Mathematics 376 – Probability and Statistics II
Solutions – Final Examination
May 13, 2004

I. A) $\overline{Y}$ is normal with mean $\mu = 4$ and variance $\sigma^2/5 = 100/5 = 20$. See Theorem 7.1 in text.

B) U has a χ^2 distribution with 3 degrees of freedom, since each $(Y_i - 4)/10$ is standard normal, and they are independent. The sum of the squares has a χ^2 distribution by Theorem 7.2 in the text.

C) V has a t-distribution with 3 degrees of freedom, since

$$V = \frac{\sqrt{3}(Y_4 - 4)}{10\sqrt{U}} = \frac{(Y_4 - 4)/10}{\sqrt{U/3}}$$

This is the quotient of a standard normal and the square root of a χ^2 random variable, divided by its number of degrees of freedom. See Definition 7.2 in the text.

II. A) We have

$$E(Y_i) = \int_0^1 y \cdot 4y^3 \, dy = 4 \int_0^1 y^4 \, dy = 4/5$$

and

$$V(Y_i) = E(Y_i^2) - (E(Y_i))^2 = \int_0^1 y^2 \cdot 4y^3 \, dy - (4/5)^2 = (4/6) - (4/5)^2 = 2/75$$

B) By the Central Limit Theorem,

$$\sqrt{n}\left(\frac{\overline{Y} - 4/5}{\sqrt{2/75}}\right) = \sqrt{n}\left(\frac{\overline{Y} - .8}{.1633}\right)$$

converges in distribution to a standard normal as $n \to \infty$. For $n = 50$, we expect the desired probability to be closely approximated by

$$P\left(Z > \sqrt{50}\left(\frac{.85 - .8}{.1633}\right)\right) = P(z > 2.16) = .0154$$

(from the standard normal table).

III. A) For any continuous random variable Y with pdf $f(y)$, the expected value of a function $g(Y)$ of Y is computed by integrating $\int_{-\infty}^{\infty} g(y)f(y)\, dy$. (Note: the g does not "interact" at all with the density here. For our log-normal Y, we will compute

$$E(\ln(Y)) = \int_0^\infty \ln(y) \frac{1}{y\sqrt{2\pi\sigma^2}} e^{-(\ln(y)-\mu)^2/2\sigma^2} \, dy$$

First, change variables letting $u = \ln(y)$. Then $du = dy/y$ and the limits of integration change to $-\infty, \infty$:

$$= \int_{-\infty}^{\infty} u \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(u-\mu)^2/2\sigma^2} \, du$$
$$= \int_{-\infty}^{\infty} u f_N(u) \, du,$$

where f_N is the pdf for a normally distributed random variable with distribution $N(\mu, \sigma^2)$. We are computing the expected value of that normal, which is just μ. Hence

$$E(\ln(Y)) = \mu$$

if Y is log-normal.

B) Given the sample values $y_1, \ldots, y_n$, the likelihood function for $\sigma^2 = 1$ is

$$\begin{aligned} L(y_1, \ldots, y_n | \mu) &= f(y_1|\mu) \cdots f(y_n|\mu) \\ &= \prod_{i=1}^{n} \frac{1}{y_i\sqrt{2\pi}} e^{-(\ln(y_i)-\mu)^2/2} \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \cdot \frac{1}{\prod_{i=1}^{n} y_i} \cdot e^{-\sum_{i=1}^{n}(\ln(y_i)-\mu)^2/2} \end{aligned}$$

Then,

$$\ln(L) = \left(\frac{-n}{2}\right) \ln(2\pi) - \ln\left(\prod_{i=1}^{n} y_i\right) - \sum_{i=1}^{n}(\ln(y_i) - \mu)^2/2$$

To find the maximum likelihood estimator, we differentiate *with respect to* μ and set equal to zero:

$$0 = \frac{d\ln(L)}{d\mu} = 0 + 0 - \sum_{i=1}^{n}(\ln(y_i) - \mu) = -\sum_{i=1}^{n} \ln(y_i) + n\mu$$

Hence, solving for μ, we have:

$$\mu = \widehat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} \ln(Y_i)$$

C) This estimator is *unbiased* since

$$E(\widehat{\mu}_{ML}) = E\left(\frac{1}{n}\sum_{i=1}^{n} \ln(Y_i)\right) = \frac{1}{n}\sum_{i=1}^{n} E(\ln(Y_i)) = \frac{1}{n} \cdot n \cdot \mu = \mu$$

by part A.

