STA 113 Name: ______________________________

Probability and Statistics in Engineering: Second Test

2:15-3:30 pm Thursday, March 21, 1996

This is a **closed-book** examination, so please do not refer to your notes, the text, or to any other books. A few pages of formulas, tables, and blank work-sheets are appended to the exam. Please do not use your own scratch paper; you may use your own calculator but may not share one.

Please ask me to explain anything you don't understand in the questions (wording, definitions, *etc.*). It is to your advantage to write your solutions as clearly as possible, and to show enough of your work to make clear how you got your answer. Good luck! Please sign the Duke Honor Code:

I have neither given nor received aid on this examination.

____

signature

1.	/20
2.	/20
3.	/20
4.	/20
5.	/20
	/100

STA 113

Probability and Statistics in Engineering: Second Test

1. Consider two discrete random variables X and Y that satisfy the following conditions:

(i) X takes values $-2, 2$;
(ii) Y takes values $-1, 1$;
(iii) $\mathsf{E}[X] = -1, \quad \mathsf{E}[Y] = 0$;
(iv) $\mathsf{Cov}(X, Y) = 1$;

a. (10pt) Find the table that gives the joint probability function of X and Y.

b. (5pt) Find the conditional expectation:

$\mathsf{E}[Y|X = -2] =$ ______________

c. (5pt) Are X and Y independent? Why?

Circle one: Yes No Reasoning:

Probability and Statistics in Engineering: Second Test

2. A taxi company is trying to decide whether to purchase brand A or brand B tires for its fleet of taxis. To estimate the difference in the two brands, an experiment is conducted using 32 tires of each brand. The tires were run until they wore out. The sample means and standard deviations were:

Brand A: $\overline{x}_1 = 38{,}025$ miles $s_1 = 5{,}400$ miles
Brand B: $\overline{x}_2 = 40{,}125$ miles $s_2 = 5{,}793$ miles

a. (8pt) Compute a 90% confidence interval for the difference of the means of the two brands.

b. (8pt) Test to see if the two brands differ. Report a P-value and also tell whether to reject or accept the null hypothesis at level $\alpha = .10$. Be sure to specify both hypotheses.

c. (4pt) Explain briefly (in at most 10 words) **in the context of this problem** what a Type I error would be. Do *not* just give a definition.

Probability and Statistics in Engineering: Second Test

3. A system has two components. The lifetime of component 1 is the random variable X and the lifetime of component 2 is the random variable Y. The joint density function is

$$f(x,y) = \begin{cases} x\exp(-(1+y)x), & \text{if } x>0,\ y>0 \\ 0 & \text{otherwise.} \end{cases}$$

a. (5pt) Compute $\mathsf{P}[X>3, Y>1] =$ ______________ (**Note** low point value)

b. (5pt) Are X and Y independent? Why?

c. (10pt) Compute the conditional expectation $\mathsf{E}[Y \mid X=2] =$ ______________

Probability and Statistics in Engineering: Second Test

4. James Bond insists that his martinis be shaken, not stirred. A skeptical bartender tests Bond with 6 martinis (using six fair coin flips to determine which drinks to shake and which to stir). Bond errs on one and correctly identifies the other 5 before passing out. Is this evidence sufficient to reject the null hypothesis that Bond can't tell the difference and is merely guessing (i.e. his probability of correct identification $p = 0.5$)?

To find out we test the hypothesis H_0 : Bond is merely guessing versus H_1 : Bond is not merely guessing, using the following decision rule: If he makes 5 or more correct guesses, reject the null hypothesis; else fail to reject. (This test has significance level $\alpha = 7/64 \approx 0.11$). Since Bond was correct five of the six times, we reject H_0.

a. (3pt) Can this possibly be a Type-I error? Why? (Explain in no more than 20 words).

b. (3pt) Can this possibly be a Type-II error? Why? (Explain in no more than 20 words).

c. (10pt) Suppose that Bond is *not* just guessing, and his probability of correct identification is actually $p = 0.8$. What is the *power* of the test? Give the answer to at least two correct decimals.

d. (4pt) Why would it be wrong to use a normal approximation to answer question c. above? (Explain in no more than 10 words).

