MAST90105: Lab and Workshop Problems for Week 10

The Lab and Workshop this week covers problems arising Module 6, Sections 3 to the end.

1 Lab

1. (Elaborating Textbook 6.5-10) The "golden" ratio is $\phi = (1+\sqrt{5})/2$. If a, b are numbers, show that they have the "golden" ratio if $\frac{a}{b} = \frac{a+b}{a}$.

   John Putz, a mathematician who was interested in music, analyzed 29 Mozart piano sonata movements which could easily be divided into 2 distinct sections, the Exposition (in which the first and second subjects, the melodies that underly the movement, are revealed) and the Development/Recapitulation (in which the first and second subjects are developed and then restated). Mozart showed interest in mathematics and Putz wondered whether the numbers of bars in the Exposition, b and Development/Recapitulation, a followed the golden ratio (the Recapitulation is often of similar length to the Exposition in Sonata movements from the "classical" period, so that the Development and Recapitulation are always longer than the Exposition).

   The data on the Mozart piano sonata movements is in the lab folder as "Mozart.xls", with a values in column 2 and b values in column 1. Import this data into R using the `read_excel()` function. You will need to install `readxl` library.

   a. Make a scatter plot of the points $a+b$ against the points a. Is this plot linear?
   b. Find the equation of the least squares regression line with and without intercept. Superimpose them on the scatter plot.
   c. On the scatter plot, superimpose the line $y = \phi x$. Compare this line with the least squares regression line.
   d. Find the sample mean of the points $(a+b)/a$. Is the mean close to ϕ?
   e. Now consider the same questions using the data on a and b. Compare and contrast your results and explain any differences.
   f. Consider the residuals from the linear models versus the response values as well as the differences between the values from $y = \phi x$ and the response values. In each linear model case, plot the residuals and the difference values on the same plot. Comment on systematic differences
   g. Do you think Mozart wrote his music thinking about the number of bars in the Development and Recapitulation being the number in the Exposition times the golden ratio? Why?

用各种 statistics 前要假定 sample data from normal distribution. $N(\mu, \sigma^2)$

$\frac{\bar{X}-\mu}{} \sim N(0,1)$

$\frac{\sigma}{\sqrt{n}}$

2 Workshop

$\bar{X}$ $\theta \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$

2. Let $X_1, \ldots, X_n$ be a random sample from a gamma distribution with $\alpha = 4$ so that

Gamma分布如何估计?

$$f(x;\theta) = \frac{1}{6\theta^4} x^3 e^{-x/\theta}, \ 0 < x < \infty, \ 0 < \theta .$$

Continuing the question from last week, give an approximate $100(1-\alpha)\%$ confidence interval for θ.

3. A random sample of size 16 from $N(\mu, 25)$ yielded $\bar{x} = 73.8$. Find a 95% confidence interval for μ. (Recall $z_{0.025} = 1.96$, $z_{0.05} = 1.645$).

(X_1, X) is a 95% CI for μ

4. A pet store sells guinea pig food in "2-pound" bags that are weighed on a an old 25-pound scale. Suppose it is known that the standard deviation of weights is $\sigma = 0.12$ pound. If a sample of 16 bags of guinea pig food were carefully weighed in a laboratory and the average weight was $\bar{x} = 2.09$ pounds, find an approximate 95% confidence interval for μ, the mean weight of gerbil food in the "2-pound" bags sold by the pet store.

$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ $\quad (\bar{X} \pm Z \cdot \frac{\sigma}{\sqrt{n}})$

5. To determine whether bacteria count was lower in the west basin of Lake Macatawa than in the east basin, $n = 37$ samples of water were taken in the west basin, and the number of bacteria colonies in 100 millilitres of water was counted. The sample characteristics were $\bar{x} = 11.95$ and $s = 11.80$, measured in hundreds of colonies. Find the approximate 95% confidence interval for the mean number of colonies, say μ_W, in 100 millilitres of water in the west basin. (Note, $t_{0.025}(36) = 2.028$, $t_{0.05}(36) = 1.688$)

$\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t(n-1)$

6. Thirteen tons of cheese is stored in some old gypsum mines, including "22-pound" wheels (label weight). A random sample of $n = 9$ of these wheels yields $\bar{x} = 20.9$ and $s = 1.858$. Assuming that the weights of the wheels is $N(\mu, \sigma^2)$ find a 95% confidence interval for μ. Is the claim these are "22 pound" wheels reasonable? ($t_{0.025}(8) = 2.306$, $t_{0.05}(8) = 1.859$)

$\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t(n-1)$

7. The length of life of brand X light bulbs is assumed to be $N(\mu_X, 784)$. The length of life of brand Y light bulbs is assumed to be $N(\mu_Y, 627)$ and these lifetimes are independent of X. If a random sample of $n = 56$ brand X light bulbs yielded $\bar{x} = 937.4$ hours and a random sample of size $m = 57$ brand Y light bulbs yielded $\bar{y} = 988.9$, find a 95% confidence interval for $\mu_X - \mu_Y$. Is it reasonable to conclude that the two brands of light bulb have the same mean lifetimes?

