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

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. (1pt) A taste test of FeverTree ginger beer versus Bundaberg ginger beer is conducted with 100 students. 75 of the students prefer FeverTree. A difference this large is too large to be due to chance If the two brands were equally preferred. To what type of error do we attribute the observed preference for FeverTree?

(A) Bias (B) Sampling (C) RPA (D) H&T (E) None

2. (1pt) Universities routinely allow students to evaluate professors at the end of a course. However, these evaluations are often voluntary for students, with the result that only a modest fraction of students in a class provides an evaluation. Furthermore, those students who provide evaluations are often motivated to do so by strong feelings (good and bad). When average evaluation received for a large number of students fails to closely match the true student attitudes about the class, what kind of error are these evaluations most likely to have experienced?

(A) Bias (B) Sampling (C) RPA (D) H&T (E) None

3. (1pt) You want to count the number of sheets in a roll of toilet paper. But you are lazy. So you measure the length of four sheets together as 16.3 inches (to the nearest 0.1 inch), then measure the total length of the roll as 2005.0 inches. Dividing, you calculate that there are 492 sheets in the roll, and your division is accurate to the nearest sheet. Although your calculation is likely to be close, it may not be exact. What type of error may plausibly underlie your use of this method?

(A) Bias (B) Sampling (C) RPA (D) H&T (E) None

4-8 (4pts). Which ideal data features are explicitly present in the paragraph?

Bull is wondering which type of tire provides the longest wear: Toyo or Schwab. He buys two tires of each kind and puts all 4 tires on his truck. (The Schwab tires are matched for size and rating with the Toyo tires.) The initial positions of the tires are Toyo on the front, Schwab on the back for the first 10,000 miles, then reversed for the next 10,000 miles. After the 20,000 miles, each tire is measured for remaining tread, measured both at Schwab and also at Bruneel (two local tire shops), although the mechanics who measure the treads are not told the intent of Bull's interest in the measurements. The Schwab and Bruneel measurements are compared and found to agree, with the most remaining tread on the 2 Toyo tires.

. **(A) = Present** (B) = absent or not described

4. (A)(B) Explicit Protocol	**7. (A)(B)** Replication
5. (A)(B) Randomization	**8. (A)(B)** Standards
6. (A)(B) Blind	

9-12 (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 it is the same across 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)	Chrissie Hynde	Chrissie Hynde	Female	B	negative
(2)	Justin Hayward	Justin Hayward	Male	?	negative
(3)	Margo Timmins	Margo Timmins	Female	A	+
(4)	#2013	Robert Plant	Male	O	?
(5)	#1869	Patsy Cline	Female	A	+
(6)	#1000	James Page	Male	O	+
(7)	Guy Clark	Guy Clark	Male	B	negative
(8)	Nanci Griffith	Nanci Griffith	Female	A	negative
(9)	#2013	Robert Plant	Male	O	?
(10)	#7193	Justin Hayward	Male	A	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

9. (A)(B)(C)(D) (E) tubes 4 and 6 analyzed for blood type

10. (A)(B)(C)(D) (E) tubes 2 & 7 analyzed for marker type

11. (A)(B)(C)(D) (E) tubes 3 & 8 analyzed for gender

12. (A)(B)(C)(D) (E) tubes 2 & 9 analyzed for gender

13-16. (4 pts) (*changed from previous years*) In an attempt to evaluate bullet lead analysis as a forensic tool, two testing companies (*Alphametrics*, *Betadynamics*) measure the trace arsenic and tin levels in boxes of ammunition from the *CCI* factory in Lewiston, ID. The levels of arsenic and tin are expected to be very small in bullet lead, but detectable with their machines. The batch of bullets used is a mix of 1 million bullets manufactured by *CCI* in August and 2 million bullets manufactured in October. The batch of lead used to make the bullets was changed between August and October, so the trace elements may well have differed between those months. *Alphametrics* measures 1000 randomly-chosen bullets from this mix, *Betadynamics* measures 8,000 randomly-chosen bullets from the same mix. What can be said about the data acquired by *Alphametrics* and *Betadynamics*?

Which are true? **(A)** = TRUE, (B) = False

13. (A)(B) RPA error in the trace-element measurements would be avoided by ensuring that the machines are properly calibrated before being used.

14. (A)(B) Randomization in the choice of samples removes any concern of a consistent difference between the *Alphametrics* and *Betadynamics* samples

15. (A)(B) Human and technical error in the determinations by each testing company is no longer a possibility if the machines were properly calibrated and tested with standards.

16. (A)(B) Regardless of whether there are consistent differences of tin and arsenic levels between August and October, sampling error of the tin and arsenic content will be **smaller** in *Alphametrics* sample (of 1,000) than in *Betadynamics* sample. (of 8,000)

17-26. Blind designs. In a taste test of dog food versus liver paté, the goal of the study is to assess the subject's possible preference for the taste or smell of canned dog food over liver paté (as specified); note that taste includes smell, but smell does not include taste. You may wish to recall class discussion for some of these. Do not infer more than is given.

You are to decide, given the design for each question, what factors might influence the subject's discrimination in ways other than taste or smell (as specified). That is, what factors still need to be addressed to avoid possible bias and other influences? Anything that allows the subject to identify the dog food or that leads them to sense a difference in the presentation of the foods should be considered to possibly affect their preference or discrimination.

17-21 (2 pts) Design for a test of subjects' preferences based on smell alone. The researcher carries out all preparations prior to the subjects' arrival. Both foods are made from meat. The two meats differ in texture: the dog food has meaty chunks, and the paté is homogenized. The two meats also differ in color.