Probability and Statistics in Engineering: Second Test

5. To see whether eyes with glaucoma have abnormally thick corneas, the corneas were measured in eight persons with glaucoma in one eye but not in the other. Here is the thickness in microns:

									$\bar{X}$	s
Healthy Eye	488	478	480	426	440	410	458	460	455	27.692
Glaucomatous Eye	484	478	492	444	436	398	464	476	459	31.314
Differences	4	0	-12	-18	4	12	-6	-16	-4	10.744

a. (15 pt) Test to see if eyes with glaucoma have typical cornea thickness or if they have abnormally thick corneas. Specify H_0 and H_1; α; the appropriate test statistic; and your final conclusion.

b. (5 pt) This data *could* be studied either as a paired-t test or as a two-sample t-test. Which would you recommend? Briefly, *Why?*

$$(n)_k = \frac{n!}{(n-k)!} = \overbrace{(n)(n-1)\cdots(n-k+1)}^{k \text{ terms}}$$

$$\text{“}n \text{ choose } k\text{”: } \binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{(n)(n-1)\cdots(n-k+1)}{(k)(k-1)\cdots(1)}$$

Distributions

Binomial	$\Pr[X=x] = \binom{n}{x} p^x q^{(n-x)},\ x = 0,1,\ldots,n$	$\mu = np,\ \sigma^2 = npq$
Poisson	$\Pr[X=k] = \lambda^k e^{-\lambda}/k!,\ k = 0,1,\ldots$	$\mu = \lambda,\ \sigma^2 = \lambda$
Geometric	$\Pr[Y=y] = pq^{y-1},\ y = 1,2,\ldots$	$\mu = 1/p,\ \sigma^2 = q/p^2$
Neg. Binomial	$\Pr[Y=y] = \binom{y-1}{r-1} p^r q^{y-r},\ y = r, r+1, \ldots$	$\mu = r/p,\ \sigma^2 = rq/p^2$
Hypergeometric	$\Pr[Y=y] = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}},\ y = 0,1,\ldots,n$	$\mu = \frac{nr}{N},\ \sigma^2 = \text{a mess}$
Exponential	$\Pr[X \le x] = 1 - e^{-x/\lambda},\ x \in (0,\infty)$ $f(x) = \frac{1}{\lambda} e^{-x/\lambda},\ x > 0;\ 0,\ x \le 0$	$\mu = \lambda,\ \sigma^2 = \lambda^2$
Uniform	$\Pr[X \le x] = (x-a)/(b-a),\ x \in (a,b)$ $f(x) = \frac{1}{b-a},\ a<x<b;\ 0,\ x \notin (a,b)$	$\mu = \frac{a+b}{2},\ \sigma^2 = \frac{(b-a)^2}{12}$

$\sum_{k=0}^n \binom{n}{k} a^k b^{n-k} = (a+b)^n$ $\qquad$ $\int_0^\infty t^n e^{-t}\,dt = n!$

$\sum_{k=0}^\infty ar^k = a/(1-r)$ $\qquad$ $\sum_{k=0}^n ar^k = (a - ar^{n+1})/(1-r)$

$\Pr[X \le x] = F(x) = \int_{-\infty}^x f(t)\,dt$ $\qquad$ $f(x) = F'(x)$

$\mathsf{E}[g(X)] = \int g(x) f(x)\,dx$ $\qquad$ $\mathsf{E}[g(X)] = \sum_x g(x) P(x)$

$\mu = \mathsf{E}[X] = \int_{-\infty}^\infty x f(x)\,dx$ $\qquad$ $\sigma^2 = \mathsf{E}[(X-\mu)^2] = \mathsf{E}[X^2] - \mu^2$

A **and** $B : A \cap B$ $\qquad$ **not** $A : A^c$ $\qquad$ A **or** $B : A \cup B$

$\mathsf{P}[A|B] = \mathsf{P}[A \cap B]/\mathsf{P}[B]$ $\qquad$ $\mathsf{P}[B|A] = \mathsf{P}[A \cap B]/\mathsf{P}[A]$

