CHAPTER 9

Section 9.1

1.

a. Yes. It is an assertion about the value of a parameter.

b. No. The sample median $\tilde{x}$ is not a parameter.

c. No. The sample standard deviation s is not a parameter.

d. Yes. It is an assertion about parameters: that the standard deviation of population #2 exceeds that of population #1.

e. No. $\overline{X}$ and $\overline{Y}$ are statistics rather than parameters, so cannot appear in a hypothesis.

f. Yes. It is an assertion about the value of a parameter.

3. In this formulation, H_0 states the welds do not conform to specification. This assertion will not be rejected unless there is strong evidence to the contrary. Thus the burden of proof is on those who wish to assert that the specification is satisfied. Using H_a: $\mu < 100$ instead results in the welds being believed in conformance unless proved otherwise, so the burden of proof is on the non-conformance claim.

5. Let σ denote the population standard deviation. The appropriate hypotheses are H_0: $\sigma = .05$ vs H_a: $\sigma < .05$. With this formulation, the burden of proof is on the data to show that the requirement has been met — the sheaths will not be used unless H_0 can be rejected in favor of H_a. Type I error: Conclude that the standard deviation is < .05 mm when it is really equal to .05 mm. Type II error: Fail to reject that the standard deviation is .05 mm when it is really < .05.

7. A type I error here involves saying that the plant is not in compliance when in fact it is. A type II error occurs when we conclude that the plant is in compliance when in fact it isn't. Reasonable people may disagree as to which of the two errors is more serious. If in your judgement it is the type II error, then the reformulation H_0: $\mu = 150$ vs H_a: $\mu < 150$ makes the type I error more serious.

9.

a. R_1 is most appropriate, because x either too large or too small contradicts $p = .5$ and supports $p \neq .5$.

b. Since $x = 6$ falls in the rejection region R_1, we would reject H_0 in favor of H_a.

c. A type I error consists of judging one of the two companies favored over the other when in fact there is a 50-50 split in the population. A type II error involves judging the split to be 50-50 when it is not.

d. When H_0 is true, X has a binomial distribution with $n = 25$ and $p = 0.5$.
$\alpha = P(\text{type I error}) = P(X \le 7 \text{ or } X \ge 18 \text{ when } X \sim \text{Bin}(25, .5)) = B(7; 25,.5) + 1 - B(17; 25,.5) = .044$.

e. $\beta(.4) = P(8 \le X \le 17 \text{ when } p = .4) = B(17; 25, .4) - B(7, 25, .4) = 0.845$, and $\beta(.6) = 0.845$ also. Similarly, $\beta(.3) = B(17;25,.3) - B(7;25,.3) = .488 = \beta(.7)$. Since power = 1 – β, power = 1 – .845 = .155 for p = .4 and p = .6, while power = 1 – .488 = .512 for p = .3 and p = .7.

11.

a. H_0: $\mu = 10$ vs H_a: $\mu \ne 10$.

b. α = P(reject H_0 when H_0 is true) = $P(\bar{X} \ge 10.1032$ or ≤ 9.8968 when $\mu = 10)$. Since $\bar{X}$ is normally distributed with standard deviation $\dfrac{\sigma}{\sqrt{n}} = \dfrac{.2}{5} = .04$, $\alpha = P(Z \ge 2.58 \text{ or } \le -2.58) = .005 + .005 = .01$.

c. When $\mu = 10.1$, $E(\bar{X}) = 10.1$, so $\beta(10.1) = P(9.8968 < \bar{X} < 10.1032 \text{ when } \mu = 10.1) =$ $P(-5.08 < Z < .08) = .5319$. Similarly, $\beta(9.8) = P(2.42 < Z < 7.58) = .0078$.

d. From part **b**, $c = \pm 2.58$.

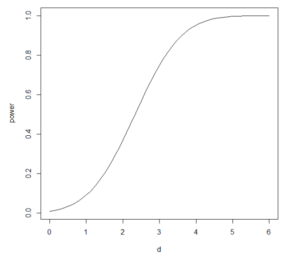
e. Regardless of sample size, Z is still normally distributed. Since $z_{.025} = 1.96$, we should now reject H_0 if $z \le -1.96$ or $z \ge 1.96$. Equivalently, with $\dfrac{\sigma}{\sqrt{n}} = \dfrac{.2}{\sqrt{10}} = .0632$, $\dfrac{\bar{x} - 10}{.0632} = \pm 1.96$ implies $\bar{x} = 10.124, 9.876$ so reject H_0 if $\bar{x} \le 9.876$ or $\bar{x} \ge 10.124$.

f. $\bar{x} = 10.020$. Since $\bar{x}$ is neither ≥ 10.124 nor ≤ 9.876, it is *not* in the rejection region. H_0 is not rejected; it is still plausible that $\mu = 10$.

