MA 486-25 (Statistics), Chernov | Midterm test #1
Show your work. | Thu, April 12, 2001

1. (12 pts) The following are quiz scores in a calculus class:

11	15	12	9	10	7	14	16	13	12
9	15	5	14	12	13	15	14	10	14

(a) Construct a frequency table for these data.
(b) Draw a histogram.
(c) Determine the mode, the median, the quartiles and the IQR.
(d) Draw a Box-and-Whisker diagram.
(e) Find the sample mean, the sample variance and the sample standard deviation.

Partial answers:

mode = 14, median = 12.5, first quartile = 10, third quartile = 14, IQR = 4

sample mean = 12, sample variance = 8.53, sample standard deviation = 2.92

2. (16 pts) Let $x_1, \ldots, x_n$ be a random sample from the distribution with probability density function

$$f(x;\theta) = \frac{|x|}{\theta^2}\, e^{x/\theta} \qquad \text{for} \quad -\infty < x < 0$$

where $\theta > 0$ is an unknown parameter (note that x is negative!).
(a) Find the maximum likelihood estimate for θ.
(b) Is this estimate unbiased? (You can use the fact that $\int_{-\infty}^{0} x^2\, e^{x/\theta}\, dx = 2\theta^3$ for $\theta > 0$.)
(c) Determine sufficient statistics.
(d) For an extra credit, find the Rao-Cramer bound.

Solutions:
(a) Likelihood function is

$$L(\theta) = \frac{|x_1 \cdots x_n|}{\theta^{2n}}\, e^{(x_1+\cdots+x_n)/\theta}$$

$$\ln L(\theta) = \ln(|x_1 \cdots x_n|) - 2n \ln\theta + (x_1 + \cdots + x_n)/\theta$$

$$\frac{d}{d\theta} \ln L(\theta) = -\frac{2n}{\theta} - \frac{x_1 + \cdots + x_n}{\theta^2} = 0$$

MLE is

$$\hat{\theta} = -(x_1 + \cdots + x_n)/2n = -\bar{x}/2$$

(b)

$$E(\hat{\theta}) = -\frac{EX}{2} = -\frac{1}{2}\int_{-\infty}^{0} x|x|\, \frac{1}{\theta^2}\, e^{x/\theta}\, d\theta = -(-\theta) = \theta$$

so the MLE is unbiased.

(c) Suffucient statistic is $x_1 + \cdots + x_n$ (alternatively, $\bar{x}$).

(d)

$$\ln f(x,\theta) = \ln|x| - 2\ln\theta + x/\theta$$

$$\frac{d}{d\theta} \ln f(x,\theta) = -2/\theta - x/\theta^2$$

$$\frac{d^2}{d\theta^2} \ln f(x,\theta) = 2/\theta^2 + 2x/\theta^3$$

$$E\left(\frac{d^2}{d\theta^2} \ln f(x,\theta)\right) = 2/\theta^2 + 2E(X)/\theta^3 = 2/\theta^2 + 2(-2\theta)/\theta^3 = -2/\theta^2$$

$$\sigma_n^2 = \frac{1}{-n \cdot (-2/\theta^2)} = \frac{\theta^2}{2n}$$

3. (12 pts) A random sample $x_1, \ldots, x_{29}$ from $N(\mu, \sigma^2)$ yielded

$$\sum_{i=1}^{29} x_i = 58 \qquad \text{and} \qquad \sum_{i=1}^{29} x_i^2 = 143$$

(a) Give a point estimate for μ.
(b) Construct a 95% confidence interval for μ.
(c) Give the lower endpoint of a 99% one-sided confidence interval for μ.

Solution:
First of all, $\bar{x} = 58/29 = 2$ and

$$s^2 = \frac{1}{28}(143 - 29 \cdot 2^2) = 0.964$$

so $s = 0.982$
(a)

$$\bar{x} \pm t_{\alpha/2}(n-1)\, s/\sqrt{n} = 2 \pm t_{.025}(28)\, 0.982/\sqrt{29} = 2 \pm 2.048 \cdot 0.982/\sqrt{29}$$

$$1.627 < \mu < 2.373$$

(b)

$$\bar{x} - t_{\alpha}(n-1)\, s/\sqrt{n} = 2 - t_{.01}(28)\, 0.982/\sqrt{29} = 2 - 2.467 \cdot 0.982/\sqrt{29}$$

$$\mu > 1.55$$

4. (14 pts) The math test scores in two classes produced the following results: $n_x = 22$, $\bar{x} = 75$ and $s_x = 4.6$ for the first class and $n_y = 26$, $\bar{y} = 72$ and $s_y = 5.1$ for the second class. Construct a 98% confidence interval for the difference $\mu_x - \mu_y$. [Note: since σ_x and σ_y are unknown and may be different, use Welsh's formula.]

Solution:
First, the number of degrees of freedom is

$$r = \frac{(s_x^2/n_x + s_y^2/n_y)^2}{(s_x^2/n_x)^2/(n_x - 1) + (s_y^2/n_y)^2/(n_y - 1)}$$

hence

$$r = \frac{(4.6^2/22 + 5.1^2/26)^2}{(4.6^2/22)^2/21 + (5.1^2/26)^2/25} = 45.8$$

Rounding off gives $r = 46$.
Now, the CI is

$$\bar{x} - \bar{y} \pm t_{\alpha/2}(r)\sqrt{s_x^2/n + s_y^2/m}$$

$$= 75 - 72 \pm t_{.01}(46)\sqrt{4.6^2/22 + 5.1^2/26}$$

$$= 3 \pm 2.326 \cdot 1.401$$

Finally,

$$-0.258 < \mu_x - \mu_y < 6.258$$

5. (10 pts) A random sample of size $n = 28$ from $N(\mu, \sigma^2)$ yielded $\bar{x} = -5.3$ and sample variance $s^2 = 6.25$.

