第 1 章

1． 年収等を誰でも答えてくれるとは考えにくく，したがって無作為標本とはとらえにくい．送り返された回答に基づく推定には偏りが生じることが予想される．

2． 一応，生産される新しいモデルの自動車全体が母集団であると考えられる．しかし，試作段階でもあり，量産体制にはいった段階で生産されるそのモデルの自動車全体からの無作為標本と考えるには慎重でなければならず，テストはさまざまなステップを踏んで継続されるべきであろう．

第 2 章

1． 100 点満点で平均点が 30 点しかないのだから，分布としては低い点に集中し右に歪んだものと予想される．その試験で平均点よりも少しよい点をとったのだから，真ん中よりある程度上位（中央値よりもある程度高い得点）であると考えられる．

2． 靴のサイズの分布が単峰で左右対称に近い（正規分布に近い）ものであれば，$(\bar{x}-2s, \bar{x}+2s)$ の外にでる割合は 5% 程度である．$\bar{x}-2s=23$，$\bar{x}+2s=27$ とすれば $s=1$ を得る．

3．（イ） 平均 $2\times 25+20=70$ 点，標準偏差 $2\times 5=10$ 点．　（ロ） $a=3$，$b=-5$．

4． ウエストが 73.0 ~ 75.5 の人は $10\times\dfrac{75.5-73.0}{75.5-70.5}=5$ 人，75.5 ~ 80.0 の人は $18\times\dfrac{80.0-75.5}{80.5-75.5}=16.2$ 人と期待される．合計 70 人中 21.2 人だから，約 30% である．

5． 数学，統計学とも得点の分布が単峰で左右対称に近い（正規分布に近い）として，偏差値で比較するのが 1 つの手である．A 君の数学，統計学の成績を偏差値に直すと

$$50+\frac{45-40}{8}\times 10=56.25,\quad 50+\frac{56-50}{6}\times 10=60$$

したがって，統計学の方が受験者全体の中での順位は上であったと考えられる．

6． $\bar{x}=170.7$，$s=5.22$.

7．（イ）

表 1: 度数分布表

階級	階級値	度数	相対度数	累積度数	相対累積度数
1.5 ~ 8.5	5.0	7	0.058	7	0.058
8.5 ~ 15.5	12.0	11	0.092	18	0.150
15.5 ~ 22.5	19.0	13	0.108	31	0.258
22.5 ~ 29.5	26.0	21	0.175	52	0.433
29.5 ~ 36.5	33.0	20	0.167	72	0.600
36.5 ~ 43.5	40.0	17	0.142	89	0.742
43.5 ~ 50.5	47.0	13	0.108	102	0.850
50.5 ~ 57.5	54.0	10	0.083	112	0.933
57.5 ~ 64.5	61.0	6	0.050	118	0.983
64.5 ~ 71.5	68.0	2	0.017	120	1.000
計	—	120	1.000	—	—

（ロ）

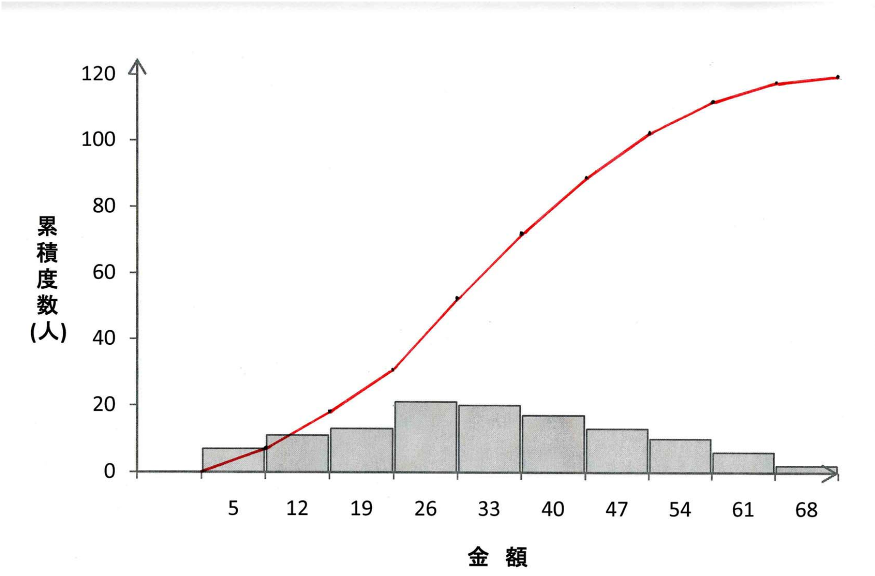

図 1　ヒストグラムと累積度数折れ線

（ハ）図 1 から，第 1 四分位数=22.2，第 2 四分位数=32.5，第 3 四分位数=44.3，四分位範囲=22.1，四分位分散係数=0.340．

（ニ）

表 2: 度数分布表

階級	階級値	度数 (f_i)	u_i	$u_i f_i$	$u_i^2 f_i$
1.5 ～ 8.5	5.0	7	−4	−28	112
8.5 ～ 15.5	12.0	11	−3	−33	99
15.5 ～ 22.5	19.0	13	−2	−26	52
22.5 ～ 29.5	26.0	21	−1	−21	21
29.5 ～ 36.5	33.0	20	0	0	0
36.5 ～ 43.5	40.0	17	1	17	17
43.5 ～ 50.5	47.0	13	2	26	52
50.5 ～ 57.5	54.0	10	3	30	90
57.5 ～ 64.5	61.0	6	4	24	96
64.5 ～ 71.5	68.0	2	5	10	50
計	—	120	—	−1	589

表 2 から $\bar{u} = \dfrac{-1}{120} = -0.01$ となるので，平均値は $\bar{x} = 33.0 - 0.01 \times 7 = 32.93$．最頻値は，最大度数を与える階級の階級値で 26.0．また，$s_u = \sqrt{\left\{589 - \dfrac{(-1)^2}{120}\right\} \dfrac{1}{120-1}} = 2.22$ となるので，標準偏差は $s_x = 7 \times 2.22 = 15.54$．そして，変動係数は，0.472.（中央値は 32.5.）よって，最頻値 $<$ 中央値 $<$ 平均値という大小関係になっている．平均値と中央値はほぼ同じであるが，最頻値については多少の差がある．左右対称に近いが，多少右に歪んでいる．

（ホ）$(\bar{x} - s,\ \bar{x} + s) = (17.39,\ 48.47)$ に含まれる観測値の数：78 個，65.0%.

$(\bar{x} - 2s,\ \bar{x} + 2s) = (1.85,\ 64.01)$ に含まれる観測値の数：118 個，98.3%.

$(\bar{x} - 3s,\ \bar{x} + 3s) = (-13.69,\ 79.55)$ に含まれる観測値の数：120 個．100%.

