STAT 513 fa 2020 Exam II

Karl B. Gregory

This is a take-home test due to COVID-19. Do not communicate with classmates about the exam until after its due date/time. You may

- *Use your notes and the lecture notes.*
- *Use books.*
- *NOT work together with others.*

Write all answers on blank sheets of paper; then take pictures and merge to a PDF. Upload a single PDF to Blackboard.

1. Copy down this sentence on your answer sheet and put your signature underneath: *I have not collaborated with any other student on this exam. The work I have presented is my own.*

2. Let $Y_1, \ldots, Y_n \overset{\text{ind}}{\sim} \text{Gamma}(\alpha, \beta)$, where α is <u>known</u>, and consider

$$H_0\colon\ \beta \leq \beta_0 \text{ versus } H_1\colon\ \beta > \beta_0.$$

   (a) Give the likelihood function $\mathcal{L}(\beta; Y_1, \ldots, Y_n)$.

   We have
   $$\mathcal{L}(\beta; Y_1, \ldots, Y_n) = (\Gamma(\alpha)\beta^\alpha)^{-n} \left(\prod_{i=1}^n Y_i\right)^{\alpha-1} \exp\left(-\frac{n\bar{Y}_n}{\beta}\right).$$

   (b) Give the log-likelihood function $\ell(\beta; Y_1, \ldots, Y_n)$.

   We have
   $$\ell(\beta; Y_1, \ldots, Y_n) = -n \log \Gamma(\alpha) - n\alpha \log \beta + (\alpha - 1) \sum_{i=1}^n \log Y_i - \frac{n\bar{Y}_n}{\beta}.$$

   (c) Find the maximum likelihood estimator $\hat{\beta}$ of β.

   Differentiating the log-likelihood with respect to β and setting this equal to zero gives
   $$\hat{\beta} = \frac{\bar{Y}_n}{\alpha}.$$

   (d) Find the restricted maximum likelihood estimator $\hat{\beta}_0$ of β defined as
   $$\hat{\beta}_0 = \underset{0 < \beta \leq \beta_0}{\text{argmax}}\ \mathcal{L}(\beta; Y_1, \ldots, Y_n).$$

   We have
   $$\hat{\beta}_0 = \begin{cases} \hat{\beta}, & \hat{\beta} \leq \beta_0 \\ \beta_0, & \hat{\beta} > \beta_0 \end{cases}$$

   (e) Give the likelihood ratio $\text{LR}(Y_1, \ldots, Y_n)$.

   We have
   $$\text{LR}(Y_1, \ldots, Y_n) = \begin{cases} 1, & \hat{\beta} \leq \beta_0 \\ \left[\left(\frac{\hat{\beta}}{\beta_0}\right) \exp\left(-\frac{\hat{\beta}}{\beta_0}\right)\right]^{\alpha n} \exp(n\alpha), & \hat{\beta} > \beta_0 \end{cases}$$

(f) State whether the likelihood ratio is monotone increasing or decreasing in $\hat{\beta}$ when $\hat{\beta} > \beta_0$.

It is monotone decreasing in $\hat{\beta}$ for $\hat{\beta} > \beta_0$.

(g) Use mgfs to find the distribution of $\hat{\beta}$.

We have $\hat{\beta} = \bar{Y}_n/\alpha$, which has mgf given by

$$\begin{aligned} M_{\hat{\beta}}(t) &= M_{\bar{Y}_n/\alpha}(t) \\ &= M_{Y_1+\cdots+Y_n}(t/(n\alpha)) \\ &= [M_{Y_1}(t/(n\alpha))]^n \\ &= [(1-\beta t/(n\alpha))^{-\alpha}]^n \\ &= (1-(\beta/(n\alpha))t)^{-\alpha n}. \end{aligned}$$

So we have

$$\hat{\beta} \sim \text{Gamma}(\alpha n, \beta/(\alpha n)).$$

(h) Give a test that is equivalent to the likelihood ratio test of size 0.05.

