**Partial answer key**

10.11

$$n=\frac{z\_{α/2}^{2}π(1-π)}{E^{2}}=\frac{2.576^{2}×0.5×(1-0.5)}{0.05^{2}}=663.58$$

The survey should include at least 664 people.

> # 10.11

> qnorm(p = 1- 0.01/2)^2\*0.5\*(1-0.5)/.05^2

[1] 663.4897

10. 13

The level of significance is α = 0.05.

Confidence interval method:

1.

Ho: π $\geq $ 0.1
Ha: π $<$ 0.1

2. Agresti-Coull C.I.: 0 < π < 0.238. If the Wilson C.I. was used instead: 0 < π < 0.235.

> #10.13

> #Agresti-Coull CI

> y <- 5

> n <- 40

> alpha <- 0.05

> pi.hat <- 5/40

> pi.tilde <-( y + qnorm(p = 1-alpha)^2 /2) / (n + qnorm(p = 1-alpha)^2)

> pi.tilde

[1] 0.1487575

> upper.AC <- pi.tilde + qnorm(p = 1-alpha) \* sqrt(pi.tilde\*(1-pi.tilde) / (n

 qnorm(p = 1-alpha)^2))

> upper.AC

[1] 0.2383252

>

> #Wilson CI

> upper.wilson <- pi.tilde + qnorm(1-alpha) \* sqrt(n) / (n + qnorm(1-

 alpha)^2) \* sqrt(pi.hat\*(1-pi.hat) + qnorm(1-alpha)^2 / (4\*n))

> upper.wilson

[1] 0.2353237

3. Because 0.1 is in this interval, do not reject Ho.

4. Therefore, we do not have sufficient evidence that less than 10% are dissatisfied.

Test statistic method

1.

Ho: π $\geq $ 0.1
Ha: π $<$ 0.1

2. z = 0.527

> pi0<-0.1

> pi.hat <- 5/40

> z.stat<-(pi.hat - pi0)/sqrt(pi0\*(1-pi0)/n)

> z.crit<-qnorm(p = alpha, mean = 0, sd = 1)

> pvalue<-pnorm(q = z.stat, mean = 0, sd = 1)

> data.frame(pi.hat, pi0, z.stat, z.crit, pvalue)

 pi.hat pi0 z.stat z.crit pvalue

1 0.125 0.1 0.5270463 -1.644854 0.7009193

3. -z0.05 = -1.645

4. Since 0.527 > -1.645, do not reject Ho.

5. Therefore, we do not have sufficient evidence that less than 10% are dissatisfied.

p-value method

1.

Ho: π $\geq $ 0.1
Ha: π $<$ 0.1

2. z=0.527. The p-value is 1 - P(Z> 0.527) = 0.7009.

(using the same R code as above)

3. α = 0.05

4. Since 0.7009 > 0.05, do not reject Ho.

5. Therefore, we do not have sufficient evidence that less than 10% are dissatisfied.

b)

The 95% Agresti-Coull C.I. is (0.0499, 0.2658). The 95% Wilson C.I. is also (0.0546, 0.2611).

> # part c

> y <- 5

> n <- 40

> alpha <- 0.05

> binom.confint(x = y, n = n, conf.level = 1-alpha, methods= "agresti-coull")

 method x n mean lower upper

1 agresti-coull 5 40 0.125 0.04993071 0.2657855

> binom.confint(x = y, n = n, conf.level = 1-alpha, methods = "wilson")

 method x n mean lower upper

1 wilson 5 40 0.125 0.054595 0.2611212

10.14

a)

Yes, we can use the normal approximation because n is fairly large and π0 is not close to 0 or 1

b)

The level of significance is α = 0.05.

Confidence interval method:

1.

Ho: π $\leq $ 0.5
Ha: π $>$ 0.5

2. Agresti-Coull C.I.: 0.5009227 < π < 1 If the Wilson C.I. was used instead: 0.5009229 < π < 1.

> # part b

> # For CI

> # Remember that mean = 0 and sd = 1 are the default argument values for
 qnorm().