区間 $(\bar{x}-s,\ \bar{x}+s)$ については，2.4.1 項で与えられている値よりも約 3 ポイント少なく，区間 $(\bar{x}-2s,\ \bar{x}+2s)$ については，約 3 ポイント多い．よって，ヒストグラムは中央部の割合が少し少なく，左右に少し広がっている形をしていると判断できる．

（へ）

表 3: 度数分布表 (層別した場合)

階級	階級値	週末 度数	ウィークデー 度数
1.5 ～ 8.5	5.0	2	5
8.5 ～ 15.5	12.0	2	9
15.5 ～ 22.5	19.0	6	7
22.5 ～ 29.5	26.0	10	11
29.5 ～ 36.5	33.0	9	11
36.5 ～ 43.5	40.0	9	8
43.5 ～ 50.5	47.0	8	5
50.5 ～ 57.5	54.0	7	3
57.5 ～ 64.5	61.0	5	1
64.5 ～ 71.5	68.0	2	0
計	—	60	60

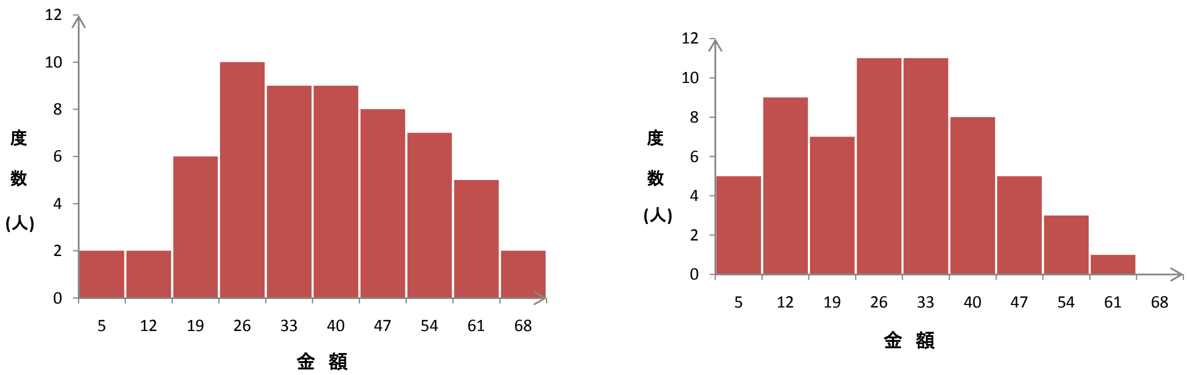

図 2 層別したヒストグラム (左が週末，右がウィークデー)

層別した結果，週末のデータは右（支払う金額が大きい側）に偏り，ウィークデーのデータは左（支払う金額が小さい側）に偏っていることがわかる．

参考までに度数分布表を用いずに計算される値を与えておく．120 個のデータすべてを用いた場合はつぎの通りである．

第 1 四分位数=22，第 2 四分位数=32，第 3 四分位数=44，四分位範囲=22，四分位分散係数=0.344，

平均値 $\bar{x}=32.95$，標準偏差 $s=15.56$，変動係数=0.472，

$(\bar{x}-s,\ \bar{x}+s)=(17.39,\ 48.51)$ に含まれる観測値の数：78 個，

$(\bar{x}-2s,\ \bar{x}+2s)=(1.83,\ 64.07)$ に含まれる観測値の数：118 個，

$(\bar{x}-3s,\ \bar{x}+3s)=(-13.73,\ 79.63)$ に含まれる観測値の数：120 個．

週末のデータだけの場合は，平均値は 37.47，標準偏差は 15.54 である．ウィークデーのデータだけの場合は，平均値は 28.43，標準偏差は 14.34 である．

8．（イ）

表 4: 度数分布表

階級	階級値	度数	相対度数	累積度数	相対累積度数
14.5 ～ 25.5	20.0	9	0.090	9	0.090
25.5 ～ 36.5	31.0	29	0.290	38	0.380
36.5 ～ 47.5	42.0	23	0.230	61	0.610
47.5 ～ 58.5	53.0	19	0.190	80	0.800
58.5 ～ 69.5	64.0	10	0.100	90	0.900
69.5 ～ 80.5	75.0	4	0.040	94	0.940
80.5 ～ 91.5	86.0	3	0.030	97	0.970
91.5 ～ 102.5	97.0	1	0.010	98	0.980
102.5 ～ 113.5	108.0	0	0.000	98	0.980
113.5 ～ 124.5	119.0	1	0.010	99	0.990
124.5 ～ 135.5	130.0	1	0.010	100	1.000
計	—	100	1.000	—	—

（ロ）

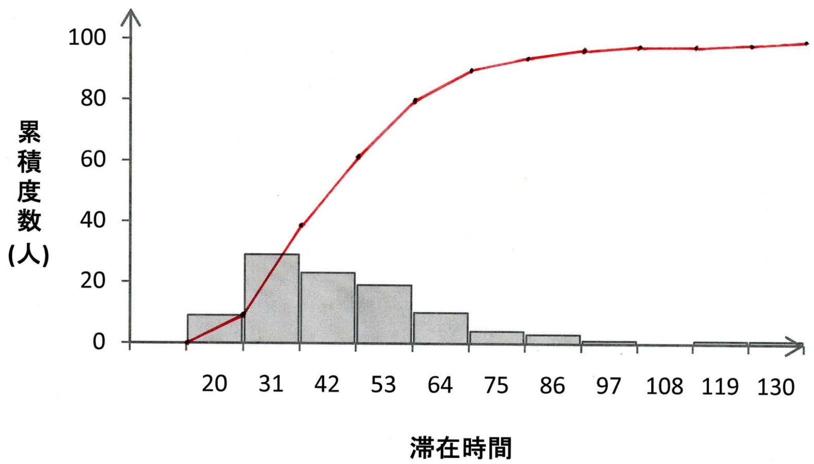

図 3　ヒストグラムと累積度数折れ線

（ハ） 図 3 から，第 1 四分位数=31.8，第 2 四分位数=42.5，第 3 四分位数=55.9，四分位範囲=24.1，四分位分散係数=0.284．

（ニ）つぎのページの表 5 から $\bar{u} = \dfrac{36}{100} = 0.36$ となるので，平均値は $\bar{x} = 42.0 + 0.36 \times 11 = 45.96$. 最頻値は，最大度数を与える階級の階級値で 31.0．また，$s_u = \sqrt{\left(346 - \dfrac{36^2}{100}\right)\dfrac{1}{100-1}} = 1.83$ となるので，標準偏差は $s_x = 11 \times 1.83 = 20.13$．そして，変動係数は，0.438.（中央値は 42.5 である.）よって，最頻値 < 中央値 < 平均値という大小関係になっている．右に歪んだ分布をしているといえる．

（ホ）つぎのページの表 5 から $\bar{u} = \dfrac{21}{98} = 0.21$ となるので，平均値は $\bar{x} = 42.0 + 0.21 \times 11 = 44.31$. 最頻値は，31.0．また，$s_u = \sqrt{\left(233 - \dfrac{21^2}{98}\right)\dfrac{1}{98-1}} = 1.53$ となるので，標準偏差は $s_x = 11 \times 1.53 = 16.83$.（そして，変動係数は，0.380 である.）中央値は 42.

