

中級統計学：宿題 1

村澤 康友

提出期限：2023 年 10 月 9 日

注意：すべての質問に解答しなければ提出とは認めない。授業の HP の解答例を正確に再現すること（乱数は除く）。グループで取り組んでもよいが、個別に提出すること。解答例をコピーしたり、他人の名前で提出した場合は、提出点を 0 点とし、再提出も認めない。すべての結果をワードに貼り付けて印刷し（A4 縦・両面印刷可・手書き不可）、2 枚以上になる場合は問題番号順に重ねて左上隅をホッチキスで留めること。

1. 教科書 p. 35 のデータ A・B・C について、Excel を利用して平均・中位数・最頻値・分散・標準偏差を求めなさい。分散・標準偏差を求める関数は各 2 種類ある (`var.p/var.s`, `stdev.p/stdev.s`)。違いを確認して適切な方を用いること。
2. gretl のサンプル・データ data2-1 は、カリフォルニア大学サンディエゴ校 1 年生の英語 (`vsat`) と数学 (`msat`) の入試成績である。このデータについて以下の分析を行いなさい。
 - (a) 各変量の基本統計量を比較しなさい。
 - (b) 各変量の度数分布表を作成しなさい。
 - (c) 各変量のヒストグラムを描きなさい。
 - (d) 散布図を描きなさい。
 - (e) 相関係数を求めなさい。※メニューの「表示」または「変数」以下から実行できる。表・グラフのコピー・保存は右クリックのメニューから実行する。
3. gretl のサンプル・データ data2-2 は、同 1 年生の大学での GPA (`colgpa`) と高校での GPA (`hsgpa`) である。このデータについて前問と同じ分析を行いなさい。

解答例

1. 分散は var.p, 標準偏差は stdev.p を用いる (教科書 p. 37 を参照).

	A	B	C
	0	0	3
	3	1	4
	3	2	4
	5	3	5
	5	5	5
	5	5	5
	5	7	5
	7	8	6
	7	9	6
	10	10	7
平均	5	5	5
中位数	5	5	5
最頻値	5	5	5
分散	6.6	10.8	1.2
標準偏差	2.569047	3.286335	1.095445

2. (a) 基本統計量

基本統計量 使用した観測: 1-427

変数	平均	中央値	最小値	最大値
vsat	501.80	500.00	200.00	700.00
msat	566.32	570.00	330.00	770.00

変数	標準偏差	変動係数	歪度	Ex. kurtosis
vsat	91.314	0.18197	-0.33536	0.10777
msat	93.119	0.16443	-0.22786	-0.54562

変数	5 百分位数.	95 百分位数.	IQ Range	欠損値数
vsat	330.00	650.00	110.00	0
msat	400.00	710.00	130.00	0

