Math 448. Introduction to Probability and Statistics II.
Spring 2000.

Schedule: M. W. F.: 08:00-09:30 a.m. SW313.

Textbook: Mathematical Statistics with Applications by D. Wackerly, W. Mendenhall lII and R. L. Scheaffer, Fifth edition. Duxbury Press.

Instructor: Miguel A. Arcones. *Office:* LN–2229. *Phone No.:* 777–2593. *E–mail:* arcones@math.binghamton.edu. *Web:* http://math.binghamton.edu/arcones.

Office hours: M: 10:30–11:30; W: 11:00–12:00; Fr: 9:40–10:30. Or by appointment. Feel free to come to my office at any time.

Grading: Homework (20%), two midterms (50%) and a final (30%). If you miss an exam, your score for that exam will be a zero.

Midterms exams: Wednesday, February 23; and Wednesday, April 5. (in the classroom).

Final exam: May 17, Wednesday, 11:00-01:00 pm S3246.

Course description: Chapters 7–13.

Homework: From time to time, I will give you a sheet of paper with problems to do at home. This is the "homework" which counts as 20 % of your grade. There will be one set of problems (between 5 to 10) for every chapter.

Course materials: You can find a copy of the materials in the courses (syllabus, notes, homeworks, tests and solutions) in the Reserve Room of the Library. You can find the syllabus, homeworks and tests in the web: http://math.binghamton.edu/arcones/448.html

pace7pt

Lective days:

	January		February				March			April				May	
M	24	31	7	14	21	28	6	13	27	3	10	17		1	8
W	26	2	9	16	23	1	8	15	29	5	12	19	26	3	10
F	28	4	11	18	25	3	10	17	31	7	14		28	5	12

Math 448. 1st Homework. Chapter 7. Due Monday, February 17, 2000.

1. A tobacco company claims that the amount of nicotine in its cigarettes is a random variable with mean 2.2 mg. and standard deviation .3 mg. however, the sample mean nicotine content of 100 randomly cigarettes was 3.1 mg. What is the approximate probability that the sample mean would have been as high or higher than 3.1 if the company's claims were true?

2. The temperature at which a thermostat foes off is normally distributed with variance σ^2. If the thermostat is to be tested five times, find $P(.85 \leq \frac{s^2}{\sigma^2} \leq 1.15)$.

3. An oil company claims that the sulfur content of its diesel fuel is at most .15 percent. To check this claim, the sulfur content of 40 randomly chosen samples were determined; the resulting sample mean and sample standard deviation were .162 and .040, respectively. What is the approximate probability that the sample mean would have been as high or higher than .162 if the company's claims were true? (Hint: use the t–distribution).

4. Students scores on exams given by a certain instructor have normal distribution with mean 70 and standard deviation 25. This instructor is about to give two exams, one for a class of size 15 and the other to a class of size 40. Approximate the probability that the average test score in the larger class exceeds that of the other class by over 4 points.

5. Twelve percent of the population is left handed. Find the probability that there are between 10 and 14 left–handers in a random sample of 100 members of this population. That is, find $P(10 \leq X \leq 14)$, where X is the number of left–handers in the sample.

Math 448. 2nd Homework. Chapter 8. Due Wednesday, February 16, 2000.

1. Let $Y_1, \dots, Y_n$ denote a random sample of size n from a uniform distribution on $(0, \theta)$, where $\theta > 0$. Find a constant a such that $a\bar{Y}$ is an unbiased estimator of θ. Find a constant b such that $bY_{(n)} = b \max(Y_1, \dots, Y_n)$ is an unbiased estimator of θ. Find the mean square error of $a\bar{Y}$ and of $bY_{(n)}$ (for the a and b found before). Which estimator is prefered?

2. A market research is interested in determining the proportion of households that are watching a particular sporting event. To accomplish this task, they plan on using a telephone poll of randomly chosen households. How large a sample is needed if they want to be 90 % percent certain that their estimate is correct to within $\pm .02$?

3. In a random sample of 300 persons eating lunch at a department store cafeteria, only 102 had dessert. If we use the sample proportion $\frac{102}{300} = 0.45$ as an estimate of the corresponding true proportion, with what confidence can we assert that out error is less than 0.05?

4. A study has been made to compare the nicotine contents of two brands of cigarettes. Ten cigarettes of Brand A had an average nicotine content of 3.1 milligrams with a standard deviation of 0.5 milligram, while eight cigarettes of Brand B had an average nicotine content of 2.7 milligram with a standard deviation of 0.7 milligrams. Assuming that the two sets of data are independent random samples from normal populations with equal variances, construct a 95 % confidence interval of the difference between the mean nicotine contents of the two brands of cigarettes. (Use the t distribution).

