COVER PAGE

STATISTICS 608 - EXAM 3

July 24/25 2017

Student's Name (PRINT) RAJAN KAPOOR

Student's Email Address (PRINT) r.kapoor@tamu.edu.

INSTRUCTIONS TO STUDENTS:

(1) Write your answers in the spaces provided on the examination paper. The last two (blank) pages can be used for rough work or as additional space for answers. You have exactly 120 minutes to complete the exam within the time frame 12:01 PM, CDT, 7/24/2017 to 12:01 PM, CDT 7/25/2017. **The exam starts AFTER you have downloaded and printed it.** If you have been granted extra time by Disability Services, your proctor will have been informed accordingly.

(2) Upon completing the exam, you have a 30-minute buffer in which to scan and upload it to Webassign. You may not work on the exam during this time.

(3) You may use your own computer together with a pocket calculator and/or any software package already loaded on your computer to do calculations.

(4) The exam is OPEN BOOK. You may make use of the textbook and any other material that you saw fit to prepare beforehand, either as hard copy or on your PC/Laptop.

(5) **You may not access the internet other than to download and to upload the exam.**

I attest that I spent no more than 120 minutes to complete the exam. I did not access the internet during the exam nor did I receive assistance from anyone during the exam. I promise not to discuss or provide any information to anyone concerning any aspect of this exam until after 3:31 PM on 7/25/2017.

Student's Signature Rajan

INSTRUCTIONS TO PROCTOR:

The exam starts only **after** the student has downloaded and printed it. **Immediately** after the exam ends, have the student scan the exam **with this cover sheet on top** to a PDF file and upload it to Webassign.

(1) I certify that the student's exam start time was 8:46AM, and the exam completion time was 10:47 AM

(2) I certify that the student has followed all the **INSTRUCTIONS TO STUDENTS** listed above.

(3) I certify that the exam was scanned into a PDF and uploaded to Webassign **in my presence**.

Proctor's Printed Name REMOTE PROCTOR

Proctor's Signature Rajan

Date 25 July 17

FG: A

Well done!

33/42

STATISTICS 608 - Examination 3

July 24/25, 2017 - Duration: **120 MINUTES**

Total points available: **48** (45 points =100%)

SHOW ALL CALCULATIONS AND EXPLANATIONS. *PARTIAL CREDIT WILL ACCRUE FOR ALL RELEVANT WORK SHOWN.*

This paper consists of ten (10) pages.

9

Question 1 [3+1+4+2+3=13 points] The data in the file *beetles.csv* show the number of insects killed after exposure to various (standardized) doses of an insecticide. Set

$$p(x) = P(killed|dose = x).$$

1.1 Fit a logistic regression model

$$log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 \times x + \beta_2 \times x^2.$$

Write the values of the three coefficient estimates here:

$\hat{\beta}_0 = 0.4999$; $\hat{\beta}_1 = 2.7262$: $\hat{\beta}_2 = 0.7116$ ✓ 3

1.2 Estimate $p(0.5)$. $\hat{\beta}_0 + \hat{\beta}_1(0.5) + \hat{\beta}_2(0.25) = 2.0409$.

1.3 Using the covariance matrix of the estimated coefficients in *1.1*, or otherwise, find an approximate 95% confidence interval for $p(0.5)$.

1.4 Use the Wald statistic to test at the 1% level of significance whether the estimated coefficient of $(dose)^2$ is statistically significant.

1.5 Assess the quality of the fit. Show *one* of your diagnostic plots (a rough *freehand* version will be acceptable) and indicate the relevant feature(s) on it that point(s) to a satisfactory or unsatisfactory fit.