> #Agresti-Coull CI

> y <- 424

> n <- 800

> alpha <- 0.05

> pi.hat <- 424/800

> pi.tilde <-( y + qnorm(p = 1-alpha)^2 /2) / (n + qnorm(p = 1-alpha)^2)

> pi.tilde

[1] 0.5298989

> lower.AC <- pi.tilde - qnorm(p = 1-alpha) \* sqrt(pi.tilde\*(1-pi.tilde) / (n qnorm(p = 1-alpha)^2))

> lower.AC

[1] 0.5009227

>

> #Wilson CI

> lower.wilson <- pi.tilde - qnorm(1-alpha) \* sqrt(n) / (n + qnorm(1-

 alpha)^2) \* sqrt(pi.hat\*(1-pi.hat) + qnorm(1-alpha)^2 / (4\*n))

> lower.wilson

[1] 0.5009229

3. Because 0.5 is not in this interval, reject Ho.

4. The standard conclusion is that sufficient evidence exists to indicate that more than half of persons suffering from chronic pain are over 50 years of age. It is important to note that 0.5 is really close to being within the interval. Thus, a better conclusion would be to say there is “marginal evidence”.

Test statistic method

1.

Ho: π $\leq $ 0.5
Ha: π $>$ 0.5

2. z = 1.697

> # For test stat and p value

> pi0<-1/2

> pi.hat <- 424/800

> z.stat<-(pi.hat - pi0)/sqrt(pi0\*(1-pi0)/n)

> z.crit<-qnorm(p = 1-alpha, mean = 0, sd = 1)

> pvalue<-(1 - pnorm(q = z.stat, mean = 0, sd = 1))

> data.frame(pi.hat, pi0, z.stat, z.crit, pvalue)

 pi.hat pi0 z.stat z.crit pvalue

1 0.53 0.5 1.697056 1.644854 0.04484301

3. z0.05 = 1.645

4. Since 1.697 > 1.645, reject Ho.

5. There is sufficient evidence to indicate that more than half of persons suffering from chronic pain are over 50 years of age. See the previous conclusion as well for an important note.

p-value method

1. Ho: π $\leq $ 0.5
 Ha: π $>$ 0.5

2. z=1.697. The p-value is P(Z> 1.697) = 0.0448.

(using the same R code as above)

3. α = 0.05

4. Since 0.0448 < 0.05, reject Ho.

5. There is sufficient evidence to indicate that more than half of persons suffering from chronic pain are over 50 years of age. See the previous conclusion as well for an important note.

c)

The 95% Agresti-Coull C.I. is (0.495, 0.564). The 95% Wilson C.I. is also (0.495, 0.564).

> library(binom)

> y <- 424

> n <- 800

> alpha <- 0.05

> binom.confint(x = y, n = n, conf.level = 1-alpha, methods= "agresti-coull")

 method x n mean lower upper

1 agresti-coull 424 800 0.53 0.4953536 0.5643597

> binom.confint(x = y, n = n, conf.level = 1-alpha, methods = "wilson")

 method x n mean lower upper

1 wilson 424 800 0.53 0.4953539 0.5643594

We are 95% confident that the interval 0.495 to 0.564 contains the true proportion of persons suffering from chronic pain that are over 50 years of age.

**10.15**

a)

$$\hat{π}=\frac{y}{n}=\frac{230}{1500}=0.1533$$

The 95% Agresti-Coull C.I. is (0.136, 0.172). The 95% Wilson C.I. is also (0.136, 0.172).

> y <- 230

> n <- 1500

> alpha <- 0.05

> binom.confint(x = y, n = n, conf.level = 1-alpha, methods= "agresti-coull")

 method x n mean lower upper

1 agresti-coull 230 1500 0.1533333 0.1359654 0.1724723

> binom.confint(x = y, n = n, conf.level = 1-alpha, methods = "wilson")

 method x n mean lower upper

1 wilson 230 1500 0.1533333 0.1359869 0.1724509

b)

As given in the book, this is how the computation should be done:



The survey should include at least 9,604 people.

> qnorm(p = 1- 0.05/2)^2\*.5\*(1-.5)/.01^2

[1] 9603.647

With the change given in the assigned problem, the answer becomes 4,147.

