

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
15	2. 集計用に縦並びにした表																						
16	小さい順に並べ替えた表																						
17	連番	賃金 (ドル/週)		境界値	度数 f			階級中央値 x								連番	賃金 (ドル/週)	順位 (小さい順)	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	累積 度数	累積 相対 度数	
18	1	49		38.5	0				最大値・最小値・四分位数(中央値を含む)などの統計値は表計算ソフトの関数でも算出できるが、「課題2]としてはまず関数を用いずに統計値を出してみる こと。 右のように、「小さい順に並べ替えた表]をつくれればその表から読み取ることができる。							93	39	1	-11.0	121.0	1	1%	
19	2	47		40.5	2			39.5									49	40	2	-10.0	100.0	2	2%
20	3	51		42.5	0			41.5									67	43	3	-7.0	49.0	3	3%
21	4	48		44.5	3			43.5									44	44	4	-6.0	36.0	4	4%
22	5	50		46.5	14			45.5									79	44	5	-6.0	36.0	5	5%
23	6	46		48.5	14			47.5									9	45	6	-5.0	25.0	6	6%
24	7	53		50.5	25			49.5									36	45	7	-5.0	25.0	7	7%
25	8	46		52.5	21			51.5									46	45	8	-5.0	25.0	8	8%
26	9	45		54.5	10			53.5									61	45	9	-5.0	25.0	9	9%
27	10	50		56.5	5			55.5									78	45	10	-5.0	25.0	10	10%
28	11	49		58.5	2			57.5								99	45	11	-5.0	25.0	11	11%	
29	12	50		60.5	2			59.5								6	46	12	-4.0	16.0	12	12%	
30	13	50		62.5	2			61.5								8	46	13	-4.0	16.0	13	13%	
31	14	47														17	46	14	-4.0	16.0	14	14%	
32	15	56		n	100											27	46	15	-4.0	16.0	15	15%	
33	16	51														31	46	16	-4.0	16.0	16	16%	
34	17	46		最小値	39											50	46	17	-4.0	16.0	17	17%	
35	18	47		第1四分位数	47											88	46	18	-4.0	16.0	18	18%	
36	19	54		中央値	50											94	46	19	-4.0	16.0	19	19%	
37	20	53		第3四分位数	52											2	47	20	-3.0	9.0	20	20%	
38	21	48		最大値	61											14	47	21	-3.0	9.0	21	21%	
39	22	50														18	47	22	-3.0	9.0	22	22%	
40	23	51														52	47	23	-3.0	9.0	23	23%	
41	24	50														56	47	24	-3.0	9.0	24	24%	
42	25	60														65	47	25	-3.0	9.0	25	25%	
43	26	51														75	47	26	-3.0	9.0	26	26%	
44	27	46														85	47	27	-3.0	9.0	27	27%	
45	28	48														4	48	28	-2.0	4.0	28	28%	
46	29	52														21	48	29	-2.0	4.0	29	29%	
47	30	52														28	48	30	-2.0	4.0	30	30%	
48	31	46														57	48	31	-2.0	4.0	31	31%	
49	32	61														77	48	32	-2.0	4.0	32	32%	
50	33	52														92	48	33	-2.0	4.0	33	33%	
51	34	49														1	49	34	-1.0	1.0	34	34%	
52	35	50														11	49	35	-1.0	1.0	35	35%	
53	36	45														34	49	36	-1.0	1.0	36	36%	
54	37	57														45	49	37	-1.0	1.0	37	37%	
55	38	54														60	49	38	-1.0	1.0	38	38%	
56	39	51														82	49	39	-1.0	1.0	39	39%	
57	40	60														87	49	40	-1.0	1.0	40	40%	
58	41	50														90	49	41	-1.0	1.0	41	41%	
59	42	56														98	49	42	-1.0	1.0	42	42%	
60	43	52														5	50	43	0.0	0.0	43	43%	
61	44	44														10	50	44	0.0	0.0	44	44%	
62	45	49														12	50	45	0.0	0.0	45	45%	
63	46	45														13	50	46	0.0	0.0	46	46%	
64	47	51														22	50	47	0.0	0.0	47	47%	
65	48	50														24	50	48	0.0	0.0	48	48%	
66	49	40														35	50	49	0.0	0.0	49	49%	
67	50	46														41	50	50	0.0	0.0	50	50%	
68	51	54														48	50	51	0.0	0.0	51	51%	
69	52	47														53	50	52	0.0	0.0	52	52%	
70	53	50														59	50	53	0.0	0.0	53	53%	
71	54	55														62	50	54	0.0	0.0	54	54%	
72	55	55														63	50	55	0.0	0.0	55	55%	
73	56	47														70	50	56	0.0	0.0	56	56%	
74	57	48														80	50	57	0.0	0.0	57	57%	
75	58	53														100	50	58	0.0	0.0	58	58%	
76	59	50														3	51	59	1.0	1.0	59	59%	
77	60	49														16	51	60	1.0	1.0	60	60%	
78	61	45														23	51	61	1.0	1.0	61	61%	
79	62	50														26	51	62	1.0	1.0	62	62%	
80	63	50														39	51	63	1.0	1.0	63	63%	
81	64	51														47	51	64	1.0	1.0	64	64%	

