Biostatistics 22S:101 **Spring 2004** **Exam I** **Answers**

Name:____**ID** ____________________

There are 16 questions. Mark answers on the answer sheet at the end of the exam and also circle the answer on the exam. You may submit your scratch paper if you expect to ask for partial credit. Tables are distributed separately.

1. Here are two probability tables. In each table compute P(A|B), P(A), and P(B) and answer the next two questions.

Table I	B^C	B
A^C	0.20	0.05
A	0.20	0.55

Table II	B^C	B
A^C	0.06	0.14
A	0.24	0.56

In which table(s) is A independent of B?

1.1 a) Table I b) Table II c) Neither d) Both

In Table I which of the following is true?

1.2 a) P(A | B) > P(A) b) P(A | B) < P(A) c) P(A | B) = P(A)

2. A diagnostic test for disease D has sensitivity P(+ | D)=0.98 and specificity P(– | ~D)=0.99. The incidence (prevalence) of the disease (D) is P(D)=0.01 . Use Bayes rule to compute P(D | +) and P(~D | –).

2.1 P(D | +) is a) ≈ 0.5 b) ≈ 0.6 c) ≈ 0.7 d) ≈ 0.8 e) ≥ 0.9

2.2 P(~D | -) is a) ≤0.97 b) 0.98 c) 0.998 d) 0.9998 e) ≥ 0.9999

3. Make a stemplot (stem-and-leaf diagram) of the following observations (do not split the stems). Compute the five number summary (use the instructor's definition of quartiles).

3.5 5.7 6.3 5.1 5.5 6.2 5.2 4.4 3.5 3.9 7.1 1.5
5.5 2.8 7.6 6.1 3.3 2.9 4.4 5.0 4.1 6.6 3.6

Stems	**Leaves**
1	**5**
2	**89**
3	**35569**
4	**144**
5	**012557**
6	**1236**
7	**16**

5 Number summary
Min 1.5
Q1 3.5
Med 5.0
Q3 6.1
Max 7.6

4. Compute the mean and standard deviation of these numbers: 63, 48, 55, 62, 52, 44. Show your work on the back.

$\bar{x} =$	**54**
$s =$	**7.56**

5. Suppose that P(A) =.4 and P(B) = .5, and P(A ∩ B) = .20.

	A	~ A	
B	**0.2**	**0.3**	**0.5**
~ B	**0.2**	**0.3**	**0.5**
	0.4	**0.6**	**1.0**

5.1 Fill in the joint and marginal probabilities in the table.

5.2 Are A and B independent? No ☐ Yes ☒ Why? **Because P(A|B) = .2/.5 = 0.4 is equal to P(A) = .4**____

{ Another justification is because P(A ∩ B) = P(A)·P(B) }

6. Mom had the frugality of most people who lived through the great depression. For years she used an old dented tin measuring cup. I told my dad that it wouldn't measure properly, but he solemnly assured me that the dent wouldn't change the capacity of the cup as long as the metal was not broken. I believed him for a long time until one day I pictured in my mind what would happen if I totally flattened the cup – it would hold next to nothing.

What is the name of this type of reasoning? What is the purpose?

6.1 Name: **Thought experiment ("Dutch book" is the wrong answer.)**

6.2 Purpose: **To identify inconsistencies in a theory. (Any response involving "fair price" or "probability" is incorrect)**

7. Give an operational definition of the concept of the subjective probability of a sentence such as,"A woman will be elected president of the U.S. before 2025."

The subjective probability of this sentence is the fair price of a bet (or futures contract) that pays $1.00 if the sentence turns out to be true.

8. As of 9/26/04 the asking prices on TradeSports.com are $6.91 to place a bet that pays $10 if Bush wins the election and $3.30 to place a bet that pays $10 if Kerry wins. Suppose you make both bets. Why is that irrational? What is the name of the type of financial situation that you would have placed yourself in?

8.1 Why irrational? **Because you are guaranteed to lose 21 cents no matter who wins.**

8.2 Name? **Dutch book**

9. Here are Jake's fair prices for various bets. Which two prices are incoherent? What should they be? What law of probability tells you this?

