MATH 376 – Mathematical Statistics
Solutions for Exam 1 Sample Problems
March 12, 2012

I.

A) This has a $\chi^2(7)$ distribution (see Theorems 6.4 and 7.2 in the text).

B) From the χ^2 table on pages 850-851 of the text, we see:

$$1.68987 = \chi^2_{.975}, \quad \text{and} \quad 14.0671 = \chi^2_{.05}$$

Hence

$$P(1.68987 \le W \le 14.0671) = .975 - .05 = .925$$

(Drawing the picture helps if this is not clear!)

C) Leaving out Z_6^2 in the sum means U has a $\chi^2(6)$ distribution. Since

$$V = \frac{6Z_7^2}{U} = \frac{Z_7^2/1}{U/6}$$

has the form given in Definition 7.3, V has an F distribution with 1 d.f. in the numerator and 6 d.f. in the denominator.

D) By the conventions used in our book, this means that $y = f_{.05}(1,6)$. From the F-table (on page 852) we find $y = 5.99$.

E) For this, we need the relation derived on Lab 1:

$$f_{1-\alpha}(n,d) = \frac{1}{f_\alpha(d,n)}$$

Here,

$$y = f_{.95}(1,6) = \frac{1}{f_{.05}(6,1)} = \frac{1}{234.0} \doteq .0043$$

F) Recalling Definition 7.2, let $X = \sum_{i=1}^5 Z_i^2$. Then

$$T = \frac{Z_6}{\sqrt{X/5}}$$

has a t-distribution with 5 degrees of freedom.

II.

A) By the formulas from §6.7, the cdf for $Y_{(n)}$ is

$$F_{(n)}(y) = (1 - e^{-y/\theta})^n$$

Hence

$$P(Y_{(n)} \ge 2\theta) = 1 - P(Y_{(n)} < 2\theta) = 1 - (1 - e^{-2\theta/\theta})^n = 1 - (1 - e^{-2})^n$$

Since $0 < 1 - e^{-2} < 1$,

$$\lim_{n\to\infty} P(Y_{(n)} \geq 2\theta) = \lim_{n\to\infty} 1 - (1 - e^{-2})^n = 1.$$

This should seem intuitively reasonable, since the exponential distributions have very long right-hand tails and the expected value is θ. If you take $n \to \infty$, the probability that the largest sample value exceeds 2θ is tending to 1.

B) The pdf for the sample minimum is

$$f_{(1)}(y) = n\left(1 - (1 - e^{-y/\theta})\right)^{n-1} \frac{e^{-y/\theta}}{\theta} = \frac{e^{-y/(\theta/n)}}{\theta/n}.$$

We recognize the form of an exponential density with mean θ/n. Therefore $E(Y_{(1)}) = \theta/n$. As a result $nY_{(1)}$ is an unbiased estimator for θ.

III.

A) The usual point estimator for μ would be the sample mean:

$$\overline{Y} = \frac{16.0 + 15.2 + 12.0 + 16.9 + 14.4 + 16.3 + 15.6 + 12.9}{8} \doteq 14.90$$

The usual point estimator for σ would be the sample standard deviation:

$$S = \sqrt{\frac{1}{7}\sum_{i=1}^{8}(Y_i - \overline{Y})^2} \doteq 1.707$$

B) Under the assumptions that the data are from a normal population, we can compute a 90% confidence interval using the small sample formulas. The endpoints are

$$\overline{Y} \pm t_{.05}\frac{S}{\sqrt{n}} = 14.9 \pm 1.895\frac{1.707}{\sqrt{8}} \doteq 14.90 \pm 1.14$$

To see whether the normality assumption is valid, by what we saw in Lab 3, we could use either a normal quantile-quantile plot and look for approximate linearity, or we could use the Shapiro-Wilk normality test.

C) No, $\mu = 16$ is contained in the interval. It is a "believable" value for μ based on this data.

D) Using the formula from page 435 (also see Lab 3), the "standard" way to get such an interval is

$$\left(\frac{7S^2}{\chi^2_{.025}(7)}, \frac{7S^2}{\chi^2_{.975}(7)}\right) \doteq (1.27, 12.08).$$

IV.

A) $\overline{p_A} = \frac{24}{642} \doteq .0373.$

B) Since $n_A = 642 > 30$, we use the large sample formula and the approximation for the standard error making use of the value from part A. We have $z_{.025} = 1.96$, so

$$p_A = .0374 \pm 1.96\sqrt{\frac{(.0374)(.9626)}{642}} \doteq .0374 \pm .0147$$

C) The interval is

$$p_A - p_B = .0374 - .0440 \pm 1.96\sqrt{\frac{(.0374)(.9626)}{642} + \frac{(.0440)(.9560)}{500}} \doteq -.0066 \pm .0232.$$

Since this interval contains both positive and negative numbers, there is not enough evidence to say that $p_A < p_B$ (i.e. not sufficient evidence to say the quality of the output from shop A is higher than the quality from shop B). The observed differences in the proportions of defectives could be explained by random variations in the sampling process.

D) In general the best way to answer a question like this is to apply the "conservative" method to estimate the standard error – use $p_A = .5$ in the standard error formula, since that gives the largest possible standard error. Then the estimator is within .01 of the true value with probability at least when

$$\frac{.01}{\sqrt{(.5)(.5)/n}} \geq 1.96, \quad \text{or}$$

$$n \geq \left(\frac{(1.96)(.5)}{.01}\right)^2 = 9604$$

Depending on what the actual value of p_A is, a much smaller n might work too. For instance, if it was true that $p_A = .0374$ exactly, then applying the standard error formula with that value of p_A, we would require only

$$\frac{.01}{\sqrt{(.0374)(.9626)/n}} \geq 1.96, \quad \text{or}$$

$$n \geq \left(\frac{(1.96)\sqrt{(.0374)(.9626)}}{.01}\right)^2 \doteq 1383.02$$

So only $n \geq 1384$ is needed.