5. Let X and let Y equal the concentration in parts per billion of chromium in the blood for healthy persons and for persons with a suspected disease, respectively. Assume that the distribution of X and Y are $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$ respectively. Using $n = 8$ observations of X

$$15 \; 23 \; 12 \; 18 \; 9 \; 28 \; 22 \; 11 \; 10$$

and $m = 10$ observations of Y

$$25 \; 20 \; 35 \; 15 \; 40 \; 16 \; 10 \; 22 \; 18 \; 32$$

(a) Give a point estimate of σ_X^2/σ_Y^2. (b) Find a 95 % confidence interval for σ_X^2/σ_Y^2.

Math 448. 3rd Homework. Chapter 9. Due Wednesday, March 1, 2000.

1. Let $Y_1, \ldots, Y_n$ denote a random sample from the density

$$f(y|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < y < \theta, \\ 0 & \text{else.} \end{cases}$$

where $\theta > 0$. Show that $\hat{\theta}_1 = \frac{n+1}{n} Y_{(n)}$ and $\hat{\theta}_2 = 2\bar{Y}$ are unbiased estimators of θ. Find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$. Which estimator is preferred?

2. Given a random sample of size n from the density

$$f(y|\theta) = \begin{cases} \frac{m y^{m-1} e^{-\frac{y^m}{\theta}}}{\theta} & \text{if } 0 < y, \\ 0 & \text{else.} \end{cases}$$

where $\theta > 0$ is unknown parameter and $m > 0$ is known. Find a sufficient statistic of θ. Find the MVUE of θ.

3. Given a random sample of size n from the density

$$f(y|\theta) = \begin{cases} \frac{\alpha y^{\alpha-1}}{\theta^\alpha} & \text{if } 0 < y < \theta, \\ 0 & \text{else.} \end{cases}$$

where $\theta > 0$ is unknown parameter and $\alpha > 0$ is known. Find a sufficient statistic of θ. Find the MVUE $\hat{\theta}_1$ of θ.

4. Let $Y_1, \ldots, Y_n$ denote a random sample from the density

$$f(y|\theta) = \begin{cases} \theta y^{\theta-1} & \text{if } 0 < y < 1, \\ 0 & \text{else.} \end{cases}$$

where $\theta > 0$. Find the method of moments estimator of θ. Show that this estimator is consistent.

5. Given a random sample of size n from the density Given a random sample of size n from the density

$$f(y|\theta) = \begin{cases} \frac{m y^{m-1} e^{-\frac{y^m}{\theta}}}{\theta} & \text{if } 0 < y, \\ 0 & \text{else.} \end{cases}$$

where $\theta > 0$ is unknown parameter and $m > 0$ is known. Find the method of moments estimator of θ. Show that this estimator is consistent.

Math 448. 4th Homework. Chapter 9. Due Wednesday, March 15, 2000.

1. Let $Y_1, \ldots, Y_n$ be a random sample from

$$f(y|\theta) = \begin{cases} \theta y^{\theta-1}, & \text{if } 0 < y < 1, \\ 0 & \text{else.} \end{cases}$$

Find the maximum likelihood estimator of θ.

2. Let $Y_1, \ldots, Y_n$ be a random sample from the density $f(y|\theta) = \frac{(\ln\theta)^y}{\theta\ y!}$, $y = 0, 1, 2, \ldots$, $\theta > 1$. Find the maximum likelihood estimator of θ.

3. Given a random sample of size n from the density

$$f(y|\theta) = \begin{cases} e^{-(y-\theta)}, & \text{if } y \geq \theta, \\ 0 & \text{else.} \end{cases}$$

where $\theta \in I\!R$ is an unknown parameter. Find a sufficient statistic of θ. Find the maximum likelihood estimator of θ.

4. Given a random sample of size n from the density

$$f(y|\theta) = \begin{cases} \frac{5y^4}{\theta^5}, & \text{if } 0 \leq y \leq \theta, \\ 0 & \text{else.} \end{cases}$$

where $\theta > 0$ is an unknown parameter. Find the maximum likelihood estimator of θ.

5. Given a random sample of size n from the density

$$f(y|\theta) = \begin{cases} \frac{my^{m-1}e^{-\frac{y^m}{\theta}}}{\theta} & \text{if } 0 < y, \\ 0 & \text{else.} \end{cases}$$

where $\theta > 0$ is unknown parameter and $m > 0$ is known. Find the method of moments estimator of θ. Show that the maximum likelihood estimator of θ is $\hat{\theta} = \frac{\sum_{i=1}^n Y_i^m}{n}$. Show that an approximate large sample $(1-\alpha)\%$ confidence interval for θ is $\hat{\theta} \pm z_{\alpha/2}\frac{\hat{\theta}}{\sqrt{n}}$.

Math 448. 5th Homework. Chapter 10. Due Monday, April 3, 2000.