表 5: 度数分布表

階級	階級値	度数 (f_i)	u_i	$u_i f_i$	$u_i^2 f_i$
14.5 ～ 25.5	20.0	9	-2	-18	36
25.5 ～ 36.5	31.0	29	-1	-29	29
36.5 ～ 47.5	42.0	23	0	0	0
47.5 ～ 58.5	53.0	19	1	19	19
58.5 ～ 69.5	64.0	10	2	20	40
69.5 ～ 80.5	75.0	4	3	12	36
80.5 ～ 91.5	86.0	3	4	12	48
91.5 ～ 102.5	97.0	1	5	5	25
102.5 ～ 113.5	108.0	0	6	0	0
113.5 ～ 124.5	119.0	1	7	7	49
124.5 ～ 135.5	130.0	1	8	8	64
計	—	100	—	36	346

（へ）$(\bar{x}-s, \bar{x}+s) = (27.48,\ 61.14)$ に含まれる観測値の数：66 個，67.3 %.

$(\bar{x}-2s, \bar{x}+2s) = (10.65,\ 77.97)$ に含まれる観測値の数：94 個，95.9 %.

$(\bar{x}-3s, \bar{x}+3s) = (-6.18,\ 94.80)$ に含まれる観測値の数：98 個，100 %.

明らかに左右対称ではなくて右に歪んだ分布をしているが，3 つの区間に含まれる観測値の割合はいずれも 2.4.1 項で与えられている値に近い.

（ト）

	度数分布表を用いない場合 100 個の場合	度数分布表を用いない場合 98 個の場合	度数分布表を用いた場合 100 個の場合	度数分布表を用いた場合 98 個の場合
第 1 四分位数	32	32	31.8	31.6
第 2 四分位数	41	40.5	42.5	42.0
第 3 四分位数	55.5	55	55.9	55.0
四分位範囲	23.5	23	24.1	23.4
四分位分散係数	0.287	0.284	0.284	0.279
最頻値			31.0	31.0
平均値 $\bar{x}$	45.79	44.15	45.96	44.31
標準偏差 s	20.14	16.69	20.13	16.83
変動係数	0.440	0.378	0.438	0.380

したがって，中央値および最頻値はほとんど影響を受けていない．平均値は 3.5% 程度小さくなっており，外れ値の影響を受けやすいことがわかる．標準偏差は 16.5% 程度減少しており，外れ値の影響を大きく受けることがわかる.

参考までに，度数分布表を用いずに 2 つの外れ値を除いた場合，それぞれの区間に含まれる観測値

の数はつぎの通りである．

$$100\text{ 個の場合} \Rightarrow 98\text{ 個の場合}$$

$(\bar{x}-s,\ \bar{x}+s) = (25.65,\ 65.93) : 79$ 個 $\Rightarrow$ $(27.46,\ 60.84) : 66$ 個

$(\bar{x}-2s,\ \bar{x}+2s) = (5.51,\ 86.07) : 96$ 個 $\Rightarrow$ $(10.77,\ 77.53) : 94$ 個

$(\bar{x}-3s,\ \bar{x}+3s) = (-14.63,\ 106.21) : 98$ 個 $\Rightarrow$ $(-5.92,\ 94.22) : 98$ 個

第 3 章

1． 夫がゴルフをする・しないと妻がゴルフをする・しないが無関係（独立）であるならば，夫婦そろってゴルフをする割合は 0.06 になることが期待される．(見方 3 である．) したがって 2 つの変数が独立であることを前提としていることになるが，現実にその前提が妥当であるかどうかは極めて疑わしい．

2． 必ずしも意味しない．仮に，理科系学部の方が文科系学部よりも上場企業への就職率が高く，かつ，B 大学の方が A 大学よりも理科系学部の学生の割合がずっと高ければ，大学全体としては B 大学の方が A 大学よりも上場企業への就職率が高いということが起こりうる．

3． 高い正の相関が見られるからといって，それが因果関係を意味するとは限らない．この場合も米の消費量と自動車登録台数の間に何らかの因果関係があるとは考えにくい．むしろ，都市の規模（人口）が大きければ，米の消費量も大きな値になり，また同時に自動車登録台数も大きな値になるため，見せかけの相関が生じている．

4． 傾きが $-45°$ の直線よりも上の部分だけが観測されることになる．したがって，全受験生では正の相関があったとしても，合格者だけについて調査すると，無相関あるいは負の相関になることが想像される．

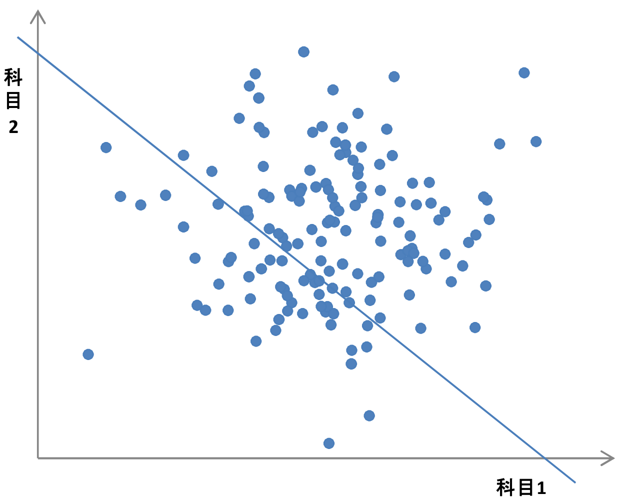

図 4　2 科目の得点の散布図

5． （イ） 100　（ロ） 20　（ハ） 150　（ニ） 30

6． （イ） a: 92% ， b: 8% ， c: 78% ， d: 22% ， e: 83% ， f: 17% （ロ） $\psi = 3.1$ ， $Q = 0.52$ ．

7. （イ）

(i)

	使っている	知っている	計
20 代	15(14.8)	20(20.2)	35
30 代	20(20.2)	28(27.8)	48
計	35	48	83

$\psi = 1.05,\ Q = 0.0244$

(ii)

	使っている・知っている	知らない	計
20 代	35(35.8)	15(14.2)	50
30 代	48(47.2)	18(18.8)	66
計	83	33	116

$\psi = 0.875,\ Q = -0.067$

(iii)

	使っている	知っている	計
20 代・30 代	35(35.6)	48(47.4)	83
40 代	13(12.4)	16(16.6)	29
計	48	64	112

$\psi = 0.897,\ Q = -0.054$

(iv)

	使っている・知っている	知らない	計
20 代・30 代	83(76.4)	33(39.6)	116
40 代	29(35.6)	25(18.4)	54
計	112	58	170

$\psi = 2.17,\ Q = 0.369$

（ロ） 4 つの 2×2 分割表のうち (i) (ii) (iii) は独立な場合のパターンに近く, (iv) だけが関連性を示している．20 代・30 代に比べ 40 代に知らない人が多く，40 代に向けて宣伝広告を行なえば，使用者が増加する可能性があることが示唆される．

8. （イ） $x = 4$ のとき $y = 7$ とする．（ロ） $(1, 2), (2, 4), (3, 5)$ が既に一直線上に乗っていないので，できない．

