ST 430
HW #1
Solution Sheet

5.2.2 $L(p) = p(1-p)(1-p)p(1-p) = p^2(1-p)^3$

$L(1/3) = \left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^3 = \frac{8}{243}$ is greater than

$L(1/2) = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$, so $\hat{p} = 1/3$.

5.2.4 $L(\theta) = \prod_{i=1}^{n} \frac{\theta^{2k_i} e^{-\theta^2}}{k_i!} = \frac{\theta^{2\sum_{i=1}^{n} k_i} e^{-n\theta^2}}{\prod_{i=1}^{n} k_i!}$.

$\ln L(\theta) = \left(2\sum_{i=1}^{n} k_i\right)(\ln \theta) - n\theta^2 + \ln \prod_{i=1}^{n} k_i!$

$\frac{d \ln L(\theta)}{d\theta} = 0$ implies $\frac{2\sum_{i=1}^{n} k_i}{\theta} - 2n\theta = \frac{2\sum_{i=1}^{n} k_i - 2n\theta^2}{\theta} = 0$

or $\hat{\theta} = \sqrt{\frac{\sum_{i=1}^{n} k_i}{n}}$

5.2.6 $L(\theta) = \prod_{i=1}^{4} \frac{\theta}{2\sqrt{y_i}} e^{-\theta\sqrt{y_i}} = \frac{\theta^4}{16\prod_{i=1}^{4}\sqrt{y_i}} e^{-\theta\sum_{i=1}^{4}\sqrt{y_i}}$

$\ln L(\theta) = 4\ln\theta - \ln\left(16\prod_{i=1}^{4}\sqrt{y_i}\right) - \theta\sum_{i=1}^{4}\sqrt{y_i}$

$\frac{d\ln L(\theta)}{d\theta} = \frac{4}{\theta} - \sum_{i=1}^{4}\sqrt{y_i}$.

$\frac{d\ln L(\theta)}{d\theta} = 0$ implies $\hat{\theta} = \frac{4}{\sum_{i=1}^{4}\sqrt{y_i}} = \frac{4}{8.766} = 0.456$

5.2.10 $L(\theta) = \prod_{i=1}^{6} \frac{2y_i}{1-\theta^2} = \frac{64\prod_{i=1}^{n} y_i}{(1-\theta^2)^6}$, if $\theta \le y_1, y_2, \ldots, y_n \le 1$ and 0 otherwise. If $\theta > y_{\min}$, then $L(\theta) = 0$. So $\hat{\theta} \le y_{\min}$. Also, to maximize $L(\theta)$, minimize the denominator, which in turn means maximize θ. Thus $\hat{\theta} \ge y_{\min}$. We conclude that $\hat{\theta} = y_{\min}$, which for these data is 0.92.

5.2.18

[redacted] For Y exponential, $E(Y) = 1/\lambda$. Then $1/\lambda = \bar{y}$ implies $\hat{\lambda} = 1/\bar{y}$.

[redacted] $E(Y) = \int_k^\infty y\theta k^\theta \left(\frac{1}{y_i}\right)^{\theta+1} dy = \theta k^\theta \int_k^\infty y^{-\theta} dy = \frac{\theta k}{\theta - 1}$

5.2.20

Setting $\frac{\theta k}{\theta - 1} = \bar{y}$ gives $\hat{\theta} = \bar{y}/(\bar{y} - k)$

[redacted] $E(Y) = \mu$, so $\hat{\mu} = \bar{y}$. $E(Y^2) = \sigma^2 + \mu^2$. Then substitute $\hat{\mu} = \bar{y}$ into the equation for $E(Y^2)$ to

5.2.22

obtain $\hat{\sigma}^2 + \bar{y}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2$ or $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2$

5.4.2 $f_{\hat{\theta}}(u) = n\frac{1}{\theta}\left(\frac{u}{\theta}\right)^{n-1} I_{(0,\theta)}(u) = \frac{n}{\theta^n} u^{n-1} I_{(0,\theta)}(u)$

$P(\hat{\theta} \le x) = F_{\hat{\theta}}(x) = \int_{-\infty}^{x} f_{\hat{\theta}}(u)\,du = \int_{-\infty}^{x} \frac{n}{\theta^n} u^{n-1} I_{(0,\theta)}(u) = \frac{u^n}{\theta^n}\Big|_0^x I_{(0,\theta)}(x) + I_{[\theta,\infty)}(x) = \frac{x^n}{\theta^n} I_{(0,\theta)}(x) + I_{[\theta,\infty)}(x)$

$\theta = 3$

$P(|\hat{\theta} - 3| < 0.2) = P(2.8 < \hat{\theta} < 3.2) = P(\hat{\theta} < 3.2) - P(\hat{\theta} \le 2.8) = P(\hat{\theta} \le 3.2) - P(\hat{\theta} \le 2.8) = F_{\hat{\theta}}(3.2) - F_{\hat{\theta}}(2.8) =$

$1 - F_{\hat{\theta}}(2.8) = 1 - \frac{2.8^n}{3^n} = 1 - \left(\frac{2.8}{3}\right)^n$

a. $n = 6$ $P(|\hat{\theta} - 3| < 0.2) = 1 - \left(\frac{2.8}{3}\right)^6 \approx 0.339$

b. $n = 3$ $P(|\hat{\theta} - 3| < 0.2) = 1 - \left(\frac{2.8}{3}\right)^3 \approx 0.187$

5.4.6 $f_{Y_{\min}}(y) = n\frac{1}{\theta}\left(1 - \frac{y}{\theta}\right)^{n-1}$, so $E(Y_{\min}) = n\frac{1}{\theta}\int_0^\theta y\left(1 - \frac{y}{\theta}\right)^{n-1} dy$

Integration by parts yields $E(Y_{\min}) = \frac{\theta}{n+1}$. An unbiased estimator would be $(n + 1)Y_{\min}$.

[redacted] a) $f_{Y_3'} = 12\left(\frac{Y}{\theta}\right)^2\left(1 - \frac{y}{\theta}\right)\frac{1}{\theta} = \frac{12}{\theta^4}[y^2(\theta - y)]$

5.4.8

$E(Y_3') = \frac{3}{5}\theta$, so the unbiased estimator is $\frac{5}{3}Y_3'$.

b) $\frac{5}{3}Y_3' = \frac{5}{3}18 = 30$

c) Suppose the sample were 10, 14, 18, 31. The estimate for θ is 30, but the largest observation 31 falls outside of the [0, 30] interval.

ST430

HW #2

Solutions

5.4.10

[redacted] $E(Y^2) = \int_0^\theta y^2 \frac{1}{\theta} dy = \frac{\theta^2}{3}$, so $3Y^2$ is unbiased.

5.4.14

[redacted] $f_{Y_{\min}}(y) = nf_Y(y)(1 - F_Y(y))^{n-1} = n\frac{1}{\theta}e^{-y/\theta}[1-(1-e^{-y/\theta})]^{n-1} = n\frac{1}{\theta}e^{-ny/\theta}$. Then

$f_{nY_{\min}}(y) = \frac{1}{n} f_{Y_{\min}}\left(\frac{y}{n}\right) = \frac{1}{n} n \frac{1}{\theta} e^{-n\frac{y}{n}/\theta} = \frac{1}{\theta}e^{-y/\theta}$. $E(nY_{\min}) = \theta$, so $nY_{\min}$ is unbiased. Also,

$\frac{1}{n}\sum_{i=1}^{n} Y_i$ is unbiased since each Y_i is (see Question 5.4.5).

5.4.15

[redacted] $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$, so $E(\hat{\theta}_n) = \frac{n-1}{n}\sigma^2$. This estimator is asymptotically unbiased since

$\lim_{n\to\infty} E(\hat{\theta}_n) = \lim_{n\to\infty} \frac{n-1}{n}\sigma^2 = \sigma^2$.

5.4.20

[redacted] $\text{Var}(\hat{\lambda}_1) = \text{Var}(X_1) = \lambda$. $\text{Var}(\hat{\lambda}_2) = \text{Var}(\bar{X}) = \lambda/n$.

$\text{Var}(\hat{\lambda}_2)/\text{Var}(\hat{\lambda}_1) = (\lambda/n)/\lambda = 1/n$

5.5.2 $\ln f_Y(Y;\theta) = -\ln \theta - Y/\theta$

$$\frac{\partial \ln f_Y(Y;\theta)}{\partial\theta} = -\frac{1}{\theta} + Y/\theta^2$$

$$\frac{\partial^2 \ln f_Y(Y;\theta)}{\partial\theta^2} = \frac{1}{\theta^2} - 2Y/\theta^3$$

$E\left[\frac{\partial^2 \ln f_Y(Y;\theta)}{\partial\theta^2}\right] = \frac{1}{\theta^2} - 2\theta/\theta^3 = \frac{-1}{\theta^2}$, so the Cramer-Rao bound is θ^2/n. Also, $\text{Var}(\hat{\theta}) = Var(\overline{Y}) = \text{Var}(Y)/n = \theta^2/n$, so $\hat{\theta}$ is a best estimator.