(b) 度数分布表

vsat の度数分布, 観測: 1-427

階級数 = 11, 平均 = 501.803, 標準偏差 = 91.3142

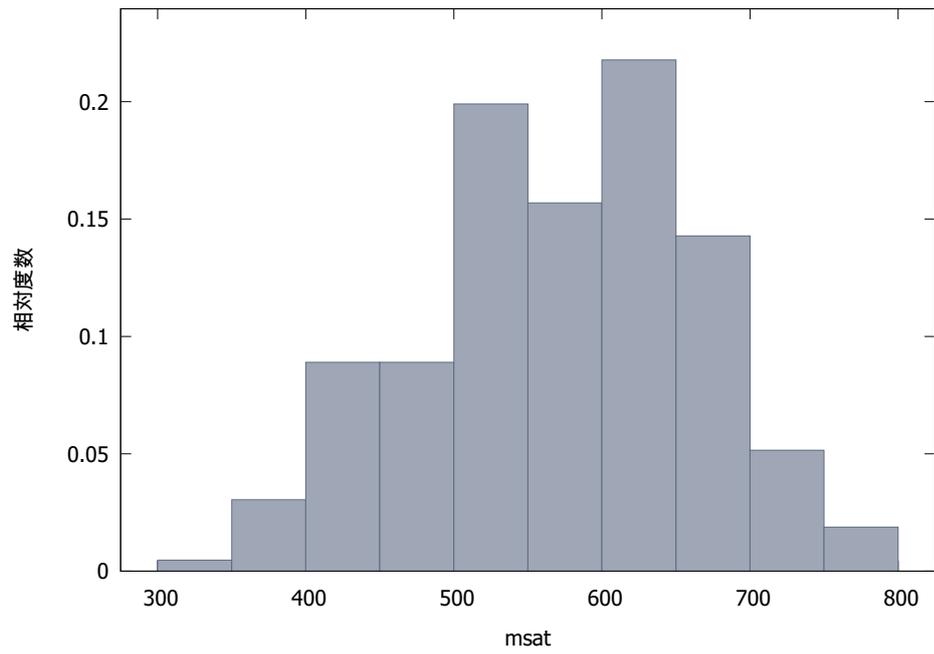
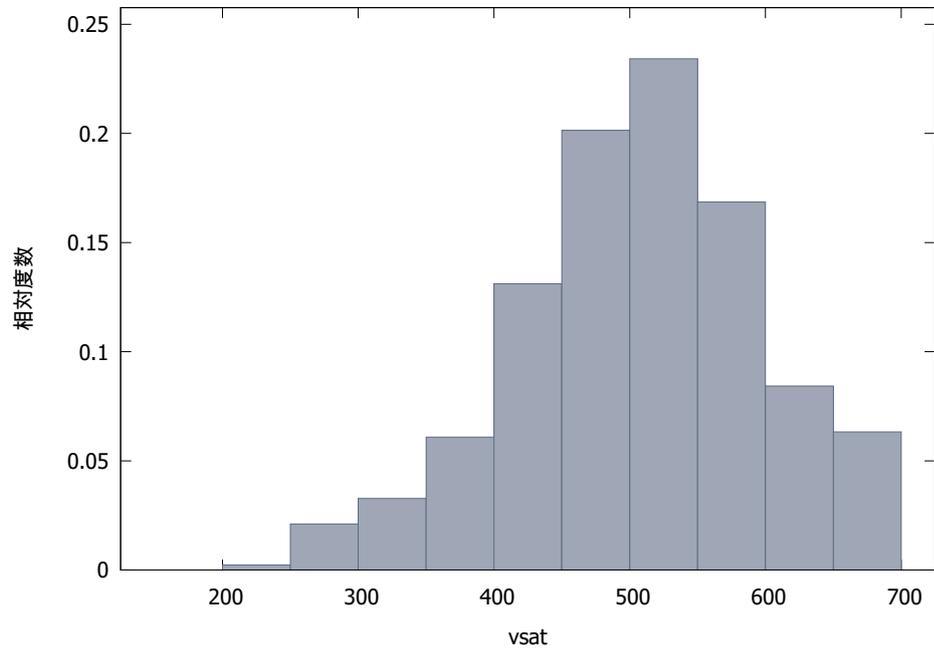
区間	中心値	度数	相対度数	累積相対度数
< 200.00	175.00	0	0.00%	0.00%
200.00 - 250.00	225.00	1	0.23%	0.23%
250.00 - 300.00	275.00	9	2.11%	2.34%
300.00 - 350.00	325.00	14	3.28%	5.62% *
350.00 - 400.00	375.00	26	6.09%	11.71% **
400.00 - 450.00	425.00	56	13.11%	24.82% ****
450.00 - 500.00	475.00	86	20.14%	44.96% *****
500.00 - 550.00	525.00	100	23.42%	68.38% *****
550.00 - 600.00	575.00	72	16.86%	85.25% *****
600.00 - 650.00	625.00	36	8.43%	93.68% ***
>= 650.00	675.00	27	6.32%	100.00% **