9. （イ） 図 5 参照．

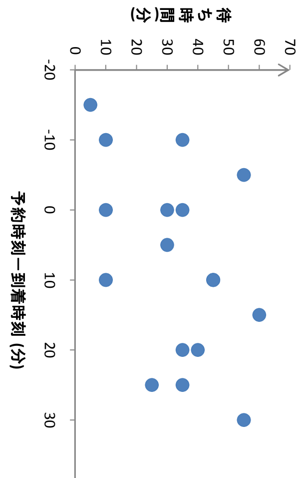

図 5　（予約時間 − 到着時間）と待ち時間

（ロ） $r = 0.414$.　　（ハ） さほど強いものではないが，早めに来る人ほど待ち時間が長いという傾向がでてしまっている．

10. （イ） 2.5.1 項の性質 I より $\bar{u} = (\bar{x} - a)c,\ \bar{v} = (\bar{y} - b)d$ なので

$$s_{uv} = \frac{1}{n-1}\sum_{i=1}^{n}(u_i - \bar{u})(v_i - \bar{v})$$

$$
\begin{aligned}
&= \frac{1}{n-1}\sum_{i=1}^{n}\{(x_i-a)c-(\bar{x}-a)c\}\{(y_i-b)d-(\bar{y}-b)d\} \\
&= \frac{cd}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = cds_{xy}
\end{aligned}
$$

（ロ） u と v の相関係数を r_{uv}，x と y の相関係数を r_{xy} とする．2.5.1 項の性質 I より，$s_u = cs_x$，$s_v = ds_y$ なので

$$
r_{uv} = \frac{s_{uv}}{s_u s_v} = \frac{cds_{xy}}{cds_x s_y} = r_{xy}
$$

第 4 章

1． （イ） $\Omega = \{$ AAAA, AAAD, AADA, AADD, ADAA, ADAD, ADDA, ADDD, DAAA, DAAD, DADA, DADD, DDAA, DDAD, DDDA, DDDD $\}$.

（ロ） $\{$ AAAA, AAAD, AADA, AADD $\}$.

（ハ） $\{$ DDAA, DADA, DAAD, ADDA, ADAD, AADD $\}$.

2． （イ） $\Omega = \{$ ABCD, ABDC, ACBD, ACDB, ADBC, ADCB, BACD, BADC, BCAD, BCDA, BDAC, BDCA, CABD, CADB, CBAD, CBDA, CDAB, CDBA, DABC, DACB, DBAC, DBCA, DCAB, DCBA $\}$.

（ロ）

	A を採用	B を採用	C を採用	D を採用	採用できない
i	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	0
ii	$\frac{11}{24}$	$\frac{5}{24}$	$\frac{1}{12}$	0	$\frac{1}{4}$
iii	$\frac{5}{12}$	$\frac{1}{12}$	0	0	$\frac{1}{2}$
iv	$\frac{1}{4}$	0	0	0	$\frac{3}{4}$

3．

枚数	1	2	$\cdots$	48	49
確率	$\frac{4}{52}$	$\frac{48}{52}\times\frac{4}{51}$	$\cdots$	$\frac{48}{52}\times\frac{47}{51}\times\cdots\times\frac{2}{6}\times\frac{4}{5}$	$\frac{48}{52}\times\frac{47}{51}\times\cdots\times\frac{2}{6}\times\frac{1}{5}$

4． （イ） $\dfrac{2}{3}$.　（ロ） $\dfrac{2}{5}$.　（ハ） 0.7 .

5． （イ） 左辺 $= 0.05+0.10+0.15+0.15+0.05 = 0.5$，右辺$=(0.15+0.15+0.05)+(0.15+0.10+0.05)-0.15 = 0.5$.　（ロ） $1-0.15 = 0.85$.

6． （イ） 0.26.　（ロ） 0.4752.

7． （イ） $\dfrac{1}{2}$.　（ロ） $\dfrac{1}{3}\times\dfrac{1}{2}+\dfrac{1}{3}\times 0+\dfrac{1}{3}\times 1 = \dfrac{1}{2}$.　（ハ） $\dfrac{P(A\cap G)}{P(G)} = \dfrac{\frac{1}{6}}{\frac{1}{2}} = \dfrac{1}{3}$.

（ニ） $P(A|G)$.

8． （イ） 0.0099 .　（ロ） 0.0297 .　（ハ） $\dfrac{1}{3}$.

9． $3!\times 4!\times 2!\times 5! = 34{,}560$,　$4!\times 2!\times 4!\times 2! = 2{,}304$.

10． $\dfrac{1}{3}$,　　$\dfrac{1}{4}$.

11．（イ）$\dfrac{{}_9\mathrm{C}_4}{{}_{12}C_4} = \dfrac{14}{55}$.　　（ロ）$\dfrac{{}_9\mathrm{C}_3 \times {}_3\mathrm{C}_1}{{}_{12}C_4} = \dfrac{28}{55}$.

第 5 章

1．

x	40	60	80	100
$p(x)$	0.10	0.27	0.34	0.29

2． $\dfrac{1}{8}$.

3．（イ）

x	2	3	4
$p(x)$	$\frac{2}{4} \times \frac{1}{3} = \frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

（ロ）$E(x) = \dfrac{10}{3}$,　$V(x) = \dfrac{5}{9}$.

4．（イ）

x	0	1	2
$p(x)$	$\frac{3}{10}$	$\frac{3}{5}$	$\frac{1}{10}$

（ロ）$E(x) = \dfrac{4}{5}$,　$\sqrt{V(x)} = \dfrac{3}{5}$.

5．（イ）

x	1	2	3	4
$p(x)$	$\frac{1}{4}$	$\frac{3}{4} \times \frac{1}{3} = \frac{1}{4}$	$\frac{3}{4} \times \frac{2}{3} \times \frac{1}{2} = \frac{1}{4}$	$\frac{1}{4}$

（ロ）$p(x) = \left(\dfrac{3}{4}\right)^{x-1} \dfrac{1}{4}$,

$x = 1, 2, 3, \ldots$

6． $\mu = 1, \quad \sigma = 1.$

7． 損失の期待値を計算して比較する．190 部の場合，損失の期待値は，

$$2{,}000 \times \frac{1}{8} + 0 \times \frac{1}{4} + 4{,}000 \times \frac{1}{2} + 8{,}000 \times \frac{1}{8} = 3{,}250$$

円である．同様に，200 部の場合 1,500 円，210 部の場合 2,750 円が損失の期待値なので，200 部にすべきである．

第 6 章

1．（イ）$\dfrac{35}{128}$.　　（ロ）$\dfrac{37}{256}$.

