

中級統計学：宿題 1

村澤 康友

提出期限：2025 年 10 月 10 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 の解答例の結果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よいが、個別に提出すること。解答例をコピー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 Word に貼り付けて印刷し（A4 縦・両面印刷可・手書き不可・写真不可・文字化け不可）、2 枚以上の場合は向きを揃えて問題番号順に重ね、左上隅をホッチキスで留めること。

準備：授業の HP 掲載資料「gretl 入門」をよく読んで、gretl を使用する環境を準備しなさい。

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 1 変量分析を行いなさい。
 - (a) 各変量の度数分布表を作成しなさい。
 - (b) 各変量のヒストグラムを描きなさい。
 - (c) 各変量の基本統計量を比較しなさい。※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 のメニューの「変数」→「ジニ係数」でローレンツ曲線の図示とジニ係数の計算ができる。前問の英語（`vsat`）と数学（`msat`）の入試成績について、ローレンツ曲線を図示し、ジニ係数を比較しなさい。
4. gretl のサンプル・データ data2-2 は、カリフォルニア大学サンディエゴ校 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以下の 2 変量分析を行いなさい。
 - (a) 散布図を描きなさい（x 軸・y 軸の変数の選び方に注意）。
 - (b) 相関係数を求めなさい。※メニューの「表示」→「変数のグラフ」→「XY 散布図」で散布図が図示できる。また「表示」→「相関行列」で相関係数が計算できる。

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. 1 変量分析

(a) 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

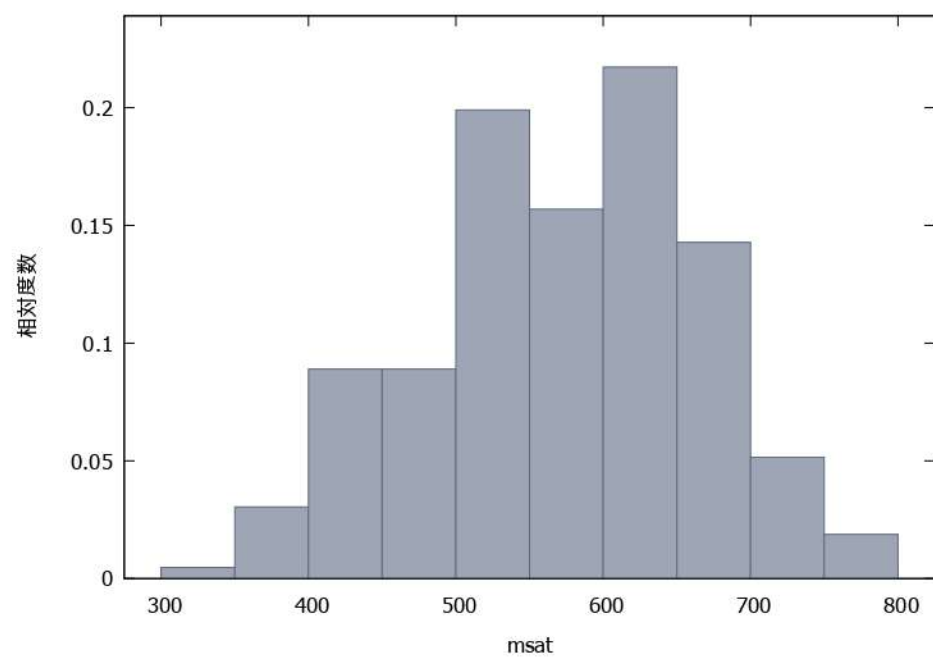
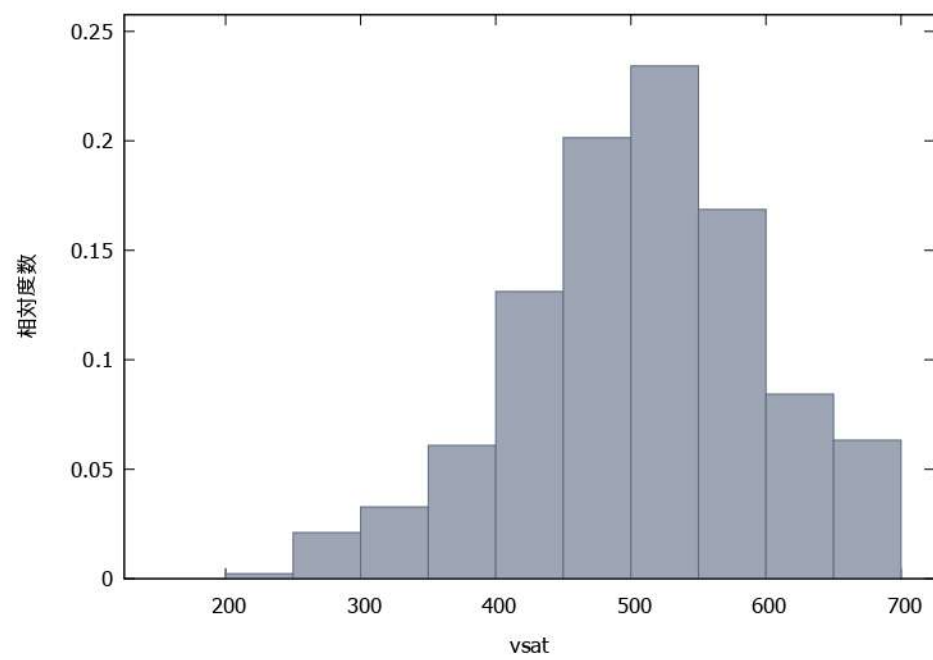
| 区間 | 中心値 | 度数 | 相対度数 | 累積相対度数 |
|-----------------|--------|-----|--------|--------------|
| < 200.00 | 175.00 | 0 | 0.00% | 0.00% |
| 200.00 - 250.00 | 225.00 | 1 | 0.23% | 0.23% |
| 250.00 - 300.00 | 275.00 | 9 | 2.11% | 2.34% |
| 300.00 - 350.00 | 325.00 | 14 | 3.28% | 5.62% * |
| 350.00 - 400.00 | 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% |