msat の度数分布, 観測: 1-427

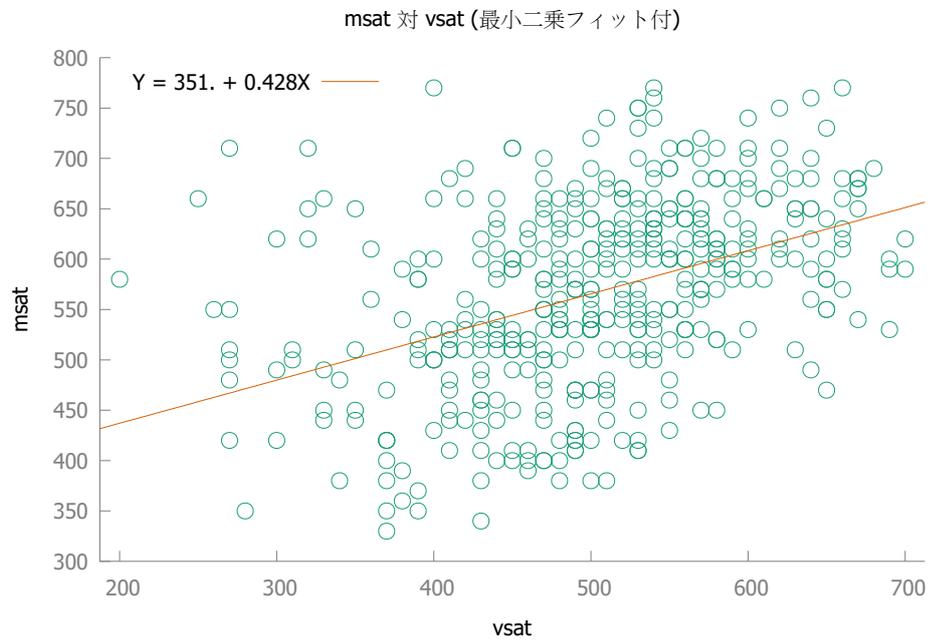
階級数 = 10, 平均 = 566.323, 標準偏差 = 93.1192

区間	中心値	度数	相対度数	累積相対度数
< 350.00	325.00	2	0.47%	0.47%
350.00 - 400.00	375.00	13	3.04%	3.51% *
400.00 - 450.00	425.00	38	8.90%	12.41% ***
450.00 - 500.00	475.00	38	8.90%	21.31% ***
500.00 - 550.00	525.00	85	19.91%	41.22% *****
550.00 - 600.00	575.00	67	15.69%	56.91% *****
600.00 - 650.00	625.00	93	21.78%	78.69% *****
650.00 - 700.00	675.00	61	14.29%	92.97% *****
700.00 - 750.00	725.00	22	5.15%	98.13% *
>= 750.00	775.00	8	1.87%	100.00%

(c) ヒストグラム



(d) 散布図



(e) 相関係数

相関係数, 使用した観測 1-427

5% 臨界値 (両側) = 0.0949 但し $n = 427$

vsat	msat	
1.0000	0.4200	vsat
	1.0000	msat

3. (a) 基本統計量

基本統計量 使用した観測: 1-427

変数	平均	中央値	最小値	最大値
colgpa	2.7855	2.7900	0.85000	3.9700
hsgpa	3.5578	3.5900	2.2900	4.5000

変数	標準偏差	変動係数	歪度	Ex. kurtosis
colgpa	0.54082	0.19416	-0.20436	-0.037888
hsgpa	0.41958	0.11793	-0.40265	-0.20704

変数	5 百分位数.	95 百分位数.	IQ Range	欠損値数
colgpa	1.9000	3.6960	0.78000	0
hsgpa	2.7540	4.2000	0.58000	0