（ハ）8 銘柄について上昇するか下降するかを観測することが，ベルヌーイ試行列と考えられるかどうかが問題である．特に，8 銘柄のそれぞれが上昇・下降することが独立に起こると考えられるかどうかがポイントである．独立であると想定するのにはかなり無理があり，その意味で（イ）,（ロ）で計算した確率にもあまり信頼は置けない．

2． $\dfrac{5}{16}$.　$\dfrac{1}{2}$ に増える．

3． 2 項分布 $B(12, 0.9)$ で 11 以上の値のでる確率を求めればよい．約 0.659 である．

4． 40m 歩くということは，5m 歩くことを 8 回繰り返すことになる．8 回中東へ歩く回数 x の確率分布は 2 項分布 $B\left(8, \dfrac{1}{2}\right)$ と考えられる．元の居酒屋へ戻るのは $x = 4$ の場合であり，確率は $\dfrac{35}{128}$ で

ある．

5． $\dfrac{7}{128}$．ランニングが肺活量を増さないという仮説が正しければ，8 人以上も肺活量が増加するという事象が起こる確率は $\dfrac{7}{128}$ である．この確率の値が大変小さくそのような事象は起こりにくいと考えるならば，起こりにくい事象が現実に起こったと考えるよりは，むしろ，仮説を否定する方が自然である．

6． 平均 2 のポアソン分布 $Po(2)$ であるから，$p(x) = \dfrac{2^x}{x!}e^{-2}$．　（イ） $p(0) = e^{-2}$, 約 0.135 である．（ロ） $p(1) = 2e^{-2}$，約 0.271 である．　（ハ） $p(2) = 2e^{-2}$，$1 - (e^{-2} + 2e^{-2} + 2e^{-2}) = 1 - 5e^{-2}$，つまり約 0.323 である．

7． 平均 2 のポアソン分布 $Po(2)$ であるから，$p(x) = \dfrac{2^x}{x!}e^{-2}$．　（イ） $p(0) + p(1) = 3e^{-2}$, 約 0.406 である．　（ロ） $1 - 5e^{-2}$，約 0.323 である．

8． $p(x) = \dfrac{3^x}{x!}e^{-3}$．　（イ） $p(0) = e^{-3}$, 約 0.050 である．　（ロ） $\dfrac{p(1)}{1 - p(0)} = \dfrac{3e^{-3}}{1 - e^{-3}}$，約 0.157 である．

9． 2 項分布 $B(100, \frac{1}{100})$ の場合：$p(x) = {}_{100}\mathrm{C}_x \left(\frac{1}{100}\right)^x \left(\frac{99}{100}\right)^{100-x}$，　$p(0) = \left(\frac{99}{100}\right)^{100} = 0.3660$,　$p(1) = \left(\frac{99}{100}\right)^{99} = 0.3697$,　$p(2) = \frac{100\times 99}{2\times 1}\left(\frac{1}{100}\right)^2\left(\frac{99}{100}\right)^{98} = 0.1849$.

ポアソン分布 $Po(1)$ の場合：$p(x) = \dfrac{1}{x!}e^{-1}$, $p(0) = e^{-1} = 0.3679$, $p(1) = e^{-1} = 0.3679$, $p(2) = \dfrac{1}{2}e^{-1} = 0.1839$.

10． $p(x) = \dfrac{2^x}{x!}e^{-2}$ とすると，

$$0 \times p(0) + 7{,}000 \times p(1) + 14{,}000 \times p(2) + 21{,}000 \times [1 - \{p(0) + p(1) + p(2)\}] - 8{,}000 = 4{,}474$$

円である．

11． （イ） $\dfrac{1{,}000 - 1{,}003}{2} = -1.5$，$Q(1.5) = 0.0668$，約 6.7 %.

（ロ） $\dfrac{998 - 1{,}003}{2} = -2.5$,　$Q(2.5) = 0.0062$,　$\dfrac{1{,}004 - 1{,}003}{2} = -0.5$,　$Q(0.5) = 0.3085$, $1 - (0.0062 + 0.3085) = 0.6853$，約 69% ．

12． （イ） $\dfrac{4.985 - 5.000}{0.01} = -1.5$,　$\dfrac{5.015 - 5.000}{0.01} = 1.5$, $1 - 2Q(1.5) = 0.8664$，約 87% ．

（ロ） $\dfrac{4.985 - 5.005}{0.01} = -2.0$,　$\dfrac{5.015 - 5.005}{0.01} = 1.0$, $1 - \{Q(1.0) + Q(2.0)\} = 0.8185$，約 82% ．

（ハ） 平均を μ とおく．$u(0.24) = 0.706$ より $\dfrac{5.015 - \mu}{0.01} = 0.706$ ，つまり，$\mu = 5.0079$ ．

13． （イ） $\dfrac{61 - 65}{5} = -0.8$, $Q(0.8) = 0.2119$, $\dfrac{63.5 - 65}{5} = -0.3$, $Q(0.3) = 0.3821$, $0.3821 - 0.2119 = 0.1702$，約 17% ．

（ロ） $u(0.1) = 1.282$,　$\dfrac{a - 65}{5} = 1.282$，つまり，$a = 71.41$，約 71.4 cm.

14． 2 項分布 $B(50, 0.8)$ を正規分布 $N(40, (\sqrt{8})^2)$ で近似する．余りがでるのは 41 名以下の場合だから，連続修正をして

$$\frac{41.5-40}{\sqrt{8}} = 0.5303, \quad Q(0.53) = 0.2981, \quad 1-0.2981 = 0.7019$$

約 70% である．同様に不足するのは

$$\frac{42.5-40}{\sqrt{8}} = 0.8839, \quad Q(0.88) = 0.1894,$$

約 19% である．

15． 200,000 は 100 に比べ十分大きいので，100 人中の黒人の数の確率分布を 2 項分布 $B(100, 0.2)$ で近似する．さらに，それを正規分布 $N(20, 4^2)$ で近似する．連続修正を用いると

$$\frac{12.5-20}{4} = -1.875, \quad Q(1.88) = 0.0301$$

約 3% である．

16． 2 項分布 $B\left(21{,}600, \dfrac{1}{25}\right)$ を正規分布 $N(864, 28.8^2)$ で近似する．

（イ） $\dfrac{899.5-864}{28.8} = 1.233, \quad Q(1.23) = 0.1093$，約 11% である．

（ロ） $u(0.01) = 2.326, \ \dfrac{a-864}{28.8} = 2.326$ より $a = 931$ を得る．

第 7 章

1．（イ） 例えば，$P\{x=0, y=0\} = 0.1, \ \ P\{x=0\} \times P\{y=0\} = 0.5 \times 0.3 = 0.15$ と両者が等しくないので，独立ではない．

（ロ） $E(x) = 0.5, \ \ V(x) = 0.25, \ \ E(y) = 1, \ \ V(y) = 0.6$.

（ハ）

$x+y$	0	1	2	3
確率	0.1	0.4	0.4	0.1

（ニ） $E(x+y) = 1.5, \quad V(x+y) = 0.65$，したがって，$E(x+y) = E(x) + E(y)$ は成り立っているが，$V(x+y) = V(x) + V(y)$ は成り立っていない．

2．（イ） $P\{x=y\} = 0.45, \ P\{y>x\} = 0.30$.

（ロ） $E(x) = 0.88, \ \ V(x) = 0.6056, \ \ E(y) = 0.95, \ \ V(y) = 0.6075$.

