)/n$ である.相対所得の総和は 1 であるから、この比は $=2+\frac{1}{n}-\frac{2}{n}\sum_{k=1}^{n} k R_k$. 1 から引くと、ジニ係数は $=\frac{2}{n}\sum_{k=1}^{n} kR_k-(1+\frac{1}{n})$ となる.

標本相関係数の性質

$$\gamma_{xy}=\frac{\text{共分散}}{\sqrt{\text{xの分散、yの分散}}}=\frac{S_{xy}}{\sqrt{S_x^2\cdot S_y^2}}=\frac{S_{xy}}{S_x\cdot S_y},$$

ベクトル$\bar{x}=(x_1-\bar{x},\cdots,x_n-\bar{x})$と$\bar{y}=(y_1-\bar{y},\cdots,y_n-\bar{y})$を用いれば、$S_x$は$\bar{x}$の大きさ（ノルム）、$S_y$は$\bar{y}$の大きさ、$S_{xy}$は$\bar{x}$と$\bar{y}$の内積である.標本相関係数は、ベクトル$\bar{x}$と$\bar{y}$の間の正弦$\cos\theta$に他ならない.従って、標本相関係数の絶対値は 1 より小になる.

変量を標準化して、 $u_1=\frac{x_1-\bar{x}}{S_x},\cdots,u_n=\frac{x_n-\bar{x}}{S_x}$, $v_1=\frac{y_1-\bar{y}}{S_y},\cdots,v_n=\frac{y_n-\bar{y}}{S_y}$,

と定義する.u と v の標本共分散$\frac{1}{n}\sum_{i=1}^{n} u_iv_i$は

$$\gamma_{xy}=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{S_x}\right)\left(\frac{y_i-\bar{y}}{S_y}\right)=\frac{1}{S_xS_y}\cdot\frac{1}{n}\sum_{i=1}^{n}\{(x_i-\bar{x})(y_i-\bar{y})\}=\frac{S_{xy}}{S_xS_y}.$$

これは x と y の標本相関係数である.

ところで $\frac{1}{n}\Sigma(u_i\pm v_i)^2=\frac{1}{n}\Sigma u_i^{\,2}\pm 2\frac{1}{n}\Sigma u_iv_i+\frac{1}{n}\Sigma v_i^2=1\pm 2\gamma_{xy}+1=2(1\pm\gamma_{xy})$

であるが、2 乗したものの合計は負になることはないから、$1 \pm \gamma_{xy} \geq 0$ である.だから、$-1 \leq \gamma_{xy} \leq 1$でなければならない.

他の証明方法：

$$\Sigma\{(x_i - \bar{x}) \pm \rho(y_i - \bar{y})\}^2 = \Sigma(x_i - \bar{x})^2 \pm 2\rho\Sigma(x_i - \bar{x})(y_i - \bar{y}) + \rho^2\Sigma(y_i - \bar{y})^2$$

が常に正であるから、ρ に関する 2 次式の判別式が負になることを利用する.これはコーシー・シュワルツと同じ証明方法である.

表現上の注意

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}(\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}) = (\frac{1}{n}\sum_{i=1}^{n} x_i y_i) - \bar{x}\,\bar{y} = \overline{xy} - \bar{x}\,\bar{y}$$

と表記されることがある.右端の等号は、「x と y の積の平均から、x の平均と y の平均の積を引く」という意味である.x と y が同じ場合は、次の表現もある.

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i)^2 - (\bar{x})^2 = \overline{x^2} - (\bar{x})^2.$$

問題解答（1 章）

1. 平均値は －8.44、分散は 743.47、だから標準偏差 27.278.従って 2 シグマ区間は -62.97 から 46.096. 2 シグマ区間の度数は 110、全体の度数は 119 で、$(110/119) > (3/4)$ なので、チェビシェフの不等式は妥当である.

2. 単純（算術）平均は、$(10.8 + 6.4 + 5.6 + 6.8 + 7.5)/5 = 7.42$ だから 7.42% となる.次に平均成長率を幾何平均で求めるため、与えられた経済成長率に 1 を加えたものを相乗する. $1.108 \times 1.064 \times 1.056 \times 1.068 \times 1.075 \approx 1.43$. 求めたい平均成長率を R とおくと、$(1+R)^5 = 1.43$. 1.43 の 5 乗根を求めて 1.07405. 7.41%. 後期については 3.4 と 3.398. 所得の変化だけを見ると、 29080/11590=2.509 だから、18 乗根を取り、1.052 となり、5.2%.

