

# データ解析基礎- Basic Data Analysis

## 最終試験- Final Examination

July 1, 2010

### Problems

When you are asked to do a calculation, you do not need to compute the decimal equivalent of a fraction or radical (square root). Fractions should be reduced to lowest terms for convenience in grading. Radicals do not need to be reduced.

計算を行うときには分数または根数のままを書いてもよい。少数にする必要はない。ただし、分数の分母と分子は互いに素にすること。

For Problems 1 to 5, use **Data Set A**. (Each student receives a different data set.) Note that the data set has the *true* standard deviation reported; you do not need to calculate it. Data Set A is a data set of the *speed* of each car passing through an intersection in one hour. This data was collected because two serious accidents have occurred recently, and the police are concerned that excessive speed is a cause. (The speed limit is 40kph.) They are considering assigning personnel to catch offenders in this particular place.

問題1～5にデータセットAを利用してください。(注意: 皆に別のデータを用意する。必ずデータセットIDを確認すること。) データセットAはある交差点を通る自動車の速度である。最近、人身事故が2つあったのでスピード違反が原因かと考え、警察官を常に異版を監督する必要の有無を判断したがる。(速度制限は40キロである。)

**Copy your data set in the space below: ここにデータを写ってください:**

1. Divide your data into **four** (4) cells. **Explain** why you chose those cells.

データを4区間に区別せよ。その区間選択の理由を説明せよ。

Name \_\_\_\_\_ Dataset# \_\_\_\_\_ ID# \_\_\_\_\_ 2

2. Construct the *absolute*, *relative*, and *cumulative relative frequency* distributions. For each cell, choose a representative value. Enter all of this information in a table.

データを「」に変換して絶対頻度分布、相対頻度分布、または（相対）累積分布を書け。1つの表にまとめること。

3. Draw the histogram of the distribution.

データを4区間に区別せよ。その区間選択の理由を説明せよ。分布のヒストグラムを描くこと。

4. Compute the *median* and the *mean* of the distribution. **Show your work or explain the process.**

中央値(メディアン) と平均値を計算せよ。ただし、計算方法を細かく書くか説明すること。

5. There are reasons to believe that this sample is not a perfectly *random sample* (independently and identically distributed observations). Explain how dependence or non-identical distributions for individual cars might be involved.

Do you think the sample is "random enough"? Explain briefly.

このサンプルはランダム(各観察の分布が独立・同一)ではないと考えられる。分布間の依存または分布の異なる理由を説明せよ。

サンプルは十分にランダムと思うか。その理由を短く説明すること。

6. You have been asked by the police to give advice on whether they should step up enforcement near this intersection. Formulate the hypothesis that the traffic through the intersection is moving at the speed limit, *i.e.*, the average speed is 40 kph, at the significance level 5%. Include the following information: What is the *null hypothesis* you test? Is your test *one-sided* or *two-sided*? What is the *alternative hypothesis*? Explain why you chose this form of the test.

あなたは警察への顧問役になっている。この交差点に観察をもっと厳しくすべきかどうかについてアドバイスが頼まれた。5%レベルで速度制限にしたがって交通が流れる仮説を作成せよ。つまり、平均速度は40キロであることだ。帰無仮説、両側か片側か、そして対立仮説を書いてその検定の形を選んだ理由を説明せよ。

7. After you graduate, what kind of work do you plan to do? Explain an example of the use of statistical estimation or hypothesis testing in your work.

あなたは経営政策プログラムを終了してからどんな仕事をしたいか？その仕事に伴う統計的推計あるいは仮説検定の利用を説明せよ。

The rest of the questions address the following R output for a regression model and the corresponding residuals. The variable  $c$  is real U.S. consumption in 2000 dollars, quarterly 1947.1 to 2010.1. The variable  $g$  is real U.S. GDP in 2000 dollars, quarterly 1947.1 to 2010.1. The variable  $g^2$  is  $g^2$ , the square of GDP.

