MATH 4720 Test 3 Name ______________________

Point totals are in parentheses next to each problem. Please show all work for partial credit.

1. Consider the simple linear regression model "through the origin".
   $Y_i = \beta_1 x_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$

   Find the maximum likelihood estimators for β_1 and σ^2.
   (Hint: $Y_i \sim N(\beta_1 x_i, \sigma^2)$)

(12)

$$f(y) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(y-\beta_1 x)^2}$$

$$L(y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(y_i-\beta_1 x_i)^2} = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2\sigma^2}\sum(y_i-\beta_1 x_i)^2}$$

$$\ln L = -n/2 \ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum(Y_i-\beta_1 x_i)^2$$

$$\frac{\partial \ln L}{\partial \beta_1} = \frac{-1}{2\sigma^2}\sum(Y_i-\beta_1 x_i)(-x_i) \overset{SET}{=} 0$$

$$\sum x_i y_i - \hat{\beta}\sum x_i^2 = 0 \Rightarrow \boxed{\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}}$$

$$\frac{\partial \ln L}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma^3}\sum(Y_i-\beta_1 x_i)^2 = 0$$

$$\Rightarrow \frac{1}{\hat{\sigma}^3}\sum(Y_i-\hat{\beta}_1 x_i)^2 = \frac{n}{\hat{\sigma}}$$

$$\Rightarrow \boxed{\hat{\sigma}^2 = \frac{1}{n}\sum(Y_i-\hat{\beta}_1 x_i)^2}$$

2. Credit scores can help determine whether an individual will qualify for a mortgage loan, and they are even used to determine the interest rates that will be charged. Six customers of a local bank are randomly selected and their credit scores (between 300 and 850) with corresponding interest charged (in percent) on a car loan are recorded. Summary statistics are listed below: (x = credit score, y = interest rate)

$\sum x = 3{,}910$ $\sum y = 72$ $\sum x^2 = 2{,}575{,}800$ $\sum y^2 = 1{,}007$ $\sum xy = 45{,}002$

a) Find S_{xx}, S_{yy}, S_{xy}

(5) $S_{xx} = 2{,}575{,}800 - \frac{3910^2}{6} = 27{,}783.33$

$S_{yy} = 1007 - \frac{72^2}{6} = 143$ $S_{xy} = 45{,}002 - \frac{(3910)(72)}{6}$

$= -1{,}918$

b) Compute r, the sample correlation coefficient

(3) $r = \frac{-1{,}918}{\sqrt{(27{,}783)(143)}} = -.96225$

c) Find the coefficient of determination and interpret the value in terms of credit scores and interest rates.

(3) $R^2 = (-.96225)^2 = .9259$ 92.6% of variation in interest rates is explained by credit scores.

d) Use Fisher's Z transformation to test $H_0: \rho = -0.80$ vs. $H_A: \rho < -0.80$ at $\alpha = 0.02$

(5) $\frac{1}{2}\ln\left(\frac{1-.9625}{1+.9625}\right) = -1.97882$ $\frac{1}{2}\ln\left(\frac{1-.8}{1+.8}\right) = -1.09861$

$z = \frac{-1.97882 - (-1.09861)}{1/\sqrt{6-3}} = -1.5246$

p-value = .0637 > α = .05 ⇒ don't reject Ho. There is not enough evidence to conclude that ρ < -.80.

3. Let $Y_1 \ldots Y_n$ be a random sample from $Y \sim EXP(\beta)$. Find the form of the Generalized Likelihood Ratio Test for $H_0: \beta = \beta_0$ vs. $H_a: \beta \neq \beta_0$.

a) Find the maximum likelihood estimators under Ω_0 and Ω.

$\Omega_0 : (\hat{\beta} = \beta_0)$

(5)

Ω :

$$L = \prod_{i=1}^{n} \frac{1}{\beta} e^{-y_i/\beta} = \frac{1}{\beta^n} e^{-\frac{1}{\beta}\Sigma y_i}$$

$$\ln L = -n \ln \beta - \frac{1}{\beta} \Sigma y_i$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{-n}{\hat{\beta}} + \frac{\Sigma y_i}{\hat{\beta}^2} \overset{SET}{=} 0 \quad \Rightarrow \frac{\Sigma y_i}{\hat{\beta}^2} = \frac{n}{\hat{\beta}} \quad \Rightarrow \boxed{\hat{\beta} = \frac{\Sigma y_i}{n} = \bar{Y}}$$

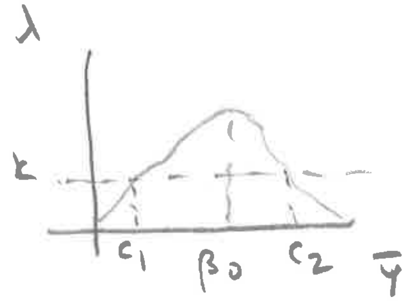
b) Evaluate $\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} \leq k$ to find the form of the GLR test. (Hint: Use $\ln(\lambda)$ to determine when λ itself is increasing or decreasing.)

