39 Questions, 4 pages Cors 236 Test 3 (9 Mar, 2022)

Your answers go on Canvas Test_3_2022.

One answer and only one answer per question. Leaving a question blank or filling in 2+ answers will be incorrect no matter what. (Canvas should not allow you to choose more than 1.)

Where relevant, the goal is underlined. *Italicized phrases are true.* Do not assume more than is given in a question.

A = True, B = False unless indicated otherwise. If any part of an answer is incorrect, treat all of it as incorrect. If different parts of an option are inconsistent with each other, consider it incorrect.

Data Quality: Errors and fixes

(RPA = rounding, precision, accuracy; H&T = human & technical; standards = knowns)

1-3. (3pt)) Which in the following list of examples describe(s) a probable basis of RPA error in making measurements or would be affected by RPA error?

(A) is RPA (B) is not.

1. (A)(B) A thermometer that has lost some mercury now reads 3° low.

2. (A)(B) The display on a piece of equipment shows only 2 decimal places

3. (A)(B) Measuring the width of a board to the nearest inch when one needs to know how many boards stacked side-by-side will be needed to fill a gap of many feet.

4. (2.5 pts) A clinical trial of a new drug is careful to ensure that neither its subjects nor its observers know which group a subject is assigned to. Unknown to them, the first subjects to enroll were assigned to the treatment group until that group filled, then the last subjects to enroll were all assigned to the control group. The study results indicated that the drug had a significant effect when compared to the control group, even though in reality there is no effect of the drug. What type of error accounts for the finding of an effect of the drug when there really is none?

(A) Sampling (B) Bias (C) Human/Technical (D) RPA (E) Faulty Data conversion/analysis (F) None

5. (2.5 pts) A policeman with a radar machine clocks you at 75.2 mph when your speedometer indicated that you were going 66, and your speedometer reads out in units of 1 mph. What type of error is indicated by the difference between the these two data of the same speed?

(A) Sampling (B) Bias (C) Human/Technical (D) RPA (E) Faulty Data conversion/analysis (F) None

6. (2.5 pts) A company gathers safety data on a new drug for which it owns the rights. It's analysis of those data indicates that the drug is safe, and it publishes the full data as part of its application for drug approval. An independent organization uses the published data in its own analysis and concludes that the drug is not safe. What type of error most likely underlies the difference in conclusions here?

(A) Sampling (B) Bias (C) Human/Technical (D) RPA (E) Faulty Data conversion/analysis (F) None

7-10 (4pts). Which ideal data features are explicitly present in the paragraph? If a feature is present in some ways but absent in others, indicate that it is present.

You are trying to determine whether chemical fertilizers yield more produce in your garden than does manure (which is organic). You thus divide your garden into 4 plots. Each plot is planted with the same strains of tomato and beans. You flip a coin to choose two of the plots for the manure and use the chemical fertilizer on the other two plots. As produce is harvested, you weigh the amounts from each plot. Periodically, you confirm that the scale is accurate by weighing a known 1-pound piece of steel. At the end of the year, you talley the total produce yield from each of the 4 plots.

(A) The feature is indicated. (B) not indicated or absent

7. (A)(B) Replication

8. (A)(B) Standards

9. (A)(B) Random

10. (A)(B) Blind at least one way

11-13 (3.5 pts) Two forms of an advertisement are tested to decide which one will get more attention. Either of the two forms to each of 200 students, assigned by drawing names from a hat. They each then fill out a questionnaire that reveals how well the advertisement worked. The questionnaire is multiple choice and scored on a machine whose performance is checked initially by putting through a form whose answers are all known in advance. Overall, the responses to the two ads are slightly different, but a statistical analysis indicates that these differences are not significant. Which of the following are true about this design?

(A) = TRUE (B) = False

11. (A)(B) The use of 2 forms is a type of replication.

12. (A)(B) The use of a multiple choice form scored by a machine rules out the possibility of bias in scoring of their responses.

13. (A)(B) The fact that a statistical test was not significant means that there was no sampling error in the responses to the two ads.