13.

a. $P(\bar{X} \ge \mu_0 + 2.33\dfrac{\sigma}{\sqrt{n}} \text{ when } \mu = \mu_0) = P\left(Z \ge \dfrac{\left(\mu_0 + 2.33\sigma/\sqrt{n}\right) - \mu_0}{\sigma/\sqrt{n}}\right) = P\left(Z \ge 2.33\right) = .01$, where Z is a standard normal rv.

b. Similarly, power = P(reject H_0) = $P\left(Z \ge \dfrac{\left(\mu_0 + 2.33\sigma/\sqrt{n}\right) - \mu}{\sigma/\sqrt{n}}\right) = P\left(Z \ge 2.33 - \dfrac{d}{\sigma/\sqrt{n}}\right) =$ $1 - \Phi\left(2.33 - \dfrac{d}{\sigma/\sqrt{n}}\right)$. A plot of power as a function of d will generally increase with d, which is consistent with H_a. To make a specific graph requires selecting values of σ and n.



c. $P(\text{reject } H_0 \text{ when } \mu = 99) = P(\bar{X} \geq 102.33 \text{ when } \mu = 99) = P\left(Z \geq \frac{102-99}{1}\right) = P(Z \geq 3.33) = .0004$.

Similarly, $\alpha(98) = P(\bar{X} \geq 102.33 \text{ when } \mu = 98) = P(Z \geq 4.33) \approx 0$. In general, we have $P(\text{type I error}) < .01$ when this probability is calculated for a value of μ less than 100. The boundary value $\mu = 100$ yields the largest α.

Section 9.2

15.

a. $\alpha = P(Z \geq 1.88 \text{ when } Z \sim N(0, 1)) = 1 - \Phi(1.88) = .0301$.

b. $\alpha = P(Z \leq -2.75 \text{ when } Z \sim N(0, 1)) = \Phi(-2.75) = .003$.

c. $\alpha = \Phi(-2.88) + (1 - \Phi(2.88)) = .004$.

17. With H_0: $\mu = .5$ vs H_a: $\mu \neq .5$ (a two-sided test), we reject H_0 if $t \geq t_{\alpha/2,n-1}$ or $t \leq -t_{\alpha/2,n-1}$.

a. $1.6 < t_{.025,12} = 2.179$, so don't reject H_0.

b. $-1.6 > -t_{.025,12} = -2.179$, so don't reject H_0.

c. $-2.6 > -t_{.005,24} = -2.797$, so don't reject H_0.

d. $-3.9 <$ the negative of all t values in the df = 24 row, so we reject H_0 in favor of H_a.

19.

a. Reject H_0 if either $z \geq 2.58$ or $z \leq -2.58$; $\frac{\sigma}{\sqrt{n}} = 0.3$, so $z = \frac{94.32-95}{0.3} = -2.27$. Since -2.27 is not in the rejection region, don't reject H_0.

b. $\beta(94) = \Phi\left(2.58 + \frac{1}{0.3}\right) - \Phi\left(-2.58 + \frac{1}{0.3}\right) = \Phi(5.91) - \Phi(.75) = .2266$.

c. $n = \left[\frac{1.20(2.58+1.28)}{95-94}\right]^2 = 21.46$, so use $n = 22$.

21. With H_0: $\mu = 750$ vs H_a: $\mu < 750$ and a significance level of .05, we reject H_0 if $z \le -1.645$. Here, $z = -2.14 \le -1.645$, so we reject the null hypothesis and do not continue with the purchase. At a significance level of .01, we reject H_0 if $z \le -2.33$; $z = -2.14 > -2.33$, so we don't reject the null hypothesis and thus continue with the purchase.

23. H_0: $\mu = 360$ vs. H_a: $\mu > 360$; $t = \dfrac{\bar{x} - 360}{s/\sqrt{n}}$; reject H_0 if $t > t_{.05,25} = 1.708$; $t = \dfrac{370.69 - 360}{24.36/\sqrt{26}} = 2.24 > 1.708$. Thus H_0 should be rejected. There appears to be a contradiction of the prior belief.

25.

a. Let μ = true mean core wood density (g/mm^3). The hypotheses are H_0: $\mu = 600$ vs H_a: $\mu \ne 600$. We will reject H_0 at the .05 level if $|t| \ge t_{.025,24} = 2.064$. Here, $\bar{x} = 570.9$, $s = 103.9$, and
$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{570.9 - 600}{103.9/\sqrt{25}} = -1.40$. Because $|-1.40| < 2.064$, H_0 is <u>not</u> rejected at the .05 level.

b. The 95% CI of (528.0, 613.8) suggests, among other things, that $\mu = 600$ is plausible because 600 falls in the interval. The hypothesis test also states that $\mu = 600$ is plausible because that hypothesis was not rejected. These two results match because both use $\alpha = .05$, a 5% chance of incorrectly rejecting a true μ value or, equivalently, a 95% chance of a CI covering μ.

