STAT 513 fa 2018 Exam II (take-home)

Karl B. Gregory

assigned: Tuesday, Nov 13th, 2018

due: Thursday, Nov 15th, 2018

Instructions:

- Looking at course notes IS allowed; Working with others IS NOT allowed.
- You will need, at most, a simple calculator.
- Write solutions on blank sheets of paper and turn these sheets in WITH A BLANK SHEET ON TOP (to cover answers) which has only your name on it. You do not need to turn in a copy of the exam itself.
- Some values of the function $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$:

z	1.282	1.645	1.96	2.576
$\Phi(z)$	0.9	0.95	0.975	0.995

- Some upper quantiles of some t distributions:

ξ	0.10	0.05	0.025	0.01	0.005
$t_{8,\xi}$	1.397	1.860	2.306	2.896	3.355
$t_{9,\xi}$	1.383	1.833	2.262	2.821	3.250
$t_{10,\xi}$	1.372	1.812	2.228	2.764	3.169

- The upper 0.05 quantile of the F_{ν_1,ν_2} distribution for come combinations of ν_1 and ν_2:

	$\nu_1 = 1$	$\nu_1 = 2$	$\nu_1 = 3$	$\nu_1 = 4$	$\nu_1 = 5$
$\nu_2 = 5$	6.61	5.79	5.41	5.19	5.05
$\nu_2 = 6$	5.99	5.14	4.76	4.53	4.39
$\nu_2 = 7$	5.59	4.74	4.35	4.12	3.97
$\nu_2 = 8$	5.32	4.46	4.07	3.84	3.69

- Some upper quantiles of some chi-squared distributions:

ξ	0.10	0.05	0.025	0.01	0.005
$\chi^2_{1,\xi}$	2.71	3.84	5.02	6.63	7.88
$\chi^2_{2,\xi}$	4.61	5.99	7.38	9.21	10.60

1. Suppose $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are independent random samples from Poisson distributions with means λ_1 and λ_2, respectively, and suppose you wish to test the hypotheses

$$H_0\colon\ \lambda_1 = \lambda_2 \text{ versus } H_1\colon\ \lambda_1 \neq \lambda_2.$$

(a) Write down the likelihood function $L(\lambda_1, \lambda_2; X_1, \ldots, X_n, Y_1, \ldots, Y_m)$.

Solution: The likelihood function is

$$\begin{aligned} L(\lambda_1, \lambda_2; X_1, \ldots, X_n, Y_1, \ldots, Y_m) &= \left(\prod_{i=1}^{n} \frac{e^{-\lambda_1}\lambda_1^{X_i}}{X_i!}\right)\left(\prod_{j=1}^{m} \frac{e^{-\lambda_2}\lambda_2^{Y_j}}{Y_j!}\right) \\ &= e^{-n\lambda_1}\lambda_1^{n\bar{X}_n} e^{-m\lambda_2}\lambda_2^{m\bar{Y}_m}\left(\prod_{i=1}^{n} X_i! \prod_{j=1}^{m} Y_j!\right)^{-1}. \end{aligned}$$

(b) Write down the log-likelihood function $\ell(\lambda_1, \lambda_2; X_1, \ldots, X_n, Y_1, \ldots, Y_m)$.

Solution: The log-likelihood function is

$$\begin{aligned} \ell(\lambda_1, \lambda_2; X_1, \ldots, X_n, Y_1, \ldots, Y_m) &= -n\lambda_1 - m\lambda_2 + n\bar{X}_n \log(\lambda_1) + m\bar{Y}_m \log(\lambda_2) \\ &\quad - \sum_{i=1}^{n} \log(X_i!) - \sum_{j=1}^{m} \log(Y_j!) \end{aligned}$$

(c) Find the maximum likelihood estimators of λ_1 and λ_2.

Solution: We get $\hat{\lambda}_1 = \bar{X}_n$ and $\hat{\lambda}_2 = \bar{Y}_m$.

(d) Under H_0 we have $\lambda_1 = \lambda_2 = \lambda_0$. Find the value of λ which maximizes $L(\lambda, \lambda; X_1, \ldots, X_n, Y_1, \ldots, Y_m)$.