(10)

$$\lambda = \frac{L(\beta_0)}{L(\bar{Y})} = \frac{\left(\frac{1}{\beta_0}\right)^n e^{-\frac{1}{\beta_0}\Sigma Y_i}}{\left(\frac{1}{\bar{Y}}\right)^n e^{-\frac{1}{\bar{Y}}\Sigma Y_i}} = \left(\frac{\bar{Y}}{\beta_0}\right)^n \frac{e^{-\frac{n\bar{Y}}{\beta_0}}}{e^{-n}} = \left(\frac{\bar{Y}}{\beta_0}\right)^n e^{-n\left[\frac{\bar{Y}}{\beta_0} - 1\right]}$$

$$\Rightarrow \ln \lambda = n\left[\ln \bar{Y} - \ln \beta_0\right] - n\left[\frac{\bar{Y}}{\beta_0} - 1\right]$$

$$\frac{\partial \ln \lambda}{\partial \bar{Y}} = \frac{n}{\bar{Y}} - \frac{n}{\beta_0} \quad \begin{matrix} > 0 \text{ if } \bar{Y} < \beta_0 \\ < 0 \text{ if } \bar{Y} > \beta_0 \end{matrix} \quad \Rightarrow \text{max at } \bar{Y} = \beta_0$$



$\boxed{\text{Reject } H_0 \text{ if } \bar{Y} \geq c_2 \text{ or } \bar{Y} \leq c_1}$

4. A farmer would like to examine the relationship between rainfall (in inches, X) and yield of wheat (in bushels per acre, Y). He collects data for 8 different harvests and records summary information below:

$\sum x = 98.1 \quad \sum y = 435.8 \quad \sum x^2 = 1299.85 \quad \sum y^2 = 25{,}705 \quad \sum xy = 5{,}772.65$

a) Determine the least squares regression line and predict the yield of wheat if the rainfall was 14 inches.

(6)

$S_{xx} = 1299.85 - \frac{98.1^2}{8} = 96.90$

$S_{yy} = 25{,}705 - \frac{435.8^2}{8} = 1{,}964.80$

$S_{xy} = 5{,}772.65 - \frac{(98.1)(435.8)}{8}$

$= 428.65$

$\hat{\beta}_1 = \frac{428.65}{96.90} = 4.424$

$\hat{\beta}_0 = \frac{435.8}{8} - 4.424\left(\frac{98.1}{8}\right)$

$= .2257$

$\hat{Y} = .2257 + 4.424(14)$

$= 62.1617$

b) Test H_0: $\beta = 3$ vs. H_0: $\beta > 3$ (at α = 0.05)

$SSE = 1{,}964.8 - 4.424(428.65) = 68.45$

(6) $s = \sqrt{\frac{68.45}{6}} = 3.38$

$t_{6,.05} = 1.943$

$T = \frac{4.424 - 3}{3.38/\sqrt{96.9}} = 4.15$ $\Rightarrow$ Reject Ho

c) Construct a 95% prediction interval for the yield of wheat if the rainfall was 14 inches.

(5)

$62.167 \pm 2.447(3.38)\sqrt{1 + \frac{1}{8} + \frac{(14-12.26)^2}{96.90}}$

62.167 ± 8.89

$\boxed{(53.27, 71.05)}$

5. A study was conducted to determine the effect of early child care on infant-mother attachment patterns. In the study, 93 infants were classified either "secure" or "anxious" using the Ainsworth strange-situation paradigm. In addition, the infants were classified according to the average number of hours per week that they spent in child care. The data appear below in the table:

Attachment Pattern	Hours in Child Care: Low (0 – 3 hrs)	Moderate (4-19 hrs)	High (20 – 54 hours)	
Secure	24 (24.09)	35 (30.98)	(8.95) 5	64
Anxious	11 (10.9)	10 (14.03)	(4.05) 8	29
	35	45	13	93

Do the data indicate a dependence between attachment patterns and the number of hours spent in child care? Test using α = 0.05.

(10)

$$E_{ij} = \frac{R_i \times C_j}{n}$$

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(24-24.09)^2}{24.09} + \cdots + \frac{(8-4.05)^2}{4.05}$$

$= 7.267$

p-value = .026

$\chi^2_{2,.05} = 5.99147$

⇒ Reject Ho i.e. dependence between attachment and day care hrs. exists.

6. The FBI claims that in Boston (the bank robbery capital of the world) there is a 40% chance of no bank robberies in a month, a 30% chance of one bank robbery each month, a 20% chance of two bank robberies each month, and a 10% chance of 3 bank robberies in a month. Using the observed data below collected for 10 years, is there evidence to reject the FBI's hypothesis on probabilities regarding bank robberies per month in Boston. Use α = 0.05.

