# Math 448–01.
# Introduction to Probability and Statistics II. Spring 1998.

*Schedule:* M. W. F.: 08:00-09:30 am. LN–2298.

*Textbook:* Introduction to Mathematical Statistics by R. V. Hogg and A. T. Craig, 1995, Fifth edition. Prentice Hall.

*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: 9:35–10:30; W: 13:10–14:10; Fr: 9:35–10:30. 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:* Friday, February 19; Wednesday, March 31 (in the classroom).

*Final exam:* May 11, Tue, 8:30–10:30 am. in AG–0007.

*Course description:* Chapters 6–11 in the textbook. We will skip the following sections: 8.4, 9.4, 10.3, 10.4, 10.8, 10.9, 11.7, 11.8 and 11.9. The main topics are: point estimation, maximum likelihood estimators, confidence intervals, tests of hypothesis, chi–square tests, sufficient statistics, complete statistics, exponential class, ancillary statistics, Rao–Cramér inequality, Bayesian estimation, uniform most powerful test, likelihood ratio test, analysis of variance, linear regression and nonparametric statistics.

*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. The following problems from the book are assigned as homework (which will no be collected):

<u>Chapter 6.</u> 1, 2, 5, 6, 7, 8, 9, 10, 14, 15, 16, 17, 18, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 35, 37, 38, 39, 40, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 60, 61, 62, 63, 64, 65, 66, 67, 69, 70, 71, 72, 73, 76, 77, 78.

<u>Chapter 7.</u> 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 38, 39, 40, 41, 42, 43, 44, 45, 47, 48, 49, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 65, 66, 68, 69, 72, 73, 75, 76, 77.

<u>Chapter 8.</u> 1, 2, 3, 4, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 26, 27, 28, 29, 31.

<u>Chapter 9.</u> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 41, 42, 47, 49, 50, 51.

<u>Chapter 10.</u> 1, 2, 3, 4, 5, 6, 7, 8, 9, 19, 20, 22, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 53, 54, 55, 56, 57, 59, 63.

<u>Chapter 11.</u> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 16, 18, 19, 26, 29, 30, 33, 34, 35, 50, 51, 53, 57.

*Lective days:*

| | January | February | | | | | March | | | | April | | | May |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| M | 25 | 1 | 8 | 15 | 22 | 1 | 8 | 22 | 29 | | 12 | 19 | 29 | 3 |
| W | 27 | 3 | 10 | 17 | 24 | 3 | 10 | 24 | 31 | 7 | 14 | 21 | 31 | 5 |
| F | 29 | 5 | 12 | 19 | 26 | 5 | 12 | 26 | | 9 | 16 | 23 | 30 | 7 |

**Math 448. 1st Homework. Chapter 6. Due Wednesday, February 24, 1999.**

1. Find the mle for $\beta$ based on a random sample $X_1, \ldots, X_n$ from $X_i \sim \text{Gamma}(\alpha = 2, \beta)$. Show that mle is consistent and unbiased. Find the mle for the mean of $X$. Find the mle for the variance of $X$.

2. Given a random sample of size $n$ from a population having the cdf

$$P(X \leq x|\alpha, \beta) = \begin{cases} 0 & \text{if } x < 0, \\ (x/\beta)^\alpha & \text{if } 0 \leq x \leq \beta. \\ 1 & \text{if } \beta < x. \end{cases}$$

Find the mle's of $\alpha$ and $\beta$. Find the method of the moments estimators of $\alpha$ and $\beta$.

3. Do 6.63 using both mle and method of the moments estimators.

4. A farm grows grapes for jelly. The following data are measurements fo sugar in the grapes of a sample take from each of 39 truckloads:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 16.0 | 15.2 | 12.0 | 16.9 | 14.4 | 16.3 | 15.6 | 12.9 | 15.3 | 15.1 |
| 15.8 | 15.5 | 12.5 | 14.5 | 14.9 | 15.1 | 16.0 | 12.5 | 14.3 | 15.4 |
| 15.4 | 13.0 | 12.6 | 14.9 | 15.1 | 15.3 | 12.4 | 17.2 | 14.7 | 14.8 |

Assume that there are observations of a $N(\mu, \sigma^2)$. Find point estimates for $\mu$ and $\sigma^2$. Find an approximate 90 % confidence interval for $\mu$. Find an approximate 95 % confidence interval for $\sigma^2$.

