MA 486-25 (Statistics), Chernov — Midterm test #1

Show your work. — Thu, April 6, 2000

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

8	4	9	2	1	6	1	4	3	10
6	3	6	7	3	8	7	5	4	3

(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 = 3, median = 4.5, first quartile = 3, third quartile = 7, IQR = 4

sample mean = 5, sample variance = 6.84, sample standard deviation = 2.616

2. (14 pts) Let $x_1, \ldots, x_n$ be a random sample from the distribution with probability density function

$$f(x;\theta) = \frac{1}{\theta}\, e^{x/\theta} \quad \text{for} \quad -\infty < x < 0$$

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

(a) Likelihood function is

$$L(\theta) = \frac{1}{\theta^n}\, e^{(x_1+\cdots+x_n)/\theta}$$

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

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

MLE is

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

(b)

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

so the MLE is unbiased.

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

(d)

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

$$\frac{d}{d\theta} \ln f(x,\theta) = -1/\theta - x/\theta^2$$

$$\frac{d^2}{d\theta^2} \ln f(x,\theta) = 1/\theta^2 + 2x/\theta^3$$

$$E\left(\frac{d^2}{d\theta^2} \ln f(x,\theta)\right) = 1/\theta^2 + 2EX/\theta^3 = 1/\theta^2 + 2(-\theta)/\theta^3 = -1/\theta^2$$

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

3. (12 pts) Colesterol level test of randomly selected $n = 25$ adult persons yielded sample mean $\bar{x} = 225$ and sample variance $s^2 = 144$.

(a) Construct a 90% confidence interval for μ, the mean cholesterol level for adults.
(b) Give the upper endpoint of a 75% one-sided confidence interval for μ.

(a)

$$\bar{x} \pm t_{\alpha/2}(n-1)\, s/\sqrt{n} = 225 \pm t_{.05}(24)\, \sqrt{144}/\sqrt{25} = 225 \pm 1.711 \cdot 12/5$$

$$220.9 < \mu < 229.1$$

(b)

$$\bar{x} + t_{\alpha}(n-1)\, s/\sqrt{n} = 225 + t_{.25}(24)\, \sqrt{144}/\sqrt{25} = 225 + 0.685 \cdot 12/5$$

$$\mu < 226.6$$

4. (12 pts) A math test in a public school produced the following results: $n = 250$ students were tested, the average score was $\bar{x} = 75$ and the sample standard deviations was $s_x = 15$. For a private school in the same city the same math test yielded, respectively, $m = 80$, $\bar{y} = 82$ and $s_y = 12$.

(a) Construct a 90% confidence interval for the difference $\mu_x - \mu_y$.
(b) Construct a 98% confidence interval for the difference $\mu_x - \mu_y$.

(a)

$$\bar{x} - \bar{y} \pm z_{\alpha/2}\sqrt{s_x^2/n + s_y^2/m} = 75 - 82 \pm z_{.05}\sqrt{15^2/250 + 12^2/80}$$
$$= -7 \pm 1.645\sqrt{225/250 + 144/80}$$

$$-9.703 < \mu_x - \mu_y < -4.297$$

(b)

$$\bar{x} - \bar{y} \pm z_{\alpha/2}\sqrt{s_x^2/n + s_y^2/m} = 75 - 82 \pm z_{.01}\sqrt{15^2/250 + 12^2/80}$$
$$= -7 \pm 2.326\sqrt{225/250 + 144/80}$$

$$-10.822 < \mu_x - \mu_y < -3.178$$

5. (12 pts) A random sample of $n = 22$ wheels of cheese yielded sample mean $\bar{x} = 17.65$ pounds and sample variance $s^2 = 2.89$ of their weights. Assume that the weight of a wheel of cheese is a normal random variable $N(\mu, \sigma^2)$.

(a) Find a 95% confidence interval for σ that cuts off 2.5% probability on the left side and 2.5% on the right side.
(b) Find the shortest 95% confidence interval for σ. How much shorter is it than the interval found in (a)?

(a)

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

$$\frac{21 \cdot 2.89}{\chi^2_{.025}(21)} < \sigma^2 < \frac{21 \cdot 2.89}{\chi^2_{.975}(21)}$$

$$\frac{21 \cdot 2.89}{35.48} < \sigma^2 < \frac{21 \cdot 2.89}{10.28}$$

$$1.71 < \sigma^2 < 5.90$$

$$1.31 < \sigma < 2.43$$

(b) From Table X

$$\frac{21 \cdot 2.89}{38.472} < \sigma^2 < \frac{21 \cdot 2.89}{11.073}$$

$$1.58 < \sigma^2 < 5.48$$

$$1.26 < \sigma < 2.34$$

Shorter by .037.

6. (14 pts) Let $x_1, \ldots, x_{25}$ and $y_1, \ldots, y_{13}$ 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 $\bar{x} = -1.3$ and $\bar{y} = 3.8$. Their sample standard deviations are $s_x = 1.5$ and $s_y = 2.5$, respectively.