14-18. (**7 pts**) Which options identify a valid "fix" for the type of error (or possible type of error) indicated; a "fix" may either reduce that error or allow you to detect that error some of the time. The (possible) error is indicated with <u>underline</u>.
A = the fix is valid; **B** = the fix is not valid

The 'Error'	Fix (choose A if the fix is valid)
14. Student <u>evaluations of instructors are not representative of the class</u> because they are voluntary, and the only students who provide evaluations out are those with strong feelings about the class.	**14. (A)(B)** Require all students in the class to fill out evaluations.
15. The <u>grading of written assignments by instructors is biased</u> because the assignments have student names visible, so the instructor can allow prior prejudices to subconciously influence grading.	**15. (A)(B)** Have the instructor grade assignments in random order.
16. A drug testing company often <u>mixes samples and thus gives the wrong result</u>.	16. (**A)(B)** Split the sample from one individual into two tubes labeled differently and send both tubes to the company. (*By labeling the two tubes differently, the company cannot know they are from the same source.*) Compare the results for both tubes.
17. You are thinking of making a large order of apples from a local supplier. You have two suppliers to which you may place the order (Abbott and Baker), and you want to be sure that you pick the supplier with the best quality apples. You have purchased 2 apples from each supplier and found that the 2 apples from Abbott were on average better than the 2 from Baker. What can you do before placing the large order to be more confident that <u>your observed difference between 2 Abbott and 2 Baker apples is representative of a real difference</u> between the two suppliers?	**17. (A)(B)** Replication: sample more apples from both suppliers before placing the large order.
18. You compare the average caffeine content of Pepsi in a sample of 100 bottles produced in Seattle with that of 200 bottles produced in Los Angeles. What can you do to be more confident that <u>sampling error cannot account for any difference you find between the Seattle bottles and the Los Angeles bottles?</u>	**18. (A)(B)** Decrease the size of the Lost Angeles sample to 100 bottles so that both samples have 100 bottles.

19-23. (5 pts) Before subjecting your employees to drug tests, you decide to assess the accuracy of the testing lab. Following the recommendations you receive from a consulting firm, you do the following test on two separate occasions. You take a sample from yourself, split it into 3 tubes, each with completely different identifying information (so that the lab thinks that all samples are different), and send all three tubes for testing to the same lab.
Which of the following features of ideal data are indicated? (A) is indicated (B) is not indicated

19 (A) (B) Explicit protocol
20 (A) (B) Replication
21 (A) (B) Standards
22 (A) (B) Random
23 (A) (B) Blind

24-27 (1 pt each) You will send pairs of tubes to a lab for analysis. For each pair of tubes, you are to decide whether replication for the characteristic indicated is present, absent or unknown to you and also whether it would be known to the lab receiving the samples. (Replication means the characteristic is the same for both samples.) You know everything given in the table. The lab only knows what is written on the tube: if a tube has a person's name on it, the lab can assume that the tube contents belong to the name of the person on the label and can infer gender but nothing else. If a tube is labeled with a number, the contents are completely unknown to the lab but known to you to the extent given in the table. However, if two tubes are labeled the same, the lab can assume the contents are the same. A question mark (?) indicates that the state of that particular sample is unknown to you. You may be able to use other information in the table to decide its property. (Gender, marker type and blood type do not change from sample to sample of the same individual, even if the assays are sometimes ambiguous.) Your options for tube contents and tube labels are:

tube	*tube label* -- what you and the lab each see	*Contents are from* – what only you see	*Gender*	Blood type	Marker type
(1)	Marsha Timmins	Marsha Timmins	Female	A	+
(2)	#45	Robert Plant	Male	O	?
(3)	#203	Patsy Cline	Female	A	+
(4)	Justin Hayward	Justin Hayward	Male	?	negative
(5)	#1973	James Page	Male	O	+
(6)	Guy Clark	Guy Clark	Male	B	negative
(7)	Nanci Griffith	Nanci Griffith	Female	A	negative
(8)	#45	Robert Plant	Male	O	?
(9)	#93	Justin Hayward	Male	A	negative
(10)	Chrissie Hynde	Chrissie Hynde	Female	B	negative

In the following questions, indicate which pairs of tubes (if any) satisfy the specified criteria.

(A) Absence of replication is known to you, and the lab cannot infer the absence

(B) Absence of replication is known to you and the lab can infer the absence

(C) Presence of replication is known to you, and the lab cannot infer the replication

(D) Presence of replication is known to you, and the lab can infer the replication

(E) Replication is unknown to you and unknown to the lab

24. (A)(B)(C)(D) (E) tubes 3 and 5 analyzed for blood type
25. (A)(B)(C)(D) (E) tubes 4 & 6 analyzed for marker type
26. (A)(B)(C)(D) (E) tubes 1 & 7 analyzed for gender
27. (A)(B)(C)(D) (E) tubes 4 & 8 analyzed for gender

28-30. (3 pts) A farmer wishes to know which of two melon strains (X, Y) produces the highest yield. He/she notes that all plants of strain X have 3 melons, but that different individual plants of strain Y have 2, 3 or 4 melons. The goal is to determine whether the average number of melon per plant is higher for strain X or Y. Which are true about making this calculation?

