75 Questions, 5 pages name (required) _____

You must turn in this hard copy (with your name on it) and your scantron to receive credit for this exam.

One answer and only one answer per question. Leaving a question blank or filling in 2+ answers will be incorrect no matter what.

A = True, B = False unless indicated otherwise

Data Quality: Errors and fixes

1-4. (5pts) There are two phases in most instances of data collection. The first (phase A) involves acquiring the subjects or objects that will be measured, as in choosing people for a study (e.g., poll or drug test), taking environmental samples, or getting forensic samples from a crime scene. Phase B involves taking the measurements themselves, as in meauring drug levels, pesticide residue levels, or a DNA profile. Which types of errors below will usually be associated with phase A and which with phase B?

1. (A) (B) Sampling error

2. (A) (B) The part of bias fixed by randomization

3. (A) (B) The part of bias fixed by blind observers

4. (A) (B) RPA error

5-11 (3pts each). In the following questions, indicate which type of error is indicated.

5. Bruce and Jim decide to investigate Charmin's claim that each roll of their toilet paper has 264 sheets. Bruce counts two rolls and gets 262 and 264 sheets in each roll. Jim counts the same two rolls and gets 264 and 264. What type of error explains the difference between their counts?

A) Sampling B) Bias C) RPA D) Human and technical E) None

6. A lab technician runs the same blood sample through a machine twice to get the level of a particular enzyme. The machine gives different results between the first and second run: the first run gives a value of 32.456, and the second run gives a value of 34.779, even though the blood sample was fully mixed and the enzyme concentration would have been the same between the runs. What type of error accounts for this difference?

A) Sampling B) Bias C) RPA D) Human and technical E) None

7. A newspaper hires a firm to compare the attitudes of Republicans and Democrats to editorials published in the newsaper. The firm takes 3 surveys. Each survey uses 100 Democrats and 100 Republicans, chosen randomly. Results of the 3 surveys are that 73%, 70% and 74% of Republicans approve of the editorials, but the corresponding numbers for Democrats is 46%, 45% and 50%. Thus there is a consistently large difference in the response between the two parties. What type of error is indicated by the consistent difference between the Republican and Democrat responses?

A) Sampling B) Bias C) RPA D) Human and technical E) None

8. A high school student is told to find the weight of an M&M piece of candy to the nearest 0.1 gram. Weighing a single piece of candy would suffice, because the weight variation between pieces is insignificant, but her scale reads only to the nearest 1 gram. So she combines 100 M&M pieces into a single jar, weighs the entire jar to the nearest gram, and (after subtracting the jar weight) calculates the average weight of a single piece to be 0.87 grams. What type of error has been reduced by her weighing the 100 pieces at once and dividing by 100?

A) Sampling B) Bias C) RPA D) Human and technical E) None

9. The use of 120 voluntary respondents to a survey instead of 120 random respondents could lead to what type of error?

A) Sampling B) Bias C) RPA D) Human and technical E) None

10. The failure to use standards can lead to what type of error going undetected?

A) Sampling B) Bias C) RPA D) Human and technical E) None

11. A teacher who grades written assignments while knowing who wrote each one is prone to what type of error in the grading?

A) Sampling B) Bias C) RPA D) Human and technical E) None

12-15. (6pts) For which below is sampling error indicated or expected to be responsible for the variation observed?

(A) = Sampling error indicated or expected, (B) = no sampling error

12. (A) (B) Variation in the number of days per year for 100 consecutive years

13. (A) (B) The different numbers of times the digit 3 occurs across different columns of a random number table

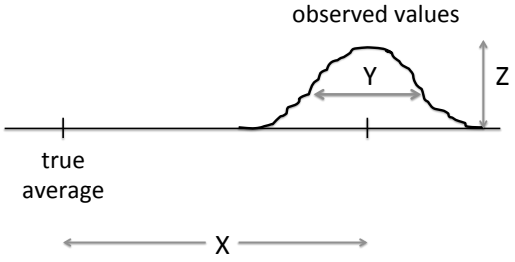
14. (A) (B) A new scale reads the weight of your fish as 2.342 pounds but an older scale reads it as 2.3 pounds.

15. (A) (B) Multiple polls using small samples are taken of the same student body to inquire about their political preferences. The poll questions and protocol are always the same. Typically, the outcome of one survey is slightly different from other surveys, sometimes with a slight excess favoring the Republican choice, other times with a slight excess favoring the Democrat choice. There is no trend over time or with any other factor as to which party is most favored in the sample.

Errors and Fixes

16-17 (4 pts). The following graph was used in a lecture on data error. Which are true of the labeled dimensions (X, Y, Z)?