(a) Give a point estimate for σ_x^2/σ_y^2.
(b) Find a 90% confidence interval for ratio of variances σ_x^2/σ_y^2.
(c) Find a 95% confidence interval for the ratio of standard deviations σ_x/σ_y.

(a) Point estimate is $s_x^2/s_y^2 = 0.36$

(b)

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

$$\frac{s_x^2/s_y^2}{F_{.05}(24, 12)} < \frac{\sigma_x^2}{\sigma_y^2} < s_x^2/s_y^2 \cdot F_{.05}(12, 24)$$

$$\frac{0.36}{2.51} < \frac{\sigma_x^2}{\sigma_y^2} < 0.36 \cdot 2.18$$

$$0.143 < \sigma_x^2/\sigma_y^2 < 0.785$$

(c)

$$\frac{s_x^2/s_y^2}{F_{.025}(24, 12)} < \frac{\sigma_x^2}{\sigma_y^2} < s_x^2/s_y^2 \cdot F_{.025}(12, 24)$$

$$\frac{0.36}{3.02} < \frac{\sigma_x^2}{\sigma_y^2} < 0.36 \cdot 2.54$$

$$0.119 < \sigma_x^2/\sigma_y^2 < 0.914$$

$$0.345 < \sigma_x/\sigma_y < 0.956$$

7. (12 pts) Let p equal the proportion of defective items manufactured in a production line. A quality test of randomly selected 600 items shows that 72 of them were defective.

(a) Find a 90% 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.

Note:

$$\bar{x} = 72/600 = 0.12$$

(a)

$$\bar{x} \pm z_{\alpha/2}\sqrt{\bar{x}(1-\bar{x})/n} = \bar{x} \pm z_{.05}\sqrt{\bar{x}(1-\bar{x})/n} = 0.12 \pm 1.645\sqrt{.12 \cdot .88/600}$$

$$0.098 < p < 0.142$$

(b)

$$\bar{x} + z_{\alpha}\sqrt{\bar{x}(1-\bar{x})/n} = \bar{x} + z_{.1}\sqrt{\bar{x}(1-\bar{x})/n} = 0.12 + 1.282\sqrt{.12 \cdot .88/600}$$

$$p < 0.137$$

(c)

$$\bar{x} - z_{\alpha}\sqrt{\bar{x}(1-\bar{x})/n} = \bar{x} - z_{.1}\sqrt{\bar{x}(1-\bar{x})/n} = 0.12 - 1.282\sqrt{.12 \cdot .88/600}$$

$$p > 0.103$$

8. (12 pts) Let p_1 equal the proportion of defective items manufactured in one production line, call it A. Let p_2 equal the proportion of defective items manufactured in another production line, call it B. Quality tests conducted on both lines, separately, yilded the following results: in line A, 72 out of randomly selected 600 items were defective; in line B, 60 out of randomly selected 400 items were defective

(a) Construct a 80% confidence interval for $p_1 - p_2$.
(b) Construct a 90% confidence interval for $p_1 - p_2$.

Note:

$$\bar{x}_1 = 72/600 = 0.12, \qquad \bar{x}_2 = 60/400 = 0.15$$

(a)

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sqrt{\bar{x}_1(1-\bar{x}_1)/n_1 + \bar{x}_2(1-\bar{x}_2)/n_2} = \bar{x}_1 - \bar{x}_2 \pm z_{.1}\sqrt{\bar{x}_1(1-\bar{x}_1)/n_1 + \bar{x}_2(1-\bar{x}_2)/n_2}$$

$$= 0.12 - 0.15 \pm 1.282\sqrt{.12 \cdot .88/600 + .15 \cdot .85/400}$$

$$-0.0585 < p_1 - p_2 < -0.0015$$

(b)

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sqrt{\bar{x}_1(1-\bar{x}_1)/n_1 + \bar{x}_2(1-\bar{x}_2)/n_2} = \bar{x}_1 - \bar{x}_2 \pm z_{.05}\sqrt{\bar{x}_1(1-\bar{x}_1)/n_1 + \bar{x}_2(1-\bar{x}_2)/n_2}$$

$$= 0.12 - 0.15 \pm 1.645\sqrt{.12 \cdot .88/600 + .15 \cdot .85/400}$$

$$-0.0666 < p_1 - p_2 < 0.0066$$

9. [Bonus] For a public opinion poll in a close presidential election, let p be the proportion of voters in favor of candidate A. How large a sample should be taken if we want the maximum error of the estimate of p to be equal 0.04 with 98% confidence?

$$n = \frac{z_{\alpha/2}^2}{4\varepsilon^2} = \frac{(2.236)^2}{4 \cdot 0.04^2} = 845.3$$

Sufficient value is $n = 846$