27.

a. We will test H_0: $\mu \ge 113$ vs. H_a: $\mu < 113$ and reject H_0 if $t < -t_{.05,5} = -2.015$. Here,
$t = \dfrac{\bar{x} - 113}{s/\sqrt{n}} = \dfrac{112.97 - 113}{4.29/\sqrt{6}} = -.02$. We fail to reject H_0 here; there is no significant evidence that the population mean isn't at least 113 grams.

b. Under these assumptions, $\bar{X}$ is normally distributed with mean $\mu = 110$ and standard error $4/\sqrt{6}$ = 1.633. In this one-sided z test, we would reject H_0 if $z < -1.645$, where $z = \dfrac{\bar{x} - 113}{4/\sqrt{6}}$. The probability of rejection is, thus, $P\left(\dfrac{\bar{X} - 113}{4/\sqrt{6}} < -1.645\right) = P\left(\dfrac{\bar{X} - 110}{4/\sqrt{6}} < -1.645 + \dfrac{113 - 110}{4/\sqrt{6}}\right) = P(Z < 0.19) = .58$.

c. Replace 6 with n in the last part of (b): we want $P\left(Z < -1.645 + \dfrac{113 - 110}{4/\sqrt{n}}\right) \ge .95$. This requires
$-1.645 + \dfrac{113 - 110}{4/\sqrt{n}} > \Phi^{-1}(.95) = 1.645$, which solving for n gives $n > 19.24$. So, at least 20 observations are required total, aka an additional 14 observations.

29. Let μ = population mean MAWL. The hypotheses are H_0: $\mu = 25$ vs H_a: $\mu > 25$. We will reject H_0 at the.05 level if $t \ge t_{.05,5-1} = 2.132$. Here, $t = \dfrac{\bar{x} - 245}{s/\sqrt{n}} = \dfrac{27.54 - 25}{5.47/\sqrt{5}} = 1.04$. Since $1.04 < 2.132$, we do not reject H_0 at the .05 significance level. It is still plausible that μ is (at most) 25.

31.

a. For $n = 8$, $n - 1 = 7$, and $t_{.05,7} = 1.895$, so H_0 is rejected at level .05 if $t \geq 1.895$. Since $\frac{s}{\sqrt{n}} = \frac{1.25}{\sqrt{8}} = .442$, $t = \frac{3.72 - 3.50}{.442} = .498$; this does not exceed 1.895, so H_0 is not rejected.

b. Use the noncentral t distribution with $\delta = \frac{4.00 - 3.50}{1.25/\sqrt{8}} = 1.13$. $\beta(4.00) = P(T < 1.895)$ when T has a noncentral t distribution with df = 8 – 1 = 7 and ncp = 1.13. From software, this probability is about .73; e.g., in R, use `pt(1.895,df=7,ncp=1.13)`.

33. The hypotheses of interest are H_0: $\mu = 7$ vs H_a: $\mu < 7$, so a lower-tailed test is appropriate. H_0 should be rejected if $t \leq -t_{.1,8} = -1.397$. $t = \frac{6.32 - 7}{1.65/\sqrt{9}} = -1.24$. Because –1.24 is not ≤ –1.397, H_0 (prior belief) is not rejected (contradicted) at level .1.

35. $\beta(\mu_o - \Delta) = \Phi\left(z_{\alpha/2} + \Delta\sqrt{n}/\sigma\right) - \Phi\left(-z_{\alpha/2} + \Delta\sqrt{n}/\sigma\right) = 1 - \Phi\left(-z_{\alpha/2} - \Delta\sqrt{n}/\sigma\right) - [1 - \Phi\left(z_{\alpha/2} - \Delta\sqrt{n}/\sigma\right)] =$
$\Phi\left(z_{\alpha/2} - \Delta\sqrt{n}/\sigma\right) - \Phi\left(-z_{\alpha/2} - \Delta\sqrt{n}/\sigma\right) = \beta(\mu_o + \Delta)$

37. We use here the fact that when $\mu = \mu'$, the test statistic T has a noncentral t distribution with n – 1 df and noncentrality parameter $\delta = \frac{\mu' - \mu_0}{\sigma/\sqrt{n}}$.

a. For the upper-tailed test, we reject H_0 if $T \geq t_{\alpha,n-1}$. When $\mu = \mu'$,

$$\text{power} = P(T \geq t_{\alpha,n-1} \text{ when } \mu = \mu' \text{ rather than } \mu_0)$$
$$= P\left(T \geq t_{\alpha,n-1} \text{ when } T \sim \text{ noncentral } t, \text{ df } = n-1, \delta = \frac{\mu' - \mu_0}{\sigma/\sqrt{n}}\right)$$
$$= 1 - F\left(t_{\alpha,n-1}; n-1, \frac{\mu' - \mu_0}{\sigma/\sqrt{n}}\right)$$

b. For the two-tailed test, we reject H_0 if $|T| \geq t_{\alpha/2,n-1}$. When $\mu = \mu'$,

$$\text{power} = P(|T| \geq t_{\alpha/2,n-1} \text{ when } \mu = \mu' \text{ rather than } \mu_0)$$
$$= P\left(|T| \geq t_{\alpha/2,n-1} \text{ when } T \sim \text{ noncentral } t, \text{ df } = n-1, \delta = \frac{\mu' - \mu_0}{\sigma/\sqrt{n}}\right)$$
$$= P\left(T \leq -t_{\alpha/2,n-1} \text{ or } T \geq t_{\alpha/2,n-1} \text{ when } T \sim \text{ noncentral } t, \text{ df } = n-1, \delta = \frac{\mu' - \mu_0}{\sigma/\sqrt{n}}\right)$$
$$= F\left(-t_{\alpha/2,n-1}; n-1, \frac{\mu' - \mu_0}{\sigma/\sqrt{n}}\right) + 1 - F\left(t_{\alpha/2,n-1}; n-1, \frac{\mu' - \mu_0}{\sigma/\sqrt{n}}\right)$$

Section 9.3

39.