IV. A) Since we only have 4, 5 samples in each group, we need to use the *small-sample* confidence interval for the difference of the means, which uses the assumption that the variances in the samples are equal (see Section 8.8). The best estimate of that common variance is the "pooled estimator"

$$S_p^2 = \frac{3s_1^2 + 4s_2^2}{7} = \frac{3(.001) + 4(.002)}{7} = .00157$$

Then the 95% confidence interval is

$$\mu_1 - \mu_2 = .25 - .17 \pm t_{.025}(7)\sqrt{.00157}\sqrt{\frac{1}{4} + \frac{1}{5}} = .08 \pm .0628$$

B) Since $0 \notin (.08 - .0628, .08 + .0628) = (.0172, .1428)$ there is sufficient evidence to see that we should reject the null hypothesis that $\mu_1 = \mu_2$ with $\alpha = .05$. (This can also be verified directly by performing a t-test: The test statistic is

$$t = \frac{.25 - .17}{\sqrt{.00157}\sqrt{\frac{1}{4} + \frac{1}{5}}} = 3.012$$

The rejection region for the level $\alpha = .05$ is

$$RR = \{|t| > t_{.025}(7) = 2.365\}$$

We reject the null hypothesis because $t = 3.012$ is in this RR.

V. A) Since we have $n = 50$ samples, a lower-tail z-test is appropriate. The rejection region is

$$RR = \{z < -z_{.01}\} = \{z < -2.33\}$$

The test statistic is

$$z = \frac{62 - 64}{8/\sqrt{50}} = -1.768$$

This is *not* in RR, so there is *not* sufficient evidence with $\alpha = .01$ to say that the mean hardness of this type of steel is less than 64 (and the manufacturer breathes a sigh of relief!!)

B) The rejection region in part A is the same as the set of $\overline{Y}$-values:

$$RR = \left\{\overline{y} < 64 - (2.33)\frac{8}{\sqrt{50}}\right\} = \{\overline{y} < 61.364\}.$$

To find β, assuming $\mu = 60$, we proceed as follows:

$$\begin{aligned}\beta = P(\overline{y} \notin RR|\mu = 60) &= P(\overline{y} \geq 61.34|\mu = 60)\\ &= P\left(\frac{\overline{y} - 60}{8/\sqrt{50}} \geq \frac{61.34 - 60}{8/\sqrt{50}}\right)\\ &= P(z \geq 1.206)\\ &= .1131\end{aligned}$$

VI. A) We use the sample variance $S_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ in each case and obtain

$$S_1^2 = 15750 \qquad S_2^2 = 10920$$

B) For testing the null hypothesis $\sigma_1^2 = \sigma_2^2$ against the alternative that $\sigma_1^2 \neq \sigma_2^2$, we use a 2-tailed F-test (see page 503 in the text). The test statistic is

$$f = S_1^2/S_2^2 = 15750/10920 = 1.44$$

Since $F_{.1}(4,5) = 3.52$ and $F_{.9}(4,5) = 1/F_{.1}(5,4) = .247$, the p-value is greater than $2(.1) = .2$. This is quite large, which means that there is extremely weak evidence to reject H_0 here; we would not reject H_0 for any $\alpha < .2$.

VII. A) The "highlights" of the computation of the regression line:

$$\begin{aligned}
\overline{x} &= 37.5 \\
\overline{y} &= 26.95 \\
S_{xx} &= 3937.5 \\
S_{xy} &= 2172.38 \\
S_{yy} &= 1202.48 \\
\widehat{\beta}_1 &= S_{xy}/S_{xx} = 2172.38/3937.5 = .5517 \\
\widehat{\beta}_0 &= \overline{y} - \widehat{\beta}_1\overline{x} = 26.95 - (.5517)(37.5) = 6.26 \\
S &= \sqrt{(S_{yy} - \widehat{\beta}_1 S_{xy})/4} = \sqrt{3.978/4} = .9972
\end{aligned}$$

B) For testing $H_0 : \beta_1 = .45$ against the alternative $\beta_1 > .45$, we use an upper-tail t-test with $n - 2 = 4$ degrees of freedom. The test statistic is

$$t = \frac{\widehat{\beta}_1 - \beta_{10}}{S/\sqrt{S_{xx}}} = \frac{.5517 - .45}{.9972/\sqrt{3937.5}} = 6.4$$

We see that $t_{.005}(4) = 4.604$, so the p-value of this test is $p < 2(.005) = .01$. Since this is quite small, we have strong evidence to reject the null hypothesis and conclude that $\beta_1 > .45$.