name (required) _____

Errors in Data

1. (4 pts) Which of the following demonstrations were used to illustrate errors in data, and which options also correctly identify the type of error illustrated? MTF

> A) coin flip: *even though the demo was and has always been conducted by having each student flip a coin 10 times and only 10 times*, it was nonetheless possible to use only the data from this year and previous years to illustrate that <u>replication reduces sampling error</u>.
>
> B) the bias illustrated in the 'choice' demo was the type that would be fixed by using a blind design
>
> C) RPA error was illustrated by measuring the width of a coin with two different instruments and showing that the instrument displaying fewer decimals produced a less accurate measure than the instrument displaying more decimals.
>
> D) human and technical error was illustrated with a form given to the class and filled out incorrectly by over 10% of the students responding

2. (4pts) Which in the following list constitute(s) bias? MTF

> A) a piece of equipment whose display shows 2 decimal places but whose sensor device reads to 4 decimal places; thus the display is not as accurate as is the sensor itself.
>
> B) a lineup procedure leads witnesses to think that a particular person in the lineup is the one they saw, even if it is the wrong person.
>
> C) a technician accidentally mixes up samples in a DNA test
>
> D) a person unconsciously feels better because they think they took a pill that should make them well

3. (6pts) For which of the following is sampling error present? MTF

> A) When several individuals are chosen at random from a population
> B) When scoring/measuring an individual for a discrete characteristic (all-or-none)
> C) When scoring/measuring an individual for a quantitative characteristic (length, weight)
> D) When a machine is zeroed incorrectly
> E) When creating a reference database that does not include the entire population
> F) When creating a reference database that does include the entire population

4-6. In the following questions, indicate which type of error is indicated. One answer per question. The underlined phrase indicates the part to be explained by (A)-(E).

4 . (4 pts) A company executive controls 10 stores in the Austin area. All 10 stores have the same inventories and product turnover rates, and their monthly profits are similar when averaged over several years (the differences in average profits are not statistically significant). (Differences in profits can be attributed to differences in the numbers of items sold.) To downsize, this executive decides to close one store. The decision of which store will be closed is based on the store with the lowest October, 2010 profit, even though the profits for that store are within its historical range. What type of error likely underlies <u>the variation in monthly profit that leads to one store being lower than the others</u>? One answer only

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

5. (4 pts) Two clinical trials of the same drug are conducted. In trial I, patients are inadvertently told that the drug is bitter and that the placebo is bland. Trial II does not convey anything about the characteristics of the drug or placebo. The parent company figures that they will merely combine the results of both trials and forget about this difference in protocol. But when the results are in, the drug has shown a much higher success rate in trial I than in II and the results cannot be combined. What type of error likely explains the reason for trial I giving a more favorable result for the drug?   (one answer only)

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

6. (4 pts) Men and women in Texas are polled for their attitudes on the death penalty. In one poll of 1000 men and 1000 women, 67% of men and only 44% of women favor it. In another poll of 1000 different men and 1000 different women, 65% of men favor it and 42% of women favor it. What type of error is attributable to the consistent difference in male/female attitudes about the death penalty?   (one answer only)