Jake's Betting Layout

All bets pay $1	H_R price: 54¢	H_D price: 46¢
S_R price: 21¢	$H_R \cap S_R$ price: 10¢	$H_D \cap S_R$ price: 7¢
S_D price: 79¢	$H_R \cap S_D$ price: 40¢	$H_D \cap S_D$ price: 39¢

`

$H_D U S_D$ price: 90¢

9.1 First wrong price $P(H_R \cap S_R)$________________ Correct Value 14¢________________

Why? Because $P(S_R) = P(H_R \cap S_R) + P(H_D \cap S_R)$ ___

9.2 Second wrong price $H_D U S_D$________________ Correct Value 86¢ ________________

Why $P(H_D U S_D) = P(H_D) + P(S_D) - P(H_D \cap S_D) = .46 + .79 - .39$__________________________

10. There are three boxes. A, B, and C. Box A contains 5 B and 15 W chips. Box B contains 10 B and 10 W chips. Box C contains 17 B and 3 W chips. Someone selects a box at random and hands it to you. You can't look in the box, but you can draw four chips from the box (with~~out~~ replacement). You made n = 4 draws and got 1 black chip and 3 white chips. Compute the posterior probabilities of the three models (A, B, and C) given the data. Label each column with a word or phrase and also with a probability expression (for example column 2 is the Prior and the probability expression is P(Model). You may use M and D to stand for the words "model" and "data." Fill in the correct word in the probability expression under the table as well.

	Prior	Likelihood	Joint	Posterior
Model	**P(Model)**	P(D|M)	$P(D \cap M)$	P(M|D)
A (5B 15W)	0.3333	0.1055	0.0352	0.6174
B (10B 10W)	0.3333	0.0625	0.0208	0.3658
C (17B 3W)	0.3333	0.0029	0.0010	0.0168
			0.0569	

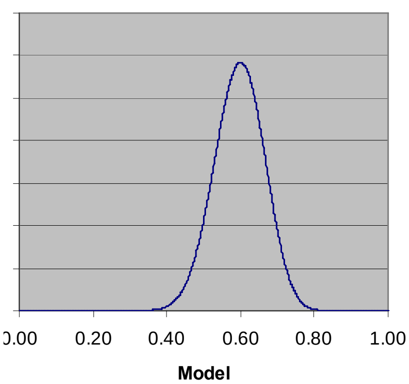
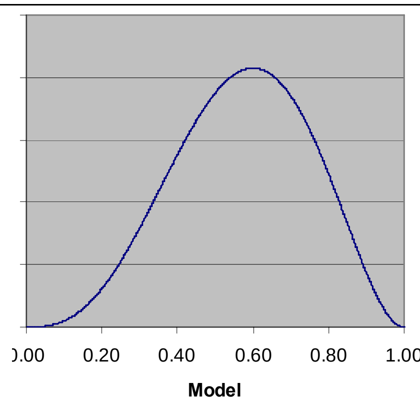
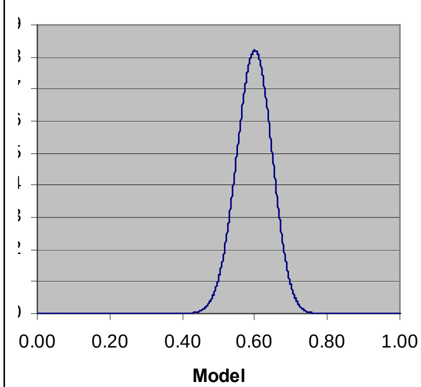
P(Data)

11. MnSOST-R is a diagnostic test to predict if a paroled sex offender will recidivate; i.e., commit another violent sexual offense (V) sometime in the future. The sensitivity of the test is P(+ | V) = 0.17 and the specificity is P(- | ~ V) = 0.97 . Assuming that the prevalence of violent sexual recidivism is P(V) = 0.30, use Bayes rule to compute P(V | +) and P(~ V | –). Show your work on the back.