1. It is claimed than 75 % of all dentists recommend a certain brand of gum for their gum chewing patients. A consumer group doubted this claim believing that the proportion is lower and decided made a sample survey. A survey of 390 dentists found that 273 recommended this brand of gum. What are the null and alternative hypothesis? Find the p value of the test. Which hypothesis would you accept if the significance level is (a) $\alpha = 0.10$? (b) $\alpha = 0.05$? (c) $\alpha = 0.01$?

2. Assume that IQ scores for a certain population are approximately normally distributed with mean μ and variance 100. To test $H_0 : \mu = 110$ against the one–sided alternative hypothesis $H_a : \mu > 110$, we take a random sample of size $n = 16$ from this population and observe $\bar{y} = 113.5$. Find the p value of the test. Which hypothesis would you accept if the significance level is (a) $\alpha = 0.10$? (b) $\alpha = 0.05$? (c) $\alpha = 0.01$? Find the rejection region for the test at the significance level $\alpha = 0.05$. What is the type II error if $\mu = 120$?

3. Let us assume that the life of a tire in miles, say Y, is normally with distributed with mean μ and standard deviation 5000. Past experience indicates that $\mu = 30,000$. The manufacturer claims that the tires made by a new process have mean $\mu > 30,000$, and it is very possible that $\mu = 35,000$. We shall observe n independent values of Y, say $y_1, \ldots, y_n$, and we shall reject H_0 if $\bar{y} \geq c$. Determine n and c so that the type I error is 0.01 and the type II error when $\mu = 35,000$ is 0.02.

4. A professor claims that the average starting salary of industrial engineering graduating seniors is greater than that of civil engineering graduates. To study this claim, samples of 16 industrial engineers and 16 civil engineers, all of whom graduate in 1993, were chosen and sample members were queried about their starting salaries. If the industrial engineers had a sample mean salary of \$47,700 and sample standard deviation of \$2,400 and the civil engineers had a sample mean salary of \$46,400 and sample standard deviation of \$2,200 has the professor's claim been verified? What are the null and alternative hypothesis? Find the appropriate p-value. Which hypothesis would you accept if the significance level is (a) $\alpha = 0.10$? (b) $\alpha = 0.05$? (c) $\alpha = 0.01$?

5. A study of chromosome abnormalities and criminality examined date on 4124 Danish males born in Copenhagen. Each man was classified as having a criminal record or not, using the penal registers maintained in the offices of the local police chiefs. Each was also classified as having the normal male XY chromosome pair or one of the abnormalities XYY or XXY. Of the 4096 men with normal chromosomes, 381 had criminal records, while 8 of the 28 men with chromosome abnormalities had criminal records. Some experts believe that chromosome abnormalities are associated with increased I criminality. Do these data lend support to this belief? What are the null and alternative hypothesis? Find the appropriate p-value. Which hypothesis would you accept if the significance level is (a) $\alpha = 0.10$? (b) $\alpha = 0.05$? (c) $\alpha = 0.01$?

Math 448. 6th Homework. Chapter 10. Due Wednesday, April 12, 2000.

1. A car company claims that its new experimental engine runs 29 minutes with one gallon of fuel. Test runs with an experimental engine it operated, respectively, for 24, 28, 21,23, 32, and 22 minutes with one gallon of fuel. Is there enough evidence at the .01 level to claim that the new experimental engine runs less than 29 minutes? What are the null and alternative hypothesis? Determine the rejection region. Estimate the p value of the test.

2. Let Y equal the number of pounds of butterfat produced by a Holstein cow during the 305–day milking period following the birth of a calf. We assume that the distribution of Y is normal with mean μ and variance σ^2. The following data was obtained:

425	710	661	664	732	714	934	761	744
653	725	657	421	573	535	602	537	
405	874	791	721	849	567	468	975	

Test the null hypothesis $H_0 : \sigma^2 = 140^2$ against the alternative $H_a : \sigma^2 > 140^2$ at the level $\alpha = 0.05$. Determine the rejection region. Do we accept or reject the null hypothesis at the level $\alpha = 0.05$. Find the p–value of the data.

3. Let Y_1 and let Y_2 denote the weights in grams of male and female gallinules, respectively. Assume that both distributions are normally distributed. Given that $n_1 = 16$, $\bar{Y}_1 = 415.16$, $s_1^2 = 1356.75$, $n_2 = 13$, $\bar{Y}_2 = 347.4$, $s_2^2 = 629.21$. (a) We test the null hypothesis $H_0 : \mu_1 = \mu_2$ against the alternative $H_a : \mu_1 > \mu_2$ at the level $\alpha = 0.01$. Determine the rejection region. Do we accept or reject the null hypothesis a the level $\alpha = 0.01$. Find the p–value of the data. (b) We test the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against the alternative $H_0 : \sigma_1^2 \neq \sigma_2^2$ at the level $\alpha = 0.05$. Determine the rejection region. Do we accept or reject the null hypothesis a the level $\alpha = 0.05$? Estimate the p–value for this test.

