

## 統計学 1 中間課題 解答

日本大学 文理学部 情報科学科 5419045 高林秀

<解答本文>

まず、数値データである「年齢」「身長」「体重」「最大血圧」「最小血圧」に関して、それぞれの、平均、標準偏差、最小値、最大値を計算したので以下にまとめる。  
(小数点以下第2位を四捨五入)

	年齢	身長	体重	最大血圧	最小血圧
平均	41.3	162.5	60.3	124.9	89.4
標準偏差	12.1	10.2	9.4	9.2	15.2
最大値	67.0	185.0	75.0	144.0	129.0
最小値	23.0	141.0	43.0	105.0	65.0

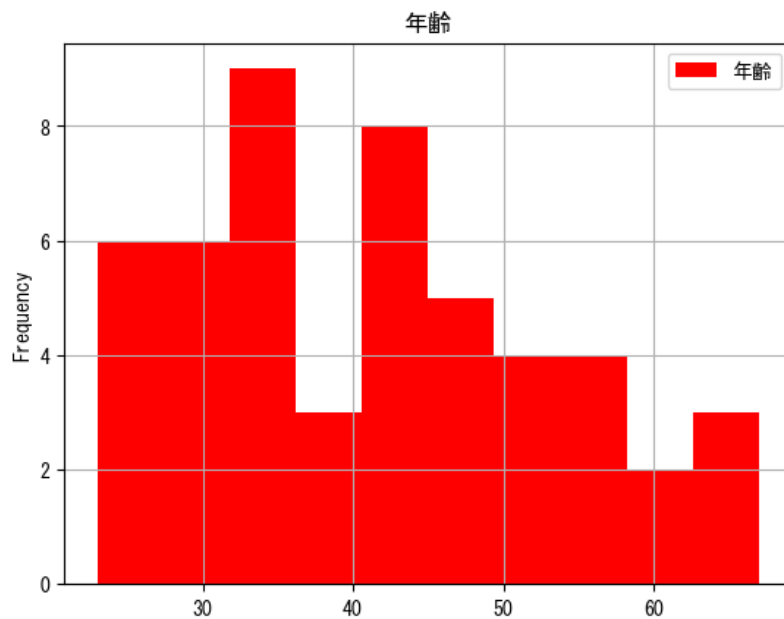
以上の平均と標準偏差の値より与えられたデータの散らばり具合に関して、次のことが言える。

- ・この母集団の年齢層は40歳を平均としており、大体28歳～50歳前後の人のデータが多い。
- ・この母集団の身長は、大体150cm～170cmの人が多い。
- ・この母集団の体重は、平均60kgで、多くのデータは50kg～70kgの範囲内に当てはまる。
- ・同じ血圧の値でも、最小血圧のデータのばらつき具合と、最大血圧のデータのばらつき具合には標準偏差の値から差があることがうかがえる。

次に、それぞれの数値データの度数分布表と対応するヒストグラムを求めてみた。

・年齢

	以上	以下	階級値	度数	相対度数	累積度数	累積相対度数
0	23.0	27.4	25.2	6	0.12	6	0.12
1	27.4	31.8	29.6	6	0.12	12	0.24
2	31.8	36.2	34.0	9	0.18	21	0.42
3	36.2	40.6	38.4	3	0.06	24	0.48
4	40.6	45.0	42.8	8	0.16	32	0.64
5	45.0	49.4	47.2	5	0.10	37	0.74
6	49.4	53.8	51.6	4	0.08	41	0.82
7	53.8	58.2	56.0	4	0.08	45	0.90
8	58.2	62.6	60.4	2	0.04	47	0.94
9	62.6	67.0	64.8	3	0.06	50	1.00