Continue answers to Question 1 here

(1.2) $\Rightarrow$ $$p(x) = \frac{d}{1+\exp(-(\beta_0+\beta_1 x+\beta_2 x^2))} = 0.885$$ ✓ 1

p-value

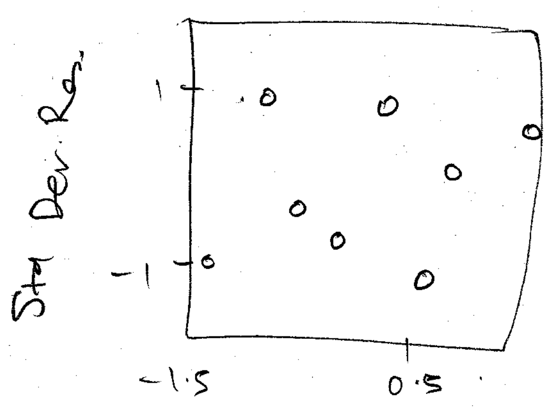
(1.4) Wald St = 0.00688 < 0.01 ✓ 2

$\therefore$ significant

(1.5) Null dev = 284.2

Res dev = 3.18 good fit

Standard ~~Deviance~~ Deviance Residual and Pearson Res are randomly distributed. (all lie -2 to 2)



✓ 3

(1.3) Prediction? Interval $\left(\frac{1}{1+\exp(-1)}, \frac{1}{1+\exp(-2.0409)}\right)$ ✗

$= (0.731, 0.8850224)$

13

Question 2 [3+3+3+5=14 points]

The mean μ of two identically normally distributed observations Y_1, Y_2 with *known variance* 1 was estimated by the sample mean $\bar{Y} = (Y_1 + Y_2)/2$. The analyst assumed that the two observations were independent and found a p-value of 0.05 in a test of the hypothesis $H_0 : \mu = 0$ against the *one-sided* alternative $H_a : \mu > 0$ based on the assumption that the statistic $\sqrt{2} \times \bar{Y}$ has a normal$(0, 1)$ distribution. However, unbeknown to the analyst, Y_1 and Y_2 were in fact the first two observations from a *moving average process*

$$Y_t = \epsilon_t + \theta\epsilon_{t-1}, \ t = 1, 2, ...$$

where the ϵ_t are independent normal$(0, 1)$ random variables and θ is **negative.**

2.1 Show that $cov(Y_1, Y_2) = \theta$.

2.2 What is the numerical value of θ? Show your calculations.

Regardless of what you found in *2.2*, take $\theta = -0.45$ in *2.3* and *2.4* below.

2.3 Show that $var(\bar{Y}) = 0.4$.

2.4 What is the *correct* p-value in the hypothesis test?

2.1

$Y_t = \epsilon_t + \theta\epsilon_{t-1}$ $\quad$ $Y_1 = \epsilon_1 + \theta\epsilon_0$