4. Let $X_1, \dots, X_n$ be random sample of size n from the density

$$f(y|\theta) = \begin{cases} \frac{10y^9 e^{-y^{10}/\theta}}{\theta} & \text{if } 0 < y, \\ 0 & \text{else.} \end{cases}$$

where $\theta > 0$ is unknown. Find the uniformly most powerful test for $H_0 : \theta = \theta_0$, versus $H_a : \theta < \theta_0$ (we just to write the rejection region in the form $\{(y_1, \dots, y_n) : T(y_1, \dots, y_n) \geq k'\}$, where $T(y_1, \dots, y_n)$ is a statistic which you need to find).

5. Let $X_1, \dots, X_n$ be a random sample from a Poisson distribution with parameter $\theta > 0$. Find the likelihood ratio test of $H_0 : \theta = \theta_0$ against $H_a : \theta \neq \theta_0$, where $\theta_0 > 0$ (we just to write the rejection region in the form $\{(y_1, \dots, y_n) : T(y_1, \dots, y_n) \geq k'\}$, where $T(y_1, \dots, y_n)$ is a statistic which you need to find).

Math 448. 7th Homework. Chapter 11. Due Monday, May 1, 2000.

1. **Robbery rate.** A criminologist studying the relationship between population and robbery rate in medium–sized U.S. cities collected the following data for a random sample of 16 cities; x is the population density of the city (number of people per area), and Y is the robbery rate last year (number of robberies per 100,000 people).

x_i	59	49	75	54	78	56	60	82
Y_i	209	180	195	192	215	197	208	189

x_i	69	83	88	94	47	65	89	70
Y_i	213	201	214	212	205	186	200	204

Fit a straight line to the data. Give estimates of β_0 and β_1. Do the data present sufficient evidence to indicate that the slope is positive? Test at at the 5 % significance level whether $\beta_1 > 0$. Estimate the p–value of the test.

Hint: You may use that $\sum x_i = 1118.00$, $\sum x_i^2 = 81452.0$, $\sum y_i = 3220.00$, $\sum y_i^2 = 649736$, $\sum x_i y_i = 225869$, $\bar{x} = 69.8750$, $\bar{y} = 201.250$ $S_{xx} = 3331.75$, $S_{xy} = 871.500$ and $S_{yy} = 1711.00$.

2. **Grade point average.** The director of admissions of a small college administered newly designed entrance test to 20 students selected at random from the new freshman class in a study to determine whether a students' grade point average at the of the freshman year (Y) can be predicted from the entrance test score (X). The results for the study follow.

x_i	5.5	4.8	4.7	3.9	4.5	6.2	6.0	5.2	4.7	4.3
Y_i	3.1	2.3	3.0	1.9	2.5	3.7	3.4	2.6	2.8	1.6

x_i	4.9	5.4	5.0	6.3	4.6	4.3	5.0	5.9	4.1	4.7
Y_i	2.0	2.9	2.3	3.2	1.8	1.4	2.0	3.8	2.2	1.5

Give estimates of β_0 and β_1. Find a 95 % confidence interal for the mean GPA of a freshpersons whose entrance test score is 4.7. Mr. Smith (a new freshman) obtained a score of 4.7 on the entrance test. Find 95 % prediction interal for the GPA of Mr. Smith.

Hint: You may use that $\sum x_i = 100$, $\sum x_i^2 = 509.12$, $\sum y_i = 50$, $\sum y_i^2 = 134.84$, $\sum x_i y_i = 257.66$, $\bar{x} = 5$, $\bar{y} = 2.5$ $S_{xx} = 9.1199$, $S_{xy} = 7.65997$ and $S_{yy} = 9.84$.

3. The following date set presents the heights of 12 male law school classmates whose law school examination scores were roughly equal. It also gives their annual salaries 5 years after graduation. Each of them went in corporate law. The height is in inches and the salary is in units of \$1,000

height (x)	64	65	66	67	69	70	72	72	74	74	75	76
salary (y)	91	94	88	103	77	96	105	88	122	102	90	104

Fit a straight line to the data (give estimates of β_0 and β_1). Do the above data establish the hypothesis that a lawyer's salary is related to this height.? Use the 5 percent level of significance. What are the null and alternative hypothesis? Estimate the p–value of the test.

Hint: You may use that $\sum x_i = 844$, $\sum x_i^2 = 59548$, $\sum y_i = 1170$, $\sum y_i^2 = 115768$, $\sum x_i y_i = 82562$, $\bar{x} = 70.3333$, $\bar{y} = 97.50$ $S_{xx} = 186.668$, $S_{xy} = 272.$ and $S_{yy} = 1693$.

