Name _____ Dataset# _____ ID# _____ 1

データ解析基礎- Basic Data Analysis

中間試験- Midterm Test

May 26, 2011

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 6, use **Data Set A**. (Each student receives a different data set. Make sure your Data Set ID is correctly entered in the space at the top of the page.) Data Set A is a data set of rice yield (productivity) in metric tonnes per hectare in Japan. Classify productivity according to the following *subjective yield* scale: $x \leq 5$ is "poor," $5 < x \leq 6.5$ is "average," $6.5 < x \leq 8$ is "good," and $x > 8$ is "excellent."

問題1～6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は日本の 米の生産性(t/ha) である。 サイズを「主観的生産性」に以下の表によって区別する: $x \leq 5$は「わる い」、$5 < x \leq 6.5$は「平均」、$6.5 < x \leq 8$は「よい」、そして $x > 8$は「優秀」。

Copy your data set here: ここにデータをここに写こと:

1. *Sort* your raw data, and write the sorted data here.
   データを順序に並べてここに書くこと。

2. Convert your data to “subjective yields,” and enter the absolute, relative, and cumulative relative frequency distributions here. What is the *median* of your data?

   データを「主観的生産性」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。中央値（メディアン）はどれですか？

3. For each “subjective yield,” choose a *representative numerical yield*. Enter the correspondence here. **Explain why you chose *each* numerical value.**

   各「主観的生産性」に代表的値を選んで、ここに記入。**それぞれの値を選んだ理由を説明せよ。**

4. Using the representative yields you chose in the previous problem, compute the *mean*, *variance*, and *standard deviation* of the *distribution* of yields. Show your work (*e.g.*, using a table like that in Homework 2's spreadsheet).

   前の問題で選んだ代表的生産性と分布を用いてサイズ分布の平均値、分散、 と標準偏差を計算せよ。計算方法を表すテーブルなどを含むこと。（たとえ、第２宿題のシートのようなもの。）

5. Pick **one** of the following three cases, and answer the question in the space provided below.

以下の状況説明から一つを選んで、下記の問aと問bを答えろ。

(a) The data set of yields was derived from the 2010 harvests in 10 different fields in one town operated by a particular farmer using the same seed in each field.
生産性データはある農家の２０１０年の収穫を田んぼ別にした。ただ し、全ての田んぼには同じ米の種類を蒔いた。

(b) The data set of yields was derived from the 2010 harvest for one representative farm in each of 10 prefectures in eastern Japan.
生産性データは関東の１０ヵ都県の代表農家の２０１０年の収穫で計 算した。

(c) The yield data was compiled for a particular farm in Tsukuba for each year from 2001 to 2010.
生産性データはあるつくば市にある農家の２００１～２０１０年度の 収穫データで計算した。

For your chosen case, give *one* example of a "hidden factor" relating different observations in the data set that could affect the way you interpret your statistics. Explain why this matters.

選択状況には「隠された要因」により観察間関係が現れ、統計量の解釈に影 響を及ぼすことがある。その要因・関係をひとつを選んで説明せよ。

For Problems 7 to 8, use **Data Set B**. (Each student receives a different data set.) Data Set B is a data set of examination scores on a 0-100 scale.

問題7～8に**データセットB**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Bはある試験の点数データで、0～100の範囲である。

Copy your data set here: ここにデータを写ってください：

7. Give the definition of *median*. Find the median of the *raw* data from Data Set B. Now, convert Data Set B to letter grades according to the usual scale, and enter a table containing the letter grade, the *scale□nterval*, the absolute frequency, the relative frequency, and the cumulative frequency distribution. What is the *median* of the distribution of letter grades? Compare it to the raw (point score) median.

   中央値(メディアン) の定義を書け。データセットBの中央値を記入せよ。デー タセットBを普通のスケールでレターグレードに変換し、レターグレード、 スケール範囲、絶対同数、相対同数、と(相対) 累積頻度分布を表に記入す ること。レターグレード分布の中央値を求めよ。点数のメディアンと比較せよ。

8. Draw a histogram for the *raw* data set. Drawing a histogram involves a choice of division into cells of values. (Recall that a *cell* is a group of values that are close to each other.) *Explain why* you chose the cells you did.

   点数データのヒストグラムを描け。ヒストグラムの作成には値の区間(セル、 仕切り) の選択が必要だ。(区間は値の範囲だ。) 区間の選択の理由を説明 せよ。

9. The Public Health Service studied the effects of smoking on health, in a large representative sample of households. They split the sample by gender and by age groups, then compared health of individuals within the same group who had different smoking histories. For both men and women, in all age groups, they found that those who had never smoked were on average somewhat healthier than current smokers. But the current smokers were

much healthier than those who had recently stopped smoking.