(b) 度数分布表

colgpa の度数分布, 観測: 1-427

階級数 = 16, 平均 = 2.7855, 標準偏差 = 0.54082

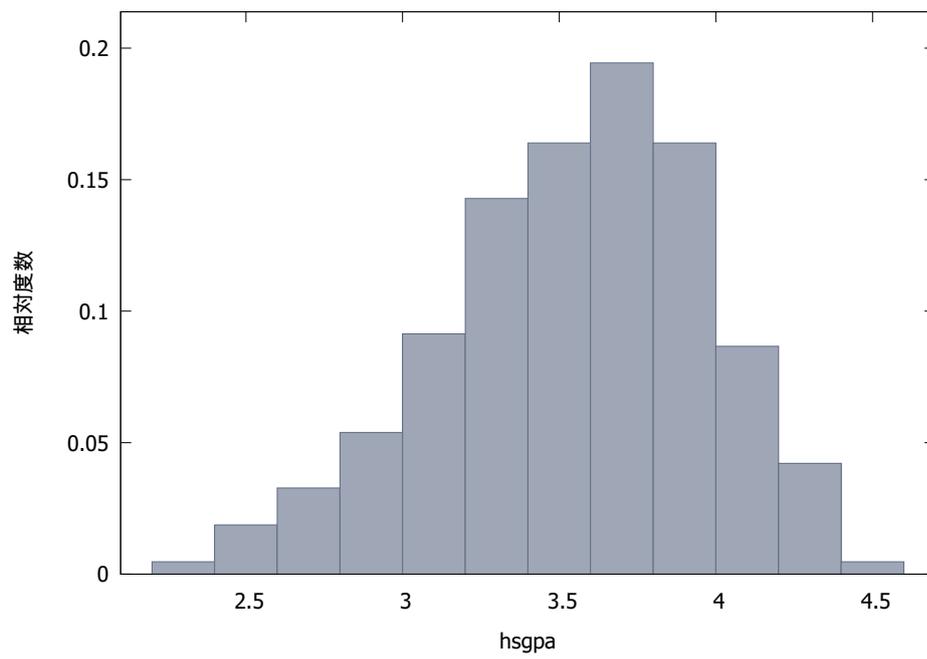
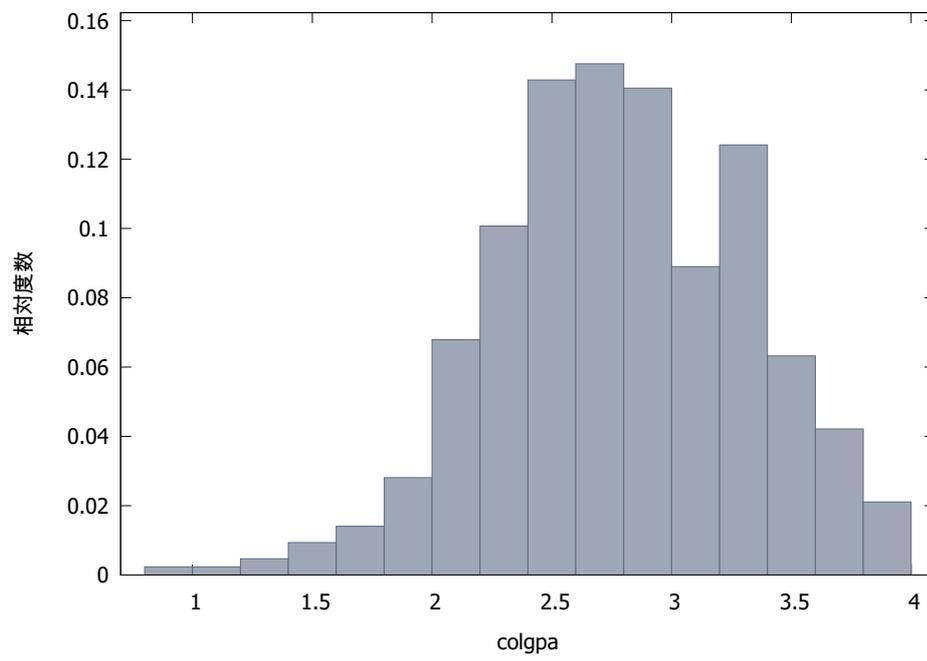
区間	中心値	度数	相対度数	累積相対度数
< 1.0000	0.90000	1	0.23%	0.23%
1.0000 - 1.2000	1.1000	1	0.23%	0.47%
1.2000 - 1.4000	1.3000	2	0.47%	0.94%
1.4000 - 1.6000	1.5000	4	0.94%	1.87%
1.6000 - 1.8000	1.7000	6	1.41%	3.28%
1.8000 - 2.0000	1.9000	12	2.81%	6.09% *
2.0000 - 2.2000	2.1000	29	6.79%	12.88% **
2.2000 - 2.4000	2.3000	43	10.07%	22.95% ***
2.4000 - 2.6000	2.5000	61	14.29%	37.24% *****
2.6000 - 2.8000	2.7000	63	14.75%	51.99% *****
2.8000 - 3.0000	2.9000	60	14.05%	66.04% *****
3.0000 - 3.2000	3.1000	38	8.90%	74.94% ***
3.2000 - 3.4000	3.3000	53	12.41%	87.35% *****
3.4000 - 3.6000	3.5000	27	6.32%	93.68% **
3.6000 - 3.8000	3.7000	18	4.22%	97.89% *
>= 3.8000	3.9000	9	2.11%	100.00%