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

Errors and Fixes

7 (5 pts). The following pair of graphs (or something similar) was shown in relation to the coin flip demo in class. Which points were illustrated by either of both graphs? The horizontal axis is the proportion heads, and both horizontal axes span 0 to 1. MTF
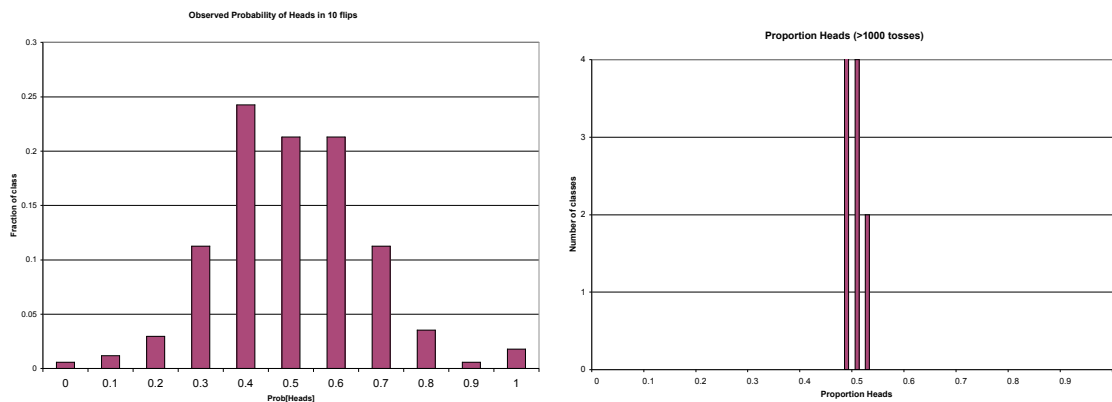
(A) Neither graph addresses RPA error, bias, or H&T error.

(B) Sampling error can be illustrated with either graph.

(C) As more data are gathered, we expect the right graph to fill out like the left graph, spanning the full axis and being shaped like the left graph.

(D)The right graph has the least RPA error.

(E) Replication is evident in both graphs.

(8-**9**). 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. all questions are MTF

(A) explicit protocol	(C) standards	(E) blind
(B) replication	(D) random	(F) none

8. (4 pts) You start your own drug testing company. The company consists of you and an expensive machine to do the testing. You set it up so that every sample is tested twice, then tested again if the two numbers do not agree. Each day's set of runs begins with a blank, then a sample spiked to a known concentration. The other samples then follow in the order received. MTF

(A)   (B)   (C)   (D)   (E)   (F)

9. (4pts) You manage a grocery store. This store has a large, refrigerated unit for storage of produce before it goes out for sale. You decide to see how long produce can be stored before it gets bad enough that customers will not buy it. Your test involves using 4 types of produce (2 fruits, 2 veggies) with 100 items of each of the 4 types. After 1 week of storage, 20 items of each produce type are chosen (by drawing numbers from a hat) and placed out for sale. Those that sell are recorded. At week 2, 20 more items for each produce type are selected similarly and put out for sale. The process continues through week 5. The customer is never told the produce age.   To control for week-week differences in consumer preferences that have nothing to do with produce quality, you find out the sales rates of like items in other stores during the same weeks. MTF

(A)   (B)   (C)   (D)   (E)   (F)

10. (4pts) Which options identify a "fix" for the type of error indicated; a "fix" may either reduce that error or at least allow you to detect/measure that error. MTF

A) error: sample mixup during testing. **Fix**: split samples when they are acquired and submit both versions of the same sample for testing

B) error: RPA error in the level of drugs detected **Fix**: design a protocol that is easier to follow when doing the test

C) error: lab falsifies results to give the prosecution its desired results. **Fix**: include standards with the suspect and victim samples.

D) error: lab occasionally declares matches that are not real, but they often go undetected. **Fix**: code samples to ensure blind analysis

11 (4 pts) Which of the following points about protocols (written procedures for data) are correct? (MTF)

A) If the protocol is followed, the types of errors likely to be present in data can be understood from reading the protocol

B) An important means of reducing errors in data is to change (improve) a protocol

C) If followed to the letter, a written procedure minimizes errors in the data. That is, most error creeps in because the data are not gathered strictly according to the protocol

D) Separate protocols may apply to gathering data, analyzing data, and reporting data.