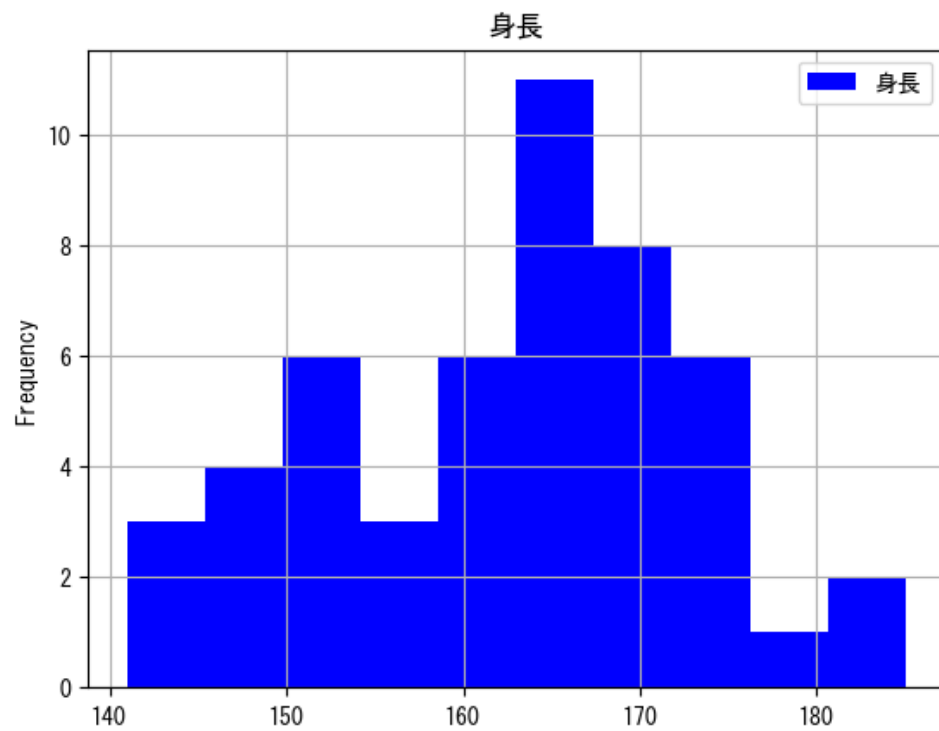


年齢のヒストグラムからも、この母集団のデータは30代から50代のデータが多いことがうかがえる。しかし、30代後半から40第にかけては、度数が3と少なくなっており、局所的にはデータ不足といえるかもしれない。

年齢の面では全体的に、偏りがみられる。

・身長

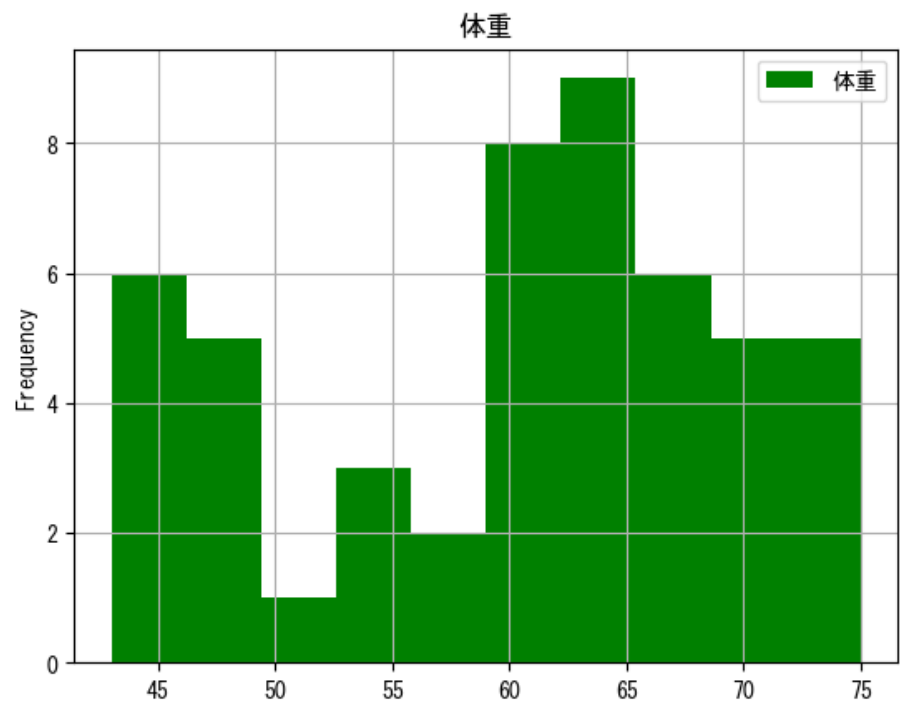
	以上	以下	階級値	度数	相对度数	累積度数	累積相对度数
0	141.0	145.4	143.2	3	0.06	3	0.06
1	145.4	149.8	147.6	4	0.08	7	0.14
2	149.8	154.2	152.0	6	0.12	13	0.26
3	154.2	158.6	156.4	3	0.06	16	0.32
4	158.6	163.0	160.8	6	0.12	22	0.44
5	163.0	167.4	165.2	11	0.22	33	0.66
6	167.4	171.8	169.6	8	0.16	41	0.82
7	171.8	176.2	174.0	6	0.12	47	0.94
8	176.2	180.6	178.4	1	0.02	48	0.96
9	180.6	185.0	182.8	2	0.04	50	1.00