5. Do Problem 6.44.

6. Do Problem 6.66.

7. Let $X_1, \ldots, X_5$ be a random sample of SAT mathematics scores, assumed to be $N(\mu_X, \sigma^2)$, and let $Y_1, \ldots, Y_8$ be an independent random sample of SAT verbal socre, asumed to be $N(\mu_Y, \sigma^2)$. If the following date are observed, find a 90 % confidence interval for $\mu_X - \mu_Y$:

$$x_1 = 644, \; x_2 = 493, \; x_3 = 532, \; x_4 = 462, \; x_5 = 565,$$

$$y_1 = 623, \; y_2 = 42, \; y_3 = 492, \; y_4 = 661, \; y_5 = 540, \; y_6 = 502, \; y_7 = 549, \; y_8 = 518.$$

Test the hypothesis $H_0 : \mu_X = \mu_Y$ versus the alternative $H_1 : \mu_X \neq \mu_Y$. Find the $p$–value of the test.

8. Do problem 6.50.

9. Do problem 6.53.

10. Do problem 6.56.

**Math 448. 2nd Homework. Chapter 7. Due Monday, March 22, 1999.**

1. Do 7.3.

2. Do 7.7.

3. Find joint sufficient statistics for a Gamma$(\alpha, \beta)$ distribution, $\alpha, \beta > 0$.

4. Do problem 7.53.

5. Do Problem 7.32.

6. Let $X_1, \ldots, X_n$ be a random sample with pdf

$$f(x, \theta) = \frac{\theta}{x^2},\ 0 < \theta < x.$$

Find a complete and sufficient statistic for $\theta$. Show that the statistic is complete and sufficient. Find the unbiased minimum variance estimator of $\theta$. Show that $\min(X_1, \ldots, X_n)$ and $\frac{\bar{x}}{\sqrt{\frac{1}{n}\sum_{j=1}^{n}(X_j-\bar{x})^2}}$ are independent.

7. Do Problem 7.38.

8. Let $X_1, \ldots, X_n$ be a random sample of size $n$ from the normal distribution $N(\theta, \sigma^2)$, where $\sigma^2$ is known. Find the unbiased minimum variance estimator of $\theta$. Find the unbiased minimum variance estimator of $\theta^2$. Show that $\bar{x}$ and $R = \max(X_1, \ldots, X_n) - \min(X_1, \ldots, X_n)$.

9. Do problem 7.59.

10. Do problem 7.77 (a)–(d). Do not do part (e).

**Math 448. 3rd Homework. Chapter 8. Due Monday, April 7, 1999.**

1. Do 8.3. Hint: you can check your solutions to some problems in the page 556 of the text.

2. Do 8.9.

3. Let $X_1, \dots, X_n$ be a random sample from the Weibull distribution with pdf $f(x, \theta, \tau) = \theta \tau x^{\tau - 1} e^{-\theta x^\tau}$ if $0 < x < \infty$, and $f(x, \theta, \tau) = 0$ else, where $0 < \theta$ is unknown and and $0 < \tau$ is known. Find the mle of $\theta$. If the parameter has a prior gamma with parameters $\alpha$ and $\beta$, find the posterior distribution. Find the Bayes' estimator when the loss function is $\mathcal{L}(\theta, \delta) = (\theta - \delta)^2$.

4. Do problem 8.17.

5. Do problem 8.18.

6. In each of the cases in problem 8.17: Find a complete and sufficient statistic for $\theta$. Find the unbiased minimun variance estimator for $\theta$. Compare the variance of the mle, the variance of the unbiased minimum variance estimator and the the Cramer–Rao lower bound. Is the Cramér–Rao lower bound attained?

7. Suppose that $X_1, \dots, X_n$ form a random sample from a exponential distribution with mean $\theta$. Determine the Cramér–Rao lower bound for $\theta^2$. Is the Cramér–Rao lower bound attained? Find a complete and sufficient statistic for $\theta^2$. Find the unbiased minimun variance estimator for $\theta^2$. Find the variance of the unbiased minimun variance estimator of $\theta^2$. Compare the variance of the mle, the variance of the unibased minimum variance estimator and the the Cramer–Rao lower bound.