厚生省が喫煙の健康への影響を調査するために多くの代表的家計を選び、 男女・年齢別のグループに分けた。各グループのメンバーの喫煙状況を調べ、 健康状態を比べた。男性でも、女性でも、各年齢でも、同じ結果が出た。平 均的に喫煙歴史のない人は現在喫煙する人よりやや健康がよかったが、現在 喫煙する人は最近タバコを止めた人より明らかに健康がよかったと言う。

(a) Is this an observational study or an experimental study? Briefly explain how you know.
この調査は観察的調査であるか、実験的調査ですか。その理由を簡単 に説明せよ。

(b) Why did they study men and women, and the different age groups, separately?
男女・年令別で調査を行った理由を説明せよ。

(c) The lesson seems to be that you shouldn't start smoking, but once you've started, you shouldn't stop. What do you think?
この研究が教えることは「タバコを始めない方がよいが、吸う習慣に なった場合には止めてはだめだ」でしょう。あなたの考えかたを説明しろ。

10. A coin is tossed six times. Two (of the many) possible sequences of results are

(i) H T T H T H (ii) H H H H H H

(The coin must land H or T in the order given; H = heads, T = tails.) Which of the following statements is correct? Explain briefly.

硬貨を6回なげる。数多くの結果の順序の中で

(i) H T T H T H　　(ii) H H H H H H

の２つがある。（硬貨は書いた通りＨまたはＴにならなければならない。Ｈ ＝表、Ｔ＝裏。）以下のケースの中から正しいのはどれであるか。その理由 を簡単に説明せよ。

(a) Sequence (i) is more likely.
順序 (i) の確立が高い。

(b) Sequence (ii) is more likely.
順序 (ii) の確立が高い。

(c) Both sequences are equally likely.
両順序の確立が等しい。

11. Suppose you pick a child at random from an elementary school. Are the events ``the child is in 2d grade'' and ``the child female'' *independent*? Are they *mutually exclusive*? Explain.

小学校の１人の生徒をランダムに選ぶ。「２年生である」と「女性である」という事象を定義する。２つの事象は「independent」ですか。「mutually exclusive」ですか。その理由を説明すること。

12. The leading general trading companies are generally considered to hire some of the best students in Japan. Take a new employee at random, and consider the events A = “the student's school was the University of Tsukuba,” and B = “the student has a TOEIC score above 750.” State as many facts as you can about $\Pr(\{\}), \Pr(A), \Pr(B), \Pr(A \cap B), \Pr(A \cup B)$, and $\Pr(\Omega)$, including comparing the probabilities of two events (*e.g.*, $\Pr(A) < \Pr(\Omega)$).

総合商社は日本のトップランナーを雇うと言われている。ランダムに辰 社員のひとりを選出し、A=「筑波大学出身」とB=「TOEICが７５０点以上」 という事象を考察しよう。$\Pr(\{\})$、$\Pr(A)$、$\Pr(B)$、$\Pr(A \cap B)$、$\Pr(A \cup B)$、$\Pr(\Omega)$ についてできるだけ 多くの事実を書け。事象の確率の比較を含む。（例：$\Pr(A) < \Pr(\Omega)$。）