(b) ヒストグラム

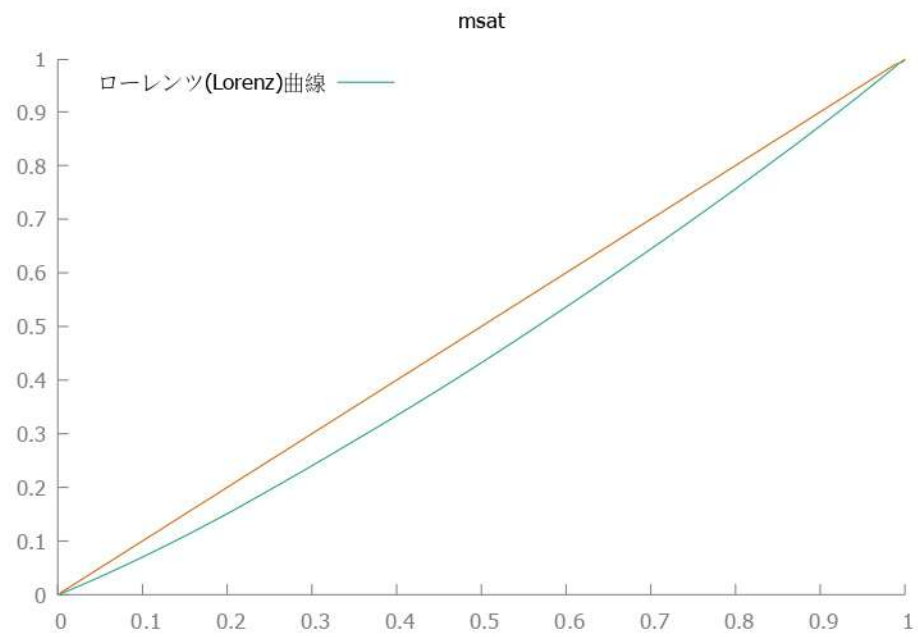
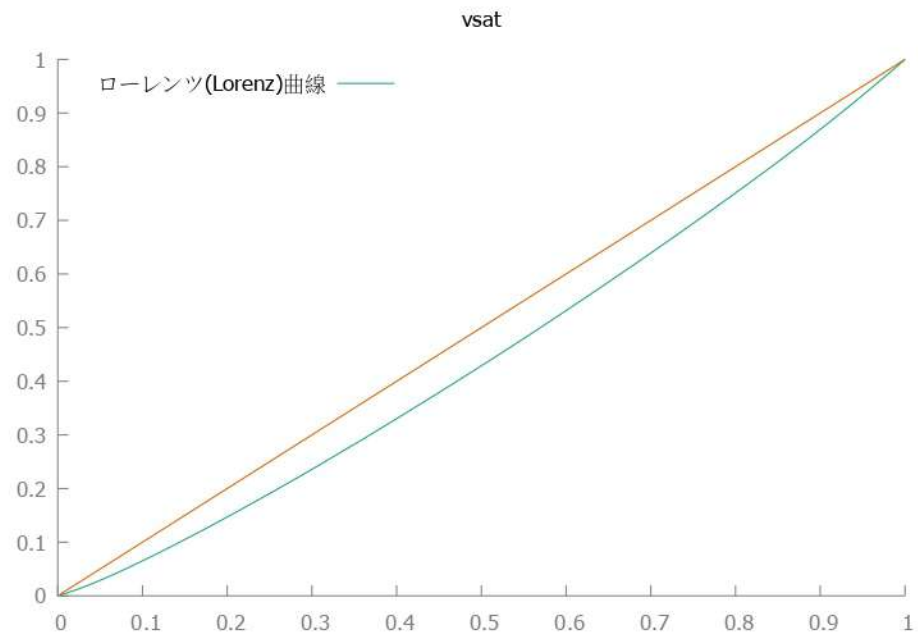