8. Let $X_1, \dots, X_n$ be a random sample with a Poisson distribution with parameter $\theta > 0$. Find the Cramer–Rao lower bound for the unbiased estimators of $\theta^2$. Find a complete and sufficient statistic of $\theta^2$. Find the unbiased minimun variance estimator of $\theta^2$. Is the Cramer–Rao lower bound attained for the unbiased estimators of $\theta^2$? Find the mle of $\theta^2$. is the mse unbiased? Compare the variance of the mle, the variance of the unbiased minimum variance estimator and the the Cramer–Rao lower bound.

9. Do problem 8.28, answering the following extra questions: find the Cramer–Rao lower bound for the unbiased estimators of $\theta$, find the unbiased minimun variance estimator of $\theta$, compare the variance of the unbiased minimum variance estimator and the the Cramer–Rao lower bound. Hint: If $X$ has the density $f(x, \theta) = \frac{\theta}{(1+x)^{\theta+1}}$, $0 < x < \infty$, then $Y = \ln(1 + X)$ has a Gamma$(1, \frac{1}{\theta})$ distribution.

10. Do problem 8.31.

**Math 448. 4th Homework. Chapter 9. Due Monday, April 21, 1999.**

1. Do problem 9.7.

2. Do problem 9.8 (write the rejection region in the form $\{(x_1, \dots, x_n) : T(x_1, \dots, x_n) \geq c\}$).

3. Do problem 9.14.

4. Let $X_1, \dots, X_n$ be random sample of size $n = 20$ from the density $N(0, \theta)$, $\theta > 0$. Find the uniformly most powerful level 95% test for $H_0 : \theta = 4$, versus $H_1 : \theta < 4$ (determine the region completely).

5. Do problem 9.33.

6. Let $X_1, \dots, X_n$ be random sample of size $n$ from a pdf from the density $f(x, \theta) = \theta^x (1-\theta)^{1-x}$, $x = 0, 1$, $0 \leq \theta \leq 1$. Show that the likelihood ratio test for $H_0 : \theta = \frac{1}{2}$, versus $H_1 : \theta \neq \frac{1}{2}$ reject $H_0$ if $|\bar{x} - \frac{1}{2}| \geq c$, where $c$ is a constant. Hint: first show that the LRT reject $H_0$ if $\bar{x}^{\bar{x}}(1-\bar{x})^{1-\bar{x}} \geq c$, where $c$ is a constant. Observe that the function $f(x) = \ln(x^x(1-x)^{1-x})$, $0 \leq x \leq 1$, satifies $f(x) = f(1-x)$ and it has a minimum at $x = \frac{1}{2}$.

7. A random sample of size $n$ is to be used to test the null hypothesis that the parameter $\theta$ of an exponential propulation equals $\theta_0$, against the alternative that it does not equal $\theta_0$. Find an expression for the likelihood ratio statistic. Use the previous result to show that the critical region of the likelihood ratio test can be written as $\bar{x}e^{-\frac{\bar{x}}{\theta_0}} \leq c$. Prove that the equation as $\bar{x}e^{-\frac{\bar{x}}{\theta_0}} \leq c$ is equivalent to $c_1 \leq \bar{x} \leq c_2$, where $c_1 < \theta_0 < c_2$. Hint: The function $f(x) = xe^{-\frac{x}{\theta_0}}$ has a maximum at $x = \theta_0$

8. To test $H_0 : \mu = 335$ against $H_1 : \mu < 335$, under normal assumptions, a random sample of size 17 yielded $\bar{x} = 324.8$ and $s = 40$. Is $H_0$ accepted at an $\alpha = .10$ significance level? Find and implement the uniformly most powerful level test.

9.