3. 標本平均を $\bar{x}$ とおく.$(1/n)\sum_{i=1}^{n} x_i = \bar{x}$ だから、

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}(x_i^2-2\bar{x}x_i+\bar{x}^2)=\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}2\bar{x}x_i+\sum_{i=1}^{n}\bar{x}^2$$

$$=\sum_{i=1}^{n}x_i^2-2\bar{x}\sum_{i=1}^{n}x_i+n\bar{x}^2=\sum_{i=1}^{n}x_i^2-2n\bar{x}^2+n\bar{x}^2=\sum_{i=1}^{n}x_i^2-n\bar{x}^2.$$

4. x の平均を $\bar{x}$、y の平均を $\bar{y}$ とおく

$$.\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}(x_iy_i-\bar{x}y_i-\bar{y}x_i+\bar{x}\,\bar{y})=\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}\bar{x}y_i-\sum_{i=1}^{n}\bar{y}x_i+\sum_{i=1}^{n}\bar{x}\,\bar{y}=$$

$$\sum_{i=1}^{n}x_iy_i-\bar{x}\sum_{i=1}^{n}y_i-\bar{y}\sum_{i=1}^{n}x_i+\sum_{i=1}^{n}\bar{x}\,\bar{y}=\sum_{i=1}^{n}x_iy_i-n\bar{x}\,\bar{y}-n\bar{x}\,\bar{y}+n\bar{x}\,\bar{y}=\sum_{i=1}^{n}x_iy_i-n\bar{x}\,\bar{y}$$

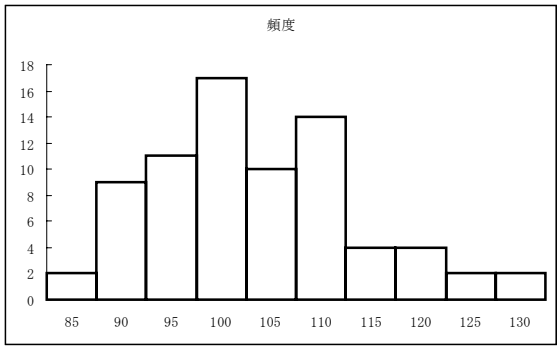
(なぜなら $\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x},\ \frac{1}{n}\sum_{i=1}^{n}y_i=\bar{y}$) （式(1.21)）

5. データの数は 75. 階級数の「目安」を知る為に Starjes の公式に数値をあ

平均	101.44
標準誤差	1.206923
中央値(ﾒｼﾞｱﾝ)	100
最頻値(ﾓｰﾄﾞ)	97
標準偏差	10.45226
分散	109.2497
範囲	50
最小	79
最大	129
合計	7608
最大値(1)	129
最小値(1)	79

ﾃﾞｰﾀ区間	頻度
85	2
90	9
95	11
100	17
105	10
110	14
115	4
120	4
125	2
130	2
次の級	0

てはめる.$1+3.3\log 75\approx 1+3.3\times 1.8751=1+6.18783\approx 7.19$.とりあえず階級数を 10 にして知能指数の度数分布表を作成してみよう.



6. -0.377.

7. ジニ係数の公式は、この問題に関して以下の様に変形できる.

$$\text{ジニ係数} = 2\times\Sigma\{(a\times 0.2)\times(b\times 0.01)\} - \frac{6}{5} = (4\times 10^{-3})\Sigma\, ab - 1.2$$

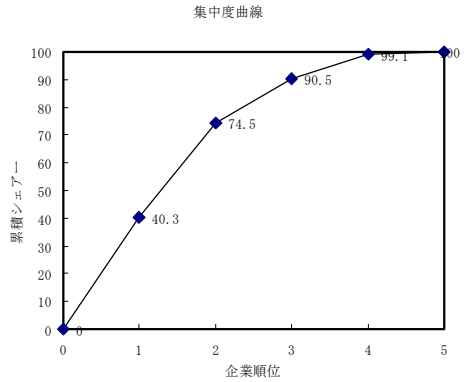
従って、日本の場合、$\Sigma\, ab = 1\times 8.7 + 2\times 13.2 + 3\times 17.5 + 4\times 23.1 + 5\times 37.5 = 367.54$ だから.ジニ係数 $= 0.273$ となる.

8. 0.825

9. 表を基に相関係数を計算する.-0.51.