> qnorm(p = 1- 0.01/2)^2\*0.5\*(1-0.5)/.02^2
[1] 4146.81

c)

Confidence interval method:

1.

Ho: π $\leq $ 0.5
Ha: π $>$ 0.5

2. Agresti-Coull C.I.: 0.139 < π < 1. If the Wilson C.I. was used instead: 0.139 < π < 1.

> # c

> #Agresti-Coull CI

> y <- 230

> n <- 1500

> alpha <- 0.05

> pi.tilde <-( y + qnorm(p = 1-alpha)^2 /2) / (n + qnorm(p = 1-alpha)^2)

> pi.tilde

[1] 0.1539575

> lower.AC <- pi.tilde - qnorm(p = 1-alpha) \* sqrt(pi.tilde\*(1-pi.tilde) / (n qnorm(p = 1-alpha)^2))

> lower.AC

[1] 0.1386436

>

> #Wilson CI

> lower.wilson <- pi.tilde - qnorm(1-alpha) \* sqrt(n) / (n + qnorm(1-

 alpha)^2) \* sqrt(pi.hat\*(1-pi.hat) + qnorm(1-alpha)^2 / (4\*n))

> lower.wilson

[1] 0.1399083

3. Because 0.5 is in this interval, do not reject Ho.

4. There is not sufficient evidence to conclude that more than half of registered voters who would favor drilling for oil in national parks.

Test statistic method

1.

Ho: π $\leq $ 0.5
Ha: π $>$ 0.5

2. z = - 26.85

> # For test stat and p value

> pi0<-1/2

> pi.hat <- 230/1500

> z.stat<-(pi.hat - pi0)/sqrt(pi0\*(1-pi0)/n)

> z.crit<-qnorm(p = 1-alpha, mean = 0, sd = 1)

> pvalue<-(1 - pnorm(q = z.stat, mean = 0, sd = 1))

> data.frame(pi.hat, pi0, z.stat, z.crit, pvalue)

 pi.hat pi0 z.stat z.crit pvalue

1 0.1533333 0.5 -26.85268 1.644854 1

3. z0.05 = 1.645

4. Since -26.85 < 1.645, do not reject Ho.

5. There is not sufficient evidence to conclude that more than half of registered voters who would favor drilling for oil in national parks.

p-value method

1. Ho: π $\leq $ 0.5
 Ha: π $>$ 0.5

2. z=-26.85. The p-value is P(Z> -26.85) = 1.0.

(using the same R code as above)

3. α = 0.05

4. Since 1 > 0.05, do not reject Ho.

5. There is not sufficient evidence to conclude that more than half of registered voters who would favor drilling for oil in national parks.

**10.16**

a)

The sampling distribution for the difference in the sample proportions $\hat{π}\_{A}-\hat{π}\_{B}$ is approximately normal with mean μ = 0.3 – 0.15 =0.15 and standard deviation σ $= \sqrt{\frac{\hat{π}\_{A}(1-\hat{π}\_{A})}{n\_{A}}+\frac{\hat{π}\_{B}(1-\hat{π}\_{B})}{n\_{B}}}=\sqrt{\frac{0.3(0.7)}{250}+\frac{0.15(0.85)}{350}}=0.0356$.

b)

The shape of the sampling distribution of $\hat{π}\_{A}-\hat{π}\_{B}$ is unimodal and bell-shaped (approximately normal).

c)

Yes, we can use the normal approximation because the sample sizes are fairly large and π’s are not close to 0 or 1. This is a little different from what we saw in class (look at contingency table). However, the difference of two random variables is also a normal random variable so this approach is used here. In actual application, one would have a sample and contingency table to work with.

**10.20**

a)

For biofeedback

The 95% Agresti-Coull C.I. is (0.529, 0.590). The 95% Wilson C.I. is also (0.529, 0.590).