10. Do 9.42.

**Math 448. 5–th Homework. Chapter 10. Due Wednesday, May 5, 1999.**

1. Four groups of three pigs each were fed individually four different feeds for a specified length of time with the observed weight gains

| | | | |
|---|---|---|---|
| $X_1$ | 194.11 | 182.8 | 187.43 |
| $X_2$ | 216.06 | 203.5 | 216.88 |
| $X_3$ | 178.1 | 189.2 | 181.33 |
| $X_4$ | 197.11 | 202.68 | 209.18 |

Find $\bar{x}_{\cdot 1}, \bar{x}_{\cdot 2}, \bar{x}_{\cdot 3}, \bar{x}_{\cdot 4}, \bar{x}_{\cdot\cdot}$ and the anova table for these data. Test the hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$, where $\mu_i$ is the mean weight gain for each of the feeds $i = 1, 2, 3, 4$, at a 95 % significance level.

2. Do problem 10.9. Find $\bar{x}_{\cdot 1}, \bar{x}_{\cdot 2}, \bar{x}_{\cdot 3}, \bar{x}_{\cdot\cdot}$ and the anova table for these data. Determine whether the null hypothesis is accepted or rejected at a 95 % significance level.

3. Do problem 10.23.

4. A research laboratory was developing a new compound for the relief of severe cases of hay fever. In an experiment with 36 volunteers, the amounts for the two active ingredients in the compound were varied at the three levels each. Randomization was used in assigning four volunteers to each of the nine treatments. The data on hours of relief follow.

| | | *Factor B* | | |
|---|---|---|---|---|
| | | *Low* | *Medium* | *High* |
| *Factor A* | | | | |
| | *Low* | 2.4 | 4.6 | 4.8 |
| | | 2.7 | 4.2 | 4.5 |
| | | 2.3 | 4.9 | 4.4 |
| | | 2.5 | 4.7 | 4.6 |
| | *Medium* | 5.8 | 8.9 | 9.1 |
| | | 5.2 | 9.1 | 9.3 |
| | | 5.5 | 8.7 | 8.7 |
| | | 5.3 | 9.0 | 9.4 |
| | *High* | 6.1 | 9.9 | 13.5 |
| | | 5.7 | 10.5 | 13.0 |
| | | 5.9 | 10.6 | 13.3 |
| | | 6.2 | 10.1 | 13.2 |

Find $\bar{x}_{ij\cdot}, \bar{x}_{i\cdot\cdot}, \bar{x}_{\cdot j\cdot}, \bar{x}_{\cdot\cdot\cdot}$ and the anova table for these data for the moel with interaction. Test at the 95 % significance level the following null hypothesis: (i) $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$, (ii) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ and (iii) $H_0 : \gamma_{ij} = 0$ for each $1 \le i, j \le 3$.

5. The cost of maintenance of shipping tractors seems to increase with the age of the tractor.

The following date were collected.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Age | 0.5 | 0.5 | 1.0 | 1.0 | 1.0 | 4.0 | 4.0 | 4.0 |
| 6 monts cost \$ | 163 | 182 | 978 | 466 | 549 | 495 | 723 | 681 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Age | 4.5 | 4.5 | 4.5 | 5.0 | 5.0 | 5.0 | 6.0 | 6.0 |
| 6 monts cost \$ | 619 | 1049 | 1033 | 890 | 1522 | 987 | 764 | 1373 |

Find $\hat{\alpha}, \hat{\beta}, s^2$ and the linear regression and the analysis of the variance table. Find a 95 % confidence interval for $\hat{\alpha}, \hat{\beta}, s^2$.

6. Consider the simple linear regression model. Let $e_i = y_i - \hat{y}_i$, $1 \leq i \leq n$, be the residuals. Prove that $\sum_{i=1}^n e_i = 0$ and $\sum_{i=1}^n x_i e_i = 0$

7. The data below give $X=$ water content of snow on April 1 and $Y$ = water yield from April to July (in inches) in the Snake River watershed in Wyoming for 17 years:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $X$ | 23.1 | 32.8 | 31.8 | 32.0 | 30.4 | 24.0 | 39.5 | 24.2 |
| $Y$ | 10.5 | 16.7 | 18.2 | 17.0 | 16.3 | 10.5 | 23.1 | 12.4 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $X$ | 52.5 | 37.9 | 30.5 | 25.1 | 12.4 | 35.1 | 31.5 | 21.2 | 27.6 |
| $Y$ | 24.9 | 22.8 | 14.1 | 12.9 | 8.8 | 17.4 | 14.9 | 10.5 | 16.1 |

Find $\hat{\alpha}, \hat{\beta}, s^2$ and the linear regression and the analysis of the variance table. Find a 95 % confidence interval for $\hat{\alpha}$. Test the null hypothesis of $H_0 : \alpha = 0$ at the 95 % significance level. This means that if there is no snow, the water yield of the Snake river is zero.