$Y_2 = \epsilon_2 + \theta\epsilon_1$

$Cov(Y_1, Y_2) = Cov(Y_1, \theta Y_1 + \epsilon_2 + \theta^2\epsilon_0)$

$= Cov(Y_1, \theta Y_1) + Cov(Y_1, \epsilon_2) + Cov(Y_1, \theta^2\epsilon_0)$

$= Cov(Y_1, \theta Y_1) + 0 + 0$

$= \theta$ ✓ 3

$Var(Y_1) = 1 = Var(Y_2)$

$Y_2 = \epsilon_2 + \theta\epsilon_1$

$Y_1 = \epsilon_1 + \theta\epsilon_0$

$Y_2 - \theta Y_1 = \epsilon_2 + \theta^2\epsilon_0$

$Y_2 = \theta Y_1 + \epsilon_2 + \theta^2\epsilon_0$

Continue answers to Question 2 here

2.3 $Var(\bar{Y}) = \frac{1}{4}(Var(Y_1) + Var(Y_2) + 2 Cov(Y_1, Y_2)]$

$= \frac{1}{4}(1 + \theta^2 + 1 + \theta^2 + 2\theta))$

$= \frac{1}{2}(1 + \theta^2 + (\theta)) = 0.5(1.2025 - 0.45)$

$\frac{1}{2}(1 + (-.45)^2 - .45)$ $= 0.376$ ✓ 3

$=$

(2.2) $Var(Y_2) = E[Y_2^2] - \cancel{E[Y_2]^2} \to 0 = 1$

$\epsilon_2 = Y_2 - \theta\epsilon_1$

$\epsilon_2 = Y_2 - \theta Y_1 + \theta^2 \epsilon_0$

$\Rightarrow E[(\epsilon_2 + \theta\epsilon_1)^2] = 1$ ✓

$(E[\epsilon_2^2] = Var[\epsilon_2])$ $\Rightarrow E[\epsilon_2^2 + \theta^2\epsilon_1^2 + 2\theta\epsilon_2\epsilon_1] = 1$ ✓

$\Rightarrow 1 + \theta^2 + 2\theta E[\epsilon_2\epsilon_1] = 1$ ✓

$\Rightarrow 1 + \theta^2 + 2\theta \cdot \theta \cdot E[\epsilon_1^2] = 1 \Rightarrow 1 + 3\theta^2 = 1$ ✗

$\Rightarrow \theta = 0.$ 2

(2.4)

$P\left(Z_{stat} > 0\right)$

$Z_{stat} = \frac{\bar{Y}}{var(\bar{Y})}$

$P\left(\frac{\bar{Y}}{var(\bar{Y})} > \frac{1.96\sigma}{\sqrt{Var(\bar{Y})}}\right)$ ✓ 5

$P\left(Z > \frac{1.96 \times \frac{1}{\sqrt{2}}}{\sqrt{0.4}}\right) = 0.0142$ ✓

$\boxed{5}$

Question 3 [6+4+6=14] The data in the file *prof_salary.csv* give the number of years experience ($= x$), the estimated third quartile of annual salary (3Quart=Y) and the sample size (n) on which each estimate was based, of full professors at a university. The objective is to relate salary to years of experience by way of a regression model

$$y_i = \beta_0 + x_i\beta_1 + \epsilon_i$$

Since the variance of y_i is inversely proportional to the sample size n_i, the error term ϵ_i is expressed in the form $e_i/\sqrt{n_i}$ with a normal$(0, \sigma^2)$ distribution for each e_i.

3.1 Find the weighted least squares estimates of β_0, β_1 and σ^2.

3.2 Does the model appear to be a valid one? Justify your response by exhibiting some appropriate plots (rough, hand drawn, plots are acceptable).

3.3 Assume, regardless of your response in *3.2*, that the model is valid. A 90% prediction interval for the third quartile of the salary of a new professor with 8 years experience has the form $(E - tS, E + tS)$ where E is an estimate of the third quartile of the salary of the new professor, t is a critical value on the Student t-distribution and S is an appropriate standard error. Write the three values

$$E = 116679.4 \;;t = 1.8595 \;;S =$$

and show how you calculated E and S.

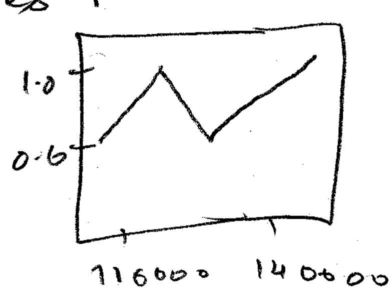
Weights are $\sqrt{n_i}$. You used $\frac{1}{n_i}$

3.1 $\hat{\beta}_0 = 103474.2$ ×

$\hat{\beta}_1 = 1650.6$ ×

4

(3.2) No. Standardized res plot is not random ?



1

Continue answers to Question 3 here

$\boxed{5}$

Question 4 [4+2=6 points] For $i = 1, 2, 3, 4$ the random variables y_i are independent and normal(μ_i, i) where

$$\mu_1 = \mu_2 = \gamma_0;\ \mu_3 = \gamma_0 - \gamma_1;\ \mu_4 = \gamma_0 + \gamma_1$$

and where γ_0 and γ_1 are unknown constants.

4.1 Write the preceding structure in general linear model form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{e},$$

identifying clearly the vectors $\mathbf{y}$ and $\boldsymbol{\beta}$ and the matrices $\mathbf{X}$ and $\mathbf{W}$.