4. Do Problem 11.6.

5. Experience with a certain type of plastic indicates that a relation exists between the hardness ((Y) measured in Brinell units) of times molded from the plastic and the elapsed time x since termination of the moding processes. Sixteen batches of the plastic were made, and from each batch one test item was moded. Each test itme was randomly assigned to one of the four predetermined time levels, and the hardness was measured after the assigned elapsed time. X is the elapsed time in hours and Y is the hardness in Brinell units.

time (x)	16	16	16	16	24	24	24	24
hardness (y)	199	205	196	200	218	220	215	223

time (x)	32	32	32	32	40	40	40	40
hardness (y)	237	234	235	230	250	248	253	246

Obtain the regression line. Obtain a point estimate of the change in mean hardness when x increases by one hour. Find a 99 % confidence interval for the change in the mean hardness when x increases by one hour. The plastic manufacturer has stated that the mean hardness should increase by 2 Brinell units per hour. Conduct a two–sided test to decide whether this hypothesis is satisfied; use $\alpha = 0.01$. State the null and alternative hypothesis and conclussion. Estimate the p-value of the test.

Hint: You may use that $\sum x_i = 448$, $\sum x_i^2 = 13824$, $\sum y_i = 3609$, $\sum y_i^2 = 819499$, $\sum x_i y_i = 103656$, $\bar{x} = 28$, $\bar{y} = 225.563$ $S_{xx} = 1280$, $S_{xy} = 2604.$ and $S_{yy} = 5443.94$.

Math 448. 8th Homework. Chapter 12. Due Wednesday, May 10, 2000.

1. A rehabilitation center research was interested in examining the relationship between physical fitness prior to surgery of persons undergoing corrective knee surgery and time require in physical therapy until successful rehabilitation patient records in the rehabilitation enter were examined, and 24 male subjects ranging in age from 18 to 30 years who had undergone similar corrective knee surgery during the past year were selected for the study. The number of days required for successful completion of physical therapy and the prior physical fitness status (below average, average, and above average) for each patient follow.

Below average	29	42	38	40	43	40	30	42		
Average	30	35	39	28	31	31	29	35	29	33
Above average	26	32	21	20	23	22				

(a) Construct the anova table for this data. (b) Test at the level 0.01 whether or not the mean number of days required for successful rehabilitation is the same of the three fitness groups. (c) Find the p–value for the previous test. (d) What appear to be the nature of the relationship between fitness status and duration of required physical therapy.

2. Five serving each of three different brands of processed meat were tested for fat content. The following data (in fat percentage per gram) resulted

Brand 1	32	34	31	35	33
Brand 2	41	32	33	29	35
Brand 3	36	37	30	28	33

(a) Construct the anova table for this data. (b) Test at the level 0.05 whether the fat content differ depends on the brand. Find the p–value for the previous test. (c) Find 95 % confidence intervals for all quantities $\mu_i - \mu_j$.

3. Show that for the oneway anova model

$$\text{Total SS} = \text{SSE} + \text{SST}.$$

4. The following data refer to the number of deaths per 10,000 adults in a large city in the different seasons for the years 1982 and 1986.

Year	Winter	Spring	Summer	Fall
1982	33.6	31.4	29.8	32.1
1983	32.5	30.1	28.5	29.9
1984	35.3	33.2	29.5	28.7
1985	34.4	28.6	33.9	30.1
1986	37.3	34.1	28.5	29.4

(a) Construct the anova table for this data. (b) Test at the 5 % level of significance the hypothesis that death rates do no depend on the season. Find the p–value for the previous test. (c) Test at the 5 % level of significance the hypothesis that there is no effect due to the year. Find the p–value for the previous test. (d) Find 95 % confidence intervals of the number of deaths in each season. (e) Find 95 % confidence interval of the difference of number of deaths between the winter and summer season.

5. Three different washing machines were employed to test four different detergents. The following data dive a coded score of the effecdtiveness of each washing

	Machine 1	Machine 2	Machine 3
Detergent 1	53	50	59
Detergent 2	54	54	60
Detergent 3	56	58	62
Detergent 4	50	45	57

(a) Construct the anova table for this data. (b) Test at the 5 % level of significance the hypothesis that the detergent used does not affect the score. Find the p–value for the previous test. (c) Test at the 5 % level of significance the hypothesis that the machine used does not affect the score. Find the p–value for the previous test.

Math 448. 9th Homework. Chapters 14 and 15. Due Wednesday, May 17, 2000 (final exam day).

1. In a large bin of crocus bulbs it is claimed that 1/4 will produce yellow crocuses, 1/4 will produce white crocuses, and 1/2 will produce purple crocuses. If 40 bulbs produce 6 yellow, 7 white, and 27 purple crocuses, would the claim be rejected at the $\alpha = .05$ significance level? Estimate the p–value of the test.