The test

$$\text{Reject } H_0 \text{ iff } \hat{\beta} > G_{\alpha n,\beta/(\alpha n),0.05},$$

where $G_{\alpha n,\beta/(\alpha n),0.05}$ is the upper 0.05 quantile of the Gamma$(\alpha n, \beta/(\alpha n))$ distribution, is equivalent to the LRT with size 0.05.

(i) Give the rejection criterion for the *asymptotic* likelihood ratio test of size 0.05 for testing

$$H_0\colon\ \beta = \beta_0 \text{ versus } H_1\colon\ \beta \neq \beta_0.$$

We have

$$-2\log \text{LR}(Y_1,\ldots,Y_n) = -2\alpha n\left[\log(\hat{\beta}/\beta_0) - \hat{\beta}/\beta_0 + 1\right].$$

The asymptotic likelihood ratio test with size 0.05 rejects H_0 if and only if this quantity exceeds $\chi^2_{0.05} = 3.841459$.

(j) Let $\alpha = 2/3$, $n = 10$, $\bar{Y}_n = 1$, and $\beta_0 = 1$.

i. Give the p-value of the asymptotic likelihood ratio test from part (i).

We have $\hat{\beta} = 3/2$, and

$$-2\log \text{LR}(Y_1,\ldots,Y_n) = 1.260465,$$

so the p-value is $P(W > 1.260465)$, $W \sim \chi^2_1$. This is

`1 - pchisq(1.260465,1)` $= 0.261563$.

ii. Give the p-value of the likelihood ratio test from part (h).

The p-value is given by

`1 - pgamma(3/2,shape = 6.6667,scale = 0.15)` $= 0.1061$.

3. Let $Y_i = 10 + \varepsilon_i$ for $i = 1, \dots, 11$, where $\varepsilon_1, \dots, \varepsilon_{11} \stackrel{\text{ind}}{\sim} \text{Normal}(0, 1)$, and let

$$(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}) = (-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5).$$

Simple linear regression is to be carried out on the data pairs $(x_1, Y_1), \dots, (x_{11}, Y_{11})$, with

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

(a) Give the following:

i. $\mathbb{E}\hat{\beta}_1$.

We have $\mathbb{E}\hat{\beta}_1 = 0$.

ii. $\text{Var}\,\hat{\beta}_1$.

We have $S_{xx} = \sum_{i=1}^{11}(x_i - \bar{x}_n)^2 = 110$ and $\sigma^2 = 1$, so $\text{Var}\,\hat{\beta}_1 = \sigma^2/S_{xx} = 1/110 = 0.00909$.

iii. $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)$.

Since $\bar{X}_n = 0$, $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -\bar{x}_n S_{xx}^{-1}\sigma^2 = 0$.

iv. $\text{Var}(\hat{\beta}_0 + 3\hat{\beta}_1)$.

We have

$$\text{Var}(\hat{\beta}_0 + 3\hat{\beta}_1) = \left[\frac{1}{11} + \frac{3^2}{110}\right] = \frac{19}{110} = 0.1727273.$$

v. $P(-1/5 < \hat{\beta}_1 < 1/5)$.

We have

$$\begin{aligned} P(-1/5 < \hat{\beta}_1 < 1/5) &= P\left(-\frac{S_{xx}}{5\sigma} < \frac{\hat{\beta}_1 - 0}{\sigma/\sqrt{S_{xx}}} < \frac{\sqrt{S_{xx}}}{5\sigma}\right) \\ &= P\left(-\sqrt{110}/5) < Z < \sqrt{110}/5\right) \\ &= P(-2.097618 < Z < 2.097618), \quad Z \sim \text{Normal}(0, 1) \\ &= 0.9640611. \end{aligned}$$

(b) Suppose we observe $\sum_{i=1}^{11} \hat{\varepsilon}_i^2 = 14.29$, $\bar{Y}_n = 10.32$, $S_{YY} = 14.47$, and $r_{xY} = 0.111$. Based on this data, give

i. The value of $\hat{\beta}_1$.