(c) 基本統計量

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 |

3. ローレンツ曲線



ジニ係数

vsat

観測数 = 427

標本ジニ (Gini) 係数 = 0.101855

母数の推定値 (estimate) = 0.102094

msat

観測数 = 427

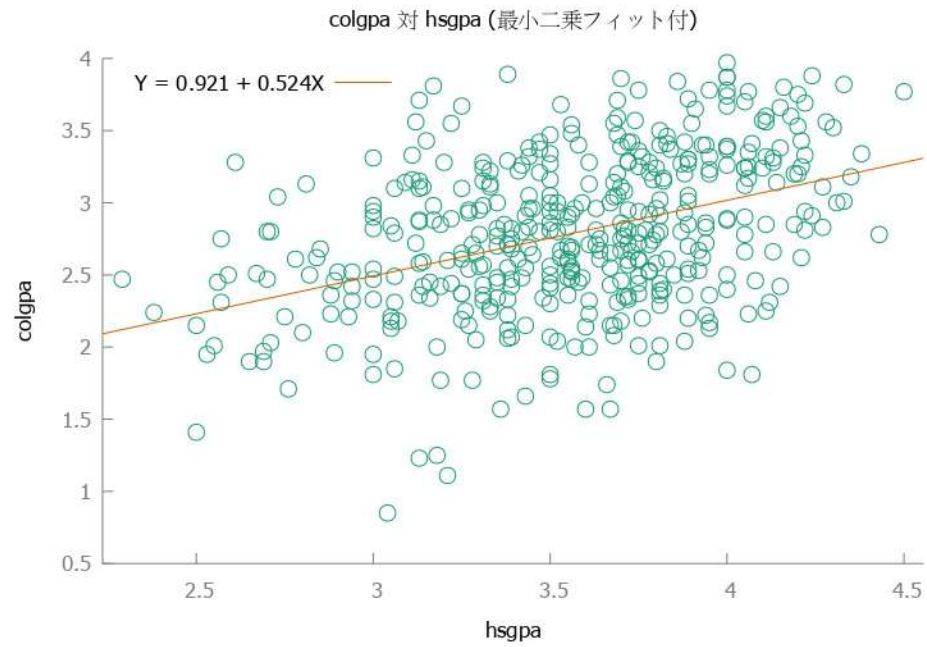
標本ジニ (Gini) 係数 = 0.0934932

母数の推定値 (estimate) = 0.0937127

英語 (vsat) の入試成績の方がジニ係数が大きく、格差が大きい。

4. 2 変量分析

(a) 散布図



(b) 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 |