ヒストグラムより、この母集団では身長 160～170cm 程度の人の方がデータが多いことが分かる。特に、身長 163～167cm の区間のデータの度数が突出して多い。

・体重

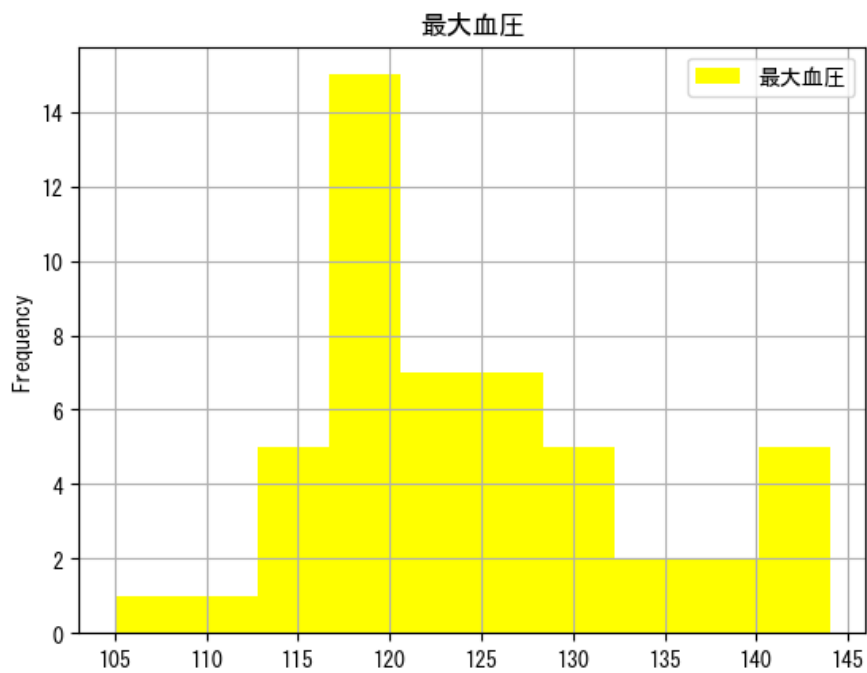
	以上	以下	階級値	度数	相対度数	累積度数	累積相対度数
0	43.0	46.2	44.6	6	0.12	6	0.12
1	46.2	49.4	47.8	5	0.10	11	0.22
2	49.4	52.6	51.0	1	0.02	12	0.24
3	52.6	55.8	54.2	3	0.06	15	0.30
4	55.8	59.0	57.4	2	0.04	17	0.34
5	59.0	62.2	60.6	8	0.16	25	0.50
6	62.2	65.4	63.8	9	0.18	34	0.68
7	65.4	68.6	67.0	6	0.12	40	0.80
8	68.6	71.8	70.2	5	0.10	45	0.90
9	71.8	75.0	73.4	5	0.10	50	1.00