8. In the simple linear regression model show that $\hat{\beta} = \frac{\sum_{j=1}^n (x_j - \bar{x}) Y_j}{\sum_{j=1}^n (x_j - \bar{x})^2}$ and $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$ have a normal distribution. Find the mean and the variance of $\hat{\beta}$ and of $\hat{\alpha}$.

9. In a college health fitness program, let $X$ be the weight in kilograms of a female freshman at the beginning of the program and let $Y$ be her weight change during the semester. We shall use the following date for $n = 16$ observations of $(x, y)$ test the null hypothesis $H_0 : \ \rho = 0$ against $H_1 : \ \rho \neq 0$ at the levels (i) $\alpha = 0.10$ and (ii) $\alpha = 0.05$.

$$(61.4, -3.2)\ (62.9, 1.4)\ (58.7, 1.3)\ (49.3, 0.6)\ (71.3, 0.2)\ (81.5, -2.2)\ (60.8, 0.9)\ (50.2, 0.2)$$

$$(60.3, 2.0)\ (54.6, 0.3)\ (51.1, 3.7)\ (53.3, 0.2)\ (81.0, -0.5)\ (67.6, -0, 8)\ (71.4, -.01)\ (72.1, -0.1)$$

10. When bowling it is possible to score well on the first game and then bowl poorly on the second game, or vice versa. The following six pairs of numbers give the score of the first and second games bowled by the same person on six consecutive Tuesday evenings. Assume a bivariate normal distribution and use these scores to test the hypothesis $H_0 : \ \rho = 0$ against $H_1 : \ \rho \neq 0$ at the levels: (i) $\alpha = 0.10$ and (ii) $\alpha = 0.05$.

| | | | | | | |
|---|---|---|---|---|---|---|
| Game 1 | 170 | 190 | 200 | 183 | 187 | 178 |
| Game 2 | 197 | 178 | 150 | 176 | 205 | 153 |

**Math 448. 6–th Homework. Chapter 11. Due Wednesday, May 5, 1999.**

1. Do problem 11.6.
2. Do problem 11.8.
3. Do problem 11.9.

4. To test whether a golf ball of brand A can be hit a greater distance off the tee than a golf ball of brand B each of 17 golfers hit a ball of each brand, eight hitting ball A before ball B and nine hitting ball B before ball A. Let $m_D = m_A - m_B$ denote the median of the differences of the distances of ball A and ball B. Use the sign test to test the null hypothesis $H_0 : m_D = 0$ against the alternative $H_1 : m_D > 0$.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Distance for Ball $A$ | 265 | 272 | 246 | 260 | 274 | 263 | 255 | 258 |
| Distance for Ball $B$ | 252 | 276 | 243 | 246 | 275 | 246 | 244 | 245 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Distance for Ball $A$ | 276 | 274 | 274 | 269 | 244 | 212 | 236 | 254 | 224 |
| Distance for Ball $B$ | 259 | 260 | 267 | 267 | 251 | 222 | 235 | 255 | 231 |

5. For the data in problem 4 use, the t–test to test the null hypothesis $H_0 : m_D = 0$ against $H_1 : m_D > 0$.

6. Do problem 11.19.

7. On any given day at a candy warehouse, 13 trucks are loaded with candy. It is claimed that the median weight $m$ of loads of candy is 40000 pounds. Test, at the level $\alpha = 0.0461$ significance level, the null hypothesis $H_0 : m = 40000$ against $H_1 : m < 40000$ using the Wilcoxon test and the following observations

| | | | | | | |
|---|---|---|---|---|---|---|
| 41195 | 39485 | 41229 | 36840 | 38050 | 40890 | 38345 |
| 34930 | 39245 | 31031 | 40780 | 38050 | 30906 | |