15	15	15	15	15	15	25	25	25	25	25	25	25	25	35
55	65	65	85	85	85	45	45	45	55	55	65	85	85	45

10.



11. L=(130×270+400×25)/(150×270+360×25)=0.911.

P=(130×320+400×28)/(150×320+360×28)=0.909. 1-(0.911/0.909)=-0.0022.

12. 年平均成長率の解をRとおくと

(i)1880 年から 1940 にかけては $(1+R)^{60} = 3.16$ より, R＝1.93%

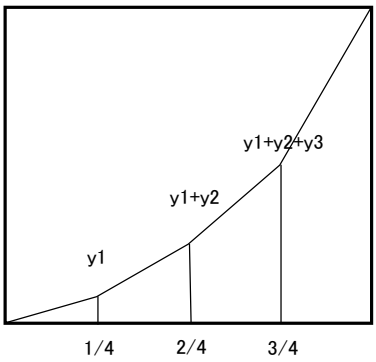
(ii) 1940 年から 1955 年にかけては $(1+R)^{15} = 0.91$ より, R＝-0.63%

(iii) 1955 年から 1990 年にかけては $(1+R)^{35} = 6.71$ より, R＝5.59%

13.　表 1.9 より、相対所得の絶対差の表は次のようになる.総和を取り、2n で割ると 2.8 になる.

0	0.05	0.09	0.15	0.3
0.05	0	0.04	0.1	0.25
0.09	0.04	0	0.06	0.21
0.15	0.1	0.06	0	0.15
0.3	0.25	0.21	0.15	0
0.59	0.44	0.4	0.46	0.91

四人の場合について証明する。



番号	1	2	3	4
相対所得	y_1	y_2	y_3	y_4
累積相対所得	y_1	y_1+y_2	$y_1+y_2+y_3$	$y_1+y_2+y_3+y_4$

図中、$y_1 \le y_2 \le y_3 \le y_4$　かつ　$y_1+y_2+y_3+y_4=1$

ローレンツ曲線下の面積＝　三角形　＋　台形が　3 個（いずれも底面は 1/4）

$$=\frac{1}{2}\times\frac{1}{4}\{y_1+(2y_1+y_2)+(2y_1+2y_2+y_3)+(2y_1+2y_2+2y_3+y_4)\}$$

$$=\frac{1}{8}\{7y_1+5y_2+3y_3+y_4\}$$

$$\textbf{ジニ係数}=1-\frac{\text{多角形}}{\text{三角形}}=1-\frac{1}{4}\{7y_1+5y_2+3y_3+y_4\}=\frac{1}{4}\{-3y_1-y_2+y_3+3y_4\}$$

他方、問 13 で与えられる式は

$$\frac{1}{4}\sum_{i=1}^{j-1}\sum_{j=2}^{n}\left|y_i-y_j\right|=\frac{1}{4}\{-3y_1-y_2+y_3+3y_4\}$$

となり一致する。ただし左辺の和は下の表の要素の和である。

	y_1	y_2	y_3	y_4
y_1	0	y_2-y_1	y_3-y_1	y_4-y_1
y_2		0	y_3-y_2	y_4-y_2
y_3			0	y_4-y_3
y_4				0

問題解答(2 章)

1. 全事象の数は 13×4＝52. 実際引いたカードがハートまたは絵札である事象(A∪B)の数は、22 である.よって確率 P(A∪B)＝22/52.

さて、引いたカードがハートである(A)事象の数は 13. 絵札である（B）事象の数は 12. ハートでかつ絵札である(A∩B)事象の数は 3. 加法定理 P(A∪B)=P(A)＋P(B)－P(A∩B)=13/52＋12/52－3/52=22/52 より先に求めた確率と等しい.

2. 全事象の数は 6×6×6=216. 目の和が 4 以下になる事象の数は（1，1，1)、（1，1、2)、(1，2，1)、(2，1，1）の 4. よって求める確率は 4/216=1/54.

3. 点数の組合せは（10，10，0)、(10，0，10)、(0，10，10)、(5，5，10)、（5，10，5）(10，5，5）の 6 通り. 各々の点数に応じて 2×2×2=8 通りの組合せがある.よって求める組合せの数は 8×6=48.

4. 全事象の数は 20×30＝600.