1 Parameter of interest: p = true proportion of cars in this particular county passing emissions testing on the first try.

2 H_0: $p = .70$

3 H_a: $p \neq .70$

4 $$z = \frac{\hat{p} - p_o}{\sqrt{p_o(1-p_o)/n}} = \frac{\hat{p} - .70}{\sqrt{.70(.30)/n}}$$

5 either $z \geq 1.96$ or $z \leq -1.96$

6 $$z = \frac{124/200 - .70}{\sqrt{.70(.30)/200}} = -2.469$$

7 Reject H_0. The data indicates that the proportion of cars passing the first time on emission testing or this county differs from the proportion of cars passing statewide.

41.

a.

1 p = true proportion of all nickel plates that blister under the given circumstances.

2 H_0: $p = .10$

3 H_a: $p > .10$

4 $$z = \frac{\hat{p} - p_o}{\sqrt{p_o(1-p_o)/n}} = \frac{\hat{p} - .10}{\sqrt{.10(.90)/n}}$$

5 Reject H_0 if $z \geq 1.645$

6 $$z = \frac{14/100 - .10}{\sqrt{.10(.90)/100}} = 1.33$$

7 Fail to Reject H_0. The data does not give compelling evidence for concluding that more than 10% of all plates blister under the circumstances.

The possible error we could have made is a Type II error: Failing to reject the null hypothesis when it is actually true.

b. $$\beta(.15) = \Phi\left[\frac{.10 - .15 + 1.645\sqrt{.10(.90)/100}}{\sqrt{.15(.85)/100}}\right] = \Phi(-.02) = .4920$$. When $n = 200$,

$$\beta(.15) = \Phi\left[\frac{.10 - .15 + 1.645\sqrt{.10(.90)/200}}{\sqrt{.15(.85)/200}}\right] = \Phi(-.60) = .2743$$

c. $$n = \left[\frac{1.645\sqrt{.10(.90)} + 1.28\sqrt{.15(.85)}}{.15 - .10}\right]^2 = 19.01^2 = 361.4$$, so use $n = 362$.

43.

a. We wish to test H_0: $p = .02$ vs H_a: $p < .02$; only if H_0 can be rejected will the inventory be postponed. The lower-tailed test rejects H_0 if $z \le -1.645$. With $\hat{p} = 15/1000 = .015$, $z = -1.01$, which is not in the rejection region. Thus, H_0 cannot be rejected, so the inventory *should* be carried out.

b. $\beta(.01) = 1 - \Phi\left[\dfrac{.02 - .01 - 1.645\sqrt{.02(.98)/1000}}{\sqrt{.01(.99)/1000}}\right] = 1 - \Phi(0.86) = .1949$.

c. $\beta(.05) = 1 - \Phi\left[\dfrac{.02 - .05 - 1.645\sqrt{.02(.98)/1000}}{\sqrt{.05(.95)/1000}}\right] = 1 - \Phi(-5.41) \approx 1$, so the chance the inventory will be *postponed* is P(reject H_0 when $p = .05$) = $1 - \beta(.05) \approx 0$. It is highly unlikely that H_0 will be rejected, and the inventory will almost surely be carried out.

45.

a. p = true proportion of current customers who qualify. H_0: $p = .05$ v. H_a: $p \ne .05$, $z = \dfrac{\hat{p} - .05}{\sqrt{.05(.95)/n}}$, reject H_0 if $z \ge 2.58$ or $z \le -2.58$. $\hat{p} = .08$, so $z = \dfrac{.03}{.00975} = 3.07 \ge 2.58$ and H_0 is rejected. The company's premise is not correct.

b. $\beta(.10) = \Phi\left[\dfrac{.05 - .10 + 2.58\sqrt{.05(.95)/500}}{\sqrt{.10(.90)/500}}\right] - \Phi\left[\dfrac{.05 - .10 - 2.58\sqrt{.05(.95)/500}}{\sqrt{.10(.90)/500}}\right] \approx \Phi(-1.85) - 0 = .0332.$

47. The hypotheses are H_0: $p = .10$ v. H_a: $p > .10$, so R has the form $\{c, \ldots, n\}$.
The values $n = 10$, $c = 3$ (i.e., $R = \{3, 4, \ldots, 10\}$) yield $\alpha = 1 - B(2; 10, .1) = .07$, while no larger R has $\alpha \le .10$. However, $\beta(.3) = B(2; 10, .3) = .383$.
The values $n = 20$, $c = 5$ yield $\alpha = 1 - B(4; 20, .1) = .043$, but again $\beta(.3) = B(4; 20, .3) = .238$ is too high.
The values $n = 25$, $c = 5$ yield $\alpha = 1 - B(4; 25, .1) = .098$ while $\beta(.7) = B(4; 25, .3) = .090 \le .10$, so $n = 25$ should be used. The rejection region is $R = \{5,\ldots,25\}$, $\alpha = .098$, and $\beta(.7) = .090$.