5.5.4 $\ln f_Y(Y;\mu) = -\ln\sqrt{2\pi}\sigma - \frac{1}{2}\frac{(Y-\mu)^2}{\sigma^2}$

$$\frac{\partial \ln f_Y(Y;\mu)}{\partial\mu} = \frac{(Y-\mu)}{\sigma^2}$$

$$\frac{\partial^2 \ln f_Y(Y;\mu)}{\partial\mu^2} = \frac{-1}{\sigma^2}$$

$E\left[\frac{\partial^2 \ln f_Y(Y;\mu)}{\partial\mu^2}\right] = \frac{-1}{\sigma^2}$, so the Cramer-Rao bound is σ^2/n. Also, $\text{Var}(\hat{\mu}) = Var(\overline{Y}) = \text{Var}(Y)/n = \sigma^2/n$, so $\hat{\mu}$ is an efficient estimator.

5.5.6 a) Y is a gamma random variable with parameters r and $1/\theta$ so $E(Y) = r\theta$.

Let $\hat{\theta} = \frac{1}{r}\overline{Y} = \frac{1}{rn}\sum_{i=1}^{n} Y_i$. Then $E(\hat{\theta}) = \frac{1}{rn}\sum_{i=1}^{n} E(Y_i) = \frac{1}{rn} nr\theta = \theta$

b) $\ln f_Y(Y;\theta) = -\ln(r-1)! - r\ln\theta + (r-1)\ln Y - Y/\theta$

$$\frac{\partial \ln f_Y(Y;\theta)}{\partial\theta} = -r/\theta + Y/\theta^2$$

$$\frac{\partial^2 \ln f_Y(Y;\theta)}{\partial\theta^2} = r/\theta^2 - 2Y/\theta^3$$

$$E\left[\frac{\partial^2 \ln f_Y(Y;\theta)}{\partial\theta^2}\right] = r/\theta^2 - 2(r\theta)/\theta^3 = -r/\theta^2$$

The Cramer-Rao bound is θ^2/rn.

$\text{Var}(\hat{\theta}) = \left(\frac{1}{rn}\right)^2 \sum_{i=1}^{n} Var(Y_i) = \left(\frac{1}{rn}\right)^2 nr\theta^2 = \theta^2/rn$, so $\hat{\theta}$ is a minimum-variance estimator.

ST 430: Homework #3 Solutions

5.6.2 (a) $F_Y(y) = 1 - e^{-(y-\theta)}$, $\theta \le y$, so

$f_{Y_{\min}}(y) = nf_Y(y)[1 - F_Y(y)]^{n-1} = ne^{-(y-\theta)}[e^{-(y-\theta)}]^{n-1} = ne^{-n(y-\theta)}$

$$\prod_{i=1}^{n} e^{-(y_i-\theta)} = e^{-\sum_{i=1}^{n} y_i} e^{n\theta}) = ne^{-n(y_{\min}-\theta)}\left[\frac{1}{n}e^{ny_{\min}}e^{-\sum_{i=1}^{n} y_i}\right].$$ Thus, $Y_{\min}$ is sufficient by Theorem 5.6.1.

(b) We need to show that the likelihood function given $y_{\max}$ is independent of θ. But the likelihood function is

$$\prod_{i=1}^{n} e^{-(y_i-\theta)} = \begin{cases} e^{\theta} e^{-\sum_{i=1}^{n} y_i} \text{ if } \theta \le y_1, y_2, \ldots y_n \\ 0 \text{ otherwise} \end{cases}$$

Regardless of the value of $y_{\max}$, the expression for the likelihood does depend on $\tilde{\theta}$. If any one of the y_i, other than $y_{\max}$, is less than θ, the expression is 0. Otherwise it is non-zero.

5.6.5 $$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{y_i^2}{\sigma^2}} = \left[(\sigma^2)^{-n/2} e^{-\frac{1}{2}\frac{1}{\sigma^2}\left(\sum_{i=1}^{n} y_i^2\right)}\right][2\pi^{-n/2}],$$ so $\sum_{i=1}^{n} Y_i^2$ is sufficient by Theorem 5.6.1.

5.6.6 $$\prod_{i=1}^{n} \frac{1}{(r-1)!\theta^r} y_i^{r-1} e^{-y_i/\theta} = \frac{1}{[(r-1!]^n} \frac{1}{\theta^{rn}} \left(\prod_{i=1}^{n} y_i\right)^{r-1} = \left[\frac{1}{\theta^{rn}} e^{-\frac{1}{\theta}\sum_{i=1}^{n} y_i}\right] \frac{1}{[(r-1)!]^n} \left(\prod_{i=1}^{n} y_i\right)^{r-1}$$

so $\sum_{i=1}^{n} Y_i$ is a sufficient statistic for θ. So also is $\frac{1}{r}\overline{Y}$. (See Question 5.6.4)

5.6.9 $\lambda e^{-\lambda y} = e^{\ln \lambda - \lambda y} = e^{y(-\lambda) + \ln\lambda}$. Take $K(y) = y$, $p(\lambda) = -\lambda$, $S(y) = 0$, and $q(\lambda) = \ln \lambda$. Then $\sum_{i=1}^{n} Y_i$ is sufficient.

5.6.10 $\theta/(1 + y)^{\theta+1} = e^{\ln \theta - (\theta+1)\ln(1+y)} = e^{[\ln(1+y)](-\theta-1) + \ln \theta}$. Take $K(y) = \ln (1 + y)$, $p(\theta) = -\theta - 1$, and $q(\theta) = \ln \theta$. Then $\sum_{i=1}^{n} K(Y_i) = \sum_{i=1}^{n} \ln(1+Y_i)$ is sufficient for θ.

ST 430: Homework #4 Solutions

5.7.2 Since $\mu = 0$, for each i, $E\left(Y_i^2\right) = \sigma^2$. By the weak law of large numbers demonstrated in Example 5.7.2, $\frac{1}{n}\sum_{i=1}^{n} Y_i^2$ is a consistent estimator of the mean of the Y_i^2, in this case σ^2. However, the proof given in the example requires that $\mathrm{Var}\left(Y_i^2\right) < \infty$. This follows from an application of the moment generating function for the normal distribution.

5.7.4 a) Let $\mu_n = E(\hat{\theta}_n)$. $E\left[(\hat{\theta}_n - \theta)^2\right] = E\left[(\hat{\theta}_n - \mu_n + \mu_n - \theta)^2\right]$

$$= E\left[(\hat{\theta}_n - \mu_n)^2 + (\mu_n - \theta)^2 + 2(\hat{\theta}_n - \mu_n)(\mu_n - \theta)\right]$$

$$= E[(\hat{\theta}_n - \mu_n)^2] + E[(\mu_n - \theta)^2] + 2(\mu_n - \theta)E[(\hat{\theta}_n - \mu_n)]$$

$$= E\left[(\hat{\theta}_n - \mu_n)^2\right] + (\mu_n - \theta)^2 + 0$$

or $E\left[(\hat{\theta}_n - \theta)^2\right] = E\left[(\hat{\theta}_n - \mu_n)^2\right] + (\mu_n - \theta)^2$

The left hand side of the equation tends to 0 by the squared-error consistency hypothesis. Since the two summands on the right hand side are non-negative, each of them must tend to zero also. Thus,

$\lim_{n\to\infty} (\mu_n - \theta)^2 = 0$, which implies $\lim_{n\to\infty} (\mu_n - \theta) = 0$, or $\lim_{n\to\infty} \mu_n = \theta$.

b) By Part (a) $\lim_{n\to\infty} \mu_n - \theta = 0$. For any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| \geq \varepsilon\right) = \lim_{n\to\infty} P\left(\left|(\hat{\theta}_n - \mu_n) - (\mu_n - \theta)\right| \geq \varepsilon\right)$$

$= \lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \mu_n\right| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(\hat{\theta}_n)}{\varepsilon^2}$ by Chebyshev's Inequality.

$\frac{\mathrm{Var}(\hat{\theta}_n)}{\varepsilon^2} = \frac{E[(\hat{\theta}_n - \mu_n)^2]}{\varepsilon^2}$, and by Part (a), $\lim_{n\to\infty} E\left[(\hat{\theta}_n - \mu_n)^2\right] = 0$,

so $\lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| \geq \varepsilon\right) = 0$. Thus, $\hat{\theta}_n$ is consistent.

5.7.5 $E\left[(Y_{\max} - \theta)^2\right] = \int_0^\theta (y - \theta)^2 \frac{n}{\theta}\left(\frac{y}{\theta}\right)^{n-1} dy$

$$= \frac{n}{\theta^n}\int_0^\theta (y^{n+1} - 2\theta y^n + \theta^2 y^{n-1})dy = \frac{n}{\theta^n}\left(\frac{\theta^{n+2}}{n+2} - \frac{2\theta^{n+2}}{n+1} + \frac{\theta^{n+2}}{n}\right)$$

$$= \left(\frac{n}{n+2} - \frac{2n}{n+1} + 1\right)\theta^2$$

Then $\lim_{n\to\infty} E\left[(Y_{\max} - \theta)^2\right] = \lim_{n\to\infty}\left(\frac{n}{n+2} - \frac{2n}{n+1} + 1\right)\theta^2 = 0$ and the estimator is squared error consistent.

5.3.2 The confidence interval is $\left(\bar{y} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{y} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) =$

$\left(70.833 - 1.96\frac{8.0}{\sqrt{6}}, 70.833 + 1.96\frac{8.0}{\sqrt{6}}\right)$ = (64.432, 77.234). Since 80 does not fall within the confidence interval, that men and women metabolize methylmercury at the same rate is not believable.