以下の問題は下記のR出力を問う。回帰モデルの中には $c$ という変数はアメリカの2000年度ドルで測った実質消費総額を指し、 $g$ はアメリカの2000年度ドルで測った実質国内総生産を指し、 $g^2$ は $g^2$ (GDPの二乗)を指す。

Call:

```
lm(formula = c ~ g + g2)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-135.416	-16.520	2.289	15.999	152.642

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1.932e+01	4.544e+00	-4.251	3.01e-05 ***
$g$	6.361e-01	2.113e-03	301.018	< 2e-16 ***
$g^2$	4.973e-06	1.591e-07	31.254	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 41.49 on 250 degrees of freedom

Multiple R-squared: 0.9998, Adjusted R-squared: 0.9998

F-statistic: 6.853e+05 on 2 and 250 DF, p-value: < 2.2e-16

8. Write the theoretical regression equation including the disturbance term.  
理論的な回帰数式を書け。誤差項を忘れずに。

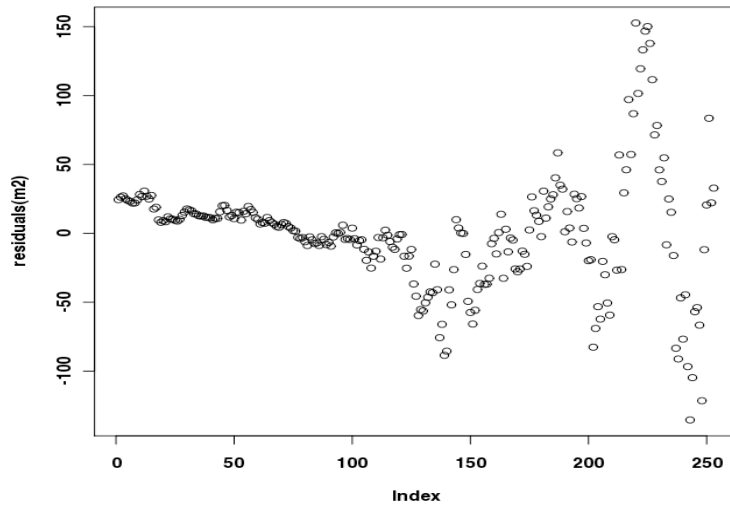


Figure 1: Residuals of the regression

9. How much of the variance of  $c$  is explained by the equation? Explain how you know.

この式には  $c$  の分散の説明できる割合はいくらですか。その理由を説明せよ。

10. Test the hypothesis that the constant term is 0 at the 5% level. Write  $H_0$  and explain briefly.

5% レベルで定数が 0 という仮説を検定せよ。  $H_0$  を書いて、 短く説明すること。

11. Test the hypothesis that the equation is *linear* in  $g$  at the 5% level. Write  $H_0$  and explain briefly.

5% レベルで定数が線形型であるという仮説を検定せよ。  $H_0$  を書いて、 短く説明すること。

1 Your data sets have data set ID #1. Be sure to enter your data set ID in the space provided.

Data Set A: 35 36 37 34 37 36 35 35 34 35

The true standard deviation of your data is 1.

2 Your data sets have data set ID #2. Be sure to enter your data set ID in the space provided.

Data Set A: 34 32 37 35 35 35 40 34 33 34

The true standard deviation of your data is 2.

3 Your data sets have data set ID #3. Be sure to enter your data set ID in the space provided.

Data Set A: 35 31 29 35 38 34 35 33 37 35

The true standard deviation of your data is 3.

4 Your data sets have data set ID #4. Be sure to enter your data set ID in the space provided.

Data Set A: 44 33 36 33 36 40 33 37 32 39

The true standard deviation of your data is 4.

5 Your data sets have data set ID #5. Be sure to enter your data set ID in the space provided.

Data Set A: 37 36 36 36 34 37 35 34 34 37

The true standard deviation of your data is 1.

6 Your data sets have data set ID #6. Be sure to enter your data set ID in the space provided.

Data Set A: 34 34 36 33 36 37 34 38 35 37

The true standard deviation of your data is 2.

7 Your data sets have data set ID #7. Be sure to enter your data set ID in the space provided.

Data Set A: 33 38 33 30 32 39 34 37 31 34

The true standard deviation of your data is 3.

8 Your data sets have data set ID #8. Be sure to enter your data set ID in the space provided.

Data Set A: 36 41 29 29 39 39 37 38 37 42

The true standard deviation of your data is 4.

9 Your data sets have data set ID #9. Be sure to enter your data set ID in the space provided.

Data Set A: 38 38 37 39 37 37 39 38 38 38

The true standard deviation of your data is 1.

10 Your data sets have data set ID #10. Be sure to enter your data set ID in the space provided.

Data Set A: 35 35 37 37 37 34 41 37 34 35

The true standard deviation of your data is 2.

11 Your data sets have data set ID #11. Be sure to enter your data set ID in the space provided.

Data Set A: 43 37 31 34 37 36 32 34 35 38

The true standard deviation of your data is 3.

12 Your data sets have data set ID #12. Be sure to enter your data set ID in the space provided.

Data Set A: 33 33 38 35 37 36 30 39 43 37