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

Data Set A: 5.7 5.1 5.4 5.9 5.0 5.5 5.9 5.4 5.6 5.1

Data Set B: 92 74 73 62 66 67 65 60 81 47

_____

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

Data Set A: 5.5 5.4 5.8 6.1 5.3 5.0 5.6 5.6 5.2 5.5

Data Set B: 67 92 74 60 62 81 65 73 66 47

_____

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

Data Set A: 6.1 5.9 5.6 5.5 6.3 5.8 5.3 5.5 5.9 5.6

Data Set B: 62 60 66 67 92 74 47 73 81 65

_____

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

Data Set A: 6.6 5.4 4.9 5.4 6.0 5.4 7.0 4.5 5.2 5.7

Data Set B: 67 81 62 60 73 74 47 65 66 92

_____

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

Data Set A: 5.1 5.4 5.5 5.6 6.3 5.8 5.9 4.4 3.8 5.7

Data Set B: 66 92 67 74 65 62 60 73 47 81

_____

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

Data Set A: 5.1 5.1 5.6 6.5 4.5 6.3 5.2 6.9 6.5 4.9

Data Set B: 92 60 65 81 62 47 66 67 73 74

_____

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

Data Set A: 5.5 6.9 5.3 5.6 3.7 5.0 6.1 5.0 5.8 6.5

Data Set B: 65 92 73 67 62 47 66 74 60 81

_____

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

Data Set A: 5.3 6.0 5.5 5.8 6.1 5.3 5.3 5.4 5.8 5.4

Data Set B: 62 67 47 60 66 65 81 92 73 74

_____

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

Data Set A: 6.1 6.2 5.4 5.1 5.7 5.5 5.7 5.8 5.4 5.6

Data Set B: 74 73 47 81 65 66 62 67 60 92

_____

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

Data Set A: 5.2 5.9 6.4 5.6 5.8 5.4 5.8 5.8 6.0 4.9

Data Set B: 67 74 92 81 66 73 60 65 47 62

_____

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

Data Set A: 5.6 5.3 4.9 5.4 4.9 5.0 5.6 5.2 6.0 5.9

Data Set B: 66 47 67 92 81 65 62 60 73 74

_____

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

Data Set A: 5.6 6.0 6.3 5.6 5.1 6.4 5.9 5.8 5.0 6.9

Data Set B: 92 47 60 66 62 67 74 73 65 81

_____

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

Data Set A: 5.1 5.9 6.3 6.1 6.0 6.0 6.7 5.1 6.3 5.1

Data Set B: 47 65 92 67 73 60 62 66 74 81

_____

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

Data Set A: 5.4 5.6 6.2 4.8 4.3 4.7 6.5 5.1 5.3 6.4

Data Set B: 92 65 67 66 60 62 81 74 47 73

_____

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

Data Set A: 5.6 5.8 6.3 5.3 5.8 6.6 5.4 6.0 6.1 5.9

Data Set B: 62 67 65 60 92 73 66 74 81 47

_____

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

Data Set A: 6.7 6.0 6.2 6.1 6.4 5.2 5.5 5.5 5.8 5.6

Data Set B: 62 67 47 66 92 74 65 73 81 60

_____

17 Your data sets have data set ID #17. Be sure to enter your data set ID in the space provided.

Data Set A: 6.0 5.7 6.2 5.4 6.6 6.3 4.7 6.6 5.6 6.5

Data Set B: 60 92 62 74 47 66 65 81 67 73

_____

18 Your data sets have data set ID #18. Be sure to enter your data set ID in the space provided.

Data Set A: 5.9 5.7 5.8 5.8 6.0 5.9 5.6 5.5 5.4 5.6

Data Set B: 73 67 66 60 81 65 74 47 92 62

_____

19 Your data sets have data set ID #19. Be sure to enter your data set ID in the space provided.

Data Set A: 5.6 5.0 6.0 6.2 5.3 7.2 6.2 6.1 7.0 6.0

Data Set B: 47 92 66 73 62 81 65 74 67 60

_____

20 Your data sets have data set ID #20. Be sure to enter your data set ID in the space provided.

Data Set A: 6.1 7.0 5.8 7.0 5.3 6.3 4.8 5.3 5.1 5.0

Data Set B: 74 47 92 81 67 66 65 73 60 62

_____

21 Your data sets have data set ID #21. Be sure to enter your data set ID in the space provided.

Data Set A: 5.4 6.4 5.4 6.1 8.0 4.4 5.1 5.6 4.6 6.6

Data Set B: 81 66 73 65 60 74 47 67 92 62

_____

22 Your data sets have data set ID #22. Be sure to enter your data set ID in the space provided.

Data Set A: 6.4 6.2 5.7 6.1 5.9 6.0 6.4 6.1 5.9 6.1

Data Set B: 67 65 74 73 66 81 62 47 92 60