A= true B = false



16. (A) (B) Z represents sampling error

17. (A) (B) X represents bias

18-20. (5 pts) A cancer researcher attempts to determine whether there are different cancer rates in different countries. He/she does this by counting the number of cancers recorded throughout each country's entire population for a one month period. (Unknown to her/him, the rates are the same in the different countries.) The population sizes of the different countries are 100,000, 1,000,000 and 10,000,000. Which are true?

18. (A) (B) The least sampling error is expected in the country with the smallest population; the largest sampling error in the sample from the largest country.

19. (A) (B) The country with the smallest population size is the ***most*** prone to bias in estimating the rate.

20. (A) (B) Given this researcher's goal, he/she should combine the numbers from the different countries to get an estimate with reduced sampling error.

21-24. (5pts) Which options identify a valid "fix" for the type of error indicated; a "fix" may either reduce that error or allow you to measure that error. **A = the fix is valid; B = the fix is not valid**

21. (A) (B) error: unintentional sample mixup during testing. **Fix**: change the protocol to institute blind testing of samples

22. (A) (B) error: RPA error that requires an extra decimal place of accuracy in the level of drug detected **Fix**: use a standard that has a similar level of drug

23. (A) (B) error: lab falsifies results to give the prosecution its desired results. **Fix**: code the samples blindly

24. (A) (B) error: lab occasionally declares matches that are not real, but they go undetected. **Fix**: report the lab error rate (LER) to the court so that the court knows not to place full emphasis on the random match probability

(**25-39).** For each of the following statements, mark the appropriate letters that describe the data design features present. Mark a data feature only if it is explicitly present at some level in the problem description.
A = present, B = absent or ambiguous

25-29 (5 pts). For a required psychology term paper, a Texas A&M student develops a questionnaire about attitudes towards gay marriage; the purpose of the survey is indicated at the top of the form. He asks 4 of his friends to get responses from their contacts. They return 73 forms to him, but when he asks them how the survey was administered and who the subjects were, his friends only give him a vague sense of what they did. Furthermore, each of the four friends apparently used different methods and solicited different groups. In the end, he has 73 filled in forms with no idea of who they represent or how the poll was administered.

25 (A) (B) Explicit protocol
26 (A) (B) Replication
27 (A) (B) Standards
28 (A) (B) Random
29 (A) (B) Blind

30-34 (5 pts). A teacher wants to know whether playing music in the classroom immediately before class starts affects group behavior of the entire class, measured as the level of noise in the classroom 1 minute after the bell rings. Without notifying the class as to why she was doing this, on Friday, October 12 she played 5 minutes of Beethoven's 5^{th} symphony to the class and secretly recorded the class noise. That noise level was compared to the noise level on the previous Friday level when no music was played. The choice of which Friday received music was made with a coin flip. 95 students were present on the first Friday, 87 present on the 12^{th}.

30 (A) (B) Explicit protocol
31 (A) (B) Replication
32 (A) (B) Standards
33 (A) (B) Random
34 (A) (B) Blind

35-39 (5pts) You are hired as a consultant for a company selling home pregnancy tests to help them market a product that will be easy to use and give accurate results. You advise them to put a picture of a woman on the front of the box and directions for use on the back. Furthermore you suggest that they provide supplies for just a single test, so that if a woman wants to test herself again, she has to purchase a second kit. Finally you suggest that they include a sample solution in the kit that will provide a definite positive result that can be used if the woman tests negative. Which aspects of the ideal data template would be satisfied by a single kit if your recommendations are followed?

35 (A) (B) Explicit protocol
36 (A) (B) Replication
37 (A) (B) Standards
38 (A) (B) Random
39 (A) (B) Blind

40-42. (4pts) In which of the following studies would the protocol need to include a design feature to ensure that **subjects are blinded** to avoid bias? Do not choose any options in which a blind design feature would would not be necessary. All options describe the subjects (underlined) as well as the purpose of the study. **A = subjects should be blinded B = not needed**

40. (A) (B) Rats fed a substance to see if it causes cancer in them.
41. (A) (B) Humans being observed after accidental exposure to a chemical; observers are worried about possible toxic effects but the exposed individuals are unaware of any possible risk.
42. (A) (B) Different car repair shops in Austin being evaluated by an independent agency to find out if they treat the average customer fairly.

43-45. (4 pts). Which of the following would constitute a standard in a drug test for evaluating lab error rates?
A = a standard, B = not a standard

43. (A) (B) A sample with no drug present.
44. (A) (B) A single sample split into two tubes, each tube labeled differently so the lab does not know they are the same
45. (A) (B) Two samples from different people in separate tubes that are labeled the same

DNA and Criminal Justice

We mentioned 4 features of an 'ideal' forensic method for matching a suspect with a forensic sample: (i) reference database, (ii) discrete characteristics, (iii) independent verification possible, (iv) labs/experts pass blind proficiency tests.