2. A random sample of $n = 1362$ persons were classified according to the respondent's education level and whether the respondent was Protestant, Catholic or Jewish. Use these data to test at an $\alpha = 0.05$ significant level the hypothesis that these attributes or classification are independent. Estimate the p–value of the test.

Education level	Protestant	Catholic	Jewish
Less than high school	359	140	5
High school or Junior College	462	200	17
Bachelor′s degree	88	39	2
Graduate degree	37	10	3

3. The administration claims that professor A is more popular than professor B. We shall test the hypothesis H_0 of no difference in popularity against the administration's claim H_a at an approximate 0.05 significance level. If 14 out of 20 students prefer professor A to professor B, is H_0 rejected and the administration's claim supported? Estimate the p–value of the test.

4. Twelve pairs of twin male lambs were selected; diet plan I was given to one twin and diet plan to the other twin in each case. The weights at eight months were as follows

Diet I	111	102	90	110	108	125	99	121	133	115	90	101
Diet II	97	90	96	95	110	107	85	104	119	98	97	104

(a) Use the sign test to test the hypothesis that there is not difference in the diets against the alternative that diet I is preferable to diet I at $\alpha = 0.10$. (b) Repeat (a) using the Wilcoxon paired–sample signed–rank test.

5. A pharmaceutical company is interested in testing the effect of humidity on the weigth of pills that are sold in aluminum packaging. Let X and Y denote the respective weights of pills and their packaging when the packaging is good and when it is defective after the pill has spent one week in a chamber containing 100 % humidity and heated to 90^{o}F. Let $p = P(X > Y)$. Use the Mann–Whitney U–test to test $H_o : p = 1/2$ against $H_a : p < 1/2$ at the

X	.7565	.7720	.7776	.7750	.7494	.7615	.7741	.7701	.7712	.7719	.7546	.7719
Y	.7870	.7750	.7720	.7876	.7795	.7972	.7815	.7811	.7731	.7613	.7816	.7815

Math 448. First Midterm. Friday, February 25, 2000.

Name: .... **Soc. Sec. No.** ....

Show all your work. No credit for lucky answers.

1. Let $Y_1, \ldots, Y_{10}$ be independent identically distributed random variables with a normal distribution with mean 3 and variance 40. Find b so that

$$P\{\bar{Y} \leq b\} = 0.95 .$$

2. Let $Y_1, \ldots, Y_{10}$ be independent identically distributed random variables with a normal distribution with mean zero and variance 4. Find b so that

$$\Pr\{\sum_{i=1}^{10} Y_i^2 \leq b\} = 0.05.$$

3. Suppose that a random sample of nine recently sold houses in a certain city resulted in a sample mean price of \$122,000, with a sample standard deviation of \$12,000. Give a 95 percent confidence interval for the mean price of all recently sold houses in this city.

4. If 132 of 200 male voters and 90 of 159 female voters favor certain candidate running for governor of Illinois. Find a 99 % confidence interval for the difference between the actual proportion of male and female who favor the candidate.

5. The capacities in ampere–hours of 10 batteries were recorded as follows:

140 136 150 144 148 152 138 141 143 151

(a) Estimate the population variance σ^2. (b) Compute a 99 percent confidence interval of σ^2.

6. Let $Y_1, \ldots, Y_n$ be a random sample from a normal distribution with mean zero and variance $\theta > 0$. (a) Find a sufficient statistic of θ. (b) Find the MVUE for θ.

7. Let $Y_1, \ldots, Y_n$ be a random sample from the density

$$f(y|\theta) = \begin{cases} \frac{2\theta^2}{y^3} & \text{if } y > \theta, \\ 0 & \text{else,} \end{cases}$$

where $\theta > 0$. (a) Find a sufficient statistic for θ. (b) Find the minimum variance unbiased estimator of θ. (c) Find the method of the moments estimator of θ.

8. Given a random sample $Y_1, \ldots, Y_n$ of size n from the density

$$f(y|\theta) = \begin{cases} \theta 4y^3 e^{-\theta y^4} & \text{if } y \geq 0, \\ 0 & \text{if } y < 0, \end{cases}$$

where $\theta > 0$ is unknown. Find the method of the moments estimator of θ.

Math 448. Second Midterm. Friday, April 5, 2000.

Name: .... **Soc. Sec. No.** ....

Show all your work. No credit for lucky answers.