(A) = true, (B) = false

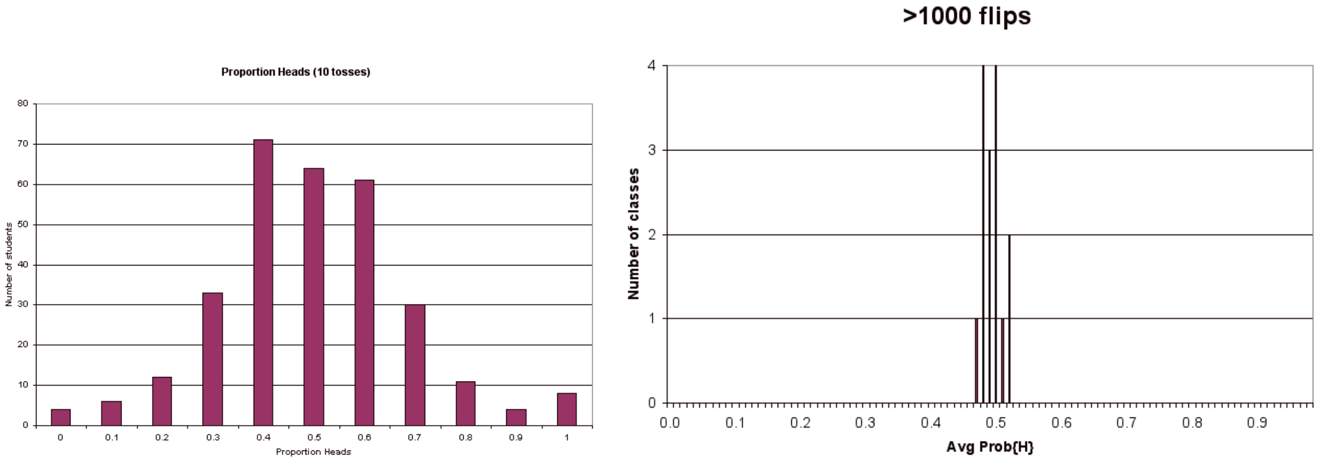
28. (A) (B) Sampling error may affect the calculated average number of melon per plant on strain Y but not X.

29. (A) (B) Measuring the number of melons per plant of strain Y will be affected by RPA error.

30. (A) (B) Once the plants have been chosen for counting, precautions are needed to avoid bias in the actual counting (for example, you would want to count melons without knowing which type of plant it was).

31-34 (4 pts). The following pair of graphs (or something similar) was shown in relation to the coin flip exercise in class. The horizontal axis is the proportion heads, and both horizontal axes span 0 to 1. The left graph is based on 10 flips per observation the right is based on over 1000 flips per observation. The graphs are thus actual, observed distributions of outcomes. The vertical axis is the number of observations.

Which points were illustrated by either or both graphs (or were made in class about these graphs)?



(A) the point is illustrated or was made in class (B) not illustrated or made in class

31. (A)(B) There is greater bias in the left graph, because the left shows that more people failed to get the right proportion of heads.

32. (A)(B) The ruggedness of both distributions (the lack of a clear peak, for example) is due to human and technical error

33. (A)(B) The point of the comparison was that replication reduces sampling error

34. (A)(B) The right graph has the least RPA error.

35-36. (2 pts) For which of the following would the described method of selecting subjects likely be free of bias for the underlined goal?

(A) = Avoids bias, **(B)** = Not

35 (A)(B) In measuring the average health of all dogs in all Moscow animal shelters (there are several Moscow shelters), you randomly choose 10 dogs from one shelter.

36 (A)(B) In attempting to measure the voting preferences of the average Moscow resident who is eligible to vote, you obtain records of all residents who have registered to vote and randomly select 98 names for your survey.

37-39. (4 pts) Which are true about possible uses and consequences of an explicit protocol? **(A)** = TRUE, (B) = False

37 (A)(B) An explicit protocol can be used by someone who did not conduct the study to identify types of errors likely to be present in the data.

38 (A)(B) By closely following an explicit protocol, bias is necessarily eliminated in the data

39 (A)(B) An explicit protocol allows the data gathering to be repeated under similar conditions should the study be repeated by someone else.