(a) Find a 98% confidence interval for σ that cuts off 1% probability on the left side and 1% on the right side.
(b) Find the shortest 99% confidence interval for σ.

Solution:
(a)

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}$$

$$\frac{27 \cdot 6.25}{\chi^2_{.01}(27)} < \sigma^2 < \frac{27 \cdot 6.25}{\chi^2_{.99}(27)}$$

$$\frac{27 \cdot 6.25}{46.96} < \sigma^2 < \frac{27 \cdot 6.25}{12.88}$$

$$3.59 < \sigma^2 < 13.10$$

$$1.90 < \sigma < 3.62$$

(b) From Table X

$$\frac{27 \cdot 6.25}{52.886} < \sigma^2 < \frac{27 \cdot 6.25}{12.501}$$

$$3.19 < \sigma^2 < 13.50$$

$$1.79 < \sigma < 3.67$$

6. (12 pts) Let $x_1, \ldots, x_{25}$ and $y_1, \ldots, y_{41}$ be two independent random samples from distributions $N(\mu_x, \sigma_x^2)$ and $N(\mu_y, \sigma_y^2)$, respectively. Their sample means are $\bar{x} = 10.4$ and $\bar{y} = -0.5$. Their sample standard deviations are $s_x = 2.1$ and $s_y = 6.4$, respectively.

(a) Give a point estimate for σ_x^2/σ_y^2.
(b) Find a 95% confidence interval for ratio of variances σ_x^2/σ_y^2.
(c) Find a 98% confidence interval for the ratio of standard deviations σ_x/σ_y.

(a) Point estimate is $s_x^2/s_y^2 = 0.108$

(b)

$$\frac{s_x^2/s_y^2}{F_{\alpha/2}(n-1, m-1)} < \frac{\sigma_x^2}{\sigma_y^2} < s_x^2/s_y^2 \cdot F_{\alpha/2}(m-1, n-1)$$

$$\frac{s_x^2/s_y^2}{F_{.025}(24, 40)} < \frac{\sigma_x^2}{\sigma_y^2} < s_x^2/s_y^2 \cdot F_{.025}(40, 24)$$

$$\frac{0.108}{2.01} < \frac{\sigma_x^2}{\sigma_y^2} < 0.108 \cdot 2.15$$

$$0.0536 < \sigma_x^2/\sigma_y^2 < 0.2315$$

(c)

$$\frac{s_x^2/s_y^2}{F_{.01}(24, 40)} < \frac{\sigma_x^2}{\sigma_y^2} < s_x^2/s_y^2 \cdot F_{.01}(40, 24)$$

$$\frac{0.108}{2.29} < \frac{\sigma_x^2}{\sigma_y^2} < 0.108 \cdot 2.49$$

$$0.047 < \sigma_x^2/\sigma_y^2 < 0.268$$

$$0.217 < \sigma_x/\sigma_y < 0.518$$

7. (8 pts) Let p equal the proportion of defective items manufactured in a production line. A quality test of randomly selected 250 items shows that 28 of them were defective.

(a) Find a 95% confidence interval for p.
(b) Give the upper endpoint of a one-sided 90% confidence interval for p.
(c) Give the lower endpoint of a one-sided 90% confidence interval for p.

Solution:
First, $\hat{p} = 28/250 = 0.112$
(a)

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} = 0.112 \pm 1.96\sqrt{.112 \cdot .888/250}$$

$$0.073 < p < 0.151$$

(b)

$$\hat{p} + z_{\alpha}\sqrt{\hat{p}(1-\hat{p})/n} = 0.112 + 1.282\sqrt{.112 \cdot .888/250}$$

$$p < 0.137$$

(c)

$$\hat{p} - z_{\alpha}\sqrt{\hat{p}(1-\hat{p})/n} = 0.112 - 1.282\sqrt{.112 \cdot .888/250}$$

$$p > 0.086$$

8. (8 pts) Let p_1 be the proportion of adult men that smoke and p_2 the proportion of adult women that smoke. In a random sample of 1940 men, 286 are smokers. In a random sample of 3515 women, 372 are smokers.

(a) Construct a 99% confidence interval for $p_1 - p_2$.
(b) Construct a 99.9% confidence interval for $p_1 - p_2$.

Note:

$$\hat{p}_1 = 286/1940 = 0.147, \qquad \hat{p}_2 = 372/3515 = 0.106$$

(a)

$$0.147 - 0.106 \pm 2.576\sqrt{0.147 \cdot 0.853/1940 + 0.106 \cdot 0.894/3515}$$

$$0.0169 < p_1 - p_2 < 0.0663$$

(b)

$$0.147 - 0.106 \pm 3.291\sqrt{0.147 \cdot 0.853/1940 + 0.106 \cdot 0.894/3515}$$

$$0.0101 < p_1 - p_2 < 0.0731$$

9. (8 pts) How large a sample should be taken from $N(\mu, \sigma^2)$ if we want the maximum error of the estimate of μ to be equal 0.01 with 98% confidence? A previous sample $x_1, \ldots, x_{40}$ from the same distribution yielded

$$\sum_{i=1}^{40} x_i = 62 \qquad \text{and} \qquad \sum_{i=1}^{40} x_i^2 = 260$$

Solution:

$$s^2 = \frac{1}{39}(260 - 62^2/40) = 4.203$$

$$n > \left(\frac{z_{\alpha/2} \cdot s}{\varepsilon}\right)^2 = \left(\frac{2.326 \cdot 2.05}{0.01}\right) = 227,371$$