Solution: We get

$$\hat{\lambda}_0 = \frac{n\bar{X}_n + m\bar{Y}_m}{n+m}.$$

(e) Show that the likelihood ratio can be written

$$\mathrm{LR}(X_1, \ldots, X_n, Y_1, \ldots, Y_m) = \frac{\left(\dfrac{n\bar{X}_n + m\bar{Y}_m}{n+m}\right)^{n\bar{X}_n + m\bar{Y}_m}}{(\bar{X}_n)^{n\bar{X}_n}(\bar{Y}_m)^{m\bar{Y}_m}}$$

Solution: The likelihood ratio is given by

$$\begin{aligned}
\mathrm{LR}(X_1,\dots,X_n,Y_1,\dots,Y_m) &= \frac{L(\hat{\lambda}_0,\hat{\lambda}_0;X_1,\dots,X_n,Y_1,\dots,Y_m)}{L(\hat{\lambda}_1,\hat{\lambda}_2;X_1,\dots,X_n,Y_1,\dots,Y_m)}\\
&= \frac{e^{-n\hat{\lambda}_0}\hat{\lambda}_0^{n\bar{X}_n}e^{-m\hat{\lambda}_0}\hat{\lambda}_0^{m\bar{Y}_m}\left(\prod_{i=1}^n X_i!\prod_{j=1}^m Y_j\right)^{-1}}{e^{-n\hat{\lambda}_1}\hat{\lambda}_1^{n\bar{X}_n}e^{-m\hat{\lambda}_2}\hat{\lambda}_2^{m\bar{Y}_m}\left(\prod_{i=1}^n X_i!\prod_{j=1}^m Y_j\right)^{-1}}\\
&= \frac{e^{-(n+m)\hat{\lambda}_0}\hat{\lambda}_0^{n\bar{X}_n+m\bar{Y}_m}}{e^{-n\hat{\lambda}_1}\hat{\lambda}_1^{n\bar{X}_n}e^{-m\hat{\lambda}_2}\hat{\lambda}_2^{m\bar{Y}_m}}\\
&= \frac{\hat{\lambda}_0^{n\bar{X}_n+m\bar{Y}_m}}{\hat{\lambda}_1^{n\bar{X}_n}\hat{\lambda}_2^{m\bar{Y}_m}}.
\end{aligned}$$

Substituting the expressions for $\hat{\lambda}_1$, $\hat{\lambda}_2$, and $\hat{\lambda}_0$ from parts (c) and (d) gives the result.

(f) Suppose for the sample sizes $n = 20$ and $m = 25$ we observe $\bar{X}_n = 3.4$, and $\bar{Y}_m = 4.56$. Compute the test statistic for the asymptotic likelihood ratio test given by $-2\log\mathrm{LR}(X_1,\dots,X_n,Y_1,\dots,Y_m)$.

Solution: We have

$$\begin{aligned}
&-2\log\mathrm{LR}(X_1,\dots,X_n,Y_1,\dots,Y_m)\\
&\qquad= -2\left[(n\bar{X}_n+m\bar{Y}_m)\log\left(\frac{n\bar{X}_m+m\bar{Y}_m}{n+m}\right) - n\bar{X}_n\log(\bar{X}_n) - m\bar{Y}_m\log(\bar{Y}_m)\right],
\end{aligned}$$

which for $n = 20$, $m = 25$, $\bar{X}_n = 3.4$, and $\bar{Y}_m = 4.56$ is equal to 3.749729.

(g) Give the critical value for the asymptotic likelihood ratio test at the 0.05 significance level.

Solution: Under the null hypotheses

$$-2\log\mathrm{LR}(X_1,\dots,X_n,Y_1,\dots,Y_m) \to \chi^2_1 \quad \text{in distribution}$$

as $n\to\infty$, where the degrees of freedom of the limiting chi-squared distribution is equal to 1 because the null hypothesis restricts the parameter space such that it is one-dimensional (only one parameter is undetermined—the common λ_0). The 0.05 critical value for the asymptotic likelihood ratio test is the upper 0.05 quantile of this distribution, which is $\chi^2_{1,0.05} = 3.84$.

(h) Give your conclusion concerning the hypotheses at the $\alpha = 0.05$ significance level based on the data from part (f).

Solution: Since the test statistic does not exceed the 0.05 critical value, we fail to reject the null hypothesis at the 0.05 significance level.