11.1 P(V | +) = 0.708

11.2 P(~ V | –) = 0.73

12. Here are three posterior distributions produced by the spreadsheet Bernoulli.xls. Each was obtained by spinning a penny n times and observing the number of heads. One graph is for n = 5 and x=3, one is for n = 20 and x = 12, and one is for n = 100 and x = 60. Say which is which, and match it to one of the credible intervals shown below

A	B	C
		

Here are the 95% credible intervals (but not in the same order as the pictures.

I	II	III
0.24 to 0.90	**0.45 to 0.74**	**0.50 to 0.69**

12.1 Results for 3 heads in 5 spins are: Graph: __**B**______ Credible Interval: **I**_______

12.2 Results for 12 heads in 20 spins are: Graph: **A**_____ Credible Interval: **II**_______

12.3 Results for 60 heads in 100 spins are: Graph: **C**_____ Credible Interval: **III**______

Numbered Equations

$$P(E|R) = \frac{P(E \cap R)}{P(R)}$$

$$P(E \cap R) = P(E|R) \cdot P(R)$$

$$P(R \cup E) = P(R) + P(E) - P(R \cap E)$$

$$P(\sim R) = 1 - P(R)$$

The sentences R and E are said to be statistically *independent* if knowing that R is true does not change the probability that E is true. Thus, the sentences R and E are independent if and only if

$$P(E|R) = P(E). \tag{2.7}$$

$$\text{fair price of a standard bet} \;=\; \$P(\text{Win})$$

$$\text{expected value} = E(V) = \sum_{i=1}^{k} V_i \cdot P_i$$

$$\begin{array}{ccccc} P(H_R \cap S_R) & + & P(H_D \cap S_R) & = & P(S_R) \\ + & & + & & \\ P(H_R \cap S_D) & + & P(H_D \cap S_D) & = & P(S_D) \\ \| & & \| & & \\ P(H_R) & & P(H_D) & & \end{array}$$

$$P(H_D \cup S_D) = P(H_D) + P(S_D) - P(H_D \cap S_D)$$

$$P(H_D \cup S_D) + P(H_D \cap S_D) = P(H_D) + P(S_D)$$

$$P(\Omega) = 1$$

$$P(\varnothing) = 0$$

$$P(S^C) = 1 - P(S)$$

$$\text{fair price of } \{\$M \text{ if } S\} = \$M \cdot P(S)$$

$$\text{bookmaker's price for} \left\{ \$1 \text{ if } S \right\} = \frac{\text{gambler's stake}}{\text{gambler's stake + bookmaker's stake}}$$

$$f = P(R \cap D) + f \cdot [1 - P(D)]$$

$$P(R \cap D) = f \cdot P(D)$$

$$P(R \cap D) = P(R|D) \cdot P(D)$$

$$P(R|D) = \frac{P(R \cap D)}{P(D)}$$

If sentences T and W are independent,

$$P(T \cap W) = P(T) \cdot P(W)$$

$$EV = \sum_{i=1}^{k} V_i \cdot P(S_i)$$

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{Sum}{n}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}} = \sqrt{\frac{CSS}{n-1}} = \sqrt{\frac{centered\ sum\ of\ squares}{n-1}}$$

$$P(M_j | D) = \frac{P(M_j \cap D)}{P(D)} = \frac{P(D|M_j) \cdot P(M_j)}{\sum_{\text{all models}} P(D|M_i) \cdot P(M_i)}$$

If D = "The patient tested positive," the likelihood is

$P(D|M_1) = P(+ \mid \text{Infected}) = \text{sensitivity}$

$P(D|M_2) = P(+ \mid \text{Not infected}) = 1 - \text{specificity}$

If D = "The patient tested negative," the likelihood is

$P(D|M_1) = P(- \mid \text{Infected}) = 1 - \text{sensitivity}$

$P(D|M_2) = P(- \mid \text{Not infected}) = \text{specificity}$

$$\begin{aligned} \text{likelihood} &= P\left(x \text{ heads in } n \text{ spins} \mid \text{heads rate} = \text{p}\right) \\ &= \text{p}^{x} \cdot (1-\text{p})^{(n-x)} \end{aligned}$$