We have
$$\hat{\beta}_1 = 0.111\sqrt{14.47/110} = 0.04025881.$$

ii. The value of $\hat{\beta}_0$.

We have
$$\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{x}_n = \bar{Y}_n = 10.32.$$

iii. A 95% confidence interval for β_1.

We have $\hat{\sigma}^2 = (1/9)\sum_{i=1}^{11} \hat{\varepsilon}_i^2 = 1.588$.
$$0.04025881 \pm \underbrace{t_{n-2,0.05/2}}_{2.262157}\sqrt{1.588}/\sqrt{110} = (-0.2316006, 0.3119642).$$

iv. A 95% confidence interval for the height of the regression function at $x_{\text{new}} = 1/2$.

$$10.32091 + (1/2)0.0403 \pm \underbrace{t_{n-2,0.05/2}}_{2.262157}\sqrt{1.588}\sqrt{1/11 + (1/2)^2/110} = (9.470872, 11.21113).$$

v. A 95% prediction interval for a new value Y_{new} of the response with $x_{\text{new}} = 1/2$.

$$10.32091 + (1/2)0.0403 \pm \underbrace{t_{n-2,0.05/2}}_{2.262157}\sqrt{1.588}\sqrt{1 + 1/11 + (1/2)^2/110} = (7.360674, 13.32133).$$

4. Suppose 12 individuals will be randomly assigned to a placebo group or to one of two treatment groups, such that 4 individuals will receive the placebo, 4 will receive treatment one, and 4 will receive treatment two. Clinical response values $Y_1, \ldots, Y_{12}$ will be recorded and the multiple linear regression model

$$Y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i, \quad i = 1, \ldots, 12,$$

where $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma^2)$, and

$$(x_{1i}, x_{2i}, x_{31}) = \begin{cases} (1,0,0), & \text{if individual } i \text{ in placebo group} \\ (0,1,0), & \text{if individual } i \text{ in treatment one group} \\ (0,0,1), & \text{if individual } i \text{ in treatment two group,} \end{cases}$$

will be fit to the data (note that the model does not have an intercept). Let $\bar{Y}_0$, $\bar{Y}_1$ and $\bar{Y}_2$ be the mean responses observed in the placebo, treatment one, and treatment two groups, respectively, and suppose the design matrix $\mathbf{X}$ of the study is given by

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

(a) Give $(\mathbf{X}^T\mathbf{X})^{-1}$.

We have

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/4 \end{bmatrix}.$$

(b) Give $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$.

We have

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/4 \end{bmatrix} \begin{bmatrix} Y_1 + Y_2 + Y_3 + Y_4 \\ Y_5 + Y_6 + Y_7 + Y_8 \\ Y_9 + Y_{10} + Y_{11} + Y_{12} \end{bmatrix} = \begin{bmatrix} \bar{Y}_0 \\ \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix}$$

(c) Consider the hypotheses

$$H_0\colon\ \beta_2 = \beta_3 \text{ versus } H_1\colon\ \beta_2 \neq \beta_3.$$

i. Give an interpretation of the null hypothesis H_0 in words in the context of the problem.

The null hypothesis states that the mean clinical response of individuals receiving treatment 1 is no different than that of individuals receiving treatment 2.

ii. What is the degrees of freedom of the chi-squared distribution from which we would find the critical value for the asymptotic likelihood ratio test?

The dimension of the entire parameter space is $d = 4$, since we have $\theta = (\beta_1, \beta_2, \beta_3, \sigma^2) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}\times[0,\infty)$. The dimension of the null space is $d_0 = 3$, since β_2 and β_3 are constrained to be the same. The degrees of freedom of the relevant chi-squared distribution is therefore $4 - 3 = 1$.