46-52. (6 pts) For which types of matching method was it claimed (in the book, web page and/or lecture) that the method clearly satisfied/satisfies the existence of and proper use of a reference database?

A = ref database exists and used properly B = not exist or not used properly

46. (A) (B) Fingerprinting (before 1990)

47. (A) (B) Fingerprinting (after 2000)

48. (A) (B) DNA typing

49. (A) (B) Dog sniffing

50. (A) (B) Hair matching (not DNA based)

51. (A) (B) Shoe print identification

52. (A) (B) Bite mark identification

53-54. (4pts) Excerpts of letters were read in class (and the full letters are in the Book) from the Chicago Police Dept to the FBI requesting DNA typing of samples. Which aspects of ideal data were specifically described in those letters (were specifically included in those requests)? A = included B = not included

53. (A) (B) Replication of the same sample

54. (A) (B) Standards

55. (A) (B) Randomization

56. (A) (B) Blind

57-59 (4pts) An eyewitness video was shown in class in which a single young male was observed. Following that video, the individuals in the class were asked to identify that person in a line-up. Which of the following is/are true as pertains to the purpose or content of that demo? You may use results from 2010 or 2011 to answer some of these as well. **A = true B = false**

57. (A) (B) Bias: the demo showed how subtle clues given about the nature of the person you saw (height, appearance) could lead you to identify the wrong person even when the right person is present in the line-up.

58. (A) (B) Bias: the demo showed how instructions about who may be present in the line-up could influence whether you identify the wrong person

59. (A) (B) Accuracy of eyewitness ID: at least 1/4 of the class identified the wrong person

60-67. New methods of sample matching have just been introduced into court, as described below. Which of the 4 features of 'ideal forensics' are indicated as being present? For the first 3 features, the problem must specifically describe their presence for it to be present. For 'independent verification' the problem must specifically describe it or describe a means by which independent verification could feasibly be performed by different labs. **A = present B= absent or incomplete**

60-63 (4 pts). The method matches the pollen spectrum – the types and abundances of different species of pollen – found at a crime scene with the pollen spectrum found on the suspect. Pollen grains are minute cells that come from plants. Pollen is often spread by wind so it is found almost everywhere, but the sources of pollen – plant species from which they originate – and amounts vary with season and location. Thus, the pollen spectrum from a crime scene is almost unique to that time and place. Each pollen grain is easily classified to a single plant species because of unique features on the surface of pollen grains (found only in that species, unambiguously present or absent), and all pollen grains from the same species have the same characteristics. There are catalogs with images of different species of pollen grains, such that one lab can easily reproduce the findings of another.

60. (A) (B) reference database

61. (A) (B) discrete characteristics

62. (A) (B) pass blind proficiency tests

63. (A) (B) Independent verification (explicitly present or the means for doing it is described)

A = present B= absent or incomplete

64-67 (4 pts). The method measures minute quantities of 12 different types of molecules on the outside of skin cells. The testing lab claims that the amounts and combinations of these molecules are unique to each individual, that no two individuals have the same profiles. Everyone has the same 12 types of molecules, and a person's profile is the set of 12 numbers representing the absolute amounts of each molecule on a scale of 1.0 to 0.00001, measured as accurately as the machine will perform. The lab's measurement method is considered 'proprietary' because it is covered by patents, and only this one lab is allowed to perform the test. Furthermore, the lab introducing this method has not yet been asked to perform any tests except by prosecution agencies sending them single suspect samples and associated, single forensic samples.

64. (A) (B) reference database

65. (A) (B) discrete characteristics

66. (A) (B) pass blind proficiency tests

67. (A) (B) Independent verification (explicitly present or the means for doing it is described)

Numbers and Data Presentation

68-71. (2 pts each) We described 4 types of data/numbers regarding their similarity to what they represent. Which type of number is indicated by each question below?

A) total counts B) simple extrapolations C) compound extrapolations (data conversions) D) fabrications

68. The evidence that N. Vietnam had launched a second attack on the U.S. in the Gulf of Tonkin in August 1964: A B C D

69. The frequent use of the number 50,000 in the media: A B C D

70. The grades that determine your GPA: A B C D

71. The cost of attending UT for 4 years based on the cost of your first 2 years: A B C D

72-74. (4pts). Which of the following points were made in the data presentation lecture?

72 (A) (B) The raw (unconverted) data should always be presented.

73 (A) (B) Different scales of presentation can change the perception of the data.

74 (A) (B) Absolute risks versus relative risks: A drug advertised as reducing heart attacks by 50% means that half the patients will be saved by the drug.

75. (4 pts.) Exam Key Code: **Fill in** bubble **(A)** on question 75 to indicate your exam code; leave the other bubbles blank. Also, fill in the correct bubbles for your name and EID on the scantron form.