最大値・最小値・四分位数(中央値を含む)などの統計値は表計算ソフトの関数でも算出できるが、「課題2]としてはまず関数を用いずに統計値を出してみる
こと。
右のように、「小さい順に並べ替えた表]をつくれればその表から読み取ることができる。

中央値は $n=100$ なので、 $(n+1)/2 = 50.5$ 番目の順位のデータの値が相当する。50番目と51番目のデータの値の平均が、中央値となる。
第1四分位数は $(n+1)/4$ 、第4四分位数は $(n+1) \times (3/4)$ の順位のデータの値となる(割り切れない場合はその前後の順位のデータの値の平均値)

最頻階級の階級値は49.5

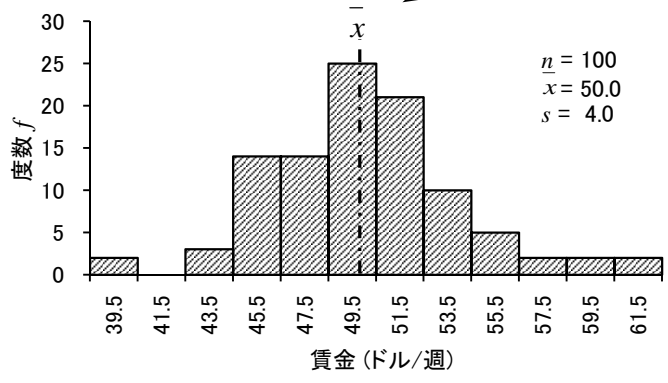


図-1. 教科書p.29, 例2"外国人労働者100人の週当たり賃金"のヒストグラム

分類したデータから平均値や標準偏差を求めると不正確な値になってしまう! 注意!

表-1. 度数分布表からの平均値の計算(教科書p.18, 表3参照)

表-2. 度数分布表からの分散と標準偏差の計算(教科書p.21, 表4参照; 左2列は表-1と重複するので省略)

i	x_i	f_i	$x_i \cdot f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
1	39.5	2	79.0	-10.5	110.3	220.5
2	41.5	0	0.0	-8.5	72.3	0.0
3	43.5	3	130.5	-6.5	42.3	126.8
4	45.5	14	637.0	-4.5	20.3	283.5
5	47.5	14	665.0	-2.5	6.3	87.5
6	49.5	25	1237.5	-0.5	0.3	6.3
7	51.5	21	1081.5	1.5	2.3	47.3
8	53.5	10	535.0	3.5	12.3	122.5
9	55.5	5	277.5	5.5	30.3	151.3
10	57.5	2	115.0	7.5	56.3	112.5
11	59.5	2	119.0	9.5	90.3	180.5
12	61.5	2	123.0	11.5	132.3	264.5
計		100	5000.0			1,6

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
82	65	47			\bar{x}	度数	相対度数	累積度数	累積相対度数							64	51	65		1.0	1.0	65	65%
83	66	54			39.5	2		2	2%							74	51	66		1.0	1.0	66	66%
84	67	43			41.5	0		2	0%							76	51	67		1.0	1.0	67	67%
85	68	53			43.5	3		5	3%							83	51	68		1.0	1.0	68	68%
86	69	55			45.5	14		19	14%							84	51	69		1.0	1.0	69	69%
87	70	50			47.5	14		33	14%							96	51	70		1.0	1.0	70	70%
88	71	53			49.5	25		58	25%							29	52	71		2.0	4.0	71	71%
89	72	52			51.5	21		79	21%							30	52	72		2.0	4.0	72	72%
90	73	52			53.5	10		89	10%							33	52	73		2.0	4.0	73	73%
91	74	51			55.5	5		94	5%							43	52	74		2.0	4.0	74	74%
92	75	47			57.5	2		96	2%							72	52	75		2.0	4.0	75	75%
93	76	51			59.5	2		98	2%							73	52	76		2.0	4.0	76	76%
94	77	48			61.5	2		100	2%							81	52	77		2.0	4.0	77	77%
95	78	45			計	100										91	52	78		2.0	4.0	78	78%
96	79	44														95	52	79		2.0	4.0	79	79%
97	80	50														7	53	80		3.0	9.0	80	80%
98	81	52														20	53	81		3.0	9.0	81	81%
99	82	49														58	53	82		3.0	9.0	82	82%
100	83	51														68	53	83		3.0	9.0	83	83%
101	84	51														71	53	84		3.0	9.0	84	84%
102	85	47														86	53	85		3.0	9.0	85	85%
103	86	53														19	54	86		4.0	16.0	86	86%
104	87	49														38	54	87		4.0	16.0	87	87%
105	88	46														51	54	88		4.0	16.0	88	88%
106	89	61														66	54	89		4.0	16.0	89	89%
107	90	49														54	55	90		5.0	25.0	90	90%
108	91	52														55	55	91		5.0	25.0	91	91%
109	92	48														69	55	92		5.0	25.0	92	92%
110	93	39														15	56	93		6.0	36.0	93	93%
111	94	46														42	56	94		6.0	36.0	94	94%
112	95	52														37	57	95		7.0	49.0	95	95%
113	96	51														97	57	96		7.0	49.0	96	96%
114	97	57														25	60	97		10.0	100.0	97	97%
115	98	49														40	60	98		10.0	100.0	98	98%
116	99	45														32	61	99		11.0	121.0	99	99%
117	100	50														89	61	100		11.0	121.0	100	100%