（ハ） $E(xy) = 0.90$ より $\mathrm{Cov}(x,y) = 0.90 - 0.88 \times 0.95 = 0.064$,

$$\rho_{xy} = \frac{0.064}{\sqrt{0.6056 \times 0.6075}} = 0.1055.$$

3． 正規分布 $N(360, (10\sqrt{2})^2)$ で，365 より小さな値のでる確率だから

$$\frac{365-360}{10\sqrt{2}} = 0.3536, \quad 1 - Q(0.35) = 0.6368$$

約 64% である．

4． 4 日間の販売数量は，平均 $30 \times 4 = 120$，標準偏差 $\sqrt{4 \times 5^2} = 10$ の正規分布である．そのとき，127 以上の値のでる確率は

$$\frac{126.5 - 120}{10} = 0.65, \quad Q(0.65) = 0.2578$$

より，約 0.26 である．

5．（イ）正規分布 $N(160, 1)$．　（ロ）100 個の丸薬入りのびんの重量の分布は正規分布 $N(260, (\sqrt{5})^2)$ である．このとき，256 以上の値のでる確率は

$$\frac{256 - 260}{\sqrt{5}} = -1.789, \quad 1 - Q(1.79) = 1 - 0.0367 = 0.9633$$

より，約 96.3% である．

（ハ） びんの重量の標準偏差を σ とすると，$u(0.01) = 2.326$ より $\dfrac{4}{\sqrt{1+\sigma^2}} = 2.326$，つまり，$\sigma^2 = -1 + \left(\dfrac{4}{2.326}\right)^2$．したがって σ の値を約 1.40 g にしなければならない．

第 8 章

1． 15 通り．a を含むのは 5 通り．したがって a を含む標本が得られる確率は $\dfrac{1}{3}$ である．

2．（イ） $\{(-15,-15), (-15,-5), (-15,5), (-5,-15), (-5,-5), (-5,5), (5,-15), (5,-5), (5,5)\}$.

（ロ）

$x_1 \backslash x_2$	-15	-5	5
-15	0.09	0.18	0.03
-5	0.18	0.36	0.06
5	0.03	0.06	0.01

（ハ）　（ニ）

$x_1 - x_2$	-20	-10	0	10	20
確率	0.03	0.24	0.46	0.24	0.03

（ホ） $E(x_1 - x_2) = 0$.　（ヘ） $V(x_1 - x_2) = 72$.　（ト） $V(x_1) = V(x_2) = 36$.

3． 正規分布 $N(2.1, (0.02 \times \sqrt{2})^2)$.

$$\frac{2.06 - 2.1}{0.02\sqrt{2}} = -1.414, \quad Q(1.41) = 0.0793$$

より，約 8% である．

4．（イ） 平均 $30 \times 100 = 3{,}000$（秒），標準偏差 $\sqrt{100} \times 2 = 20$（秒）．

（ロ） 中心極限定理から正規分布 $N(3{,}000,\ 20^2)$ で近似されると考えられるので，

$$\frac{3{,}020 - 3{,}000}{20} = 1.0, \quad Q(1.0) = 0.1587.$$

約 0.16 である．

5．（イ） 略.　　（ロ） 平均 $4.5 \times 36 = 162$，分散 $36 \times 8.25 = 297$.

（ハ） 中心極限定理から正規分布 $N(162{,}297)$ で近似されると考えられるので，

$$\frac{180 - 162}{3\sqrt{33}} = 1.044, \quad Q(1.04) = 0.1492$$

約 0.15 である．

6．

投げる回数	表のでる回数		表のでる割合	
	平均	標準偏差	平均	標準偏差
100	50	5	0.5	0.05
2,500	1,250	25	0.5	0.01
10,000	5,000	50	0.5	0.005
1,000,000	500,000	500	0.5	0.0005

第 9 章

1． μ の値は，人間にはわからないが 1 つの固定された値であると考えるので，“μ が区間 $(10.2, 12.8)$ に含まれる確率” というものは考えられない．95% 信頼区間の 95% の意味は，標本をとり信頼区間を計算することを数多く繰り返したとすれば，100 回中平均的に 95 回は計算された信頼区間が平均 μ の値を含むという意味である．標本分布を基礎にして信頼度を評価しているのである．

2．（イ） $1{,}050 \pm t(15, 0.05) \times \dfrac{100}{4} = (997,\ 1{,}103)$ ．

（ロ） $n = \dfrac{1.96^2 \times 100^2}{10^2} = 384.16$ ． 385 程度である．

3．（イ） 64 人についての平均値 $\bar{x}$ が正規分布 $N\left(\mu, \left(\dfrac{1}{4}\right)^2\right)$ にしたがうので

$$P\{-0.3 < \bar{x} - \mu < 0.3\} = 2 \times \{0.5 - Q(1.2)\} = 0.7698$$

約 0.77 である．

（ロ） $1.645 \times \dfrac{2.0}{\sqrt{n}} \times 2 \leq 0.4$ であるためには $n \geq 270.6$ つまり $n \geq 271$ でなければならない．

4．（イ） $11.2 \pm 1.96 \times \dfrac{3.0}{\sqrt{12}} = (9.50, 12.90)$.

（ロ） 標本標準偏差 $s = 3.01$, $t(11, 0.05) = 2.201$ なので $11.2 \pm 2.201 \times \dfrac{3.01}{\sqrt{12}} = (9.29, 13.11)$.（イ）での σ の値 3.0 と（ロ）での s の値 3.01 はほとんど同じだが,（ロ）の場合の方が信頼区間の幅は広い.

5． (iii) が正しいと考えられる．両市とも人口は標本の大きさ 1,000 に比べると大きいので，1,000 人中の賛成する人数の確率分布は 2 項分布 $B(1{,}000,\ p)$ であると考えてよい．ここで p は（両市でその値がほぼ等しい）賛成する人の割合である．このように考えると，もはや人口の違いは何も精度に影響しない．

6．（イ） $n = \dfrac{1.96^2 pq}{0.02^2} \leq \dfrac{1.96^2}{0.02^2 \times 4} = 2{,}401$ である．

（ロ） 95% 信頼区間の幅は $1.96 \times \sqrt{\dfrac{\hat{p}\hat{q}}{2500}} \times 2 \leq \dfrac{3.92}{50} \times \dfrac{1}{2} = 0.0392$.