_____

23 Your data sets have data set ID #23. Be sure to enter your data set ID in the space provided.

Data Set A: 5.8 6.0 5.5 6.1 6.1 6.7 5.8 6.1 5.9 6.2

Data Set B: 92 62 60 74 66 81 47 67 65 73

_____

24 Your data sets have data set ID #24. Be sure to enter your data set ID in the space provided.

Data Set A: 5.7 6.0 6.1 6.8 6.5 5.5 5.7 6.5 6.4 5.7

Data Set B: 60 47 81 74 73 62 66 67 65 92

_____

25 Your data sets have data set ID #25. Be sure to enter your data set ID in the space provided.

Data Set A: 5.6 6.0 5.7 5.1 6.7 6.8 7.1 6.5 5.5 6.1

Data Set B: 92 47 73 74 62 60 67 65 81 66

_____

26 Your data sets have data set ID #26. Be sure to enter your data set ID in the space provided.

Data Set A: 5.8 6.2 7.6 5.8 5.9 5.5 6.0 5.4 5.3 6.8

Data Set B: 66 62 73 47 92 74 60 65 67 81

_____

27 Your data sets have data set ID #27. Be sure to enter your data set ID in the space provided.

Data Set A: 6.0 6.3 5.6 6.3 5.2 6.4 5.7 6.1 6.3 5.2

Data Set B: 81 47 92 62 65 74 60 73 67 66

_____

28 Your data sets have data set ID #28. Be sure to enter your data set ID in the space provided.

Data Set A: 5.7 5.5 7.3 5.3 6.2 5.3 5.3 5.9 5.7 5.8

Data Set B: 66 92 47 65 60 81 74 67 73 62

_____

29 Your data sets have data set ID #29. Be sure to enter your data set ID in the space provided.

Data Set A: 6.1 6.0 6.5 6.7 6.1 6.3 6.4 6.5 6.1 6.5

Data Set B: 62 65 67 74 92 60 47 66 81 73

_____

30 Your data sets have data set ID #30. Be sure to enter your data set ID in the space provided.

Data Set A: 5.7 5.4 5.7 6.8 6.3 6.7 6.8 7.0 6.7 6.2

Data Set B: 74 67 65 92 62 47 66 60 81 73

_____

31 Your data sets have data set ID #31. Be sure to enter your data set ID in the space provided.

Data Set A: 6.1 5.9 6.0 6.5 7.1 5.8 6.2 6.7 6.5 6.6

Data Set B: 92 67 81 73 66 60 65 47 74 62

_____

32 Your data sets have data set ID #32. Be sure to enter your data set ID in the space provided.

Data Set A: 6.7 6.0 6.9 6.7 6.5 5.8 6.9 6.7 6.2 6.6

Data Set B: 60 92 74 67 65 62 81 73 47 66

_____

33 Your data sets have data set ID #33. Be sure to enter your data set ID in the space provided.

Data Set A: 6.4 5.6 7.4 6.2 6.7 6.4 5.5 5.2 6.2 6.8

Data Set B: 73 62 47 60 81 74 67 65 66 92

_____

34 Your data sets have data set ID #34. Be sure to enter your data set ID in the space provided.

Data Set A: 6.8 6.0 7.8 5.6 7.6 7.0 5.2 5.2 7.0 4.5

Data Set B: 73 81 74 47 66 92 67 65 60 62

_____

35 Your data sets have data set ID #35. Be sure to enter your data set ID in the space provided.

Data Set A: 6.4 7.0 5.0 6.9 6.5 5.7 6.4 6.8 7.0 7.3

Data Set B: 73 92 81 74 67 47 60 62 66 65

_____

36 Your data sets have data set ID #36. Be sure to enter your data set ID in the space provided.

Data Set A: 5.7 6.2 6.3 6.3 6.7 6.4 5.9 7.2 6.7 6.6

Data Set B: 74 81 62 67 47 66 65 92 60 73

_____

37 Your data sets have data set ID #37. Be sure to enter your data set ID in the space provided.

Data Set A: 6.2 7.1 6.8 7.2 6.5 6.6 6.2 7.0 6.1 7.1

Data Set B: 47 60 73 65 81 62 66 92 74 67

_____

38 Your data sets have data set ID #38. Be sure to enter your data set ID in the space provided.

Data Set A: 6.7 6.2 6.4 6.2 6.8 5.6 5.7 6.6 6.2 7.4

Data Set B: 65 66 92 81 74 62 60 67 73 47

_____

39 Your data sets have data set ID #39. Be sure to enter your data set ID in the space provided.

Data Set A: 6.1 5.6 7.1 7.3 6.2 6.5 6.4 5.2 6.3 5.6

Data Set B: 73 74 60 62 47 92 66 65 67 81

_____

40 Your data sets have data set ID #40. Be sure to enter your data set ID in the space provided.

Data Set A: 6.7 7.1 5.1 5.1 6.0 6.8 7.1 6.4 6.4 5.5

Data Set B: 65 62 60 67 92 74 47 73 66 81

_____

41 Your data sets have data set ID #41. Be sure to enter your data set ID in the space provided.