$\mathsf{P}[A \cap B] = \mathsf{P}[A|B]\,\mathsf{P}[B]$ $\qquad$ $\mathsf{P}[A \cup B] = \mathsf{P}[A] + \mathsf{P}[B] - \mathsf{P}[A \cap B]$

Independent $\Leftrightarrow \mathsf{P}[A \cap B] = \mathsf{P}[A]\,\mathsf{P}[B]$; $\qquad$ Exclusive $\Leftrightarrow \mathsf{P}[A \cap B] = 0$

Bayes: $\mathsf{P}[A_i|B] = \frac{\mathsf{P}[B|A_i]\mathsf{P}[A_i]}{\mathsf{P}[B|A_1]\mathsf{P}[A_1] + \ldots + \mathsf{P}[B|A_k]\mathsf{P}[A_k]}$ $\quad$ $A_i \cap A_j = \emptyset,\quad \mathsf{P}[A_1] + \ldots + \mathsf{P}[A_k] = 1$

		Continuous	**Discrete**
Mean	$\mu = \mathsf{E}[X] =$	$\int_{-\infty}^{\infty} x\, f(x) dx$	$= \sum xp(x)$
Variance	$\sigma^2 = \mathsf{V}[X] =$	$\int_{-\infty}^{\infty} (x-\mu)^2\, f(x) dx$	$= \sum (x-\mu)^2 p(x)$
		$= \mathsf{E}[X^2] - \mathsf{E}[X]^2$	$= \mathsf{E}[X^2] - \mathsf{E}[X]^2$
Expectation	$\mathsf{E}[g(X,Y)] =$	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f(x,y) dx\, dy$	$= \sum_x \sum_y g(x,y)\, p(x,y)$
Covariance	$\mathsf{Cov}[X,Y] =$	$\iint_{-\infty,-\infty}^{\infty,\infty} (x-\mu_x)(y-\mu_y)\, f(x,y) dx\, dy$	$= \sum_x \sum_y (x-\mu_x)(y-\mu_y)\, p(x,y)$
		$= \mathsf{E}[XY] - \mathsf{E}[X]\mathsf{E}[Y]$	$= \mathsf{E}[XY] - \mathsf{E}[X]\mathsf{E}[Y]$
Correlation	$\rho[X,Y] =$	$\mathsf{Cov}[X,Y]/\sigma_X \sigma_Y$	
Conditional	$f(x\|y) =$	$f(x,y)/f_2(y)$	$= p(x,y)/p_2(y)$
Marginals	$f_1(x) =$	$\int_{-\infty}^{\infty} f(x,y) dy$	$= \sum_y p(x,y)$
	$f_2(y) =$	$\int_{-\infty}^{\infty} f(x,y) dx$	$= \sum_x p(x,y)$
Independent		$\Leftrightarrow f(x,y) = f_1(x)\, f_2(y)$	$\Leftrightarrow p(x,y) = p_1(x)\, p_2(y)$
		$\Leftrightarrow f(x\|y) = f_1(x)$	$\Leftrightarrow p(x\|y) = p_1(x)$
		$\Leftrightarrow f(y\|x) = f_2(y)$	$\Leftrightarrow p(y\|x) = p_2(y)$

Conf. Intervals	$\mu = \bar{x} \pm z_{\alpha/2} S/\sqrt{n}$	Large Sample
	$= \bar{x} \pm t_{\alpha/2} S/\sqrt{n}$	Small Sample (normal)
	$p = \hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}\,\hat{q}\,/n}$	Population Proportion
Diff. of Means	$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	Large Samples
	$= (\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	Small Samples (normal, same σ^2)
	$= \overline{x_1 - x_2} \pm z_{\alpha/2} s_\Delta/\sqrt{n}$	Matched Pairs, large n
	$= \overline{x_1 - x_2} \pm t_{\alpha/2} s_\Delta/\sqrt{n}$	Matched Pairs, small n, normal
Sample Mean	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	
Sample Variance	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$	
Pooled Estimate	$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$	