5.3.4 a) $P(-1.64 < Z < 2.33) = 0.94$, a 94% confidence level.

b) $P(-\infty < Z < 2.58) = 0.995$, a 99.5% confidence level.

a) $P(-1.64 < Z < 0) = 0.45$, a 45% confidence level.

5.3.8 From Theorem 5.3.1, the confidence interval is

$$\left(\frac{179}{220} - 1.64\sqrt{\frac{(179/220)(1-179/220)}{220}}, \quad \frac{179}{220} + 1.64\sqrt{\frac{(179/220)(1-179/220)}{220}}\right)$$

= (0.771, 0.857)

5.3.10 Let p be the probability that a viewer would watch less than a quarter of the advertisements during Super Bowl XXIX. The confidence interval for p is

$$\left(\frac{281}{1015} - 1.64\sqrt{\frac{(281/1015)(1-281/1015)}{1015}}, \quad \frac{281}{1015} + 1.64\sqrt{\frac{(281/1015)(1-281/1015)}{1015}}\right)$$

= (0.254, 0.300)

5.3.12
$$\frac{x}{n} - 0.67\sqrt{\frac{(x/n)(1-x/n)}{n}} = 0.57$$

$$\frac{x}{n} + 0.67\sqrt{\frac{(x/n)(1-x/n)}{n}} = 0.63$$

Adding the two equations gives $2\frac{x}{n} = 1.20$ or $\frac{x}{n} = 0.60$

Substituting the value for $\frac{x}{n}$ into the first equation above gives

$$0.60 - 0.67\sqrt{\frac{(0.60)(1-0.60)}{n}} = 0.57.$$

Solving this equation for n gives $n = 120$.

5.3.16 $g(p) = p - p^2$. $g'(p) = 1 - 2p$. Setting $g'(p) = 0$ gives $p = 1/2$.
$g''(p) = -2$. Since the second derivative is negative at $p = 1/2$, a maximum occurs there. The maximum value of $g(p)$ is $g(1/2) = 1/4$.

ST 430
Spring 2007
Homework #5

Solutions

5.3.18 From Definition 5.3.1, $d = \dfrac{1.96}{2\sqrt{202}} = 0.069$. The sample proportion is 86/202 = 0.426. The largest believable value is 0.426 + 0.069 = 0.495, so we should not accept the notion that the true proportion is as high as 50%.

5.3.24 Take n to be the smallest integer $\geq \dfrac{z_{.005}^2 p(1-p)}{(0.05)^2} = \dfrac{2.58^2(0.40)(0.60)}{(0.05)^2} = 639.01$, so $n = 640$.

5.3.26 a) Take n to be the smallest integer $\geq \dfrac{z_{.075}^2}{4(0.03)^2} = \dfrac{1.44^2}{4(0.03)^2} = 576$.

b) Take n to be the smallest integer $\geq \dfrac{z_{.075}^2 p(1-p)}{(0.03)^2} = \dfrac{1.44^2(0.10)(0.90)}{(0.03)^2} = 207.36$, so let $n =$ 208.

6.2.2 Let μ = true average IQ of students after drinking Brain-Blaster. To test H_0: $\mu = 95$ versus H_1: $\mu \neq 95$ at the $\alpha = 0.06$ level of significance, the null hypothesis should be rejected if $z = \dfrac{\bar{y} - 95}{15/\sqrt{22}}$ is either 1) ≤ -1.88 or 2) ≥ 1.88. Equivalently, H_0 will be rejected if $\bar{y}$ is either **1)** $\leq 95 - (1.88)\dfrac{15}{\sqrt{22}} = 89.0$ or **2)** $\geq 95 + (1.88)\dfrac{15}{\sqrt{22}} = 101.0$.

6.2.4 Assuming there is no reason to suspect that the polymer would shorten a tire's lifetime, the alternative hypothesis should be H_1: $\mu > 32{,}500$. At the $\alpha = 0.05$ level, H_0 should be rejected if the test statistic exceeds $z_{.05} = 1.64$. But $z = \dfrac{33{,}800 - 32{,}500}{4000/\sqrt{15}} = 1.26$, implying that the observed mileage increase is not statistically significant.

6.2.6 By definition, $\alpha = P(29.9 \le \bar{Y} \le 30.1 \mid H_0 \text{ is true}) = P\left(\frac{29.9-30}{6.0/\sqrt{16}} \le \frac{\bar{Y}-30}{6.0/\sqrt{16}} \le \frac{30.1-30}{6.0/\sqrt{16}}\right) =$ $P(-0.07 \le Z \le 0.07) = 0.056$. The interval (29.9, 30.1) is a poor choice for C because it rejects H_0 for the $\bar{y}$ - values that are most compatible with H_0 (that is, closest to $\mu_0 = 30$). Since the alternative is two-sided, H_0 should be rejected if $\bar{y}$ is either

1) $\le 30 - 1.91 \cdot \frac{6.0}{\sqrt{16}} = 27.1$ or **2)** $\ge 30 + 1.91 \cdot \frac{6.0}{\sqrt{16}} = 32.9$.

6.2.10 Let μ = true average blood pressure when taking statistics exams. Test H_0: $\mu = 120$ versus H_1: $\mu > 120$. Given that $\sigma = 12$, $n = 50$ and $\bar{y} = 125.2$, $z = \frac{125.2-120}{12/\sqrt{50}} = 3.06$. The corresponding P-value is approximately 0.001 ($= P(Z \ge 3.06)$), so H_0 would be rejected for any usual choice of α.

6.3.4 The null hypothesis would be rejected if $z = \frac{k - 200(0.45)}{\sqrt{200(0.45)(0.55)}} \ge 1.08$ ($= z_{.14}$). For that to happen, $k \ge 200(0.45) + 1.08 \cdot \sqrt{200(0.45)(0.55)} \doteq 98$.

6.3.6 Let $p = P$(person dies if month preceding birthmonth). Test H_0: $p = \frac{1}{12}$ versus H_1: $p < \frac{1}{12}$. Given that $\alpha = 0.05$, H_0 should be rejected if $z \le -1.64$. In this case, $z = \frac{16 - 348(1/12)}{\sqrt{348(1/12)(11/12)}} =$ -2.52, which suggests that people do not necessarily die randomly with respect to the month in which they were born. More specifically, there appears to be a tendency to "postpone" dying until the next birthday has passed.

6.4.4 For $n = 16$, $\sigma = 4$, and $\alpha = 0.05$, H_0: $\mu = 60$ should be rejected in favor of a two-sided H_1 if either **1)** $\bar{y} \le 60 - 1.96 \cdot \frac{4}{\sqrt{16}} = 58.04$ or **2)** $\bar{y} \ge 60 + 1.96 \cdot \frac{4}{\sqrt{16}} = 61.96$. Then, for arbitrary μ, $1 - \beta = P(\bar{Y} \le 58.04 | \mu) + P(\bar{Y} \ge 61.96 | \mu)$. Selected values of ($\mu$, $1 - \beta$) that would lie on the power curve are listed in the accompanying table.

μ	$1-\beta$
56	0.9793
57	0.8508
58	0.5160
59	0.1700
60	0.05 ($=\alpha$)
61	0.1700
62	0.5160
63	0.8508
64	0.9793

ST 430
Homework #6: Solutions

6.4.6 a) In order for α to be 0.07, $P(60 - \bar{y}^* \le \bar{Y} \le 60 + \bar{y}^* \mid \mu = 60) = 0.07$. Equivalently,

$$P\left(\frac{60-\bar{y}^*-60}{8.0/\sqrt{36}} \le \frac{\bar{Y}-60}{8.0/\sqrt{36}} \le \frac{60+\bar{y}^*-60}{8.0/\sqrt{36}}\right) = P(-0.75\bar{y}^* \le Z \le 0.75\bar{y}^*) = 0.07.$$ But

$P(-0.09 \le Z \le 0.09) = 0.07$, so $0.75\bar{y}^* = 0.09$, which implies that $\bar{y}^* = 0.12$.

b) $1 - \beta = P(\text{reject } H_0 \mid H_1 \text{ is true}) = P(59.88 \le \bar{Y} \le 60.12 \mid \mu = 62) =$

$$P\left(\frac{59.88-62}{8.0/\sqrt{36}} \le Z \le \frac{60.12-62}{8.0/\sqrt{36}}\right) = P(-1.59 \le Z \le -1.41) = 0.0793 - 0.0559 =$$

0.0234.

c) For $\alpha = 0.07$, $\pm z_{\alpha/2} = \pm 1.81$ and H_0 should be rejected if $\bar{y}$ is either

1) $\le 60 - 1.81 \cdot \dfrac{8.0}{\sqrt{36}} = 57.50$ or **2)** $\ge 60 + 1.81 \cdot \dfrac{8.0}{\sqrt{36}} = 62.41$. Suppose $\mu = 62$. Then

$1 - \beta = P(\bar{Y} \le 57.59 \mid \mu = 62) + P(\bar{Y} \ge 62.41 \mid \mu = 62) = P(Z \le -3.31) + P(Z \ge 0.31) =$
$0.0005 + 0.3783 = 0.3788.$

6.4.8 If $n = 45$, H_0 will be rejected when $\bar{y}$ is either **1)** $\le 10 - 1.96 \cdot \dfrac{4}{\sqrt{45}} = 8.83$ or **2)** $\ge 10 + 1.96 \cdot \dfrac{4}{\sqrt{45}} = 11.17$. When $\mu = 12$, $\beta = P(\text{accept } H_0 \mid H_1 \text{ is true}) = P(8.83 \le \bar{Y} \le 11.17 \mid \mu = 12) = P\left(\dfrac{8.83-12}{4/\sqrt{45}} \le Z \le \dfrac{11.17-12}{4/\sqrt{45}}\right) = P(-5.32 \le Z \le -1.39) = 0.0823$. It follows that a sample of size $n = 45$ <u>is</u> sufficient to keep β smaller than 0.20 when $\mu = 12$.