No. of bank robberies in the month	0	1	2	3	
Count	57	36	15	12	n = 120

(10)

Ho: $p_0 = .40$, $p_1 = .30$, $p_2 = .20$, $p_3 = .10$

Ha: at least one $p_i \neq p_{i0}$

$E_0 = 120(.40) = 48$

$E_1 = 36$, $E_2 = 24$,

$E_3 = 12$

$$\chi^2 = \frac{(57-48)^2}{48} + \cdots + \frac{(12-12)^2}{12} = 5.0625$$

$\chi^2_{3,.05} = 7.815$ ⇒ don't reject Ho

Not enough evidence to reject FBI's claim.

7. Let $Y_1 \ldots Y_n$ be a random sample from $Y \sim POI(\lambda)$ with prior distribution for $\lambda \sim GAM(\alpha, \beta)$ where α, β are known constants.

a) Find the joint density for $Y_1 \ldots Y_n, \lambda$.

(3)

$$f(y_1, \ldots, y_n, \lambda) = \prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta}$$

$$= \boxed{\frac{\lambda^{\Sigma y_i} e^{-n\lambda}}{\prod_{i=1}^{n} y_i!} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta}}$$

b) Use your result from part (a) to find the marginal density for $Y_1 \ldots Y_n$

$$g(y_1, \ldots, y_n) = \int f(y_1 \ldots y_n, \lambda)\, d\lambda$$

(8)

$$= \int_0^\infty \frac{\lambda^{\Sigma y_i} e^{-n\lambda}}{\prod_{i=1}^{n} y_i!} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta}\, d\lambda$$

$$= \frac{1}{\prod_{i=1}^{n} y_i!\, \Gamma(\alpha)\beta^\alpha} \cdot \int_0^\infty \lambda^{(\Sigma y_i + \alpha)-1} e^{-\lambda(n + 1/\beta)}\, d\lambda$$

$$= \frac{1}{\prod_{i=1}^{n} y_i!\, \Gamma(\alpha)\beta^\alpha} \cdot \int_0^\infty \lambda^{(\Sigma y_i + \alpha)-1} e^{-\lambda/\left(\frac{\beta}{n\beta+1}\right)}\, d\lambda$$

$$= \boxed{\frac{\Gamma(\Sigma y_i + \alpha)\left(\frac{\beta}{n\beta+1}\right)^{\Sigma y_i + \alpha}}{\left(\prod_{i=1}^{n} y_i!\right) \cdot \Gamma(\alpha) \cdot \beta^\alpha}} \quad \text{or} \quad \frac{\Gamma(\Sigma y_i + \alpha)\, \beta^{\Sigma y_i}}{\left(\prod_{i=1}^{n} y_i!\right) \Gamma(\alpha)\, (n\beta+1)^{\Sigma y_i + \alpha}}$$

(7 cont.)

c) Use your results from parts (a) and (b) to show that the posterior density for $\lambda \mid Y_1 \ldots Y_n$ is $GAM(\alpha^*, \beta^*)$ and identify α^* and β^*.

(6)

$$h(\lambda \mid y_1 \cdots y_n) = \frac{f(y_1 \cdots y_n, \lambda)}{g(y_1 \cdots y_n)}$$

$$= \frac{\lambda^{\Sigma y_i} e^{-n\lambda}}{\cancel{\prod_{i=1}^{n} y_i!}} \cdot \frac{1}{\cancel{\Gamma(\alpha)\beta^{\alpha}}} \lambda^{\alpha-1} e^{-\lambda/\beta} \cdot \frac{\cancel{\left(\prod_{i=1}^{n} y_i!\right)\Gamma(\alpha)\beta^{\alpha}}}{\Gamma(\Sigma y_i+\alpha)\left(\frac{\beta}{n\beta+1}\right)^{\Sigma y_i+\alpha}}$$

$$= \frac{\lambda^{\Sigma y_i+\alpha-1} e^{-\lambda(n+1/\beta)}}{\Gamma(\Sigma y_i+\alpha)\left(\frac{\beta}{n\beta+1}\right)^{\Sigma y_i+\alpha}}$$

$$= \boxed{\frac{\lambda^{(\Sigma y_i+\alpha-1)} e^{-\lambda/\left(\frac{\beta}{n\beta+1}\right)}}{\Gamma(\Sigma y_i+\alpha)\left(\frac{\beta}{n\beta+1}\right)^{\Sigma y_i+\alpha}}} \sim GAM\left(\alpha^* = \Sigma y_i+\alpha,\ \beta^* = \frac{\beta}{n\beta+1}\right)$$

d) Use your result in part (c) to find the Bayes estimator for λ. (Hint: the Bayes estimator is the mean of the posterior distribution)

(3)

$$\hat{\lambda} = E[\lambda \mid y_1 \cdots y_n] = \alpha^* \cdot \beta^*$$

$$= \boxed{(\Sigma y_i+\alpha)\left(\frac{\beta}{n\beta+1}\right)}$$