The true standard deviation of your data is 4.

13 Your data sets have data set ID #13. Be sure to enter your data set ID in the space provided.

Data Set A: 38 39 38 37 38 38 38 38 37 39

The true standard deviation of your data is 1.

14 Your data sets have data set ID #14. Be sure to enter your data set ID in the space provided.

Data Set A: 40 37 36 37 39 39 36 41 38 40

The true standard deviation of your data is 2.

15 Your data sets have data set ID #15. Be sure to enter your data set ID in the space provided.

Data Set A: 39 38 34 43 41 35 34 35 36 31

The true standard deviation of your data is 3.

16 Your data sets have data set ID #16. Be sure to enter your data set ID in the space provided.

Data Set A: 40 33 39 33 40 41 39 33 42 43

The true standard deviation of your data is 4.



- 17 Your data sets have data set ID #17. Be sure to enter your data set ID in the space provided.  
Data Set A: 39 39 38 40 38 38 41 40 39 37  
The true standard deviation of your data is 1.
- 18 Your data sets have data set ID #18. Be sure to enter your data set ID in the space provided.  
Data Set A: 41 40 39 39 41 40 38 41 38 40  
The true standard deviation of your data is 2.
- 19 Your data sets have data set ID #19. Be sure to enter your data set ID in the space provided.  
Data Set A: 40 43 37 36 42 39 35 40 37 35  
The true standard deviation of your data is 3.
- 20 Your data sets have data set ID #20. Be sure to enter your data set ID in the space provided.  
Data Set A: 37 46 32 39 41 41 36 38 40 36  
The true standard deviation of your data is 4.
- 21 Your data sets have data set ID #21. Be sure to enter your data set ID in the space provided.  
Data Set A: 40 41 41 40 41 41 41 40 41 40  
The true standard deviation of your data is 1.
- 22 Your data sets have data set ID #22. Be sure to enter your data set ID in the space provided.  
Data Set A: 41 43 37 40 39 40 42 44 41 42  
The true standard deviation of your data is 2.
- 23 Your data sets have data set ID #23. Be sure to enter your data set ID in the space provided.  
Data Set A: 34 37 37 43 37 40 46 44 41 33  
The true standard deviation of your data is 3.
- 24 Your data sets have data set ID #24. Be sure to enter your data set ID in the space provided.  
Data Set A: 36 40 38 45 38 39 35 39 46 31  
The true standard deviation of your data is 4.

- 25 Your data sets have data set ID #25. Be sure to enter your data set ID in the space provided.  
Data Set A: 41 41 42 40 40 42 40 41 42 40  
The true standard deviation of your data is 1.
- 26 Your data sets have data set ID #26. Be sure to enter your data set ID in the space provided.  
Data Set A: 43 43 43 41 39 41 41 39 44 41  
The true standard deviation of your data is 2.
- 27 Your data sets have data set ID #27. Be sure to enter your data set ID in the space provided.  
Data Set A: 39 44 40 39 40 40 40 44 41 44  
The true standard deviation of your data is 3.
- 28 Your data sets have data set ID #28. Be sure to enter your data set ID in the space provided.  
Data Set A: 36 38 38 33 48 47 41 40 33 48  
The true standard deviation of your data is 4.
- 29 Your data sets have data set ID #29. Be sure to enter your data set ID in the space provided.  
Data Set A: 36 35 34 36 37 34 37 36 35 34  
The true standard deviation of your data is 1.
- 30 Your data sets have data set ID #30. Be sure to enter your data set ID in the space provided.  
Data Set A: 37 33 36 37 31 34 31 36 38 39  