2. Let $Y_1, \ldots, Y_{10}$ be random variables such that

$$Y_i = 1 + (1/2)x_i + \varepsilon_i, \text{ for } i = 1, \ldots, 10,$$

where $x_1 = 1$, $x_2 = 2$, $x_3 = 3$, $\ldots$, and $x_{10} = 10$ and $\varepsilon_1, \ldots, \varepsilon_{10}$ are independent Normal$(0, \sigma^2 = 4)$ random variables. In addition, let

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\text{argmin}} \sum_{i=1}^{n} [Y_i - (\beta_0 + \beta_1 x_i)]^2.$$

(a) Compute $\bar{x} = (1/10)\sum_{i=1}^{10} x_i$ and quantity $S_{xx} = \sum_{i=1}^{10}(x_i - \bar{x})^2$.

Solution: We get

$$\bar{x} = (1/10)10(11)/2 = 5.5$$
$$S_{xx} = \sum_{i=1}^{10}(x_i - \bar{x})^2 = \sum_{i=1}^{10} x_i^2 - 10^2 = 10(11)(21)/6 - 10(5.5)^2 = 82.5$$

(b) Give the following (each answer is a number):

i. $\mathbb{E}\hat{\beta}_0$
ii. $\mathbb{E}\hat{\beta}_1$
iii. $\text{Var}\, \hat{\beta}_0$
iv. $\text{Var}\, \hat{\beta}_1$
v. $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)$
vi. $\mathbb{E}(1/8)\sum_{i=1}^{10}[Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$

Solution:

i. $\mathbb{E}\hat{\beta}_0 = 1$

ii. $\mathbb{E}\hat{\beta}_1 = 1/2$

iii. $\text{Var}\, \hat{\beta}_0 = (1/10 + (5.5)^2/82.5)4 = 1.866667$

iv. $\text{Var}\, \hat{\beta}_1 = 4/82.5 = 0.04848485$

v. $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -(5.5)/82.5(4) = -0.2666667$

vi. $\mathbb{E}(1/8)\sum_{i=1}^{10}[Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 = 4$

(c) Give the matrix $\mathbf{X}$ such that

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y},$$

where $\mathbf{Y} = (Y_1, \ldots, Y_{10})^T$.

Solution: Use

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \\ 1 & 8 \\ 1 & 9 \\ 1 & 10 \end{bmatrix}$$

(d) Which of the following is true? *Hint: Begin by writing down the distribution of* $\hat{\beta}_1$.

A. $0 < P(\hat{\beta}_1 < 0) < 0.01$

B. $0.01 < P(\hat{\beta}_1 < 0) < 0.025$

C. $0.025 < P(\hat{\beta}_1 < 0) < 0.05$

D. $0.05 < P(\hat{\beta}_1 < 0) < 0.10$

E. $0.10 < P(\hat{\beta}_1 < 0) < 1$

Solution: We have $\hat{\beta}_1 \sim \text{Normal}(1/2, 4/82.5)$, so

$$\begin{aligned} P(\hat{\beta}_1 < 0) &= P((\hat{\beta}_1 - 1/2)/\sqrt{4/82.5} < -(1/2)/\sqrt{4/82.5}) \\ &= \Phi(-(1/2)/\sqrt{4/82.5}) \\ &= \Phi(-2.270738). \end{aligned}$$

Since $-2.576 < -2.270738 < -1.96$, we have $0.01 < P(\hat{\beta}_1 < 0) < 0.025$.

(e) Using the true regression function, give an interval within which the random variable Y_3 will fall with probability 0.95.

Solution: The random variable Y_3 has the $\text{Normal}(1 + (1/2)3, \sigma^2 = 4)$ distribution, so Y_3 will fall in the interval $2.5 \pm 1.96(2) = (-1.42, 6.42)$ with probability 0.95.

(f) Suppose you don't know the true relationship between $Y_1, \dots, Y_{10}$ and $x_1, \dots, x_{10}$, but you believe that

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \text{ for } i = 1, \dots, 10,$$

for some β_0 and β_1 and you wish to test H_0: $\beta_1 = 0$ versus H_1: $\beta_1 \neq 0$ at the $\alpha = 0.01$ significance level.

i. Suppose you obtain $\hat{\beta}_1 = 0.4$ and $(1/8)\sum_{i=1}^{10}[Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 = 3$. What is your conclusion?

ii. Sketch a power curve for this test, where the power is a function of β_1. Draw a horizontal line at α and a vertical line at the null value of β_1.