Section 9.4

49. Using $\alpha = .05$, H_0 should be rejected whenever P-value < .05.

a. P-value = .001 < .05, so reject H_0.

b. .021 < .05, so reject H_0.

c. .078 is not < .05, so don't reject H_0.

d. .047 < .05, so reject H_0 (a close call).

e. .148 > .05, so H_0 can't be rejected at level .05.

51. In each case, the P-value equals $P(Z > z) = 1 - \Phi(z)$.

a. .0778
b. .1841
c. .0250
d. .0066
e. .5438

53. Use Table A.7.

a. $P(T > 2.0)$ at 8 df = .040.
b. $P(T < -2.4)$ at 11 df = .018.
c. $2P(T < -1.6)$ at 15 df = 2(.065) = .130.
d. By symmetry, $P(T > -.4) = 1 - P(T > .4)$ at 19 df = 1 – .347 = .653.
e. $P(T > 5.0)$ at 5 df < .005.
f. $2P(T < -4.8)$ at 40 df ≈ 2(.000) = .000 to three decimal places.

55. The P-value is greater than the level of significance $\alpha = .01$, therefore fail to reject H_0. The data does not indicate a statistically significant difference in average serum receptor concentration between pregnant women and all other women.

57.

a. For testing H_0: $p = .2$ v. H_a: $p > .2$, an upper-tailed test is appropriate. The computed Z is $z = .97$, so the P-value = $1 - \Phi(.97) = .166$. Because the P-value is rather large, H_0 would not be rejected at any reasonable α (it can't be rejected for any $\alpha < .166$), so no modification appears necessary.

b. With $p = .5$, $1-\beta(.5) = 1-\Phi\left[\left(-.3+2.33(.0516)\right)/.0645\right] = 1-\Phi(-2.79) = .9974$.

59. Let μ = the true average percentage of organic matter in this type of soil, and the hypotheses are H_0: $\mu = 3$ v. H_a: $\mu \neq 3$. With $n = 30$, and assuming normality, we use the t test:
$t = \dfrac{\bar{x}-3}{s/\sqrt{n}} = \dfrac{2.481-3}{.295} = \dfrac{-.519}{.295} = -1.759$. The P-value = $2[P(T > 1.759)] = 2(.041) = .082$. At significance level .10, since .082 ≤ .10, we would reject H_0 and conclude that the true average percentage of organic matter in this type of soil is something other than 3. At significance level .05, we would not have rejected H_0.

61.

a. The appropriate hypotheses are H_0: $\mu = 10$ v. H_a: $\mu < 10$.

b. P-value = $P(T > 2.3) = .017$, which is ≤ .05, so we would reject H_0. The data indicates that the pens do not meet the design specifications.

c. P-value = $P(T > 1.8) = .045$, which is not ≤ .01, so we would not reject H_0. There is not enough evidence to say that the pens don't satisfy the design specifications.

d. P-value = $P(T > 3.6) \approx .001$, which gives strong evidence to support the alternative hypothesis.

63. With H_0: $\mu = .60$ v. H_a: $\mu \neq .60$,and a two-tailed P-value of .0711, we fail to reject H_0 at levels .01 and .05 (thus concluding that the amount of impurities need not be adjusted), but we would reject H_0 at level .10 (and conclude that the amount of impurities does need adjusting).

Section 9.5

65.

a. The likelihood function here is $f(\mathbf{x};\sigma^2) = (2\pi\sigma^2)^{-n/2}\exp(-\Sigma x_i^2/2\sigma^2)$, so the most powerful test rejects H_0 when $\dfrac{(2\pi(3)^2)^{-n/2}\exp(-\Sigma x_i^2/2(3)^2)}{(2\pi(2)^2)^{-n/2}\exp(-\Sigma x_i^2/2(2)^2)} \geq k$, i.e., $\exp(\Sigma x_i^2/2(2)^2 - \Sigma x_i^2/2(3)^2) \geq k'$.
Taking logarithms to solve for $\sum x_i^2$ gives a solution of the form $\sum x_i^2 \geq c$.

b. Under normality with μ known, $\sum X_i^2 \sim \sigma^2\chi^2_{10}$ when $n = 10$. Hence, we want $.05 = \alpha = P(\sum X_i^2 \geq c$ when $\sigma^2 = 2) = P(2\chi^2_{10} \geq c) = P(\chi^2_{10} \geq c/2) \rightarrow c = 2\chi^2_{.05,10} = 2(18.307) = 36.614$.

c. Yes – the Neyman-Pearson method in (a) would yield the same test form (i.e., reject the null if $\sum X_i^2 \geq c$) for any alternative value $\sigma_a^2 > 2$, not just 3.