ヒストグラムからも明らかな様に、母集団全体では、体重 60kg 以上の人のデータが多いことがうかがえる。

・最大血圧

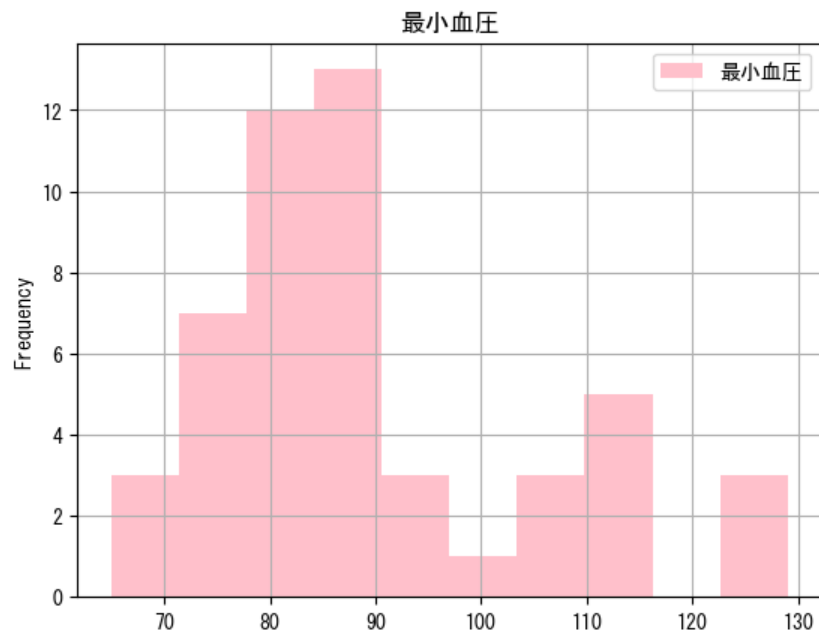
	以上	以下	階級値	度数	相対度数	累積度数	累積相対度数
0	105.0	108.9	106.95	1	0.02	1	0.02
1	108.9	112.8	110.85	1	0.02	2	0.04
2	112.8	116.7	114.75	5	0.10	7	0.14
3	116.7	120.6	118.65	15	0.30	22	0.44
4	120.6	124.5	122.55	7	0.14	29	0.58
5	124.5	128.4	126.45	7	0.14	36	0.72
6	128.4	132.3	130.35	5	0.10	41	0.82
7	132.3	136.2	134.25	2	0.04	43	0.86
8	136.2	140.1	138.15	2	0.04	45	0.90
9	140.1	144.0	142.05	5	0.10	50	1.00



最大血圧に関しては、115~120 の間のデータが多く、ほかの区間に関しては極端にデータが少ないことがうかがえる。

・最小血圧

	以上	以下	階級値	度数	相対度数	累積度数	累積相対度数
0	65.0	71.4	68.2	3	0.06	3	0.06
1	71.4	77.8	74.6	7	0.14	10	0.20
2	77.8	84.2	81.0	12	0.24	22	0.44
3	84.2	90.6	87.4	13	0.26	35	0.70
4	90.6	97.0	93.8	3	0.06	38	0.76
5	97.0	103.4	100.2	1	0.02	39	0.78
6	103.4	109.8	106.6	3	0.06	42	0.84
7	109.8	116.2	113.0	5	0.10	47	0.94
8	116.2	122.6	119.4	0	0.00	47	0.94
9	122.6	129.0	125.8	3	0.06	50	1.00



最小血圧に関しては、80~90 の区間のデータが多く、それ以上のデータが少ないようにうかがえる。特に、116~122 間のデータが 1 つもないことが度数分布表、ヒストグラムからも明らか。

ここからは、男女別に数値データを分析してみる。

各数値データの平均、標準偏差、最大値、最小値を計算したので、表にまとめた。

・男性

(小数点以下第2位を四捨五入)

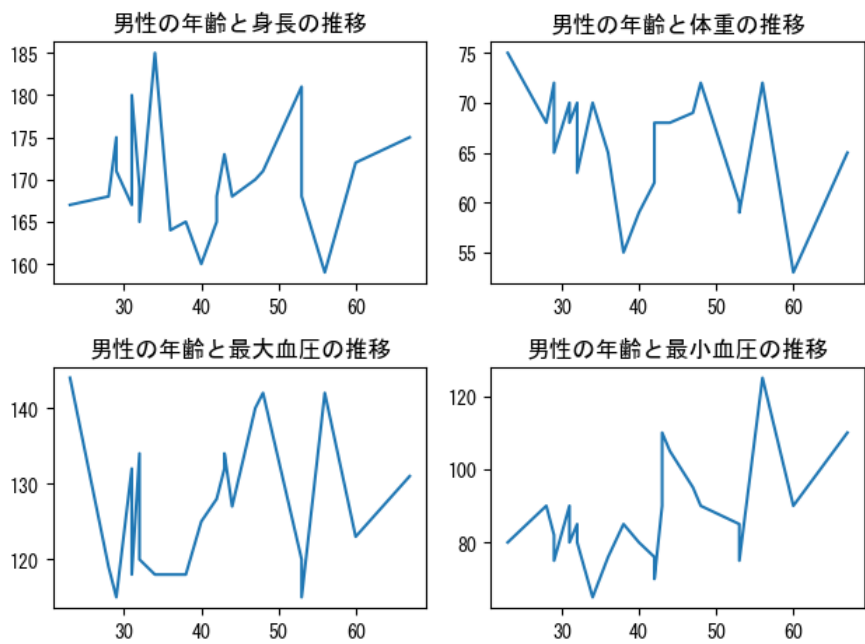
	年齢	身長	体重	最大血圧	最小血圧
平均	40.0	170.0	66.0	126.6	87.0
標準偏差	11.2	6.2	5.6	9.3	13.9
最大値	67.0	185.0	75.0	75.0	125.0
最小値	23.0	159.0	53.0	115.0	65.0