hsgpa の度数分布, 観測: 1-427

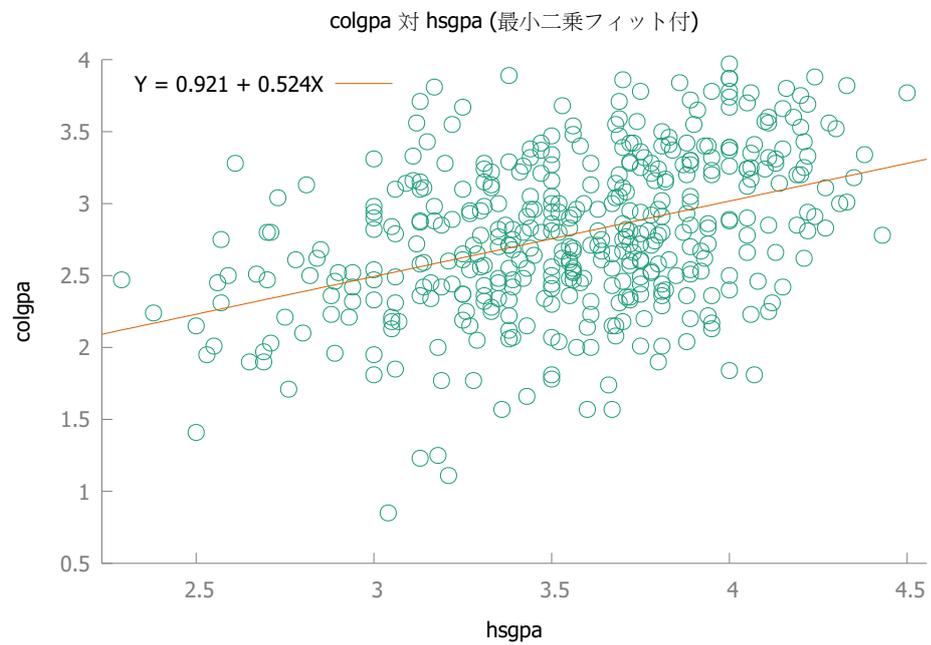
階級数 = 12, 平均 = 3.55785, 標準偏差 = 0.419577

区間	中心値	度数	相対度数	累積相対度数
< 2.4000	2.3000	2	0.47%	0.47%
2.4000 - 2.6000	2.5000	8	1.87%	2.34%
2.6000 - 2.8000	2.7000	14	3.28%	5.62% *
2.8000 - 3.0000	2.9000	23	5.39%	11.01% *
3.0000 - 3.2000	3.1000	39	9.13%	20.14% ***
3.2000 - 3.4000	3.3000	61	14.29%	34.43% *****
3.4000 - 3.6000	3.5000	70	16.39%	50.82% *****
3.6000 - 3.8000	3.7000	83	19.44%	70.26% *****
3.8000 - 4.0000	3.9000	70	16.39%	86.65% *****
4.0000 - 4.2000	4.1000	37	8.67%	95.32% ***
4.2000 - 4.4000	4.3000	18	4.22%	99.53% *
>= 4.4000	4.5000	2	0.47%	100.00%

(c) ヒストグラム



(d) 散布図



(e) 相関係数

相関係数, 使用した観測 1-427

5% 臨界値 (両側) = 0.0949 但し n = 427

colgpa	hsgpa	
1.0000	0.4067	colgpa
	1.0000	hsgpa