67.

a. The likelihood function here is $f(\mathbf{x};\theta) = \lambda^n\exp(-\lambda\Sigma x_i)$, so the most powerful test rejects H_0 when $\dfrac{(.5)^n\exp(-(.5)\Sigma x_i)}{1^n\exp(-1\Sigma x_i)} \geq k$, i.e., $\exp(.5\Sigma x_i) \geq k'$. Taking logarithms gives a solution of the form $\sum x_i \geq c$.

b. Yes – the Neyman-Pearson method in (a) would yield the same test form (i.e., reject the null if $\sum x_i \geq c$ for some c) for any alternative value $\lambda_a < 1$, not just 0.5.

69. Rewrite the likelihood as $C\theta^{2x_1+x_2}(1-\theta)^{x_2+2x_3} = C\theta^{2x_1+x_2}(1-\theta)^{2n-[2x_1+x_2]} = C\theta^y(1-\theta)^{2n-y}$, where $y = 2x_1 + x_2$. Then the most powerful test rejects H_0 when $\dfrac{C(.8)^y(1-.8)^{2n-y}}{C(.5)^y(1-.5)^{2n-y}} \geq k$, i.e., $\left(\dfrac{.8}{1-.8}\right)^y \geq k'$. Solving for y gives a solution of the form $y \geq c$. Yes, the test is UMP for the alternative $H_a : \theta > .5$ because the tests for $H_0 : \theta = .5$ vs. $H_a : \theta = p_0$ all have the same form for $p_0 > .5$.

71.

a. The usual test rejects H_0 at $\alpha = .05$ if $|z| > z_{.025} = 1.96$, where $z = \dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \dfrac{\bar{x}-0}{4/\sqrt{16}} = \bar{x}$.

b. $\pi(0) = P(\bar{X} \geq 2.17$ or $\bar{X} \leq -1.81$ when $\mu = 0) = P(Z \geq 2.17$ or $Z \leq -1.81) = 1 - \Phi(2.17) + \Phi(-1.81) = .0502$.

c. $\pi(.1) = P(\bar{X} \geq 2.17$ or $\bar{X} \leq -1.81$ when $\mu = .1) = P(Z \geq 2.07$ or $Z \leq -1.91) = 1 - \Phi(2.07) + \Phi(-1.91) =$ .04345. Similarly, $\pi(-.1) = 1 - \Phi(2.27) + \Phi(-1.71) = .05826$. This test is <u>not</u> unbiased, since $\pi(.1) < \pi(0)$. By definition, a test is unbiased only if its power is greater on the alternative than on the null.

d. $\pi(.1) = P(\bar{X} \geq 1.96$ or $\bar{X} \leq -1.96$ when $\mu = .1) = 1 - \Phi(1.86) + \Phi(-2.06) = .05114$. Similarly, $\pi(-.1) = 1 - \Phi(2.06) + \Phi(-1.86) = .05114$. This test is not UMP, since it has less power for the alternative $\mu = -.1$ than the test in (b) and (c): .05114 < .05286.

73.

a. From the algebra already presented in the text, $\Lambda = \left(\dfrac{1}{1+n(\bar{x}-\mu_0)^2/\sum(x_i-\bar{x})^2}\right)^{n/2} =$

$$\left(1+\frac{n(\bar{x}-\mu_0)^2}{\sum(x_i-\bar{x})^2}\right)^{-n/2} = \left(1+\frac{n(\bar{x}-\mu_0)^2}{(n-1)s^2}\right)^{-n/2} = \left(1+\frac{[(\bar{x}-\mu_0)/(s/\sqrt{n})]^2}{n-1}\right)^{-n/2} = \left(1+\frac{t^2}{n-1}\right)^{-n/2}.$$

The approximate chi-square statistic is then $-2\ln(\Lambda) = -2(-n/2)\ln\left(1+\dfrac{t^2}{n-1}\right) = n\ln\left(1+\dfrac{t^2}{n-1}\right)$.

b. In Exercise 59, we found $t = -1.759$. Substitute t and $n = 30$ into the above expression to get $-2\ln(\Lambda) = 3.041$. $-2\ln(\Lambda)$ has a chi-square distribution with df = 2 – 1 = 1, so the P-value is .081, compared to .089 for Exercise 59.

Section 9.6

75.

a. Here $\beta = \Phi\left(\dfrac{-.01+.9320/\sqrt{n}}{.4073/\sqrt{n}}\right) = \Phi\left(\dfrac{-.01\sqrt{n}+.9320}{.4073}\right) = .9793, .8554, .4325, .0944$, and 0 for n = 100, 2500, 10,000, 40,000, and 90,000, respectively.

b. Here $z = .025\sqrt{n}$ which equals .25, 1.25, 2.5, and 5 for the four n's, whence P-value = .4213, .1056, .0062, .0000, respectively.

c. No — even the slightest deviation from the null hypothesis would cause H_0 to be rejected, regardless of its practical importance. The α level for a huge sample should be incredibly small.

