STAT 513 fa 2019 Exam II

Karl B. Gregory

Do not open this test until told to do so; no calculators allowed; no notes allowed; no books allowed; show your work so that partial credit may be given.

Some upper quantiles of some t-distributions:

ξ	0.10	0.05	0.025	0.01	0.005
$t_{28,\xi}$	1.3125	1.7011	2.0484	2.4671	2.7633
$t_{29,\xi}$	1.3114	1.6991	2.0452	2.4620	2.7564
$t_{30,\xi}$	1.3104	1.6973	2.0423	2.4573	2.7500
$t_{31,\xi}$	1.3095	1.6955	2.0395	2.4528	2.7440

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

For any $\mathbf{a} \in \mathbb{R}^{p+1}$, a $(1-\alpha)100\%$ confidence interval for $\mathbf{a}^T\hat{\boldsymbol{\beta}}$ is given by

$$\mathbf{a}^T\hat{\boldsymbol{\beta}} \pm t_{n-p-1,\alpha/2}\hat{\sigma}\sqrt{\mathbf{a}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{a}},$$

and the t-test statistic for comparing $\mathbf{a}^T\boldsymbol{\beta}$ to a null value a^* is given by

$$T_n = \frac{\mathbf{a}^T\hat{\boldsymbol{\beta}} - a^*}{\hat{\sigma}\sqrt{\mathbf{a}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{a}}}.$$

Let $X_1, \dots, X_n$ be a random sample with likelihood function $L(\theta; X_1, \dots, X_n)$. Then for hypotheses of the form H_0: $\theta = \theta_0$ versus H_1: $\theta \neq \theta_0$ the likelihood ratio is

$$\text{LR}(X_1, \dots, X_n) = \frac{L(\theta_0; X_1, \dots, X_n)}{L(\hat{\theta}; X_1, \dots, X_n)},$$

where $\hat{\theta}$ is the MLE of θ.

1. Each of n randomly selected subjects is given m opportunities to predict the outcome of a coin toss. Let p denote the probability that a subject makes the correct prediction. The numbers of correct predictions $Y_1, \ldots, Y_n$ made by the subjects are assumed to be independent Binomial(m, p) random variables. Researchers from L'Institut de la Clairvoyance wish to test the hypotheseses

$$H_0\colon\ p = 1/2 \text{ versus } H_1\colon\ p \neq 1/2.$$

The likelihood function is given by

$$L(p; Y_1, \ldots, Y_n) = \prod_{i=1}^{n} \binom{m}{Y_i} p^{Y_i} (1-p)^{m-Y_i}.$$

(a) Write down the log-likelihood function $\ell(p; Y_1, \ldots, Y_n)$. It will be convenient to use $\sum_{i=1}^{n} Y_i = n\bar{Y}_n$.

Solution: The log-likelihood function is

$$\ell(p; Y_1, \ldots, Y_n) = \sum_{i=1}^{n} \log \binom{m}{Y_i} + n\bar{Y}_n \log p + (nm - n\bar{Y}_n) \log(1-p).$$

(b) Find the maximum likelihood estimator of p.

Solution: Setting the partial derivative of $\ell(p; Y_1, \ldots, Y_n)$ with respect to p equal to zero and solving for p gives

$$\hat{p} = \bar{Y}_n/m.$$

(c) Write down the likelihood ratio $\mathrm{LR}(Y_1, \ldots, Y_n)$ and simplify it as much as seems reasonable.

Solution: We have

$$\begin{aligned}
\mathrm{LR}(Y_1, \ldots, Y_n) &= \frac{\prod_{i=1}^{n} \binom{m}{Y_i} (1/2)^{Y_i} (1-1/2)^{m-Y_i}}{\prod_{i=1}^{n} \binom{m}{Y_i} (\bar{Y}_n/m)^{Y_i} (1-\bar{Y}/m)^{m-Y_i}} \\
&= \prod_{i=1}^{n} \left[\frac{1/2}{\bar{Y}_n/m}\right]^{Y_i} \left[\frac{1-1/2}{1-\bar{Y}_n/m}\right]^{m-Y_i} \\
&= \left[\frac{1/2}{\bar{Y}_n/m}\right]^{n\bar{Y}_n} \left[\frac{1-1/2}{1-\bar{Y}_n/m}\right]^{nm-n\bar{Y}_n}.
\end{aligned}$$

(d) Give the form of the likelihood ratio test for the hypotheses of interest.

Solution: The LRT rejects H_0 if and only if

$$\left[\frac{1/2}{\bar{Y}_n/m}\right]^{n\bar{Y}_n}\left[\frac{1-1/2}{1-\bar{Y}_n/m}\right]^{nm-n\bar{Y}_n} < c$$

for some $c \in [0,1]$.

(e) Get an expression for $-2\log \mathrm{LR}(Y_1,\dots,Y_n)$ and state the rejection rule of the asymptotic likelihood ratio test of size $\alpha = 0.05$.

Solution: The asymptotic likelihood ratio test is based on the quantity

$$-2\log \mathrm{LR}(Y_1,\dots,Y_n) = -2\left[n\bar{Y}_n \log\left(\frac{1/2}{\bar{Y}_n/m}\right) + (nm - n\bar{Y}_n)\log\left(\frac{1-1/2}{1-\bar{Y}_n/m}\right)\right].$$

The size-0.05 asymptotic LRT rejects H_0 when this quantity exceeds $\chi^2_{1,0.05} = 3.84$.

2. Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

for the data $(Y_1,x_1),\dots,(Y_n,x_n)$, where $\varepsilon_1,\dots,\varepsilon_n$ are independent Normal$(0,\sigma^2)$ error terms and $x_1,\dots,x_n$ are fixed constants. Suppose we will reject H_0: $\beta_1 = 0$ in favor of H_1: $\beta_1 \neq 0$ if and only if

$$\frac{|\hat{\beta}_1|}{\hat{\sigma}\sqrt{S_{xx}^{-1}}} > t_{n-2,\alpha/2},$$

where $\hat{\beta}_1$ is the least-squares estimator of β_1 and $S_{xx} = \sum_{i=1}^n (x_i - \bar{x}_n)^2$. Fill in the blanks.

(a) A __**smaller**__ (larger/smaller) error term variance leads to __**larger**__ (larger/smaller) power of the test when $\beta_1 \neq 0$.

(b) Values of $x_1,\dots,x_n$ with a __**larger**__ (larger/smaller) spread lead to __**larger**__ (larger/smaller) power of the test when $\beta_1 \neq 0$.

(c) Values of β_1 of __**larger**__ (larger/smaller) magnitude lead to __**larger**__ (larger/smaller) power of the test.

(d) A __**smaller**__ (larger/smaller) value of α leads to __**smaller**__ (larger/smaller) power of the test when $\beta_1 \neq 0$.

3. Suppose $\mathrm{Var}(\hat{\beta}_0) = 2$, $\mathrm{Var}(\hat{\beta}_1) = 1$, and $\mathrm{Cov}(\hat{\beta}_0,\hat{\beta}_1) = -1/2$. Then give the variance of $\hat{Y}_{\text{new}} = \hat{\beta}_0 + x_{\text{new}}\hat{\beta}_1$ for $x_{\text{new}} = 3$.

Solution: We have

$$\mathrm{Var}\,\hat{Y}_{\text{new}} = \mathrm{Var}(\hat{\beta}_0 + x_{\text{new}}\hat{\beta}_1) = \mathrm{Var}\,\hat{\beta}_0 + x_{\text{new}}^2\,\mathrm{Var}(\hat{\beta}_1) + 2x_{\text{new}}\,\mathrm{Cov}(\hat{\beta}_0,\hat{\beta}_1).$$

For $x_{\text{new}} = 3$, the above is equal to $2 + 3^2(1) + 2(3)(-1/2) = 8$.

4. Consider the multiple linear regression model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

for $i = 1, \dots, 31$, where $Y_1, \dots, Y_n$ are volumes of timber harvested from trees, $x_{11}, \dots, x_{n1}$ are the girths of the trees, and $x_{12}, \dots, x_{n2}$ are the heights of the trees, and $\varepsilon_1, \dots, \varepsilon_n$ are assumed to be independent random variables with the Normal$(0, \sigma^2)$ distribution. Let $\mathbf{X}$ be the corresponding design matrix and suppose that

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 4.9519 & 0.0287 & -0.0697 \\ 0.0287 & 0.0046 & -0.0012 \\ -0.0697 & -0.0012 & 0.0011 \end{bmatrix}, \quad \hat{\boldsymbol{\beta}} = \begin{bmatrix} -57.9877 \\ 4.7082 \\ 0.3393 \end{bmatrix}, \quad \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|_2^2 = 421.9214.$$

In addition $\sum_{i=1}^n (Y_i - \bar{Y}_n)^2 = 8106.084$. *For the following, you do not need to evaluate your answers.*

(a) Give the value of $\hat{\sigma}^2$.

> **Solution:** We have
>
> $$\hat{\sigma}^2 = 421.9214/28.$$

(b) Give a 95% confidence interval for β_1.

> **Solution:**
>
> $$4.7082 \pm 2.0484\sqrt{421.9214/28}\sqrt{0.0046}.$$

(c) Suppose you wish to test H_0: $\beta_2 = 0$ versus H_1: $\beta_2 \neq 0$ using the t-test.

i. Give the test statistic.

> **Solution:** The test statistic is
>
> $$T = \frac{0.3393}{\sqrt{421.9214/28}\sqrt{0.0011}}.$$

ii. Give the rejection rule (give the critical value) under which the test will have size $\alpha = 0.01$.

> **Solution:** We would reject H_0 if and only if the absolute value of the test statistic were to exceed $t_{28,0.005} = 2.7633$.

(d) i. State the null and alternate hypotheses which would be tested by the overall F-test of significance for this model.

> **Solution:** The overall F-test of significance would test
>
> H_0: $\beta_1 = \beta_2 = 0$ versus H_1: Either β_1 or β_2 or both are not equal to zero.

ii. Give the value of the test statistic for the overall F-test of significance,

$$F = \frac{\text{MSM}}{\text{MSE}} = \frac{\text{SSM}\,/p}{\text{SSE}\,/(n-p-1)}.$$

Solution: We obtain SSM as SST − SSE = 8106.084 − 421.9214. So we have

$$F = \frac{(8106.084 - 421.9214)/2}{421.9214/(31-2-1)}.$$

iii. Explain how you would obtain a p-value for the overall F-test using the statistic in the previous part. You must give the numerator and denominator degrees of freedom of the relevant F-distribution.

Solution: The p-value is the area under the pdf of the $F_{2,28}$ distribution to the right of the value of the F-statistic.