Data Set A: 7.5 6.2 5.5 6.1 7.4 5.6 5.9 6.5 6.3 6.5

Data Set B: 47 60 81 62 66 65 74 92 73 67

_____

42 Your data sets have data set ID #42. Be sure to enter your data set ID in the space provided.

Data Set A: 6.1 7.5 6.2 6.3 5.9 7.3 7.1 5.6 7.0 4.8

Data Set B: 65 92 67 47 81 66 74 62 60 73

_____

43 Your data sets have data set ID #43. Be sure to enter your data set ID in the space provided.

Data Set A: 6.8 7.0 7.4 6.6 6.7 6.5 6.7 6.9 6.4 6.8

Data Set B: 81 73 92 65 47 60 67 74 62 66

_____

44 Your data sets have data set ID #44. Be sure to enter your data set ID in the space provided.

Data Set A: 7.3 6.2 6.9 6.5 6.0 6.0 6.7 6.8 6.5 6.5

Data Set B: 47 65 62 74 73 92 67 66 60 81

_____

45 Your data sets have data set ID #45. Be sure to enter your data set ID in the space provided.

Data Set A: 7.0 6.7 7.3 7.0 5.8 7.2 6.2 7.2 6.6 6.7

Data Set B: 47 67 81 65 92 62 74 60 73 66

_____

46 Your data sets have data set ID #46. Be sure to enter your data set ID in the space provided.

Data Set A: 7.4 7.0 7.5 7.4 6.8 6.5 6.7 7.5 6.4 6.5

Data Set B: 67 66 74 73 92 62 60 47 65 81

_____

47 Your data sets have data set ID #47. Be sure to enter your data set ID in the space provided.

Data Set A: 6.6 5.8 5.9 6.7 6.1 6.3 6.6 6.9 5.4 6.6

Data Set B: 66 47 81 92 67 65 62 74 60 73

_____

48 Your data sets have data set ID #48. Be sure to enter your data set ID in the space provided.

Data Set A: 7.5 6.4 6.4 6.1 6.3 6.4 7.0 6.6 6.9 6.0

Data Set B: 47 60 66 92 65 81 67 62 74 73

_____

49 Your data sets have data set ID #49. Be sure to enter your data set ID in the space provided.

Data Set A: 6.7 7.9 5.3 4.0 6.3 6.6 6.6 6.2 7.4 6.9

Data Set B: 66 73 47 67 60 92 81 74 65 62

_____

50 Your data sets have data set ID #50. Be sure to enter your data set ID in the space provided.

Data Set A: 6.7 6.9 7.0 6.8 6.8 7.0 7.4 6.4 6.5 6.6

Data Set B: 74 92 66 73 47 65 67 62 81 60

_____

51 Your data sets have data set ID #51. Be sure to enter your data set ID in the space provided.

Data Set A: 6.3 6.5 6.6 6.7 6.8 6.8 6.1 6.8 7.0 6.8

Data Set B: 73 92 67 60 81 65 47 62 74 66

_____

52 Your data sets have data set ID #52. Be sure to enter your data set ID in the space provided.

Data Set A: 6.9 7.0 7.2 6.7 6.5 7.2 6.0 6.1 6.9 7.9

Data Set B: 65 60 92 73 66 47 67 74 81 62

_____

53 Your data sets have data set ID #53. Be sure to enter your data set ID in the space provided.

Data Set A: 6.6 7.2 6.5 7.2 6.1 6.6 6.1 6.7 5.8 7.3

Data Set B: 62 47 73 74 92 67 66 81 60 65

_____

54 Your data sets have data set ID #54. Be sure to enter your data set ID in the space provided.

Data Set A: 6.9 7.1 6.2 7.4 6.5 7.5 6.9 6.5 6.3 6.5

Data Set B: 62 81 74 65 92 67 60 73 47 66

_____

55 Your data sets have data set ID #55. Be sure to enter your data set ID in the space provided.

Data Set A: 7.1 8.4 6.3 5.9 7.9 6.1 8.4 5.8 6.8 5.7

Data Set B: 73 67 47 65 66 60 81 62 74 92

_____

56 Your data sets have data set ID #56. Be sure to enter your data set ID in the space provided.

Data Set A: 7.8 6.0 6.3 7.5 6.1 6.6 7.6 7.2 5.8 6.0

Data Set B: 65 81 62 92 74 60 47 66 67 73

_____

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

Data Set A: 6.9 6.8 7.3 6.8 7.5 7.2 6.8 7.2 7.0 7.2

Data Set B: 66 67 81 74 92 60 62 47 65 73

_____

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

Data Set A: 7.1 6.6 7.0 7.0 7.3 6.8 6.8 7.1 6.4 7.2

Data Set B: 66 67 74 92 81 62 60 47 73 65

_____

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

Data Set A: 6.6 7.7 7.1 7.0 7.1 7.0 7.2 6.9 7.7 7.4

Data Set B: 47 66 73 74 65 81 92 67 62 60

_____

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

Data Set A: 7.0 6.9 7.7 7.0 6.5 6.6 7.4 6.9 6.2 6.8

Data Set B: 47 60 66 81 65 92 74 67 62 73

_____