77.

a. Using the information provided, when H_0 is true, $\dfrac{S^2-\sigma_0^2}{\sqrt{2\sigma_0^4/(n-1)}}$ has an approximately standard normal distribution.

b. From the information provided, the test statistic value is $\dfrac{.07^2 - .1^2}{\sqrt{2(.1)^4/(100-1)}} = -3.59$. Using the standard normal distribution, the lower-tailed *P*-value is $P(Z \le -3.59) \approx .0002$, so we strongly reject H_0 at the .05 level and conclude that weight variability has indeed decreased.

79.

a. $\bar{x} - t_{.05,n-1}\dfrac{s}{\sqrt{n}} = 17.986 - 1.667\dfrac{5.937}{\sqrt{70}} = 16.803$.

b. At the 95% confidence level, $\mu > 16.803$. Since 15 is *not* > 16.803, we reject H_0: $\mu = 15$ at the $\alpha = 1 - .95 = .05$ level and conclude at that level that μ is greater than 15 percent. (Equivalently, since we are convinced that μ exceeds 16.803, we are in particular convinced that μ exceeds 15.)

c. No: The equivalent to a two-tailed hypothesis test ($\mu \ne 15$) is a regular two-sided CI. In particular, we'd require the .025 critical value in each direction to make a two-sided decision at $\alpha = .05$, so we'd need a 97.5% CI.

d. Since the LCB in part **a** tells us to reject H_0 at the .05 level, the same bound also tells us to reject H_0 at the .10 level. But we cannot make a rejection decision at the .01 level: rejecting H_0 at $\alpha = .05$ implies that we might or might not reject H_0 at $\alpha = .01$.

81. The following R code performs the bootstrap simulation described in this section.

```
mu0 = 113; N = 5000
x = c(117.6,109.5,111.6,109.2,119.1,110.8)
w = x - mean(x) + mu0
wbar = rep(0,N)  # allocating space for bootstrap means
for (i in 1:N){
       resample = sample(w,length(w),replace=T)
       wbar[i] = mean(resample)
}
```

The result of the simulation is 5000 $\bar{w}_i^*$ values. The *P*-value is estimated by the proportion of these $\bar{w}_i^*$ values that are at or below the observed $\bar{x}$ value of 112.9667. In one run of this code, that proportion was .5018. Such a large (estimated) *P*-value suggests that we would not reject H_0: $\mu = 113$ in favor of H_a: $\mu < 113$ at any reasonable significance level. This is consistent with the fact that the sample mean weight of the bagels is only barely below 113 grams, certainly not convincing evidence for H_a.

Supplementary Exercises

83.

a. Since the burden falls on the pharma company to convince the FDA, the null hypothesis is H_0: the reformulated drug *is no safer than* the original, recalled drug. The alternative hypothesis is H_a: the reformulated drug is safer than the recalled drug.

b. Type I error: The FDA rejects H_0 and concludes the new drug is safer, when in fact it isn't.
Type II error: The FDA fails to recognize that H_a is true and the new drug is indeed safer.

c. A type I error results in an equally-dangerous drug hitting the market, while a type II error results in a safer drug being denied FDA approval. One can argue which is worse. If we decide the FDA's primary consideration should be safety, then a type I error is worse, and so the FDA should use an especially low α level to minimize the chance of committing a type I error.

85. With $n = 50$, the t test can be approximated by a z test. The sample size required to achieve the stated goals is $n = \left[\frac{.30(z_{.05} + z_{.05})}{3.20 - 3.00}\right]^2 = \left[\frac{.30(1.645 + 1.645)}{.20}\right]^2 = 24.35$, so $n = 50$ was unnecessarily large.

87. H_0: $\mu = 15$ vs H_a: $\mu > 15$. Because the sample size is less than 40, and we can assume the distribution is approximately normal, the appropriate statistic is $t = \frac{\bar{x} - 15}{s/\sqrt{n}} = \frac{17.5 - 15}{2.2/\sqrt{32}} = \frac{2.5}{.390} = 6.4$. This t value is "off the chart" at 31 df, from which we may conclude P-value < .05 and reject H_0 in favor of the conclusion that the true average time exceeds 15 minutes.

89.

a. No, the distribution does not appear to be normal. It appears to be skewed to the right, since 0 is less than one standard deviation below the mean. It is not necessary to assume normality if the sample size is large enough due to the central limit theorem. This sample size is large enough so we can conduct a hypothesis test about the mean.

b.

1 Parameter of interest: μ = true daily caffeine consumption of adult women.

2 H_0: $\mu = 200$

3 H_a: $\mu > 200$

4 $t = \frac{\bar{x} - 200}{s/\sqrt{n}}$

5 Reject H_0 if $t \geq t_{.10,46} = 1.300$ or if P-value $\leq .10$

6 $t = \frac{215 - 200}{235/\sqrt{47}} = .44$; P-value $\approx .33$

7 Fail to reject H_0. The data does not indicate that mean daily consumption of all adult women exceeds 200 mg.

91. A t test is appropriate. H_0: $\mu = 1.75$ is rejected in favor of H_a: $\mu \neq 1.75$ if the P-value < .05. The computed test statistic is $t = \frac{1.89 - 1.75}{.42/\sqrt{26}} = 1.70$. Since the P-value is $2P(T > 1.7) = 2(.051) = .102 > .05$, do not reject H_0; the data does not contradict prior research.
We assume that the population from which the sample was taken was approximately normally distributed.