8. Let $X$ and $Y$ equal the percentages of body fat for freshman women and men respectively, with distributions $F(x)$ and $G(x)$. Use the median test to test the null hypothesis $H_0 : \text{median}(F) = \text{median}(G)$ against $H_1 : \text{median}(F) > \text{median}(G)$.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| women | 16.6 | 16.7 | 18.5 | 19.2 | 21.5 | 22.4 | 22.6 | 23.2 | 24.2 | 26.3 |
| men | 9.4 | 9.7 | 11.3 | 11.8 | 13.3 | 15.6 | 16.1 | 16.5 | 18.2 | 21.7 |

9. For the data in problem 8 use, the run test to test the null hypothesis $H_0 : F(z) = G(z)$ for each $z$ against $H_1 : F(z) \leq G(z)$ for each $z$ with strict equality for some $z$.

10. For the data in problem 8 use, the Mann-Whitney–Wilcoxon test to test the null hypothesis $H_0 : F(z) = G(z)$ for each $z$ against $H_1 : F(z) \leq G(z)$ for each $z$ with strict equality for some $z$.

**Math 448. First Midterm. Friday, February 19, 1999.**

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

Show all your work. No credit for lucky answers.

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

$$f(x|\theta) = \frac{m x^{\alpha m - 1} e^{-\frac{x^m}{\theta}}}{\Gamma(\alpha)\theta^{\alpha}}, x > 0, \theta > 0,$$

where $m, \alpha > 0$ are known. Find the maximum likelihood estimator of $\theta$. Find a sufficient statistic of $\theta$. Show that the mle is a function of the sufficient statistic.

2. Let $X_1, \dots, X_n$ be a random sample with pdf

$$f(x, \theta) = \frac{1}{\theta},\ 0 < x < \theta.$$

Find the method of the moments estimator of $\theta$. Is this estimator unbiased? What about consistent? Find the mle of $\theta$. Is this estimator unbiased? What about consistent?

3. The assembly time in a plant is a normal random variable with mean 18.5 seconds and standard deviation 2.4 seconds. A random sample of 10 assembly times gave $\bar{x} = 19.6$. Is this evidence that $H_0 : \mu = 18.5$ should be rejected in favor of the alternative $H_1 : \mu > 18.5$ at the significance level 0.05. Find the probability that $H_0$ is accepted if $\mu = 19$. It is very important that the assembly time not exceed 20 seconds. How large a sample is necessary to reject $H_0 : \mu = 18.5$ with probability 0.95 if $\mu = 20$.

4. Suppose that a random sample from a normal population with the known variance $\sigma^2$ is to be used to test the null hypothesis $\mu = \mu_0$ against the alternative hypothesis $\mu = \mu_1$, where $\mu_1 > \mu_0$, and that the probabilities of type I and and type II errors are to have the preassigned values $\alpha$ and $\beta$. Show that the require size of the sample is given by $n = \frac{\sigma^2(z_\alpha + z_\beta)^2}{(\mu_1 - \mu_0)^2}$.

5. How large a sample must be selected from a normal distribution with standard deviation 12 in order to estimate $\mu$ to within 2 units (plus or minus two units) with probability 0.95?

6. In a random sample of 24 Black Angus steers of certain age have a standard deviation of 238 pounds. Assuming that the weights constitute a random sample from a normal population, find a 95 % confidence interval for $\sigma^2$.

7. 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. Test at the level 0.01 that the proportion for males and females are equal.

8. A die was cast $n = 120$ independent times and the following data resulted

| Spots up | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Frequency | $b$ | $b$ | 20 | 20 | $20 - b$ | $20 - b$ |

If we use a chi–square test, for what values of $b$ would the hypothesis that the die is unbiased be rejected at the 0.05 significance level.

**Math 448. Second Midterm. Wednesday, March 31, 1999.**

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

Show all your work. No credit for lucky answers.