4.2 State the distributions of the components of the error vector $\mathbf{e}$.

$$y_1 = \mathcal{N}(\gamma_0, 1) + \mathcal{N}(\gamma_1, 1)\times 0 = \gamma_0 + \mathcal{N}(0,1)$$
$$y_2 = \mathcal{N}(\gamma_0, 2) + \mathcal{N}(\gamma_1, 2)\times 0 = \gamma_0 + \mathcal{N}(0,2)$$
$$y_3 = \mathcal{N}(\gamma_0, 3) + \mathcal{N}(\gamma_1, 3)\times(-1) = \gamma_0 - \gamma_1 + \mathcal{N}(0,3)$$
$$y_4 = \mathcal{N}(\gamma_0, 4) + \mathcal{N}(\gamma_1, 4) = \gamma_0 + \gamma_1 + \mathcal{N}(0,4)$$

$$\mathcal{N}(0,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-x}{2\pi\sigma^2}\right)$$

$$\bar{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix},\quad \bar{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & -1 \\ 1 & 1 \end{bmatrix},\quad \bar{\beta} = \begin{bmatrix} \gamma_0 \\ \gamma_1 \end{bmatrix},\quad \bar{W} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{4}} \end{bmatrix}$$

✓ ✓ ✓ 4

4.2 $\bar{e}$ a components are normally ✓ distributed with mean 0 and variance 1, 2, 3, 4 respectively. ✗ 1

for e_1, e_2, e_3, e_4

1

Question 5 [1] An experimenter tells you he wishes to fit the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e$$

by least squares, subject to the restriction that $\beta_1 = 1$. He asks you if he can just fit

$$Y - x_1 = \beta_0 + \beta_2 x_2 + \epsilon$$

by least squares to get what he wants. Can he? Answer yes or no.

Yes. ✓

Additional space for rough work or answers

$Y.$ θ

Additional space for rough work or answers

$$\mathrm{Cov}(Y_1, Y_2) = E[Y_1 Y_2] - E[Y_1]E[Y_2]$$

$$= E[(\epsilon_1 + \theta\epsilon_0)(\epsilon_2 + \theta\epsilon_1)] - 0$$

$$= E[\epsilon_1\epsilon_2 + \theta^2\epsilon_0\epsilon_1 + \theta(\epsilon_0\epsilon_2 + \epsilon_1^2)]$$

$$= \cancel{\epsilon} 1 + \theta^2 + \theta$$

$$Y_1 = \epsilon_1 + \theta\epsilon_0 \Rightarrow \theta Y_1 = \theta\epsilon_1 + \theta^2\epsilon_0$$

$$Y_2 = \epsilon_2 + \theta\epsilon_1$$

$$\mathrm{Var}[\epsilon_t] = E[\epsilon_t^2]$$

$$Y_2 - \theta Y_1 = \epsilon_2 + \theta^2\epsilon_0$$

$$Y_2 = \theta Y_1 + \epsilon_2 + \theta^2\epsilon_0$$

$$Y_t = \epsilon_t + \theta\epsilon_{t-1}$$

$$\mathrm{Var}(F) = 1 +$$

$$\frac{Y_1 + Y_2}{2}.$$

$$E[Y_t] = 0 + \theta \times 0 = 0$$

$$\mathrm{Var}(Y_t) = E[Y_t^2] - E[Y_t]^2$$

$$1 + \theta^2 + 1 + \theta^2 + 2(-0.45).$$

$$= E[\epsilon_t^2 + \theta^2\epsilon_{t-1} + \epsilon_{t-1}^2]$$

$$2 - 0.9 + \theta^2$$

$$= 1 + \theta^2 + 1$$

$$1.1 + 0.2025$$

$$= 2 + \theta^2$$

$$\frac{4 + 2\theta^2 + 2\theta}{4}$$

$$= \frac{4.405 - 0.9}{4} = \frac{3.505}{4}$$