93. Let p = the true proportion of mechanics who could identify the problem. Then the appropriate hypotheses are H_0: $p = .75$ vs H_a: $p < .75$, so a lower-tailed test should be used. With $p_0 = .75$ and $\hat{p} = 42/72 = .583$, $z = -3.28$ and P-value $= \Phi(-3.28) = .0005$. Because this P-value is so small, the data argues strongly against H_0, so we reject it in favor of H_a.

95. The 20 df row of Table A.6 shows that $\chi^2_{.99,20} = 8.26 < 8.58$ (H_0 not rejected at level .01) and $8.58 < 9.591 = \chi^2_{.975,20}$ (H_0 rejected at level .025). Thus .01 < P-value < .025 and H_0 cannot be rejected at level .01 (the P-value is the smallest α at which rejection can take place, and this exceeds .01).

97.

a. When H_0 is true, $2\lambda_0 \Sigma X_i = 2\sum \frac{X_i}{\mu_0}$ has a chi-squared distribution with df = $2n$. If the alternative is H_a: $\mu > \mu_0$, large test statistic values (large Σx_i, since $\bar{x}$ is large) suggest that H_0 be rejected in favor of H_a, so rejecting when $2\sum \frac{X_i}{\mu_0} \ge \chi^2_{\alpha,2n}$ gives a test with significance level α. If the alternative is H_a: $\mu < \mu_0$, rejecting when $2\sum \frac{X_i}{\mu_0} \le \chi^2_{1-\alpha,2n}$ gives a level α test. The rejection region for H_a: $\mu \ne \mu_0$ is $2\sum \frac{X_i}{\mu_0} \ge \chi^2_{\alpha/2,2n}$ or $2\sum \frac{X_i}{\mu_0} \le \chi^2_{1-\alpha/2,2n}$.

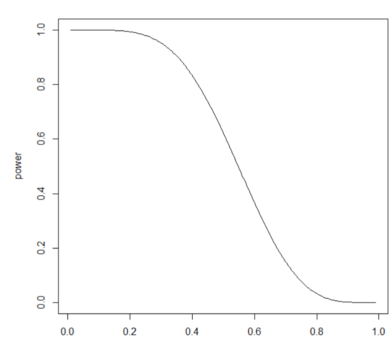
b. H_0: $\mu = 75$ v. H_a: $\mu < 75$. The test statistic value is $\frac{2(737)}{75} = 19.65$. At level .01, H_0 is rejected if $2\sum \frac{X_i}{\mu_0} \le \chi^2_{.99,20} = 8.260$. Clearly 19.65 is not in the rejection region, so H_0 should not be rejected. The sample data does not suggest that true average lifetime is less than the previously claimed value.

99.

a. $\alpha = P(X \le 5$ when $p = .9) = B(5; 10, .9) = .002$, so the rejection region {0, 1, …, 5} does specify a level .01 test.

b. The first value to be placed in the upper-tailed part of a two tailed region would be the maximum, 10, but $P(X = 10$ when $p = .9) = .349$, so whenever 10 is in the rejection region, $\alpha \ge .349$.

c. power(p') = $P(X \le R$ when $p = p') = B(5;10, p')$. The test has almost no ability to detect a false null hypothesis when $p > .90$ (see the graph for $.90 < p' < 1$). This is a by-product of the unavoidable one-sided rejection region (see **a** and **b**). The test also has undesirably low power for medium-to-large p', a result of the small sample size.



101. Rewrite $a \cdot \alpha + b \cdot \beta$ as follows:

$$a \cdot \alpha + b \cdot \beta = a \sum_R f(\mathbf{x}; \theta_0) + b\left[1 - \sum_R f(\mathbf{x}; \theta_\text{a})\right] = b + \sum_R \left[af(\mathbf{x}; \theta_0) - bf(\mathbf{x}; \theta_\text{a})\right]$$

The expression in brackets can be positive or negative. Among all possible test procedures (equivalently, all possible rejection regions R), $a \cdot \alpha + b \cdot \beta$ is minimized by choosing R to be exactly the set where the expression in brackets is negative (or zero). That is, $a \cdot \alpha + b \cdot \beta$ is minimized by the rejection region

$R^* = \{\mathbf{x} : af(\mathbf{x}; \theta_0) - bf(\mathbf{x}; \theta_\text{a}) \le 0\} = \left\{\mathbf{x} : \dfrac{f(\mathbf{x}; \theta_\text{a})}{f(\mathbf{x}; \theta_0)} \ge \dfrac{a}{b}\right\}$, as claimed.