6.4.10 a) $P(\text{Type I error}) = P(\text{reject } H_0 \mid H_0 \text{ is true}) = P(Y \ge 3.20 \mid \lambda = 1) =$

$$\int_{3.20}^{\infty} e^{-y}\,dy = 0.04.$$

b) $P(\text{Type II error}) = P(\text{accept } H_0 \mid H_1 \text{ is true}) =$

$$P\left(Y < 3.20 \middle| \lambda = \frac{4}{3}\right) = \int_0^{3.20} \frac{3}{4} e^{-3y/4}\,dy = \int_0^{2.4} e^{-u}\,du = 0.91.$$

6.4.16 If H_0 is true, $X = X_1 + X_2$ has a binomial distribution with $n = 6$ and $p = \dfrac{1}{2}$. Therefore,

$$\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true}) = P\left(X \ge 5 \middle| p = \frac{1}{2}\right) = \sum_{k=5}^{6} \binom{6}{k} \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{6-k} = 7/2^6 = 0.11.$$

6.4.18 a) $\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true}) = P(X \leq 2 \mid \lambda = 6) = \sum_{k=0}^{2} \frac{e^{-6}6^k}{k!} = 0.062.$

b) $\beta = P(\text{accept } H_0 \mid H_1 \text{ is true}) = P(X \geq 3 \mid \lambda = 4) = 1 - P(X \leq 2 \mid \lambda = 4) =$
$1 - \sum_{k=0}^{2} \frac{e^{-4}4^k}{k!} = 1 - 0.238 = 0.762.$

6.4.20 $\beta = P(\text{accept } H_0 \mid H_1 \text{ is true}) = P(Y < \ln 10 \mid \lambda) = \int_0^{\ln 10} \lambda e^{-\lambda y} dy = 1 - e^{-\lambda \ln 10} = 1 - 10^{-\lambda}.$

6.4.21 $\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true}) = P(Y_1 + Y_2 \leq k \mid \theta = 2)$. When H_0 is true, Y_1 and Y_2 are uniformly distributed over the square defined by $0 \leq Y_1 \leq 2$ and $0 \leq Y_2 \leq 2$, so the joint pdf of Y_1 and Y_2 is a plane parallel to the Y_1Y_2-axis at height $\frac{1}{4}\left(= f_{Y_1}(y_1) \cdot f_{Y_2}(y_2) = \frac{1}{2} \cdot \frac{1}{2}\right)$. By geometry, α is the volume of the triangular wedge in the lower left-hand corner of the square over which Y_1 and Y_2 are defined. The hypotenuse of the triangle in the Y_1Y_2-plane has the equation $y_1 + y_2 = k$. Therefore, α = area of triangle × height of wedge = $\frac{1}{2} \cdot k \cdot k \cdot \frac{1}{4} = k^2/8$. For α to be 0.05, $k = \sqrt{0.04} = 0.63$.

6.4.22 $\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true}) = P(Y_1Y_2 \leq k^* \mid \theta = 2)$. If $\theta = 2$, the joint pdf of Y_1 and Y_2 is the horizontal plane $f_{Y_1,Y_2}(y_1, y_2) = \frac{1}{4}$, $0 \leq y_1 \leq 2$, $0 \leq y_2 \leq 2$. Therefore, $\alpha = P(Y_1Y_2 \leq k^* \mid \theta = 2) =$
$2 \cdot \frac{k^*}{2} \cdot \frac{1}{4} + \int_{k^*/2}^{2} \int_0^{k^*/y_1} \frac{1}{4} dy_2 dy_1 = \frac{k^*}{4} + \int_{k^*/2}^{2} \frac{k^*}{4y_1} dy_1 = \frac{k^*}{4} + \left(\frac{k^*}{4} \ln y_1 \Big|_{k^*/2}^{2} \right) =$
$\frac{k^*}{4} + \frac{k^*}{4} \ln 2 - \frac{k^*}{4} \ln \frac{k^*}{2}$. By trial and error, $k^* = 0.087$ makes $\alpha = 0.05$.

Note: The k* value in 6.4.22 is incorrect. The correct value is approximately 0.0349.

ST 430
Homework #7 Solutions

6.5.2 Let $y = \sum_{i=1}^{10} y_i$. Then $L(\hat{\omega}) = \prod_{i=1}^{10} \lambda_o e^{-\lambda_o y_i} = \lambda_0^{10} e^{-\lambda_0 \sum_{i=1}^{10} y_i} = \lambda_0^{10} e^{-\lambda_0 y}$. Also, $L(\lambda) = \prod_{i=1}^{10} \lambda e^{-\lambda y_i} = \lambda^{10} e^{-\lambda y}$, so $\ln L(\lambda) = 10 \ln \lambda - \lambda y$ and $\frac{d \ln L(\lambda)}{d\lambda} = \frac{10}{\lambda} - y$. Setting the latter equal to 0 implies that the maximum likelihood estimate for λ is $\lambda_e = \frac{10}{y}$. Therefore,

$L(\hat{\Omega}) = \left(\frac{10}{y}\right)^{10} e^{-\left(\frac{10}{y}\right)y} = (10/y)^{10} e^{-10}$. The generalized likelihood ratio, then, is the quotient $\lambda_0^{10} e^{-\lambda_0 y} / (10/y)^{10} e^{-10} = (\lambda_0 e/10)^{10} y^{10} e^{-\lambda_0 y}$. It follows that H_0 should be rejected if $\lambda = y^{10} e^{-\lambda_0 y} \le \lambda^*$, where λ^* is chosen so that $\int_0^{\lambda^*} f_\lambda(\lambda \mid H_0 \text{ is true}) d\lambda = 0.05$.

6.5.3 $L(\hat{\omega}) = \prod_{i=1}^{n} (1/\sqrt{2\pi}) e^{-\frac{1}{2}(y_i - \mu_0)^2} = (2\pi)^{-n/2} e^{-\frac{1}{2}\sum_{i=1}^{n}(y_i - \mu_0)^2}$. Since $\bar{y}$ is the maximum likelihood estimate for μ (recall the first derivative taken in Example 5.2.4),

$L(\hat{\Omega}) = (2\pi)^{-n/2} e^{-\frac{1}{2}\sum_{i=1}^{n}(y_i - \bar{y})^2}$. Here the generalized likelihood ratio reduces to

$\lambda = L(\hat{\omega}) / L(\hat{\Omega}) = e^{-\frac{1}{2}((\bar{y} - \mu_0)/(1/\sqrt{n}))^2}$. The null hypothesis should be rejected if

$e^{-\frac{1}{2}((\bar{y} - \mu_0)/(1/\sqrt{n}))^2} \le \lambda^*$ or, equivalently, if $|(\bar{y} - \mu_0)|/(1/\sqrt{n}) > \lambda^{**}$, where values for λ^{**} come from the standard normal pdf, $f_Z(z)$.

6.5.4 To test H_0: $\mu = \mu_0$ versus H_1: $\mu = \mu_1$, the "best" critical region would consist of all those samples for which $\prod_{i=1}^{n} (1/\sqrt{2\pi})^n e^{-\frac{1}{2}(y_i - \mu_0)^2} / \prod_{i=1}^{n} (1/\sqrt{2\pi})^n e^{-\frac{1}{2}(y_i - \mu_1)^2} \le k$. Equivalently, H_0 should be rejected if $\sum_{i=1}^{n} (y_i - \mu_0)^2 - \sum_{i=1}^{n} (y_i - \mu_1)^2 > 2\ln k$. Simplified, the latter becomes

$2(\mu_1 - \mu_0) \sum_{i=1}^{n} y_i > 2\ln k + n\left(\mu_1^2 - \mu_0^2\right)$. Consider the case where $\mu_1 < \mu_0$. Then $\mu_1 - \mu_0 < 0$, and the decision rule reduces to rejecting H_0 when $\bar{y} < \frac{2\ln k + n\left(\mu_1^2 - \mu_0^2\right)}{2n(\mu_1 - \mu_0)}$.