Solution:

i. The test statistic is

$$\frac{\hat{\beta}_1}{\hat{\sigma}/\sqrt{S_{xx}}} = (0.4)/\sqrt{3/82.5} = 2.097618.$$

The upper $0.01/2 = 0.005$ quantile of the t_8 distribution is 3.355, so we fail to reject the null hypothesis.

ii. Sketch should have an inverted bell shape with a minimum value of 0.01 occurring at $\beta_1 = 0$.

3. Let $Y_1, \ldots, Y_{10}$ be random variables such that

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 \mathbf{1}(x_i \geq 5) + \beta_3(x_i \cdot \mathbf{1}(x_i \geq 5)) + \varepsilon_i, \text{ for } i = 1, \ldots, 10,$$

for some $\beta_0, \beta_1, \beta_2, \beta_3$, where $x_1 = 1$, $x_2 = 2$, $x_3 = 3$, $\ldots$, and $x_{10} = 10$ and $\varepsilon_1, \ldots, \varepsilon_{10}$ are independent Normal$(0, \sigma^2 = 4)$ random variables. The function $\mathbf{1}(x_i \geq 5)$ is equal to 1 if $x_i \geq 5$ and equal to 0 otherwise. In addition, let $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)^T$, where

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = \underset{\beta_0,\beta_1,\beta_2,\beta_3}{\operatorname{argmin}} \sum_{i=1}^{n} [Y_i - (\beta_0 + \beta_1 x_i + \beta_2 \mathbf{1}(x_i \geq 5) + \beta_3(x_i \cdot \mathbf{1}(x_i \geq 5)))]^2.$$

(a) Give the design matrix $\mathbf{X}$ such that

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y},$$

where $\mathbf{Y} = (Y_1, \ldots, Y_{10})^T$.

Solution: Use

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ 1 & 5 & 1 & 5 \\ 1 & 6 & 1 & 6 \\ 1 & 7 & 1 & 7 \\ 1 & 8 & 1 & 8 \\ 1 & 9 & 1 & 9 \\ 1 & 10 & 1 & 10 \end{bmatrix}$$

(b) Give the vector $\mathbf{a}$ such that $\mathbb{E}\mathbf{a}^T\hat{\boldsymbol{\beta}} = \beta_3$.

Solution: Use the vector $\mathbf{a} = [0, 0, 0, 1]^T$.

(c) Four different models were fit to the data $(x_1, Y_1), \ldots, (x_{10}, Y_{10})$, resulting in four different sums of squared residuals equal to 29.35, 52.23, 29.13, and 34.86.

Model fit to the data	Sum of squared residuals
$Y_i = \beta_0 + \varepsilon_i$	
$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$	
$Y_i = \beta_0 + \beta_1 x_i + \beta_2 \mathbf{1}(x_i \geq 5) + \varepsilon_i$	
$Y_i = \beta_0 + \beta_1 x_i + \beta_2 \mathbf{1}(x_i \geq 5) + \beta_3 (x_i \cdot \mathbf{1}(x_i \geq 5)) + \varepsilon_i$	

Place each of the numbers 29.35, 52.23, 29.13, and 34.86 in the correct row of the table above.

Solution: Numbers should be placed in decreasing order.

(d) Compute the test statistic for the full-reduced model F-test for testing

$$H_0\colon\ \beta_2 = \beta_3 = 0 \text{ versus } H_1\colon\ \beta_j \neq 0 \text{ for some } j \in \{2, 3\}.$$

Solution: The test statistic is

$$\frac{(34.86 - 29.13)/(3 - 1)}{29.13/(10 - 3 - 1)} = 0.59.$$

(e) State your conclusion about the hypotheses in part (d) at the $\alpha = 0.05$ significance level.

Solution: We compare 0.59 to the critical value $F_{2,6,0.05} = 5.14$. Since $0.59 < 5.14$ we fail to reject the null hypothesis.

(f) Compute the test statistic for the overall F-test of significance for testing

$$H_0\colon\ \beta_1 = \beta_2 = \beta_3 = 0 \text{ versus } H_1\colon\ \beta_j \neq 0 \text{ for some } j \in \{1, 2, 3\}.$$

Solution: The test statistic is

$$\frac{(52.23 - 29.13)/3}{29.13/(10 - 3 - 1)} = 1.585994.$$

(g) State your conclusion about the hypotheses in part (f) at the $\alpha = 0.05$ significance level.

Solution: We compare 1.585994 to the critical value $F_{3,6,0.05} = 4.76$. Since $1.585994 < 4.76$ we fail to reject the null hypothesis.