（2 枚目が 1 枚目より大きな値をとる場合。）1 枚目に引いたカードが 1 の場合、2 枚目は 11 から 30 までであればよいので事象の数は 20. 1 枚目に引いたカードが 2 の場合、2 枚目は 12 から 30 までであればよいから、事象の数は 19.同様に 1 枚目に引いたカードの値が増えると条件を満たす事象の数は減る. 事象の数は、20+19+18+…+1＝210.

（2 枚目が 1 枚目より小さい値をとる場合.）1 枚目に引いたカードが 11 のとき、2 枚目は 1 であればよいので、事象の数は 1.一枚目に引いたカードが 12 のとき、2 枚目は 1 か 2 であればよいから、事象の数は 2．同様にして、1 枚目のカードが 20 の場合、10 である． 事象の総数は

1＋2＋3＋・・・＋10=55. 両方合わせると、確率は 265/600.

5. 目の和が 6 である事象の数．それは（赤、青、緑）が（1，2，3）（1，1，4)、（2, 2, 2）の各組み合わせの中における 3 つの数の順列の総数. 6＋3＋1=10. この条件下で 3 個のサイの目が等しくなるのは（2，2，2）の時だけなのでその事象の数は 1．よって求める条件つき確率は 1/10.

目の和が 9 である事象の数: それは（赤、青、緑）が（1、2，6）（1，3，5)、（1，4，4)、(2，2，5)（2，3，4)（3，3，3）の各組み合わせの中における 3 つの数の順列の総数. 6＋6＋3＋3＋6＋1=25. この条件下で 3 個のサイの目が等しくなるのは（3，3，3）の時だけなのでその事象の数は 1. よって求める条件つき確率は 1/25.

6. a）全事象の数: (男子学生の数)＋(女子学生の数)=(1325＋1200＋950＋1100）＋(1100＋950＋775＋950）=4575＋3775＝8350.

3 年生である事象の数は 950＋775＝1725 であるから､求める確率は 1725/8350.

b）全事象の数は 8350．女子学生でかつ 2 年生である事象の数は 950．よって求める確率は 950/8350=0.114.

c）男子学生である事象の総数は 4575．男子学生でかつ 2 年生である事象の数は 1200 よって求める条件付確率は 1200/4575.

d）独立性の条件から女子学生である条件のもとの 22 歳以上である確率と、一般に 22 歳以上である確率と等しい．このことから、女子学生でありかつ 22 歳以上である確率は女子学生である確率と 22 歳以上である確率の積に等しい.

よって求める確率は

(3775/8350)×(85＋125＋350＋850)/8350=(3775/8350)×(1410/8350)

=0.07634・・.つまりおよそ 7.6%である.

7. a） 1 : P(X∩P) =P(X | P)×P(P) =0.2×0.3=0.06.

4: P(Y∩P)=P(Y | P)×P(P)=(1-P(X | P))×P(P)=(1-0.2)×0.3=0.8×0.3=0.24.

b）ベイズの定理によるべきだが、ここでは 2、5、3、6 の計算を先にする．a と同様にして 2: 0.8×0.5=0.4、5: (1－0.8)×0.5=0.1、3: 0.7×0.2＝0.14、

6: (1－0.7)×0.2=0.06. P(Q | X)は 2/(1,2,3 の総和) だから、

P(Q | X) =0.4/(0.06＋0.4＋0.14)=2/3. また、P(X∪P）は 1，2，3，4 の確率の総和だから、P(X∪P)=0.06＋0.4＋0.14＋0.24=0.84.

c) 独立でない．たとえば、P(X∩P)は 1 の確率だから、0.06．独立ならばこれは P(X)と P(P)の積に等しくなるが、P(X)P(P)=0.6×0.3=0.18.（P(X）は 1，2，3 の確率の総和；0.06＋0.4＋0.14＝0.6）等しくないので独立でない．**独立でないことを示すには**、等号が成立しないことを一つのセルについて示せばよい。

２×２の場合では、一つのセルで等号が成立すれば 4 個の全てのセルについて等号が成立する。次の表では、２と３のセルは行和が x、列和が q になることから容易に求めることができる。４のセルについても同様である。

	Q	R	
X	xq	2	P(X)=x
Y	3	4	P(Y)=y
	P(Q)=q	P(R)=r	1

8. ベイズ定理により $=\dfrac{0.95\times0.3}{0.95\times0.3+0.99\times0.7}\fallingdotseq0.29.$

9. P(A | B)=0.7,P(A | B^C)=0.8. ベイズの定理により

=0.7×0.05/(0.7×0.05＋0.8×0.95)≒0.044.