DNA and Criminal Justice

(12-14). Do-it-yourself protocol. You are conducting an external review/test of a genotyping lab. Your job is to send **two tubes** to the lab, with labels. There are several options for the content of and label on a tube. You must decide which contents to send and how to label the tubes so that the features of ideal data requested in the question are present. 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. If a tube is labeled with a number, the contents are unknown to the lab but known to you. **A ? indicates that you do not know the individual's status for that characteristic.** Your options for tube contents and tube labels are:

option	_tube label_	_Contents in the tube are from_	_Blood type_	_Gender_	_Marker status_
(A)	Laura Baker	Laura Baker	B	Female	+
(B)	Darin Rokyta	Harry Wichman	A	Male	?
(C)	Rachael Springman	Rachael Springman	AB	Female	+
(D)	#101	Harry Wichman	A	Male	?
(E)	#218	Patsy Cline	A	Female	negative
(F)	#10	Pam Hines	O	Female	negative
(G)	Jerry Allison	Jerry Allison	A	Male	+
(H)	#101	Brent Iverson	AB	Male	+
(I)	No combination of tubes can satisfy the protocol				

In the following questions, choose two letters among options (A)-(H) to describe the two tubes that will be sent to the lab. *The tube labels are the only information the lab receives about the samples, and the lab does not have prior information about the individuals.* **If it is possible to satisfy the protocol, the question will require exactly <u>two</u> letters and only two letters** -- one for each tube. Thus, the answer for a question might be (A) & (B), or it might be (D) & (F). If more than one pair of options are possible correct answers, fill in only one correct pair of options. Thus, if (A) & (B) is one acceptable answer, and (C) & (D) is another acceptable answer, fill in either (A)&(B) **or** (C)&(D), but not both. If a factor (such as identity, blood type, gender, etc.) is not specified in the protocol, then that factor will be ignored in grading the answer.

Alternatively, **if a protocol cannot be satisfied** with two from (A)-(H)**, fill in (I).**

12. (3 pts) Choose two tubes to achieve <u>replication of blood type and marker status but not gender</u>. You should know that both are replicated and that nothing else is replicated, and the <u>replication should be blind to the lab</u> (you can assume the lab will know gender from the name on the tube).

two tubes or I:　　　(A)　　(B)　　(C)　　(D)　　(E)　　(F)　　(G)　　(H)　　(I)

13. (3 pts) Make the tubes <u>replicated for marker, gender and blood type,</u> but <u>all replication is blind to the lab</u>. You should know that the replication is present, even if you don't know the marker, gender, and blood type.

two tubes or I:　　　(A)　　(B)　　(C)　　(D)　　(E)　　(F)　　(G)　　(H)　　(I)

14. (3pts) Choose two tubes so that <u>none of blood type, gender, marker status is replicated,</u> and <u>absence of replication should be blind to the lab</u>.

two tubes or I:　　　(A)　　(B)　　(C)　　(D)　　(E)　　(F)　　(G)　　(H)　　(I)

15-17.l This statement applies to the following 3 questions. *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.*

15. (5pts) Eyewitness mis-identification is a common cause of wrongful conviction. Which of the following points were made?

A) A photo catalog of possible suspects may be considered as a reference database for eyewitness ID, used in the same fashion as reference databases for other methods.

B) The class demo showed that there is a high rate of eyewitness error.

C) The class demo showed that the instructions given before a lineup have a large influence on misidentification by eyewitnesses

D) Eyewitness identification fails on at least 3 of the 4 ideal features of a forensic method.

16. (6pts) Which of the following are true? MTF

A) A proficiency test serves no purpose if the other 3 features of a method are satisfied.

B) Independent verification requires that the characteristics being measured are permanent

C) The claim that a match is unique –that the suspect must be the source of the sample - is an indication that a reference database is missing or not used properly

D) The use of discrete characteristics eliminates errors in the scoring of individuals

E) When a method satisfies the independent verification criterion, another lab/expert should be able to take the same samples and reach the same conclusions about a match.

F) A reference database must include the entire population to allow a RMP calculation

17. (7pts) In the past, several forensics methods have been used in courts ostensibly to identify the perpetrators of crimes. Which of the following are true about those methods? *The table on line may be used to answer some of these questions.*

A) The only methods employing the measure of discrete characteristics were DNA, fingerprints (before and after 1990) and hair matching.

B) A reference database for fingerprints existed throughout much of the last century (pre-1990) but was unusable because there was no easy way to screen it.