The true standard deviation of your data is 2.
- 31 Your data sets have data set ID #31. Be sure to enter your data set ID in the space provided.  
Data Set A: 37 33 36 27 34 41 34 32 36 38  
The true standard deviation of your data is 3.
- 32 Your data sets have data set ID #32. Be sure to enter your data set ID in the space provided.  
Data Set A: 33 32 39 35 28 39 34 41 37 40  
The true standard deviation of your data is 4.

- 33 Your data sets have data set ID #33. Be sure to enter your data set ID in the space provided.  
Data Set A: 35 37 36 36 36 35 37 37 36 36  
The true standard deviation of your data is 1.
- 34 Your data sets have data set ID #34. Be sure to enter your data set ID in the space provided.  
Data Set A: 38 37 34 37 34 36 34 36 33 34  
The true standard deviation of your data is 2.
- 35 Your data sets have data set ID #35. Be sure to enter your data set ID in the space provided.  
Data Set A: 34 38 37 37 32 35 35 39 34 33  
The true standard deviation of your data is 3.
- 36 Your data sets have data set ID #36. Be sure to enter your data set ID in the space provided.  
Data Set A: 40 41 36 41 37 37 34 37 38 34  
The true standard deviation of your data is 4.
- 37 Your data sets have data set ID #37. Be sure to enter your data set ID in the space provided.  
Data Set A: 37 36 38 37 39 37 37 36 37 36  
The true standard deviation of your data is 1.
- 38 Your data sets have data set ID #38. Be sure to enter your data set ID in the space provided.  
Data Set A: 39 37 39 38 38 39 39 39 38 38  
The true standard deviation of your data is 2.
- 39 Your data sets have data set ID #39. Be sure to enter your data set ID in the space provided.  
Data Set A: 34 39 35 40 38 35 42 36 41 38  
The true standard deviation of your data is 3.
- 40 Your data sets have data set ID #40. Be sure to enter your data set ID in the space provided.  
Data Set A: 28 38 45 40 38 42 37 45 40 36  
The true standard deviation of your data is 4.

- 41 Your data sets have data set ID #41. Be sure to enter your data set ID in the space provided.  
Data Set A: 38 36 39 38 37 38 39 38 39 39  
The true standard deviation of your data is 1.
- 42 Your data sets have data set ID #42. Be sure to enter your data set ID in the space provided.  
Data Set A: 39 39 36 35 39 35 40 40 38 37  
The true standard deviation of your data is 2.
- 43 Your data sets have data set ID #43. Be sure to enter your data set ID in the space provided.  
Data Set A: 37 41 36 38 39 33 36 36 37 28  
The true standard deviation of your data is 3.
- 44 Your data sets have data set ID #44. Be sure to enter your data set ID in the space provided.  
Data Set A: 36 31 30 41 32 32 34 31 44 32  
The true standard deviation of your data is 4.
- 45 Your data sets have data set ID #45. Be sure to enter your data set ID in the space provided.  
Data Set A: 38 40 40 39 39 38 39 40 41 39  
The true standard deviation of your data is 1.
- 46 Your data sets have data set ID #46. Be sure to enter your data set ID in the space provided.  
Data Set A: 42 39 40 42 37 37 37 38 39 38  
The true standard deviation of your data is 2.
- 47 Your data sets have data set ID #47. Be sure to enter your data set ID in the space provided.  
Data Set A: 39 42 35 35 43 43 41 42 40 38  
The true standard deviation of your data is 3.
- 48 Your data sets have data set ID #48. Be sure to enter your data set ID in the space provided.  
Data Set A: 38 44 40 40 40 38 36 35 40 40  