7.（イ）

$n \setminus p$	0.1	0.2	0.3	0.4	0.5
5,000	0.8	1.1	1.3	1.4	1.4
2,500	1.2	1.6	1.8	1.9	2.0
1,000	1.9	2.5	2.8	3.0	3.1
500	2.6	3.5	4.0	4.3	4.4
100	5.9	7.8	9.0	9.6	9.8

ただし，単位は% である．

（ロ）

$e \setminus p$	0.1	0.2	0.3	0.4	0.5
0.05	139	246	323	369	385
0.02	865	1,537	2,017	2,305	2,401
0.01	3,458	6,147	8,068	9,220	9,604

8.（イ）p_1： $\dfrac{1{,}050}{1{,}500} \pm 1.96 \times \sqrt{\left(\dfrac{1{,}050}{1{,}500} \times \dfrac{450}{1{,}500}\right) \times \dfrac{1}{1{,}500}} = (0.677,\ 0.723)$

p_2： $\dfrac{990}{1{,}500} \pm 1.96 \times \sqrt{\left(\dfrac{990}{1{,}500} \times \dfrac{510}{1{,}500}\right) \times \dfrac{1}{1{,}500}} = (0.636,\ 0.684)$

（ロ） $\dfrac{1{,}050-990}{1{,}500} \pm 1.96 \times \sqrt{\dfrac{1{,}050}{1{,}500} \times \dfrac{450}{1{,}500} \times \dfrac{1}{1{,}500} + \dfrac{990}{1{,}500} \times \dfrac{510}{1{,}500} \times \dfrac{1}{1{,}500}} = (0.007,\ 0.073)$

（イ）では 2 つの 95% 信頼区間に重なり合う部分があるが,（ロ）では 95% 信頼区間に 0 が含まれないことに注意する．

第 10 章

1. 第 1 種の誤りは，ロットが良質で受け入れるべきものであるにもかかわらずそのロットを受け入れないという誤りである．これは生産者にとって不利益を生じる誤りであり，生産者危険とよぶのが適切である．一方，第 2 種の誤りは，ロットが良質ではなくて受け入れるべきものでないにもかかわらずそのロットを受け入れてしまうという誤りである．これは消費者にとって不利益を生じる誤りであり，消費者危険とよぶのが適切である．

2.（イ） これは標準正規分布 $N(0,1)$ において $1.645-\sqrt{9}=-1.355$ よりも小さい値のでる確率である．$Q(1.36)=0.0869$ より，約 9% である．

（ロ） $u(0.01)=2.326$ より $1.645-\sqrt{n} \leq -2.326$ を n について解けば $n \geq 16$ を得る．

3.（イ）$\mu_0 - t(n-1, 0.05)\dfrac{s}{\sqrt{n}} \leq \bar{x} \leq \mu_0 + t(n-1, 0.05)\dfrac{s}{\sqrt{n}}$ という不等式を μ_0 についての不等式に書き直すと

$$\bar{x} - t(n-1, 0.05)\frac{s}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + t(n-1, 0.05)\frac{s}{\sqrt{n}}$$

つまり

$$\text{区間}\left[\ \bar{x} - t(n-1, 0.05)\frac{s}{\sqrt{n}},\ \bar{x} + t(n-1, 0.05)\frac{s}{\sqrt{n}}\ \right]$$

である．

（ロ） 両端点 $\bar{x} \pm t(n-1, 0.05)\dfrac{s}{\sqrt{n}}$ が含まれることを除けば,（イ）で得られた結果は μ の 95% 信頼区間と一致している．つまり，両側検定を行なって棄却されないような μ_0 の値の全体が信頼区間を構成しているのである．

4． 正規分布 $N(\mu, 13.5^2)$ をしている母集団から大きさ 40 の無作為標本をとり，その結果 $\bar{x} = 52.9$ が得られたと考える．$H_0 : \mu = 60.2$ を $H_1 : \mu \neq 60.2$ に対して有意水準 0.05 で検定してみると

$$\bar{x} = 52.9 < 56.0 = 60.2 - 1.96 \times \frac{13.5}{\sqrt{40}}$$

となり棄却される．したがって平均点がこんなにも低くなるのは，とても偶然とは考えにくく，A 先生に反省を求めるべきであるという結論になる．

5． 100 組の兄弟についてのデータであり，$\bar{x}_1$ と $\bar{x}_2$ が独立に分布していると考えてはいけない状況である．したがって平均値の差の検定を適用できないので，各兄弟についてその差をとり，$H_0 : \mu = 0$ の検定を行なう方が適切である．

6．（イ） 第 1 種の誤りは，生産工程に異常が生じていないにもかかわらず，異常が生じていると判断してしまうという誤りであり，第 2 種の誤りは，生産工程に異常が生じているにもかかわらず，それを見過ごすという誤りである．

（ロ） 第 1 種の誤りを犯す確率を求めればよい．それは $\bar{x}$ が正規分布 $N(50,\ (1/5)^2)$ にしたがうとき，49.5 未満または 50.5 を超える値のでる確率である．

$$\frac{49.5 - 50}{1/5} = 2.5, \quad \frac{50.5 - 50}{1/5} = 2.5, \quad 2 \times Q(2.5) = 0.0124$$

より，0.0124 である．

（ハ） $\mu = 49.2$ のとき：正規分布 $N(49.2,\ (1/5)^2)$ において区間 $[49.5,\ 50.5]$ に属する値のでる確率を求めればよい．

$$\frac{49.5 - 49.2}{1/5} = 1.5, \quad \frac{50.5 - 49.2}{1/5} = 6.5,$$

$Q(6.5)$ はほとんど 0 なので，$Q(1.5) - Q(6.5) \fallingdotseq Q(1.5) = 0.0668$, 約 0.07．$\mu = 49.4$ のときも同様にして求めると，約 0.31 であり，$\mu = 49.6$ のときは，約 0.69 である．

7． 適合度検定を行なってみる．

$$Z = \frac{(15 - 24)^2}{24} + \frac{(39 - 40)^2}{40} + \frac{(26 - 16)^2}{16} = 9.65$$

$\chi^2(2, 0.05) = 5.99$, $\chi^2(2, 0.01) = 9.21$, $\chi^2(2, 0.005) = 10.60$ より，このような食い違いの起こる確率は 0.01 未満 0.005 以上であることがわかる．

8． $n_1 = n_2 = 600$, $\hat{p}_1 = \dfrac{320}{600} = \dfrac{8}{15}$, $\hat{p}_2 = \dfrac{280}{600} = \dfrac{7}{15}$, $\hat{p} = \dfrac{320 + 280}{600 + 600} = 0.5$ である．

$$|\hat{p}_1 - \hat{p}_2| = 0.0667 > 0.0566 = 1.96 \times \sqrt{0.5^2 \times \left(\frac{1}{600} + \frac{1}{600}\right)}$$

より，仮説を棄却する．父親の学歴により子供の大学進学率に差があるといえる．

$$p_1\text{の 95\%信頼区間：}\frac{8}{15} \pm 1.96 \times \sqrt{\left(\frac{8}{15} \times \frac{7}{15}\right) \times \frac{1}{600}} = (0.493, 0.573)$$

$$p_2\text{の 95\%信頼区間：}\frac{7}{15} \pm 1.96 \times \sqrt{\left(\frac{7}{15} \times \frac{8}{15}\right) \times \frac{1}{600}} = (0.427, 0.507)$$

であり，2 つの区間は重なり合うが，$H_0 : p_1 = p_2$ は有意水準 0.05 で棄却される．

9．（イ）

有意水準 α	0.1	0.05	0.02	0.01
c	0.0582	0.0693	0.0822	0.0911

（ロ）

α	0.1	0.05	0.02	0.01
$x = 113$	する	しない	しない	しない
$x = 117$	する	する	する	しない

（ハ） 有意水準 α が 0.0768 より大きい場合には棄却し，小さい場合には棄却しないので矛盾はしていない．$x \geq 117$ を $x \geq 116.5$ と考えれば，$\hat{p} \geq 0.585$ は $\hat{p} \geq 0.5825$ と書き直されるので

$$\frac{0.5825 - 0.5000}{1/(20\sqrt{2})} = 2.333, \quad Q(2.33) = 0.0099, \quad 0.0099 \times 2 = 0.0198$$