・女性

(小数点以下第2位を四捨五入)

	年齢	身長	体重	最大血圧	最小血圧
平均	41.7	155.7	55.0	123.3	91.6
標準偏差	13.0	8.2	9.2	8.9	16.3
最大値	65.0	172.0	75.0	143.0	129.0
最小値	23.0	141.0	43.0	105.0	65.0

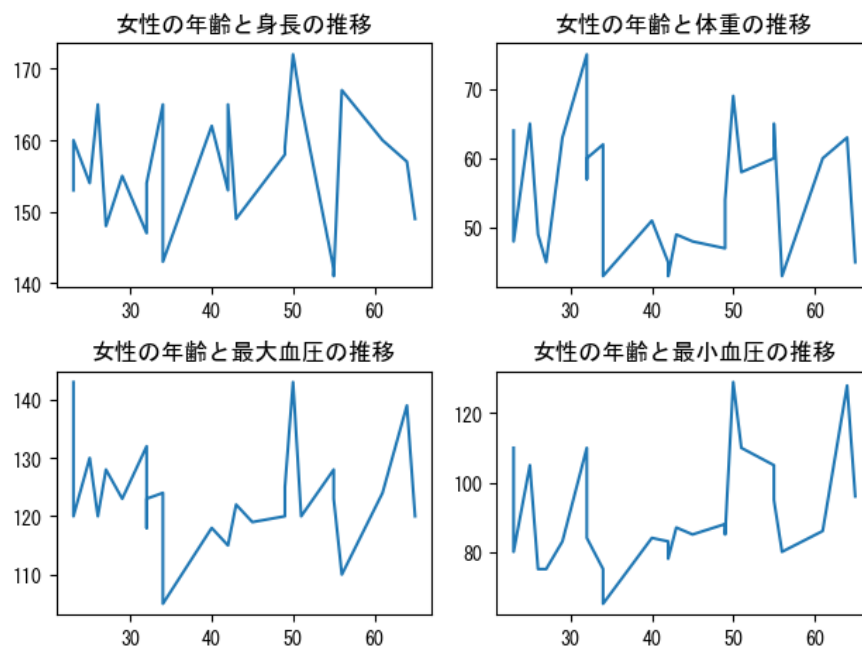
次に、各性別の年齢を横軸にした各数値の推移を図示する。



最小血圧に関しては、年齢が上がるにつれて、若干ではあるが上昇しているようにも見え

る。しかし、実際に相関係数を算出してみると、最小血圧と年齢の相関係数は約0.4であるので、ほとんど両者の間に相関はないことがうかがえる。

次に女性の場合で同様の分析を試みた。



こちらも男性同様、年齢によって数値が大きく開いているせいか、グラフの形状は右肩上がりでも、左肩上がりでもない。相関係数の面では、年齢と身長が-0.6と中程度の負の相関がみられた。男性の場合では、年齢と身長の相関係数は-0.4だったので、女性の方が男性よりも、年齢が上がるにつれ身長が短くなる傾向にあるように思える。

最後に、年齢と心電図、血圧判定の関係について分析する。

年齢と文字データである心電図の関係性はクラメールの関連係数を使って算出することが出来る。

クラメールの関連係数値は0以上1以下の値をとり、1に近いほど2つのデータの関連性が深いことになる。

なお、今回は文字データとして扱いやすくするため、以下のように値を変更している。

- ・ 1 → A
- ・ 2 → B
- ・ 3 → C
- ・ 4 → D



まず、性別を区別せずに、年齢と心電図の  $\chi^2$  乗値を計算すると、86.119 となる。  
よって、年齢と心電図判定のクラメールの関連係数値は 0.054 となり、これは 0.2 よりも小さい値なので、一般的には年齢と心電図判定の結果には関連性が無いと判断される。  
したがって、心電図判定は年齢に大きく依存せずに、個人差のある判定になると言える。

次に、年齢と血圧判定のクラメールの関連係数を計算すると、 $\chi^2$  乗値が 96.218 となり、年齢と血圧判定のクラメールの関連係数値は 0.273 となる、こちらは 0.2 をわずかに上回るので、一般的には年齢と血圧判定には弱い関連性があるという事になる。  
したがって、血圧判定は、年齢と若干の依存関係があり、年齢が血圧判定に影響を及ぼすことがあると考えられる。