$(\bar{x}-\bar{y} \pm Z_{0.975} \times \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}) \sim N(0,1)$

8. A test was conducted to determine if a wedge on the end of a plug designed to hold a seal onto that plug was operating correctly. The data were the force required to remove a seal from the plug with the wedge in place (X) and without the wedge (Y). Assume the distributions of X and Y are $N(\mu_X, \sigma^2)$ and $N(\mu_Y, \sigma^2)$ respectively. Samples of size 10 on each variable yielded:

$$Pr\left(t_{\frac{\alpha}{2}, m+n-2} < \frac{\bar{x}-\bar{y}-(\mu_x-\mu_y)}{\sigma_p\sqrt{\frac{1}{n}+\frac{1}{m}}} < t_{1-\frac{\alpha}{2}, m+n-2}\right) = 1-\alpha$$

$\sigma_p^2 = \frac{(n-1)S_x^2}{\ldots} + \frac{(m-1)S_y^2}{\ldots}$

$m+n-2$ $m+n-2$

$\left(\bar{X}-\bar{Y} \pm \sigma_p\sqrt{\frac{1}{n}+\frac{1}{m}} \cdot t_{m+n-2}^{1-\frac{\alpha}{2}}\right)$

Variable	n	$\bar{x}$	s
X	10	2.548	0.323
Y	10	1.564	0.210

a. Find a 95% confidence interval for $\mu_X - \mu_Y$. ($t_{0.025}(18) = 2.101$, $t_{0.05}(18) = 1.734$)

b. Do you think the wedge is operating correctly?

本周任务：学Hypothesis testing 时回头，把两道题做完、学T.

9. Let X be the length in centimeters of a species of fish when caught in the spring. A random sample of 13 observations yielded the sample variance $s^2 = 37.751$. Find a 95% confidence interval for σ. ($\chi^2_{0.025}(12) = 4.404$, $\chi^2_{0.975}(12) = 23.337$).

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ (左右不对称 不能用士)

$\frac{(n-1)S^2}{\chi^2_{0.975}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{0.025}(n-1)}$

不对称

10. Let X be the length of a male grackle (a type of bird). Suppose $X \sim N(\mu, 4.84)$. Find the sample size that is needed if we are to be 95% confident the maximum error (ie. $z_{\alpha/2}(\sigma/\sqrt{n})$) of the estimate of μ is 0.4. ($z_{0.025} = 1.96$) ??? $(X \pm 1.96\sqrt{\frac{4.84}{n}})$

11. For a public opinion poll for a close election, let p denote the proportion of votes who favour candidate A. How large a sample should be taken if we want the maximum error of the estimate of p to be equal to

$(0.4 \pm 1.96 \cdot \frac{4.84}{\sqrt{n}})$

a. 0.03 with 95% confidence?

b. 0.02 with 95% confidence?

c. 0.03 with 90% confidence? ($z_{0.05} = 1.645$).

最小值 $> t$

12. Let $Y_1 < \cdots < Y_5$ be the order statistics of 5 independent observations from an exponential distribution that has a mean of $\theta = 3$.

最值相关

a. Find the p.d.f. of the sample minimum Y_1

b. Compute the probability that $Y_5 < 5$

c. Determine $P(1 < Y_1)$ $\frac{1}{\theta}e^{-x/\theta}$

$P(T \le t) = 1 - P(T > t)$

$= 1 - P(\min(Y_1, \cdots, Y_5) > t)$

$T = \min(Y_1, \cdots, Y_5) = 1 - [1-(1-e^{-\frac{x}{\theta}})]^5$

$= 1 - [e^{-\frac{x}{\theta}}]^5$

$= 1 - e^{-\frac{x}{\theta}5}$

$= \frac{5}{3}e^{-\frac{x}{\theta}}$

13. In a clinical trial, let the probability of a successful outcome have a prior distribution that is uniform over $[0, 1]$. Suppose that the first patient has a successful outcome. Find the Bayes estimate of θ that would be obtained for the squared error loss. Also find the Bayes estimate with absolute loss. In both cases, find a 85% posterior probability interval that is symmetric around the Bayes estimate.

$Pr(Y_5 < 5) = 1 - e^{-\frac{x}{\theta}} (5) = 1 - e^{-\frac{5}{3}}$

$Y_5 = \max\{X_1, \cdots, X_5\} = (1 - e^{-\frac{x}{\theta}})^5 =$

$1 - P(Y_1 \le 1)$

Y_1

$f(x, \theta) = f(\theta) \cdot f(x_i | \theta)$

$m(x) = \int$

Y_2 (第2小) $(X_1, \cdots, X_5)$
2nd min

$P(Y_2 < y)$ 5个有两小

$\binom{5}{2}(1-e^{-\frac{2}{3}y})$

2. Let $X_1, \ldots, X_n$ be a random sample from a gamma distribution with $\alpha = 4$ so that

Gamma分布如何估计?

$$f(x;\theta) = \frac{1}{6\theta^4} x^3 e^{-x/\theta},\ 0 < x < \infty,\ 0 < \theta\,.$$

Continuing the question from last week, give an approximate $100(1-\alpha)\%$ confidence interval for θ.