1. Given a random sample of size n from the density

$$f(y|\theta) = \begin{cases} \theta y^{\theta-1} & \text{if } 0 < y, \\ 0 & \text{else.} \end{cases}$$

where $\theta > 0$ is unknown parameter. Find the maximum likelihood estimator of θ

2. A company chip manufacturer claims that no more than 2 percent of the is sends out are defective. An electronics company, impressed by this claim, has purchased a large quantity of such ships. To determine if the manufacturer's claim can be taken literally, the company has decided to test a sample of 300 of these chips. If 10 of these 300 chips are found to be defective, should the manufacturer's claim rejected against the alternative hypothesis that the proportion of defective chips is higher than .02? What are the null and alternative hypothesis? Estimate the p–value. Would you reject the null hypothesis if the significance level is (a) $\alpha = 0.10$? (b) $\alpha = 0.05$? (c) $\alpha = 0.01$?

3. Do urban and rural households that display Christmas trees have the same preference for natural versus artificial trees? Let p_1 be the proportion of natural trees in urban households displaying a Christmas tree. Let p_2 be the proportion of natural trees in rural households displaying a Christmas tree. Test the hypothesis that rural household are more likely to choose natural Christmas trees. the survey responses show that 89 of the 261 urban households and 64 of the 160 rural households that displayed a tree chose a natural tree. What are the null and alternative hypothesis? Estimate the p–value. Would you reject the null hypothesis if the significance level is (a) $\alpha = 0.10$? (b) $\alpha = 0.05$? (c) $\alpha = 0.01$?

4. The breaking strengths of cables produced by a manufacturer have mean 1800 lb. It is claimed that a new manufacturing process will increase the mean breaking strength of the cables. To test this hypothesis, a sample of 30 cables manufactured using the new process is tested, giving $\bar{Y} = 1850$ and $s = 100$. What are the null and alternative hypothesis? Estimate the p–value. Would you reject the null hypothesis if the significance level is (a) $\alpha = 0.10$? (b) $\alpha = 0.05$? (c) $\alpha = 0.01$?

5. To find out whether the inhabitants of two South Pacific islands may be regarded as having the same racial ancestry, an anthropologist determines the cephalic indices of six adult males from each island, getting $\bar{Y}_1 = 77.4$, $s_1 = 3.3$, $\bar{Y}_2 = 72.2$ and $s_2 = 2.1$. Assume the populations sampled are normal and have equal variances. Test whether the difference between the two sample means can reasonable be attributed to chance. What are the null and alternative hypothesis? Estimate the p–value. Would you reject the null hypothesis if the significance level is (a) $\alpha = 0.10$? (b) $\alpha = 0.05$? (c) $\alpha = 0.01$?

6. A gun–like apparatus has recently been designed to replace needles in administering vaccines. The apparatus can be set to inject different amounts of the serum, but because of random fluctuations the actual amount injected is normally distributed with a mean equal to the setting and with an unknown variance σ^2. It has been decided that the apparatus

would be too dangerous to use if σ exceeds .10. If a random sample of 50 injections resulted in a sample standard deviation of .12, should use the new apparatus be discontinued? What are the null and alternative hypothesis? Estimate the p–value. Would you reject the null hypothesis if the significance level is (a) $\alpha = 0.10$? (b) $\alpha = 0.05$? (c) $\alpha = 0.01$?

7. Let $X_1, \ldots, X_n$ be random sample of size n from a normal density with mean zero and variance σ^2. Find the uniformly most powerful test for $H_0 : \sigma^2 = \sigma_0^2$, versus $H_a : \sigma^2 < \sigma_0^2$, where $\sigma_0^2 > 0$.

8. Let $X_1, \ldots, X_n$ be a random sample from the density

$$f(y|\theta) = \begin{cases} \frac{4y^3 e^{-\frac{y^4}{\theta}}}{\theta} & \text{if } 0 < y, \\ 0 & \text{else.} \end{cases}$$

where $\theta > 0$ is unknown parameter. Find the maximum likelihood estimator of θ. Find the rejection region for the likelihood ratio test of $H_0 : \theta = \theta_0$ against $H_a : \theta \neq \theta_0$, where $\theta_0 > 0$.

Math 448. Final. Wednesday, May 17, 2000.

Name: .... **Soc. Sec. No.** ....

Show all your work. No credit for lucky answers.

1. An oil company claims that less than 20 % of all car owners have not tried its gasoline. Test this claim at the 0.01 level of significance if a random check reveals that 22 of 200 car owners have not tried the oil company's gasoline.

2. Let us compare the failure times of a certain type of light bulb produced by two different manufactures, X and Y, by testing 10 bulbs selected at random from each of the outputs. The data, in hundred of hours used before failure, are

X	5.6	4.6	6.8	4.9	6.1	5.3	4.5	5.8	5.4	4.7
Y	7.2	8.1	5.1	7.3	6.9	7.8	5.9	6.7	6.5	7.1

Do the data present sufficient evidence to indicate a different in the location of the distributions of failure times of the two types of bulbs. Test with $\alpha = 0.05$. Estimate the p–value of the test.