7.3.2 Subsituting $\frac{n}{2}$ and $\frac{1}{2}$ for r and λ, respectively, in the moment-generating function for a gamma pdf gives $M_{\chi_n^2}(t) = (1-2t)^{-n/2}$. Also, $M_{\chi_n^2}^{(1)}(t) = (-n/2)(1-2t)^{-n/2-1}(-2) =$ $n(1-2t)^{-n/2-1}$ and $M_{\chi_n^2}^{(2)}(t) = \left(-\frac{n}{2}-1\right)(n)(1-2t)^{-n/2-2}(-2) = (n^2+2n)\cdot(1-2t)^{-n/2-2}$, so $M_{\chi_n^2}^{(1)}(0) = n$ and $M_{\chi_n^2}^{(2)}(0) = n^2+2n$. Therefore, $E(\chi_n^2) = n$ and $\text{Var}\left(\chi_n^2\right) =$ $n^2+2n-n^2 = 2n$.

7.3.4 Let $Y = \frac{(n-1)S^2}{\sigma^2}$. Then $\text{Var}(Y) = \text{Var}\left(\chi_{n-1}^2\right) = 2(n-1) = \frac{(n-1)^2\text{Var}(S^2)}{\sigma^4}$. If follows that $\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$.

7.3.8 $P\left(2.51 < \frac{V/7}{U/9} < 3.29\right) = P(2.51 < F_{7,9} < 3.29) = P(F_{7,9} < 3.29) - P(F_{7,9} \leq 2.51) = 0.95 - 0.90 =$ 0.05. But $P(3.29 < F_{7,9} < 4.20) = 0.975 - 0.95 = 0.025$.

7.3.11 $F = \frac{V/m}{U/n}$, where U and V are independent χ^2 random variables with m and n degrees of freedom, respectively. Then $\frac{1}{F} = \frac{U/n}{V/m}$, which implies that $\frac{1}{F}$ has an F distribution with n and m degrees of freedom.

7.3.12 If $P(a \leq F_{m,n} \leq b) = q$, then $P\left(a \leq \frac{1}{F_{n,m}} \leq b\right) = q = P\left(\frac{1}{b} \leq F_{n,m} \leq \frac{1}{a}\right)$. From Appendix Table A.4, $P(0.052 \leq F_{2,8} \leq 4.46) = 0.95$. Also, $P(0.234 \leq F_{8,2} \leq 19.4) = 0.95$. But $\frac{1}{4.46} = 0.224$ and $\frac{1}{0.052} = 19.23 \doteq 19.4$.

ST 430
Homework #8 Solutions

7.4.2 a) 2.508 b) −1.079 c) 1.7056 d) 4.3027

7.4.4 Since $\frac{\bar{Y}-27.6}{S/\sqrt{9}}$ is a Student t random variable with 8 df, $P\left(-1.397 \le \frac{\bar{Y}-27.6}{S/\sqrt{9}} \le 1.397\right) =$ 0.80 and $P\left(-1.8595 \le \frac{\bar{Y}-27.6}{S/\sqrt{9}} \le 1.8595\right) = 0.90$ (see Appendix Table A.2).

7.4.6 $P(90.6 - k(S) \le \bar{Y} \le 90.6 + k(S)) = 0.99 =$
$P\left(\frac{90.6-k(S)-90.6}{S/\sqrt{20}} \le \frac{\bar{Y}-90.6}{S/\sqrt{20}} \le \frac{90.6+k(S)-90.6}{S/\sqrt{20}}\right) = P\left(\frac{k(S)}{S/\sqrt{20}} \le T_{19} \le \frac{k(S)}{S/\sqrt{20}}\right) =$
$P(-2.8609 \le T_{19} \le 2.8609)$, so $\frac{k(S)}{S/\sqrt{20}} = 2.8609$, implying that $k(S) = \frac{2.8609 \cdot S}{\sqrt{20}}$.

7.4.8 Given that $n = 7$, $t_{\alpha/2,n-1} = t_{.025,6} = 2.4469$. Here $\sum_{i=1}^{n} y_i = 12{,}808$ and $\sum_{i=1}^{n} y_i^2 = 26{,}540{,}436$ so $\bar{y}$ $= \frac{1}{7}(12{,}808) = 1829.71$ and $s = \sqrt{\frac{7(26{,}540{,}436)-(12{,}808)^2}{7(6)}} = 719.43$.
The confidence interval is $\left(1829.71 - 2.4469\frac{719.43}{\sqrt{7}}, 1829.71 + 2.4469\frac{719.43}{\sqrt{7}}\right)$
= ($1164.35, $2495.07).

7.4.10 Let μ = true average daily fat intake of males in the age group 25 to 34. Since $\bar{y} = \frac{1}{10}(1101.3) = 110.13$, $s = \sqrt{\frac{10(128{,}428.67)-(1101.3)^2}{10(9)}} = 28.17$, and $t_{.05,9} = 1.8331$, the 90% confidence interval for μ is $\left(110.13 - 1.8331 \cdot \frac{28.17}{\sqrt{10}}, 110.13 + 1.8331 \cdot \frac{28.17}{\sqrt{10}}\right)$, which reduces to (93.80, 126.46).

7.4.12 Given that $n = 16$, $t_{\alpha/2,n-1} = t_{.025,15} = 2.1315$, so $\left(\bar{y} - 2.1315 \cdot \frac{s}{\sqrt{16}}, \bar{y} + 2.1315 \cdot \frac{s}{\sqrt{16}}\right) =$ (44.7, 49.9). Therefore, $49.9 - 44.7 = 5.2 = 2(2.1315) \cdot \frac{s}{\sqrt{16}}$, implying that $s = 4.88$. Also, because the confidence interval is centered around the sample mean, $\bar{y} = \frac{44.7+49.9}{2} = 47.3$.

7.4.19 Let μ = true average GMAT increase earned by students taking the review course. The hypotheses to be tested are H_0: $\mu = 40$ versus H_1: $\mu < 40$. Here, $\sum_{i=1}^{15} y_i = 556$ and $\sum_{i=1}^{15} y_i^2 = 20{,}966$, so $\bar{y} = \frac{556}{15} = 37.1$, $s = \sqrt{\frac{15(20{,}966) - (556)^2}{15(14)}} = 5.0$, and $t = \frac{37.1 - 40}{5.0/\sqrt{15}} = -2.25$. Since $-t_{.05,14} = -1.7613$, H_0 should be rejected at the $\alpha = 0.05$ level of significance, suggesting that the MBAs 'R Us advertisement may be fraudulent.

7.5.2 a) 0.95 b) 0.90 c) $0.975 - 0.025 = 0.95$
d) 0.99

7.5.6 $P\left(\frac{S^2}{\sigma^2} < 2\right) = P\left(\frac{(n-1)S^2}{\sigma^2} < 2(n-1)\right) = P\left(\chi^2_{n-1} < 2(n-1)\right)$. Values from the 0.95 column in a χ^2 table show that for each $n < 8$, $P\left(\chi^2_{n-1} < 2(n-1)\right) < 0.95$. But for $n = 9$, $\chi^2_{.95,8} = 15.507$, which means that $P\left(\chi^2_8 < 16\right) > 0.95$.

7.5.8 If $n = 19$ and $\sigma^2 = 12.0$, $\frac{18S^2}{12.0}$ has a χ^2 distribution with 18 df, so

$$P\left(8.231 \le \frac{18S^2}{12.0} \le 31.526\right) = 0.95 = P(5.49 \le S^2 \le 21.02).$$

ST 430
Homework #9: Solutions

9.2.2 $s_p = \sqrt{\dfrac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}} = \sqrt{\dfrac{3(267^2)+3(224^2)}{4+4-2}} = 246.44$

$t = \dfrac{\bar{x}-\bar{y}}{s_p\sqrt{1/n+1/m}} = \dfrac{1133.0-1013.5}{246.44\sqrt{1/4+1/4}} = 0.69$

Since $-t_{.025,6} = -2.4469 < t = 0.69 < t_{.025,6} = 2.4469$, accept H_0.

9.2.4 $s_p = \sqrt{\dfrac{5(15.1^2)+8(8.1^2)}{6+9-2}} = 11.317$

$t = \dfrac{70.83-79.33}{11.317\sqrt{1/6+1/9}} = -1.43$

Since $-t_{.005,13} = -3.0123 < t = -1.43 < t_{.005,13} = 3.0123$, accept H_0.

9.2.8 The solution given in the manual is incorrect. The means and standard deviations are given in different units, which must be adjusted so that the t statistic is unitless. This solution converts all the units to minutes. Alternately, all units could be converted to hours.

H_0: $u_x - 1 = u_y$
H_1: $u_x - 1 < u_y$

$s_p = \text{sqrt}((10 * 12^2 + 10 * 16^2)/(10 + 10 - 2)) = \text{sqrt}(200) = 14.1421$

$t = ((2.1 - 1 - 1.6) * 60)/(14.1421 * \text{sqrt}(1/10 + 1/10)) = -4.743$

Reject H_0 if $t < -t_{0.05,18} = -1.7341$ $\quad t < -t_{0.05,18}$ $\quad$ Reject H_0 and conclude H_1.