Each food is taken from the original container and put directly in the serving containers without modification; serving containers are labeled with a letter (A, B, C, …), chosen randomly. All subjects are familiar with the appearance of unmodified canned dog food and with that of paté. The researcher knows which serving containers have the dog food. The researcher engages in casual conversation with the subjects during the tests but avoids intentionally conveying information to subjects. Smelling involves no physical contact between the subject and the food, just having the subject smell 4" above a serving dish held by the researcher. All smelling tests are conducted before any of the taste tests. Foods are presented in random order to subjects. The subjects are blindfolded during the smell test.

Given the design, what factors should be considered as possibly influencing the subjects' preferences in ways other than smell?

(A) the design does not adequately avoid the effects of this factor

(B) the design adequately avoids the effects of this factor

17 (A)(B) taste

18 (A)(B) labeling of the container

19 (A)(B) texture

20 (A)(B) color

21 (A)(B) unconscious influence by the researcher

22-26 (2 pts) Design for a test of subjects' preferences based on taste (which includes smell). The researcher carries out all preparations prior to the subjects' arrival. Both foods are made from meat. The two meats differ in texture: the dog food has meaty chunks, and the paté is homogenized. The two meats also differ in color.

Each food is taken from the original container and put directly in the serving containers without modification; serving containers are labeled with a letter (A, B, C, …), chosen randomly. All subjects are familiar with the appearance of unmodified canned dog food and with that of paté. The researcher knows which serving containers have the dog food. The researcher engages in casual conversation with the subjects during the tests but avoids intentionally conveying information to subjects. Tasting involves the subject eating one or more spoonsful, and subjects are not blindfolded. Foods are presented in random order to subjects.

Given the design, what factors should be considered as possibly affecting the subjects' preferences in ways other than taste?

(A) the design does not adequately avoid the effects of this factor

(B) the design adequately avoids the effects of this factor

22 (A)(B) smell

23 (A)(B) labeling of the container

24 (A)(B) texture

25 (A)(B) color

26 (A)(B) unconscious influence by the researcher

27-28. (2 pts) For which of the following would the described method of selecting subjects avoid bias for the underline goal? You may have to rely on common knowledge about car ownership for one of the questions. **(A)** = Avoids bias, **(B)** = Not

27 (A)(B) In measuring the average health of all dogs in the Moscow animal shelter, you randomly choose 10 dogs from that shelter.

28) (A)(B) In attempting to measure the voting preferences of the average Moscow resident, you obtain records of all residents who have filed car registrations and randomly select 200 names for your survey.

29-31. (2 pts) Which are true about possible uses and consequences of an explicit protocol? **(A)** = TRUE, (B) = False

29 (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.

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

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

Criminal Justice (RMP is random match probability)

We mentioned 5 features of an 'ideal' forensic method for matching a suspect with a forensic sample:

(i) labs/experts pass blind proficiency tests, (ii) independent verification possible, (iii) reference database, (iv) discrete characteristics, and (v) all procedures blind. Discrete characteristics are all-or-none, not subject to RPA error.

32-34. (3 pts) Which of the following would constitute a legitimate reference database for the goal indicated (underlined)? You want the database to allow calculating the RMP of a match between a suspect and a crimescene sample. A legitimate reference database should not have major flaws. In all cases listed below, consider the numbers listed to be sufficient if there are no other problems.

(A) = is a legitimate reference database for the case in question (B) = Has major flaws

32. (A)(B) For a shoe-print match obrtained from a crime committed in 2015, prints of 5000 different shoes of all types found in stores and online from 2005- 2015.

33. (A)(B) For a match between the bite mark on a murder victim and the suspect's own bite mark, bite marks from 10,000 different individuals in the same area where the murder victim and suspect lived.

34. (A)(B) For a hair match, the characteristics of 1500 hairs from both the victim's and suspect's head

35-38. (3 pts) Our class was subjected to a test of eyewitness identification, and similar tests conducted in previous classes were discussed and one set of results was shown. You were also supposed to watch a video of an audience subjected to an eyewitness test,. From these demonstrations and discussions, which are true? **(A)** = TRUE (B) = false

35. (A)(B) In a past demonstration done in class, instructions given immediately before the lineup were found to have a major effect on the results.

36 (A)(B) More than half of those who responded in our class correctly identified the perpetrator in the lineup.

37. (A)(B) In the audience-participation video online, less than half the audience correctly identified the perpetrator in the lineup.

38. (A)(B) Eyewitness identification satisfies 3 of the 4 criteria of an ideal matching method.

39-42. (4 pts) A company develops a forensic method to determine whether the composition of grains of sand on a suspect's shoes matches that left in footprints at a crime scene. Sand grains can come from distinctly different rock types (granite, limestone, marble, …). The company's machine records the different rock types in the sand sample but also accounts for the relative amounts of each rock type in the grains (e.g., whether 11.2% of the grains are of a particular type) and also accounts for the distribution of grain weights of each rock type (which can range from 0.000000001 grams to 0.01 grams) . Using proprietary software kept secret by the company, the crime scene sample is compared to the suspect's sample to determine if there is a match. If a match is declared, the company calculates the RMP from samples it has obtained from 5,000 different geographical regions, spanning all possible sources of sand that are relevant to crimes committed locally. Lab personnel have consistently shown their ability to provide useful data for prosecutors that have led to convictions. **(A)** = TRUE (B) = false

39. (A) (B) Evidence is presented that lab personnel have passed proficiency tests

40. (A) (B) Independent verification of a declared match is indicated or is described as being straightforward

41. (A) (B) an appropriate reference database that can be screened for a RMP is indicated

42. (A) (B) at least some of the characteristics used are not discrete

43. (-1 pt. for getting it wrong) Exam Key Code **A**: Answer **(A)** on question **43** to indicate your exam code.

Format: **last name first name** , last 4 digits of your V00 number , 1 answer 2 answer …