3. In a poll, 1000 individuals are classified by gender and by whether they favor or oppose on a complete ban on smoking in public places. Test the null hypothesis that gender and opinion on smoking public places are independent. Estimate the p–value of the test.

Gender	Favor	Oppose	Total
Male	262	231	493
Female	302	205	507
Total	564	436	1000

4. A purification process for a chemical involves passing it, in solution, through a resin on which impurities are adsorbed. A chemical engineer wishing to test the efficiency of 3 different resins took a chemical solution and broke it into 15 batches. She tested each resin 5 times and then measured the concentration of impurities after passing through the resins. Her data were as follows.

Resin I	46	25	14	17	43
Resin II	38	35	31	22	12
Resin III	31	42	20	18	39

Construct the anova table for this data. Test the hypothesis that there is no difference in the efficiency of the resins. Estimate the p–value of the test. Hint SST=15, SSE=1789, Total SS=1804, $\bar{Y}_{1,\cdot} = 29.00$, $\bar{Y}_{2,\cdot} = 29.00$, $\bar{Y}_{3,\cdot} = 29.00$ and $\bar{Y} = 28.8667$.

5. Twelve patients having high albumin content in their blood were treated with a medicine. Their blood content of albumin was measured before and after treatment. The measured values are shown in the table.

Patient	1	2	3	4	5	6	7	8	9	10	11	12
Before treatment	5.02	5.08	4.75	5.25	4.80	5.77	4.85	5.09	6.05	4.77	4.85	5.24
After treatment	4.66	5.15	4.30	5.07	5.38	5.10	4.80	4.91	5.22	4.50	4.85	4.56

Is the effect of the medicine significant at the 5 % level? Estimate the p–value of the test?

6. A person's muscle mass is expected to decrease with age. To explore this relationship in women, a nutritionist randomly selected four women from each 10 year age group, beginning with age 40 and ending with age 79. The results follow X is age, and Y is a measure of muscle mass. Assume the simple linear repression model. Obtain the estimate regression function. Test at the 5 % level whether muscle mass decreases with age. Estimate the p–value of the test.

Age	71	64	43	67	56	73	68	56	76	65	45	58	45	53	49	78
Muscle	82	91	100	68	87	73	78	80	65	84	116	76	97	100	105	77

Is the effect of the medicine significant at the 5 % level? Estimate the p–value of the test. Hint: You may use that $\sum_{i=1}^n x_i = 967.$, $\sum_{i=1}^n x_i^2 = 60409.0$, $\sum_{i=1}^n y_i = 1379.00$, $\sum_{i=1}^n y_i^2 = 121887$, $\sum_{i=1}^n x_i y_i = 81331.0$, $\bar{x} = 60.4375$, $\bar{y} = 86.1875$ $S_{xx} = 1965.94$, $S_{xy} = -2012.31$ and $S_{yy} = 3034.44$.

7. A National Science Foundation survey of 1985 salaries of men and women scientist in the United States reported summary of statistics for scientist in different fields. Some of the results appears in the following table, which gives mean yearly salaries in thousand of dollars. Test at the 5 % level whether the mean salary of men and women is the same. Estimate the p–value of the test.

Field	Women	Men	Mean
Physics	41.2	48.6	44.900
Math	34.7	42.3	38.500
Biology	34.5	42.0	38.250
Mean	36.8	44.3	40.550

8. A market research firm supplies manufacturers with estimates of the retails sales of their products form samples of retail stores. Marketing managers are prone to look at the estimate and ignore sampling error. Suppose that a random sample of 75 stores this month shows mean sales of 52 units of a small appliance, with a standard deviation of 13 units. During the same month last year, a random sample of 53 stores gave mean sales of 49 units, with a standard deviation of 11 units. An increase from 49 to 52 is a rise of 6 %. The marketing is happy because sales are up 6 %. Give a 95 % confidence interval for the difference in mean number of units sold at all retail stores. Test at the 5 % level whether the mean number of units sold in a month is the same in this year and in last year. Estimate the p–value of the test. Should the manager be happy for the data?

9. Let $Y_1, \ldots, Y_n$ be a random sample from the density

$$f(y|\theta) = \begin{cases} \frac{3y^5 e^{-\frac{y^3}{\theta}}}{\theta^2} & \text{if } y > 0, \\ 0 & \text{else,} \end{cases}$$

where $\theta > 0$. (a) Find a sufficient statistic for θ. (b) Find the minimum variance unbiased estimator of θ.

10. Let $Y_1, \ldots, Y_n$ be a random sample from the density

$$f(y|\theta) = \begin{cases} \frac{4y^3}{\theta^4} & \text{if } 0 \leq y \leq \theta, \\ 0 & \text{else,} \end{cases}$$

where $\theta > 0$ is unknown. (a) Find the method of the moments estimator of θ. (b) Find the maximum likelihood estimator of θ.