1. Let $X_1, \dots, X_n$ be a random sample with pdf

$$f(x, \theta) = \frac{x^2 e^{-x/\theta}}{2\theta^3}, \text{ if } 0 < x < \infty.$$

Find a complete and sufficient statistic for $\theta$. Find the unbiased minimum variance estimator of $\theta$. Show that $\bar{x}$ and $\frac{\min(X_1,\dots,X_n)}{\max(X_1,\dots,X_n)}$ are independent.

2. Let $X_1, \dots, X_n$ be a random sample from $N(\theta, 100)$. Find a minimal sufficient statistic for $\theta$ (prove that this statistic is a minimal sufficient statistic for $\theta$).

3. Let $X_1, \dots, X_n$ be a random sample from $N(\theta, 100)$. Find a complete and sufficient statistic for $\theta$. Find the unbiased minimum variance estimator of $\theta^2$. Find the Cramer–Rao lower bound for the unbiased estimators of $\theta^2$. Compare the variance of the unbiased minimum variance estimator and the Cramer–Rao lower bound.

4. Let $X_1, \dots, X_n$ be a random sample from a Poisson distribution wth parameter $\theta > 0$. If the parameter has a prior gamma with parameters $\alpha$ and $\beta$, find the posterior distribution. Find the Bayes' estimator when the loss function is $\mathcal{L}(\theta, \delta) = (\theta - \delta)^2$.

5. Let $X_1, \dots, X_n$ be a random sample from a Poisson distribution with parameter $\theta > 0$. Find a complete and sufficient statistic for $\theta$. Find the unbiased minimun variance estimator for $\theta$. Find the Cramer–Rao lower bound for the unbiased estimators of $\theta$. Compare the variance of the unbiased minimum variance estimator and the Cramer–Rao lower bound.

6. Let $X_1, \dots, X_n$ be a random sample with pdf Poisson($\theta$). Find the mle $\hat{\theta}$. Find the asymptotic distribution of $\hat{\theta}$. Use this to find an asymptotic 95 % level confidence interval for $\theta$.

7. Let $X_1, \dots, X_n$ be random sample of size $n$ from the density $f(x|\theta) = \frac{1}{\theta} 10 x^9 e^{-x^{10}/\theta}$, $x > 0$, $f(x|\theta) = 0$, else, where $\theta > 0$ is unknown. Find the uniformly most powerful level $\alpha$ test for $H_0 : \theta = 1$, versus $H_1 : \theta > 1$ (we just to write the rejection region in the form $\{(x_1, \dots, x_n) : T(x_1, \dots, x_n) \geq c\}$, where $T(x_1, \dots, x_n)$ is a statistic which you need to find).

8. Let $X_1, \dots, X_n$ be a random sample from a normal distribution with mean $\mu$ and variance 9. Show that the likelihood ratio test of $H_0 : \mu = 5$ against $H_1 : \mu \neq 5$, has a rejection region of the form

$$\{(x_1, \dots, x_n) \in I\!R^n : |\bar{x} - 5| \geq c\},$$

where $c > 0$. Find $n$ and $c$ so that the power of the previous test satisfies $K(5) = 0.05$ and $K(10) = 0.90$.

**Math 448. Final Exam. Tuesday, May 11, 1999.**

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

Show all your work. No credit for lucky answers.

1. Let $X_1, \ldots, X_n$ be a random sample from the density $f(x;\theta) = e^{-(x-\theta)}$, $x > \theta$, $f(x;\theta) = 0$, else, where $\theta > 0$. Find the method of the moments estimator for $\theta$. Find the mle for $\theta$. Are these estimators consistent? Are they unbiased? Compare the mean square error of these two estimators? Hint: The MSE is $E[(\hat{\theta} - \theta)^2]$.

2. In a random sample of 250 viewers in a large city, 190 had seen a certain controversial program. Construct a 95 % confidence interval of the true proportion of people that saw that program.

3. Twenty pilots were tested in a flight simulator, and the time for each to complete a certain corrective action was measured in seconds, with the following results:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 5.2 | 5.6 | 7.6 | 6.8 | 4.8 | 5.7 | 9.0 | 6.0 | 4.9 | 7.4 |
| 6.5 | 7.9 | 6.8 | 4.3 | 8.5 | 3.6 | 6.1 | 5.8 | 6.4 | 4.0 |

Find a 95 % confidence interval for the mean time to take corrective action. Hint: $\sum_{i=1}^n x_i = 122.9$ and $\sum_{i=1}^n x_i^2 = 796.11$.