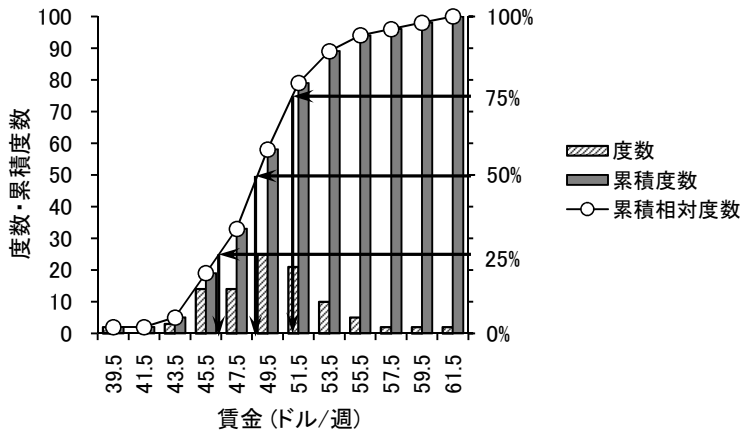


図-2. 度数分布図(ヒストグラム)と累積度数分布図

注) 累積相対度数の図から、中央値と四分位数を読み取ることができる。
 中央値は累積相対度数50%の値で約49.0ドル/週、
 第1四分位数は累積相対度数25%の値で約46.5ドル/週、
 第3四分位数は累積相対度数75%の値で約51.0ドル/週となる。
 (度数分布から求める方法は素データから求める場合に比べて誤差が生じる)

		↓この列の合計		↓この列の合計
合計	5000		偏差平方和	1578.0
$n = 100$				↓偏差平方和をnで割る
平均値 \bar{x}	50.0		分散	15.94
				↓分散のルート
			標準偏差 s	3.99

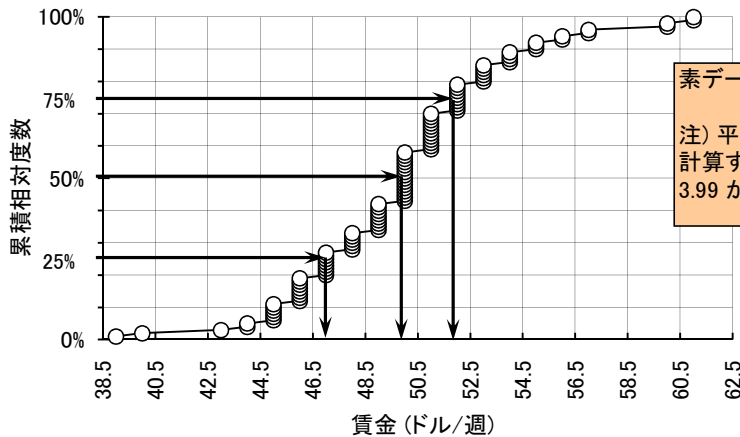


図-3. 素データからの累積相対度数

注) 素データから累積相対度数の図を作成することもできる。
 この場合、図から中央値や四分位数を正確に読み取ることができる。

素データ(分類しないデータ)で計算した平均値と標準偏差
 注) 平均値はたまたま同じだが、標準偏差は分類したデータで計算すると4.02と異なる値になる(分類しないデータで計算した3.99が正しい値)

	$\bar{x} - s$	46.0	「平均値±標準偏差」の範囲内に入る度数は、該当範囲の値のデータ数を数え上げることでわかる。教科書の例では分類したデータから計算しているの71と不正確な値となっている。
	$\bar{x} + s$	54.0	
	範囲内度数	78	
	範囲内相対度数	78.0%	
	$\bar{x} - 2s$	42.0	
	$\bar{x} + 2s$	58.0	
	範囲内度数	94	
	範囲内相対度数	94.0%	