より，0.0198 を得る．これも，もちろん,（ロ）の結果に矛盾していない．

第 11 章

1． 直線回帰モデルでは，説明変数 x は確率変数とは考えていない．その意味で，x 軸方向には偶然変動はなく，y 軸方向にだけ偶然変動がはいると想定している．そのため，y 軸に平行な方向に距離を測るのが自然なのである．

2．

$$n\hat{\alpha} + \sum x_i \hat{\beta} = \sum y_i \tag{1}$$

$$\sum x_i \hat{\alpha} + \sum x_i^2 \hat{\beta} = \sum x_i y_i \tag{2}$$

$(2) - (1) \times \dfrac{\sum x_i}{n}$ により

$$\left\{ \sum x_i^2 - \frac{(\sum x_i)^2}{n} \right\} \hat{\beta} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

$$\sum (x_i - \bar{x})^2 \hat{\beta} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

したがって

$$\hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2}$$

また，(1) 式から

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

を得る．

3. （イ） 散布図は図 6 のとおり．

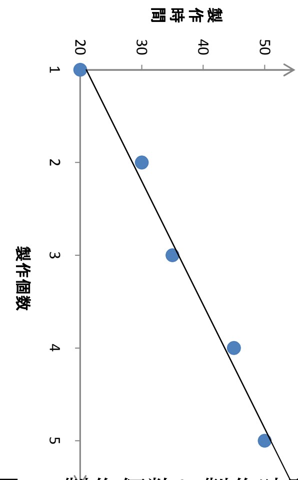

図 6　製作個数と製作時間

（ロ）

	x_i	y_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$	x_i^2	x_iy_i
	1	20	-2	4	-16	256	32	1	20
	2	30	-1	1	-6	36	6	4	60
	3	35	0	0	-1	1	0	9	105
	4	45	1	1	9	81	9	16	180
	5	50	2	4	14	196	28	25	250
計	15	180	–	10	–	570	75	55	615

$\bar{x}=3,\ \bar{y}=36$

（ハ）

$$5\hat{\alpha}+15\hat{\beta}=180,\quad 15\hat{\alpha}+55\hat{\beta}=615$$

これを解いて $\hat{\alpha}=13.5,\ \hat{\beta}=7.5$ を得る．

（ニ）

$$\hat{\beta}=\frac{s_{xy}}{s_x^2}=\frac{75}{10}=7.5,\qquad \hat{\alpha}=\bar{y}-\hat{\beta}\bar{x}=36-7.5\times 3=13.5\ .$$

（ホ） 回帰式は $\hat{y}=13.5+7.5x$ となる．$x=6$ を代入すれば，$\hat{y}=13.5+7.5\times 6=58.5$ より 58.5 と推定される．

（ヘ）

	y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$
	20	21	-1	1
	30	28.5	1.5	2.25
	35	36	-1	1
	45	43.5	1.5	2.25
	50	51	-1	1
計	–	–	0	7.5

残差平方和は，$S_e=7.5$.（ロ）において，$S_T=570$ から，

決定係数は，$r^2=1-\dfrac{S_e}{S_T}=1-\dfrac{7.5}{570}=\dfrac{75}{76}=0.9868$.

σ^2 の推定量は，$s^2=\dfrac{7.5}{5-2}=2.5$.

（ト） $7.5\pm t(3,0.05)\times\dfrac{\sqrt{2.5}}{\sqrt{4}\times\sqrt{\frac{10}{4}}}=(5.91,9.09)$

（チ）$|\hat{\beta}| = 7.5 > 1.591 = t(3,\ 0.05) \times \dfrac{\sqrt{2.5}}{\sqrt{4} \times \sqrt{\frac{10}{4}}}$ より仮説を棄却する．

4．（イ）散布図は図 7 のとおり．

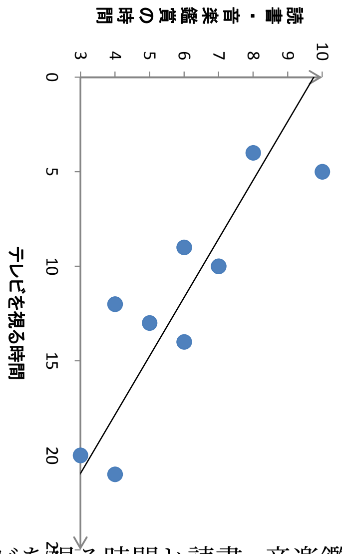

図 7　テレビを視る時間と読書 · 音楽鑑賞の時間

（ロ）$\hat{y} = 9.758 - 0.3225x$　（ハ）$S_e = 10.19,\ r^2 = 0.738$　（ニ）$(-0.494,\ -0.151)$

（ホ）$|\hat{\beta}| = 0.3225 > 0.1717 = t(7,\ 0.05) \times 0.07262$ より仮説を棄却する．

5．（イ）散布図は図 8 のとおり．

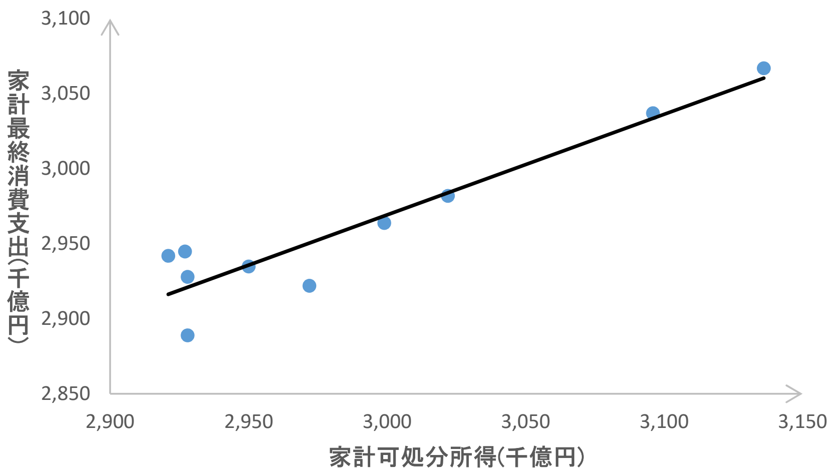

図 8　家計可処分所得と家計最終消費支出

（ロ）$\hat{y} = 959.5 + 0.6699x$．家計可処分所得が増加すると，およそ 3 分の 2 が家計最終消費支出にまわる．

（ハ）$r^2 = 0.8782,\ s^2 = 404.2$　（ニ）$0.6699 \pm t(8, 0.05) \times 0.0882 = (0.467, 0.873)$

（ホ）$|\hat{\alpha}| = 959.5 > 607.9 = t(8, 0.05) \times 263.63$ より仮説は棄却される．検定の立場からは，原点を通る直線を当てはめること，言い換えると，家計最終消費支出は家計可処分所得の一定の割合であるというモデルは支持されないことになる．