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

$$f(x;\theta) = \begin{cases} \frac{3x^5 e^{-\frac{x^3}{\theta}}}{\theta^2} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0, \end{cases}$$

where $\theta > 0$. Find a complete sufficient statistic for $\theta$. Find the uniform minimum variance estimator of $\theta$. Find the CRLB for the unbiased estimators for $\theta$. Compare the CRLB for the unbiased estimators for $\theta$ with the variance of the UMVUE of $\theta$.

5. Let $p$ equal the fraction defective of a certain manufactured item. To test $H_0 : p = \frac{1}{26}$ against $H_1 : p \neq \frac{1}{26}$, we inspect $n$ items in the sample. We reject $H_0$, if the observed number of defective items $y$ satisfies $y \geq c$. Find $n$ and $c$ so that $\alpha = K(\frac{1}{26}) = 0.05$ and $K(\frac{1}{10}) = 0.90$ approximately, where $K(p) = P_p(Y \geq c)$.

6. Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with parameter $\theta > 0$. Find the uniformly most powerful level $\alpha$ test for $H_0 : \theta = 1$, versus $H_1 : \theta > 1$ (we just to write the rejection region in the form $\{(x_1, \ldots, x_n) : T(x_1, \ldots, x_n) \geq c\}$, where $T(x_1, \ldots, x_n)$ is a statistic which you need to find).

7. Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ are unknown parameters. Find the rejection region for the likelihood ratio test $H_0 : \mu = \mu_0$, versus $H_1 : \mu \neq \mu_0$.

8. A rehabilitation center researcher was interested in examining the relationship between physical fitness prior to surgery of persons undergoing corrective knee surgery and time required in physical therapy until successful rehabilitation. Patient records in the rehabilitation center were examined, and 24 males 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 fitness status (below average, average, 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 | | | | |

Test at the 0.05 significance level whether or not the mean number of days required for successful rehabilitation is the same for the three fitness groups. Complete the following anova table:

| Source | degrees of freedom | Sums of squares | Mean square error | $F$ value |
|---|---|---|---|---|
| factor | | 672 | | |
| error | | 416 | | |
| total | | 1088 | | |

9. The number of pounds of steam used per month as a plant is though to be related to the average monthly ambient temperature. The last year's usages and temperatures are shown below

| Month | Jan | Feb | Mar | Apr | May | Jun |
|---|---|---|---|---|---|---|
| Temper. | 21 | 24 | 32 | 47 | 50 | 59 |
| Usage/1000 in pounds | 185.79 | 214.47 | 288.03 | 424.84 | 454.68 | 539.03 |

| Month | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|
| Temper. | 68 | 74 | 62 | 50 | 41 | 30 |
| Usage/1000 in pounds | 621.55 | 675.06 | 562.03 | 452.93 | 369.95 | 273.98 |

Fit linear regression model $\hat{y} = \hat{\alpha} + \hat{\beta}x$ to the data, that is find $\hat{\alpha}$ and $\hat{\beta}$. Hint: $\sum_{i=1}^{n} x_i = 558$, $\sum_{i=1}^{n} x_i^2 = 29256$, $\sum_{i=1}^{n} y_i = 5056.2$ and $\sum_{i=1}^{n} x_i y_i = 265873$.

10. The following data in tons are the amount of sulfur of oxides emitted by a large industrial plant in 40 days.

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 17 | 15 | 20 | 29 | 19 | 18 | 22 | 25 | 27 | 9 | 24 | 20 | 17 | 6 | 24 | 14 | 15 | 23 | 24 | 26 |
| 19 | 23 | 28 | 19 | 16 | 22 | 24 | 17 | 20 | 13 | 19 | 10 | 23 | 18 | 31 | 13 | 20 | 17 | 24 | 14 |

Use the sign test to test $H_0 : \xi = 21.5$ against the alternative hypothesis $H_1 : \xi \neq 21.5$ at the 0.05 significance level, where $\xi$ is the median of the distribution of sulfur oxides emitted by the industrial plant.