9.2.9 a) Reject H_0 if $t > t_{.005,15} = 2.9467$, so we seek the smallest value of $|\bar{x}-\bar{y}|$ such that

$t = \dfrac{|\bar{x}-\bar{y}|}{s_p\sqrt{1/n+1/m}} = \dfrac{|\bar{x}-\bar{y}|}{15.3\sqrt{1/6+1/11}} > 2.9467$, or $|\bar{x}-\bar{y}| > (15.3)(0.508)(2.9467)$
$= 22.90$

b) Reject H_0 if $t > t_{.05,19} = 1.7291$, so we seek the smallest value of $\bar{x}-\bar{y}$ such that

$t = \dfrac{\bar{x}-\bar{y}}{s_p\sqrt{1/n+1/m}} = \dfrac{\bar{x}-\bar{y}}{214.9\sqrt{1/13+1/8}} > 1.7291$, or $\bar{x}-\bar{y} > (214.9)(0.44936)(1.7291)$
$=166.97$

9.3.4 The observed $F = 3.18^2/5.67^2 = 0.315$. Since $F_{.025,9,9} = 0.248 < 0.315 < 4.03 = F_{.975,9,9}$, we can accept H_0 that the variances are equal.

9.3.6 The observed $F = 398.75/274.52 = 1.453$. Let $\alpha = 0.05$. The critical values are $F_{.025,13,11}$ and $F_{.975,13,11}$. These values are not in Table A.4, so approximate them by $F_{.025,12,11} = 0.301$ and $F_{.975,12,11} = 3.47$. Since $0.301 < 1.453 < 3.47$, accept H_0 that the variances are equal. Theorem 9.2.2 is appropriate.

9.4.2 $\hat{p} = \dfrac{x+y}{n+m} = \dfrac{66+93}{423+423} = 0.188$

$$z = \frac{\frac{x}{n} - \frac{y}{m}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{\hat{p}(1-\hat{p})}{m}}} = \frac{\frac{66}{423} - \frac{93}{423}}{\sqrt{\frac{0.188(0.812)}{423} + \frac{0.188(0.812)}{423}}} = -2.38$$

For this experiment, H_0: $p_X = p_Y$ and H_1: $p_X < p_Y$. Since $z = -2.38 < -1.64 = -z_{.05}$, reject H_0.

9.4.4 $\hat{p} = \dfrac{53+705}{91+1117} = 0.627$

$$z = \frac{\frac{53}{91} - \frac{705}{1117}}{\sqrt{\frac{0.627(0.373)}{91} + \frac{0.627(0.373)}{1117}}} = -0.92$$

Since $-2.58 < z = -0.92 < 2.58 = z_{.005}$, accept H_0 at the 0.01 level of significance.

9.4.6 $\hat{p} = \dfrac{2915+3086}{4134+4471} = 0.697$

$$z = \frac{\frac{2915}{4134} - \frac{3086}{4471}}{\sqrt{\frac{0.697(0.303)}{4134} + \frac{0.697(0.303)}{4471}}} = 1.50$$

Since $-1.96 < z = 1.50 < 1.96 = z_{.025}$, accept H_0 at the 0.05 level of significance.

9.4.8 $\hat{p} = \dfrac{78+50}{300+200} = 0.256$

$$z = \frac{\frac{78}{300} - \frac{50}{200}}{\sqrt{\frac{0.256(0.744)}{300} + \frac{0.256(0.744)}{200}}} = 0.25.$$ In this situation, H_1 is $p_X > p_Y$.

Since $z = 0.25 < 1.64 = z_{.05}$, accept H_0. The player is right.

9.5.2 The center of the confidence interval is $\bar{x} - \bar{y} = 6.7 - 5.6 = 1.1$. $s_p = \sqrt{\dfrac{8(0.54^2) + 6(0.36^2)}{14}} = 0.47$. The radius is $t_{\alpha/2, n+m-2} s_p \sqrt{\dfrac{1}{n} + \dfrac{1}{m}} = 1.7613(0.47)\sqrt{\dfrac{1}{9} + \dfrac{1}{7}} = 0.42$. The confidence interval is (1.1 – 0.42, 1.1 + 0.42) = (0.68, 1.52). Since 0 is not in the interval, we can reject the null hypothesis that $\mu_X = \mu_Y$.

9.5.8 The confidence interval is $\left(\dfrac{s_X^2}{s_Y^2} F_{.025,5,7}, \dfrac{s_X^2}{s_Y^2} F_{.975,5,7}\right) = \left(\dfrac{137.4}{340.3}(0.146), \dfrac{137.4}{340.3}(5.29)\right)$
= (0.06, 2.14)
Since the confidence interval contains 1, we can accept H_0 that the variances are equal, and Theorem 9.2.1 applies.

9.5.12 The center of the confidence interval is $\dfrac{x}{n} - \dfrac{y}{m} = \dfrac{106}{3522} - \dfrac{13}{115} = -0.083$. The radius is

$$z_{.025}\sqrt{\frac{\left(\frac{x}{n}\right)\left(1-\frac{x}{n}\right)}{n} + \frac{\left(\frac{y}{m}\right)\left(1-\frac{y}{m}\right)}{m}} = 1.97\sqrt{\frac{\left(\frac{106}{3522}\right)\left(1-\frac{106}{3522}\right)}{3522} + \frac{\left(\frac{13}{115}\right)\left(1-\frac{13}{115}\right)}{115}} = 0.058$$

The 95% confidence interval is (–0.083 – 0.058, –0.083 + 0.058) = (–0.141, –0.025)
Since the confidence interval lies to the left of 0, there is statistical evidence that the suicide rate among women members of the American Chemical Society is higher.

ST 430
Homework #10 Solutions

10.2.2 Let X_1 = number of round and yellow phenotypes, X_2 = number of round and green phenotypes, and so on. Then $P(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1) =$
$$\frac{4!}{1!1!1!1!}\left(\frac{9}{16}\right)^1\left(\frac{3}{16}\right)^1\left(\frac{3}{16}\right)^1\left(\frac{1}{16}\right)^1 = 0.0297.$$

10.2.4 Let Y denote a recruit's IQ and let X_i denote the number of recruits in class i, $i = 1, 2, 3$. Then $p_1 = P(\text{class I}) = P(Y < 90) = P\left(Z < \frac{90-100}{16}\right) = 0.2643$, $p_2 = P(\text{class II}) = P(90 \le Y \le 110) = P\left(\frac{90-100}{16} \le Z \le \frac{110-100}{16}\right) = 0.4714$, and $p_3 = P(\text{class III}) = P(Y > 110) = 1 - p_1 - p_2 = 0.2643$. From Theorem 10.2.1, $P(X_1 = 2, X_2 = 4, X_3 = 1) = \frac{7!}{2!4!1!}(0.2643)^2(0.4714)^2(0.2643)^1 = 0.0957$.

10.2.8 $M_{X_1,X_2,X_3}(t_1,t_2,t_3) = \sum\sum\sum e^{t_1k_1+t_2k_2+t_3k_3} \cdot \frac{n!}{k_1!k_2!k_3!} \cdot p_1^{k_1}p_2^{k_2}p_3^{k_3} =$
$\sum\sum\sum \frac{n!}{k_1!k_2!k_3!}\left(p_1e^{t_1}\right)^{k_1}\left(p_2e^{t_2}\right)^{k_2}\left(p_3e^{t_3}\right)^{k_3}$, where the summation extends over all the values of (k_1, k_2, k_3) such that $k_i \ge 0$, $i = 1, 2, 3$ and $k_1 + k_2 + k_3 = n$. Recall Newton's binomial expansion. Applied here, it follows that the triple sum defining the moment-generating function for (X_1, X_2, X_3) can also be written $\left(p_1e^{t_1} + p_2e^{t_2} + p_3e^{t_3}\right)^n$.

10.2.10 The log of the likelihood vector $(k_1, k_2, \ldots, k_t)$ is $\log L = \log\ p_1^{k_1}p_2^{k_2}\ldots p_t^{k_t} = k_1 \log p_1 + k_2 \log p_2 + \ldots + k_t \log p_t$, where the p_i's are constrained by the condition that $\sum_{i=1}^{t} p_i = 1$. Finding the MLE for the p_i's can be accomplished using Lagrange multipliers. Differentiating $\log L - \lambda\sum_{i=1}^{t} p_i$ with respect to each p_i gives $\frac{\partial}{\partial p_i}\left[\log L - \lambda\sum_{i=1}^{t} p_i\right] = \frac{k_i}{p_i} - \lambda$, $i = 1, 2, \ldots, t$. But these derivatives equal 0 only if $\frac{k_i}{p_i} = \lambda$ for all i. The latter equations, together with the fact that $\sum_{i=1}^{t} p_i = 1$, imply that $\hat{p}_i = \frac{k_i}{n}$, $i = 1, 2, \ldots, t$.

10.3.2 If the hypergeometric model applies, $\pi_1 = P(0 \text{ whites are drawn}) = \binom{4}{0}\binom{6}{2}\bigg/\binom{10}{2} = \frac{15}{45}$, $\pi_2 = P(1 \text{ white is drawn}) = \binom{4}{1}\binom{6}{1}\bigg/\binom{10}{2} = \frac{24}{45}$, and $\pi_3 = P(2 \text{ whites are drawn}) = \binom{4}{2}\binom{6}{0}\bigg/\binom{10}{2} = \frac{6}{45}$. Let p_1, p_2, and p_3 denote the actual probabilities of drawing 0, 1, and 2 white chips, respectively. To test H_0: $p_1 = \frac{15}{45}$, $p_2 = \frac{24}{45}$, $p_3 = \frac{6}{45}$ versus H_1: at least one $p_i \neq \pi_i$, reject H_0 if $d \geq \chi^2_{1-\alpha,k-1} = \chi^2_{.90,2} = 4.605$.
Here, $d = \frac{(35 - 100(15/45))^2}{100(15/45)} + \frac{(55 - 100(24/45))^2}{100(24/45)} + \frac{(10 - 100(6/45))^2}{100(6/45)} = 0.96$, so H_0 (and the hypergeometric model) would not be rejected.