C) The methods listed included dog sniffing, bite marks, and shoe print ID.

D) The video on bullet lead analysis claimed that the significance of a match (between a bullet found at a crime scene and in a box of shells found in the suspect's possession) was grossly overstated in testimony. There was no suggestion that the chemical analysis of lead itself was flawed, only that the signatures of an individual bullet were shared by thousands or millions of other bullets.

E) Fingerprint experts who submitted to voluntary proficiency tests in the 1990s were found to have error rates typically under 2%.

F) Hair matching was discredited because its reference database was biased toward miniorities.

G) In some of the recent dog sniffing cases overturned, the person had been convicted by supposed dog identification even when other samples did not place the suspect at the crime scene.

18. (5pts) Combining sources of error in wrongful matches. Lecture described an example in which the RMP was given and in which it was reported that the lab falsely declared matches a certain fraction of the time. Use the principle revealed in that discussion to solve this new problem. A match has been declared between a forensic sample and a suspect. The RMP (random match probability) is 1/1000 (= 0.001). It has further been revealed that, through its own mistakes, the lab wrongfully declares matches 1% of the time. In addition, it has been revealed that the crime scene specialists sometimes mix up samples even before sending them to the lab; this mixup by itself leads to an erroneous match 0.5% of the time. There are thus 3 separate reasons why a suspect may not be the source of a sample when a match is declared: the match is 'random,' the lab made a mistake, or the specialists mixed up the samples. Given these data, what is the chance that the suspect is not the source of the sample? Choose the answer closest to correct, if there is indeed a correct answer in this case. One answer only.

A) The largest of the three: 1%

B) The smallest of the three: 0.001

C) The product of the 3: 1% x 0.5% x 0.001.

D) The sum of the 3: 1% + 0.5% + 0.001 = 0.016

E) The sum of the lab error rate and the RMP: 1% + 0.001 = 0.011

F) You cannot combine error rates from different sources

19. (5 pts). A new method of sample-suspect matching has just been introduced into court, as described below. Which of the 4 features of 'ideal forensics' are indicated? MTF

The method measures minute quantities of 12 different types of molecules on the outside of skin cells. The lab claims that the 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 0 to 10^{-3}, measured as accurately as the machine will perform. Their measurement method is considered 'proprietary' because it is covered by patents, and only this one lab is allowed to perform the test. Other labs are therefore not available to do the analysis. 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.

A) Reference Database B) Discrete characteristics C) Independent verification D) blind proficiency tests

Numbers

20-24 (2 pts each).

We described 4 types of data/numbers regarding their similarity to what they represent:

A) actual counts   B) simple extrapolations   C) data conversions   D) fabrications

Which type of number is indicated by each question below?

20. The votes counted in an election to determine the winner (one only): A B C D

21. The win-loss record of a baseball pitcher during his career (one only): A B C D

22. A poll to predict the fraction of voters likely to choose the Democratic candidate (one only): A B C D

23. The readout of the scantron machine as it scans your form (one only): A B C D

24. The frequent use of the number 50,000 in the media (one only): A B C D

Data Presentation

25. (4pts). Which of the following points about the data presentation lecture and chapter are true? MTF

   A)   A major theme of this chapter/lecture (and the reason we spent so little time on it) is that, although the presentation of data affects how well people remember the data, the presentation does not affect how people interpret or understand the data

   B)   When presenting data on benefits of a medication, the use of <u>absolute</u> risks (lives saved per 1000 individuals) gives a more accurate view of the benefits than does <u>relative</u> risks.

   C)   More people got the right answer about the chance a positive HIV test was a 'false positive' when the data were presented as numbers per 100 people in different categories than when presented as 'conditional' probabilities.

   D)   The demonstration using graphs (histograms) showed that the use of different vertical scales impacted the impression drawn from the graph.

26. (4 pts.) Exam Key Code: Fill in bubbles (AB) on question 26 to indicate your exam code; leave the other bubbles blank. Also, fill in the correct bubbles for your name and EIDon the scantron form.

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