The true standard deviation of your data is 4.

- 49 Your data sets have data set ID #49. Be sure to enter your data set ID in the space provided.  
Data Set A: 39 40 39 38 39 41 39 39 42 40  
The true standard deviation of your data is 1.
- 50 Your data sets have data set ID #50. Be sure to enter your data set ID in the space provided.  
Data Set A: 42 44 37 39 39 41 43 38 42 39  
The true standard deviation of your data is 2.
- 51 Your data sets have data set ID #51. Be sure to enter your data set ID in the space provided.  
Data Set A: 40 41 43 41 43 41 39 42 44 44  
The true standard deviation of your data is 3.
- 52 Your data sets have data set ID #52. Be sure to enter your data set ID in the space provided.  
Data Set A: 37 41 40 46 41 46 36 36 43 47  
The true standard deviation of your data is 4.
- 53 Your data sets have data set ID #53. Be sure to enter your data set ID in the space provided.  
Data Set A: 42 44 41 42 42 39 41 40 43 41  
The true standard deviation of your data is 1.
- 54 Your data sets have data set ID #54. Be sure to enter your data set ID in the space provided.  
Data Set A: 42 39 42 40 42 38 41 42 43 43  
The true standard deviation of your data is 2.
- 55 Your data sets have data set ID #55. Be sure to enter your data set ID in the space provided.  
Data Set A: 47 41 41 41 36 42 39 41 43 37  
The true standard deviation of your data is 3.
- 56 Your data sets have data set ID #56. Be sure to enter your data set ID in the space provided.  
Data Set A: 35 42 45 40 44 39 40 34 37 42  
The true standard deviation of your data is 4.

57 Your data sets have data set ID #57. Be sure to enter your data set ID in the space provided.

Data Set A: 38 36 34 35 34 36 34 35 35 33

The true standard deviation of your data is 1.

58 Your data sets have data set ID #58. Be sure to enter your data set ID in the space provided.

Data Set A: 35 36 35 35 35 33 38 35 37 37

The true standard deviation of your data is 2.

59 Your data sets have data set ID #59. Be sure to enter your data set ID in the space provided.

Data Set A: 32 42 36 36 32 38 36 38 37 37

The true standard deviation of your data is 3.

60 Your data sets have data set ID #60. Be sure to enter your data set ID in the space provided.

Data Set A: 34 41 34 35 35 41 35 33 27 35

The true standard deviation of your data is 4.

61 Your data sets have data set ID #61. Be sure to enter your data set ID in the space provided.

Data Set A: 36 37 36 34 36 36 35 36 36 36

The true standard deviation of your data is 1.

62 Your data sets have data set ID #62. Be sure to enter your data set ID in the space provided.

Data Set A: 37 34 39 34 36 37 37 34 37 35

The true standard deviation of your data is 2.

63 Your data sets have data set ID #63. Be sure to enter your data set ID in the space provided.

Data Set A: 38 33 44 36 32 36 39 34 35 39

The true standard deviation of your data is 3.

64 Your data sets have data set ID #64. Be sure to enter your data set ID in the space provided.

Data Set A: 36 42 30 35 45 37 30 36 31 37

The true standard deviation of your data is 4.

- 65 Your data sets have data set ID #65. Be sure to enter your data set ID in the space provided.  
Data Set A: 38 38 37 38 36 37 36 39 37 36  
The true standard deviation of your data is 1.
- 66 Your data sets have data set ID #66. Be sure to enter your data set ID in the space provided.  
Data Set A: 38 40 35 37 38 35 36 38 38 37  
The true standard deviation of your data is 2.