10.3.4 If births occur randomly in time, then $\pi_1 = P(\text{baby is born between midnight and 4 A.M.}) = \frac{1}{6}$ and $\pi_2 = P(\text{baby is born at a “convenient” time}) = 1 - \pi_1 = \frac{5}{6}$. Let p_1 and p_2 denote the actual probabilities of birth during those two time periods. The null hypothesis to be tested is H_0: $p_1 = \frac{1}{6}$, $p_2 = \frac{5}{6}$. At the $\alpha = 0.05$ level of significance, H_0 should be rejected if $d \geq \chi^2_{.95,1} = 3.841$. Given that $n = 2650$ and that X_1 = number of births between midnight and 4 A.M. = 494, it follows that $d = \frac{(494 - 2650(1/6))^2}{2650(1/6)} + \frac{(2156 - 2650(5/6))^2}{2650(5/6)} = 7.44$. Since the latter exceeds 3.841, we reject the hypothesis that births occur uniformly in all time periods.

10.3.6 In the terminology of Theorem 10.3.1, $X_1 = 1383$ = number of schizophrenics born in first quarter and X_2 = number of schizophrenics born after the first quarter. By assumption, $n\pi_1 = 1292.1$ and $n\pi_2 = 3846.9$ (where $n = 5139$). The null hypothesis that birth month is unrelated to schizophrenia is rejected if $d \geq \chi^2_{.95,1} = 3.841$. But $d = \frac{(1383 - 1292.1)^2}{1292.1} + \frac{(3756 - 3846.9)^2}{3846.9} = 8.54$, so H_0 is rejected, suggesting that month of birth may, indeed, be a factor in the incidence of schizophrenia.

10.3.10 Let $p_i = P(\text{horse starting in post position } i \text{ wins})$, $i = 1, 2, \ldots, 8$. One relevant null hypothesis to test would be that p_i is not a function of i—that is, H_0: $p_1 = p_2 = \ldots = p_8 = \frac{1}{8}$ versus H_1: at least one $p_i \neq \frac{1}{8}$. If $\alpha = 0.05$, H_0 should be rejected if $d \geq \chi^2_{.95,7} = 14.067$. Each $E(X_i)$ in this case is $144 \cdot \frac{1}{8} = 18.0$, so $d = \frac{(32 - 18.0)^2}{18.0} + \frac{(21 - 18.0)^2}{18.0} + \ldots + \frac{(11 - 18.0)^2}{18.0} = 18.72$. Since $18.72 \geq \chi^2_{.95,7}$, we reject H_0 (which is not surprising because faster horses are typically awarded starting positions close to the rail).

10.3.12 Let the random variable Y denote the prison time served by someone convicted of grand theft auto. In the accompanying table is the frequency distribution for a sample of 50 y_i's, together with expected frequencies based on the null hypothesis that $f_Y(y) = \frac{1}{9}y^2$, $0 \le y \le 3$. For example, $E(X_1) = 50 \cdot \pi_1 = 50\int_0^1 \frac{1}{9}y^2 dy = 1.85$. Combining the first two intervals (because $E(X_1) < 5$) yields $k = 2$ final classes, so H_0: $f_Y(y) = \frac{1}{9}y^2$, $0 \le y \le 3$ should be rejected if $d \ge \chi^2_{.95,1} = 3.841$. But $d = \frac{(24-14.81)^2}{14.81} + \frac{(26-35.19)^2}{35.19} = 8.10$, implying that the proposed quadratic pdf does not provide a good model for describing prison time.

Prison time, y	Freq.	π_i	$E(X_i)$	
$0 \le y < 1$	8	1/27	1.85	$\}$ 14.81
$1 \le y < 2$	16	7/27	12.96	
$2 \le y < 3$	26	19/27	35.19	
	50	1	50.00	

10.4.2 For the Poisson pdf, $\hat{\lambda} = \frac{59(0)+27(1)+9(2)+1(3)}{96} = 0.50$ so the hypotheses being tested are H_0: $P(i \text{ vacancies}) = e^{-0.50}(0.50)^i/i!$, $i = 0, 1, 2, \ldots$ vs. H_1: $P(i \text{ vacancies}) \ne e^{-0.50}(0.50)^i/i!$, $i = 0, 1, 2, \ldots$ As the table indicates, the original frequency distribution needs to have several classes combined because the expected frequencies are too small.

No. of vacancies, i	No. of years	$\hat{p}_i$	$96 \cdot \hat{p}_i$	
0	59	0.607	58.27	
1	27	0.303	29.09	
2	9	0.076	7.30	$\}$ 8.65
3	1	0.013	1.25	
4+	0	0.001	0.10	
	96	1.000	96.00	

10.4.4 Let $\hat{\lambda} = \frac{109(0)+65(1)+22(2)+3(3)+4(1)}{200} = 0.61$. Then the model to be fit under H_0 is the Poisson pdf, $p_X(i) = e^{-0.61}(0.61)^i/i!$, $i = 0, 1, 2, \ldots$ Using $t = 4$ final classes (the combined "4.8" is close enough to 5 for the χ^2 approximation to be adequate), we should reject H_0 if $d_1 \ge \chi^2_{.99,4-1-1} = 9.210$. In the table, the observed and expected frequencies are in excellent agreement, so d_1 will be very small (and the Poisson model will not be rejected). Specifically, $d_1 = \frac{(109-108.7)^2}{108.7} + \frac{(65-66.3)^2}{66.3} + \frac{(22-20.2)^2}{20.2} + \frac{(4-4.8)^2}{4.8} = 0.32$.

No. of Deaths, i	Freq.	$\hat{p}_i$	$200 \cdot \hat{p}_i$	
0	109	0.5434	108.7	
1	65	0.3314	66.3	
2	22	0.1011	20.2	
3	3	0.0206	4.1	$\}$ 4.8
4+	1	0.0035	0.7	
	200	1.0000	200.0	

10.3.9 Let the random variable X denote the length of a World Series. Then $P(X = 4) = \pi_1 = P($AL wins in 4$) + P($NL wins in 4$) = 2 \cdot P($AL wins in 4$) = 2\left(\frac{1}{2}\right)^4 = \frac{1}{8}$. Similarly, $P(X = 5) = \pi_2 = 2 \cdot P($AL wins in 5$) = 2 \cdot P($AL wins exactly 3 of first 4 games$) \cdot P($AL wins 5th game$) = 2 \cdot \binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1 \cdot \frac{1}{2} = \frac{1}{4}$. Also, $P(X = 6) = \pi_3 = 2 \cdot P($AL wins exactly 3 of first 5 games$) \cdot P($AL wins 6th game$) = 2 \cdot \binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right) = \frac{5}{16}$, and $P(X = 7) = \pi_4 = 1 - P(X = 4) - P(X = 5) - P(X = 6) = \frac{5}{16}$. Listed in the table is the information necessary for calculating the goodness-of-fit statistic d. The "Bernoulli model" is rejected if $d \geq \chi^2_{.90,3} = 6.251$. For these data, $d = \frac{(9-6.25)^2}{6.25} + \frac{(11-12.50)^2}{12.50} + \frac{(8-15.625)^2}{15.625} + \frac{(22-15.625)^2}{15.625} = 7.71$, so H_0 is rejected.

Number of games	Number of years	$50 \cdot \pi_i$
4	9	6.25
5	11	12.50
6	8	15.625
7	22	15.625
	50	50.000

ST 430
Homework #11

10.4.6 Below is the set of observed and expected frequencies, the latter based on the null hypothesis that the states' SAT scores are normally distributed with $\bar{y}$ = 949.4 and s = 68.4. With t = 4 classes and two estimated parameters, H_0 should be rejected if $d_1 \geq \chi^2_{.95,4-1-2}$ = 3.841. For these data,

$$d_1 = \frac{(18-12.2)^2}{12.2} + \frac{(10-13.7)^2}{13.7} + \frac{(6-13.5)^2}{13.5} + \frac{(17-11.6)^2}{11.6} = 10.44,$$ suggesting that the normality assumption is unwarranted.

Range	Frequency	Probability	Expected Frequency
≤ 900	18	0.2389	12.2
901-950	10	0.2691	13.7
950-1000	6	0.2654	13.6
≥1001	17	0.2266	11.6
	51	1.0000	51.0

10.4.8 The table below gives the observed frequencies for 100 supposedly random choices from the [0, 1] interval, as well as the expected values of 10 for each category. With 10 classes and no parameters estimated, H_0 should be rejected if $d_1 \geq \chi^2_{.95,10-1}$ = 16.919. For these data,

$$d_1 = \frac{(12-10)^2}{10} + \frac{(9-10)^2}{10} + \cdots + \frac{(8-10)^2}{10} = 1.8$$

We can accept the null hypothesis that the data come from a uniform pdf over [0, 1].