☆ $\hat{\theta}_{MLE}$ and $I(\theta)$ FI ??? what condition we can use?

How about MM

考 ☆ $\hat{\theta} \overset{d}{\approx} N(\theta, I(\theta))$???

$I(\theta) = \frac{4n}{\theta^2}$ $\hat{\theta} = \frac{\bar{X}}{4}$ is MLE

$$\hat{\theta} \overset{d}{\approx} N\left(\theta, \frac{\theta^2}{4n}\right) \Rightarrow \frac{\hat{\theta} - \theta}{\sqrt{\frac{\theta^2}{4n}}} = \frac{2\sqrt{n}\left(\frac{\bar{X}}{4} - \theta\right)}{\theta} \approx N(0,1)$$

$$-\underset{a}{Z_{1-\frac{\alpha}{2}}} < \frac{2\sqrt{n}\left(\frac{\bar{X}}{4} - \theta\right)}{\theta} < \underset{a}{Z_{1-\frac{\alpha}{2}}}$$

$$-a\theta < 2\sqrt{n}\left(\frac{\bar{X}}{4} - \theta\right) < a\theta$$

$$-a\theta < \frac{2\sqrt{n}\bar{x}}{4} - 2\sqrt{n}\theta < a\theta \quad = -a\theta < \frac{1}{2}\sqrt{n}\bar{x} - 2\sqrt{n}\theta < a\theta$$

$$P\left(-a\theta < \frac{1}{2}\sqrt{n}\bar{x} - 2\sqrt{n}\theta, \quad \frac{1}{2}\sqrt{n}\bar{x} - 2\sqrt{n}\theta < a\theta\right)$$

$$\Rightarrow P\left(-\theta(a-2\sqrt{n}) < \frac{2\sqrt{n}\bar{x}}{4}, \quad \frac{2\sqrt{n}\bar{x}}{4} < \theta(1+2\sqrt{n})\right)$$

$$\Rightarrow P\left(\theta > -\frac{2\sqrt{n}\bar{x}}{4(a-2\sqrt{n})}, \theta > \frac{2\sqrt{n}\bar{x}}{4(1+2\sqrt{n})}\right)$$

$$\Rightarrow \left(\frac{\sqrt{n}\bar{x}}{4\sqrt{n}+2z_{1-\frac{a}{2}}}, \frac{\sqrt{n}\bar{x}}{4\sqrt{n}-2z_{1-\frac{a}{2}}}\right)$$

$\hat{\theta}_{ML} \approx N(\theta, I^{-1}(\theta))$ $\quad I(\theta)$: FI

$Y_1 < \cdots < Y_5$ order statistics from exponential distribution $\theta = 3$.

PDF minimum Y_1

CDF: $P(Y_1 \le y) = P(\min(X_1, \cdots, X_5) \le y) = 1 - P(\min(X_1, \cdots, X_5) \ge y)$

$= 1 - P(X_1 \ge y) \cdot P(X_2 \ge y) \cdot P(X_3 \ge y) \cdots P(X_5 \ge y)$

$$\left(1 - 1 + e^{-\frac{x}{3}}\right)$$

$$= 1 - \left(e^{-\frac{y}{3}}\right)^5 = 1 - e^{-\frac{5y}{3}}$$

b. Compute the probability that $Y_5 < 5$

c. Determine $P(1 < Y_1)$ $\quad \frac{1}{\theta}e^{-x/\theta}$

$$P(Y_5 < 5) \Rightarrow P(\max(X_1, \ldots, X_5) < 5) \Rightarrow P(X_1 < 5) \cdots P(X_5 < 5)$$

$$= \left(1 - e^{-\frac{5}{3}}\right)^5 = $$

$$P(Y_1 > 1) = 1 - P(Y_1 \le 1) = 1 - \left(1 - e^{-\frac{5}{3}}\right) = e^{-\frac{5}{3}}$$

13. In a clinical trial, let the probability of a successful outcome have a prior distribution that is uniform over [0, 1]. Suppose that the first patient has a successful outcome. Find the Bayes estimate of θ that would be obtained for the squared error loss. Also find the Bayes estimate with absolute loss. In both cases, find a 85% posterior probability interval that is symmetric around the Bayes estimate.

$$f(\theta|x) = \frac{\theta^y (1-\theta)^{1-y} \times 1}{\int_0^1 \theta^y (1-\theta)^{1-y} d\theta} =$$

$$f(\theta|y=1) = \frac{\theta}{\int_0^1 \theta \, d\theta} = \frac{\theta}{\frac{1}{2}\theta^2 \Big|_0^1} = 2\theta = 2\theta \qquad 0 \le \theta \le 1$$

$$\int_0^1 2\theta \cdot \theta = \frac{2}{3}\theta^3 \Big|_0^1 = \frac{2}{3}$$