- 67 Your data sets have data set ID #67. Be sure to enter your data set ID in the space provided.  
Data Set A: 37 33 35 42 37 37 37 38 37 31  
The true standard deviation of your data is 3.
- 68 Your data sets have data set ID #68. Be sure to enter your data set ID in the space provided.  
Data Set A: 39 38 42 44 33 46 40 33 36 33  
The true standard deviation of your data is 4.
- 69 Your data sets have data set ID #69. Be sure to enter your data set ID in the space provided.  
Data Set A: 38 38 38 38 38 37 39 38 38  
The true standard deviation of your data is 1.
- 70 Your data sets have data set ID #70. Be sure to enter your data set ID in the space provided.  
Data Set A: 37 39 39 35 38 36 39 34 40 39  
The true standard deviation of your data is 2.
- 71 Your data sets have data set ID #71. Be sure to enter your data set ID in the space provided.  
Data Set A: 39 35 39 38 36 30 39 40 41 44  
The true standard deviation of your data is 3.
- 72 Your data sets have data set ID #72. Be sure to enter your data set ID in the space provided.  
Data Set A: 38 30 39 38 33 43 38 35 40 35  
The true standard deviation of your data is 4.

- 73 Your data sets have data set ID #73. Be sure to enter your data set ID in the space provided.  
Data Set A: 38 40 38 38 40 40 39 41 38 38  
The true standard deviation of your data is 1.
- 74 Your data sets have data set ID #74. Be sure to enter your data set ID in the space provided.  
Data Set A: 42 39 37 37 36 39 42 40 38 41  
The true standard deviation of your data is 2.
- 75 Your data sets have data set ID #75. Be sure to enter your data set ID in the space provided.  
Data Set A: 39 36 44 37 41 42 39 40 41 37  
The true standard deviation of your data is 3.
- 76 Your data sets have data set ID #76. Be sure to enter your data set ID in the space provided.  
Data Set A: 35 37 33 39 38 35 41 41 43 40  
The true standard deviation of your data is 4.
- 77 Your data sets have data set ID #77. Be sure to enter your data set ID in the space provided.  
Data Set A: 39 39 40 39 40 41 40 42 40 42  
The true standard deviation of your data is 1.
- 78 Your data sets have data set ID #78. Be sure to enter your data set ID in the space provided.  
Data Set A: 38 41 40 42 38 40 36 39 37 37  
The true standard deviation of your data is 2.
- 79 Your data sets have data set ID #79. Be sure to enter your data set ID in the space provided.  
Data Set A: 40 37 42 41 43 39 39 44 42 44  
The true standard deviation of your data is 3.
- 80 Your data sets have data set ID #80. Be sure to enter your data set ID in the space provided.  
Data Set A: 44 41 33 39 38 38 42 42 36 41  
The true standard deviation of your data is 4.



81 Your data sets have data set ID #81. Be sure to enter your data set ID in the space provided.

Data Set A: 42 42 41 39 42 40 41 41 42 41

The true standard deviation of your data is 1.

82 Your data sets have data set ID #82. Be sure to enter your data set ID in the space provided.

Data Set A: 42 41 43 40 42 43 36 40 38 43

The true standard deviation of your data is 2.

83 Your data sets have data set ID #83. Be sure to enter your data set ID in the space provided.

Data Set A: 43 39 41 43 44 40 41 36 42 36

The true standard deviation of your data is 3.

84 Your data sets have data set ID #84. Be sure to enter your data set ID in the space provided.

Data Set A: 42 41 39 42 43 45 42 31 37 40

The true standard deviation of your data is 4.