Interval	Observed Frequency	Expected Frequency
.000-.099	12	10
.100-.199	9	10
.200-.299	11	10
.300-.399	8	10
.400-.499	11	10
.500-.599	10	10
.600-.699	11	10
.700-.799	9	10
.800-.899	11	10
.900-.999	8	10
	100	100

10.4.10 Take $\hat{\lambda}$ to be the mean of the data or 0.363. The model to be fit, then, is the Poisson pdf with parameter 0.363. The table gives the observed frequencies, the estimated probabilities and the estimated frequencies. Note that the last three classes should be collapsed, giving a total of three classes. With one parameter estimated, we should reject H_0 if $d_1 \geq \chi^2_{.95,3-1-1} = 3.841$. The data gives

$$d_1 = \frac{(82-78.6)^2}{78.6} + \frac{(25-28.5)^2}{28.5} + \frac{(6-5.9)^2}{5.9} = 0.58$$

and we can accept the Poisson model for these data.

No. of years	Frequency	$\hat{p}_i$	$113 \cdot \hat{p}_i$
0	82	0.6956	78.6
1	25	0.2525	28.5
2	4	0.0458	5.2
3	0	0.0055	0.6
4	2	0.0006	0.1
		1.0000	113.0

10.5.2 At the $\alpha = .05$ level, H_0: Type of company and importance of work force are independent is rejected if $d_2 \geq \chi^2_{.95,(2-1)(2-1)} = 3.841$. But $d_2 = \dfrac{(168-163.79)^2}{163.79} + \ldots + \dfrac{(26-21.79)^2}{21.79} = 1.54$, so H_0 is not rejected.

	Manufacturing	Other	
Important	168	73	241
	(163.79)	(77.21)	
Not Important	42	26	68
	(46.21)	(21.79)	
	210	99	309

10.5.6 Let $\alpha = 0.05$. To test H_0: Children's blood pressures are independent of their parent's blood pressures versus H_1: Children's blood pressures are not independent of their parent's blood pressures, reject the null hypothesis if $d_2 \geq \chi^2_{.95,(3-1)(3-1)} = 9.488$. Here,

$d_2 = \frac{(14-11.12)^2}{11.12} + \ldots + \frac{(12-8.83)^2}{8.83} = 3.81$, so H_0 would not be rejected. Based on these data, attempts to use one group to screen for high-risk individuals in the other group are not likely to be successful.

		Child's blood pressure			
		Lower	Middle	Upper	
Father's blood Pressure	Lower	14 (11.12)	11 (11.48)	8 (10.40)	33
	Middle	11 (10.45)	11 (10.78)	9 (9.77)	31
	Upper	6 (9.43)	10 (9.74)	12 (8.83)	28
		31	32	29	92

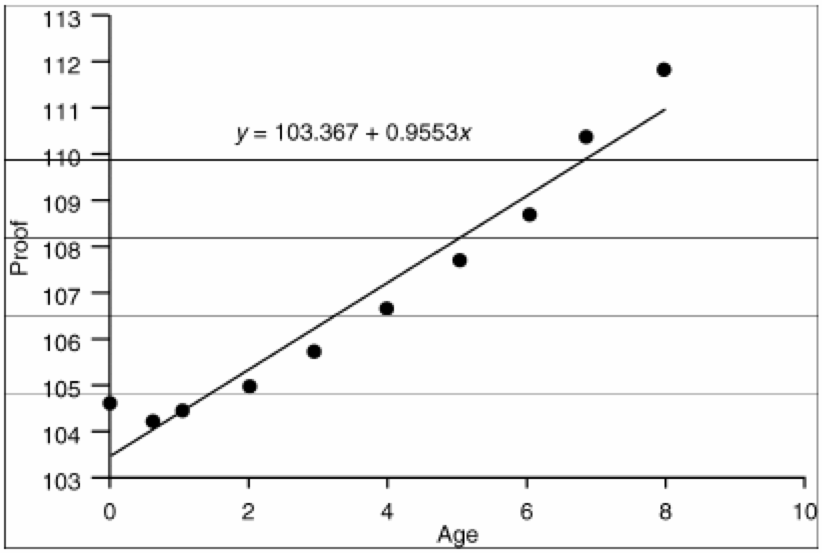
10.5.8 The null hypothesis that enrollment rates are independent of racial groups is rejected at the $\alpha = 0.05$ level if $d_2 \geq \chi^2_{.95,(4-1)(2-1)} = 7.815$. For these data,

$d_2 = \frac{(2592-2622.49)^2}{2622.49} + \ldots + \frac{(399-379.63)^2}{379.63} = 10.29$, implying that the differences in enrollment rates from race to race are statistically significant.

	Admitted	Enrolled	
White	2592 (2622.49)	1481 (1450.51)	4073
Af.-Amer.	159 (152.60)	78 (84.40)	237
Hispanic	800 (756.55)	375 (418.45)	1175
Asian	667 (686.37)	399 (379.63)	1066
	4218	2333	6551

11.2.2 $b = \dfrac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} = \dfrac{10(3973.35) - (36.5)(1070)}{10(204.25) - (36.5)^2} = 0.9953$

$a = \dfrac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n} = \dfrac{1070 - 0.9953(36.5)}{10} = 103.367$



11.2.4 In the first graph, all of the residuals are positive. The residuals in the second graph alternate from positive to negative. Neither graph would normally occur from linear models.

11.2.6. The problem here is the gap in x values, leaving some doubt as to the x–y relationship.

11.2.12 Using Cramer's rule we obtain

$$b = \frac{\begin{vmatrix} n & \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i y_i \end{vmatrix}}{\begin{vmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{vmatrix}} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i\right)}$$

which is essentially the form of b in Theorem 11.2.1. The first row of the matrix equation is $na + \left(\sum_{i=1}^{n} x_i\right)b = \sum_{i=1}^{n} y_i$. Solving this equation for a in terms of b gives the expression in Theorem 11.2.1 for a.

ST 430
Homework #12

11.3.1 $\beta_1 = \dfrac{4(93) - 10(40.2)}{4(30) - 10^2} = -1.5$

$\beta_0 = \dfrac{(40.2) - (-1.5)(10)}{4} = 13.8$

Thus, $y = 13.8 - 1.5x$. $t = \dfrac{\hat{\beta}_1 - \beta_1^0}{s \Big/ \sqrt{\sum_{i=1}^{4}(x_i - \bar{x})^2}} = \dfrac{-1.5 - 0}{2.114/\sqrt{5}} = -1.59$

Since $-t_{.025,2} = -4.3027 < t = -1.59 < 4.3027 = t_{.025,2}$, accept H_0.

11.3.2 (a) The radius of the confidence interval =

$$t_{.025,11}\frac{s}{\sqrt{\sum_{i=1}^{13}(x_i - \bar{x})^2}} = 2.2010\frac{42.745}{\sqrt{4602525.692}} = 0.044$$

The center is $\beta_1 = 0.055$, and the confidence interval is (0.011, 0.099)

(b) Since 0 is not in the confidence interval, we can reject H_0 at the 0.05 level of significance.

(c) See the solution to Question 11.2.7. The linear fit for x values less than \$4300 is not very good, suggesting a search for other contributing variables in the x range of \$3500 to \$4200.

11.3.3 $t = \dfrac{\beta_1 - \beta_1^0}{s \Big/ \sqrt{\sum_{i=1}^{15}(x_i - \bar{x})^2}} = \dfrac{3.291 - 0}{3.829/\sqrt{40.55733}} = 5.47.$

Since $t = 5.47 > t_{0.005,13} = 3.0123$, reject H_0.

11.3.9 $t = \dfrac{\hat{\beta}_1 - \beta_1^0}{s \Big/ \sqrt{\sum_{i=1}^{11}(x_i - \bar{x})^2}} = \dfrac{0.84 - 0}{2.404/\sqrt{156.909}} = 4.38$

Since $t = 4.38 > t_{.025,9} = 2.2622$, reject H_0.

11.3.10 $E(\bar{Y}) = \dfrac{1}{n}\sum_{i=1}^{n} E(Y_i | x_i) = \dfrac{1}{n}\sum_{i=1}^{n}(\beta_0 + \beta_1 x_i) = \dfrac{1}{n}n\beta_0 + \beta_1\dfrac{1}{n}\sum_{i=1}^{n} x_i = \beta_0 + \beta_1\bar{x}$

11.3.16 (a) The radius of the confidence interval is $t_{.25,16}s\sqrt{\frac{1}{n}+\frac{(x-\overline{x})^2}{\sum_{i=1}^{18}(x_i-\overline{x})^2}}$

$= 2.1199(0.202)\sqrt{\frac{1}{18}+\frac{(14.0-15.0)^2}{96.38944}} = 0.110$

The center is $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = -0.104 + 0.988(14.0) = 13.728$.

The confidence interval is (13.62, 13.84)

(b) The radius of the prediction interval is

$t_{.025,16}s\sqrt{1+\frac{1}{n}+\frac{(x-\overline{x})^2}{\sum_{i=1}^{18}(x_i-\overline{x})^2}} = 2.1199(0.202)\sqrt{1+\frac{1}{18}+\frac{(14.0-15.0)^2}{96.38944}} = 0.442$

The center is $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = -0.104 + 0.988(14.0) = 13.728$.

The confidence interval is (13.29, 14.17)