> # Biofeedback

> y1 <- 560

> n1 <- 1000

> alpha <- 0.05

> binom.confint(x = y1, n = n1, conf.level = 1-alpha, methods= "agresti-

 coull")

 method x n mean lower upper

1 agresti-coull 560 1000 0.56 0.5290618 0.590479

> binom.confint(x = y1, n = n1, conf.level = 1-alpha, methods = "wilson")

 method x n mean lower upper

1 wilson 560 1000 0.56 0.5290626 0.5904781

For NSAID

The 95% Agresti-Coull C.I. is (0.650, 0.708). The 95% Wilson C.I. is also (0.650, 0.708).

> # NSAID

> y2 <- 680

> n2 <- 1000

> alpha <- 0.05

> binom.confint(x = y2, n = n2, conf.level = 1-alpha, methods= "agresti-coull")

 method x n mean lower upper

1 agresti-coull 680 1000 0.68 0.6504382 0.7081842

> binom.confint(x = y2, n = n2, conf.level = 1-alpha, methods = "wilson")

 method x n mean lower upper

1 wilson 680 1000 0.68 0.6504464 0.708176

b)

The level of significance is α = 0.05.

Confidence interval method:

1.

Ho: π1 – π2 = 0
Ha: π1 – π2 ≠ 0

2.

The 95% Agresti and Caffo C.I. is (-0.162, -0.078).

> # b

> alpha <- 0.05

> #CI

> y1<- 560

> n1<- 1000

> y2<- 680

> n2<- 1000

> pi.tilde1<-(y1+1)/(n1+2)

> pi.tilde2<-(y2+1)/(n2+2)

> var.AC<-pi.tilde1\*(1-pi.tilde1)/(n1+2)+pi.tilde2\*(1-pi.tilde2)/(n2+2)

> lower.AC<-pi.tilde1 - pi.tilde2 - qnorm(p = 1-alpha/2)\*sqrt(var.AC)

> upper.AC<-pi.tilde1 - pi.tilde2 + qnorm(p = 1-alpha/2)\*sqrt(var.AC)

> data.frame(lower.AC, upper.AC)

 lower.AC upper.AC

1 -0.1619438 -0.07757719

3. Because 0 is not in this interval, reject Ho.

4. We reject Hoand conclude that there is significant evidence of a difference in the two treatments

relative to the proportions of patients who experienced a significant reduction in pain.

Test statistic method:

1.

Ho: π1 – π2 = 0
Ha: π1 – π2 ≠ 0

2. z=-5.53.

> pi.hat1 <- y1/n1

> pi.hat2 <- y2/n2

> pi.hat.p<-(y1 + y2)/(n1 + n2)

> z.stat<-(pi.hat1 - pi.hat2)/sqrt(pi.hat.p\*(1-pi.hat.p)\*(1/n1 + 1/n2))

> z.crit<-qnorm(p = 1-alpha/2, mean = 0, sd = 1)

> pvalue<-2\*(1 - pnorm(q = abs(z.stat), mean = 0, sd = 1))

> data.frame(pi.hat1, pi.hat2, z.stat, z.stat^2, z.crit, pvalue)

 pi.hat1 pi.hat2 z.stat z.stat.2 z.crit pvalue

1 0.56 0.68 -5.528135 30.56027 1.959964 3.236538e-08

> prop.test(x = c(y1, y2), n = c(n1, n2), conf.level = 0.95, correct = FALSE, alternative = "two.sided")

 2-sample test for equality of proportions without continuity correction

data: c(y1, y2) out of c(n1, n2)

X-squared = 30.56, df = 1, p-value = 3.237e-08

alternative hypothesis: two.sided

95 percent confidence interval:

 -0.16221892 -0.07778108

sample estimates:

prop 1 prop 2

 0.56 0.68

3. ±z0.025 = ±1.96

4. Because -5.53 < -1.96, reject Ho.

5. We reject Hoand conclude that there is significant evidence of a difference in the two treatments

relative to the proportions of patients who experienced a significant reduction in pain.

P-value method:

1.

Ho: π1 – π2 = 0
Ha: π1 – π2 ≠ 0

2. z=-5.53. The p-value is 2\*P(Z > |-5.53|) $≈0$.

(Using same R code as above)

3. α = 0.05

4. Since 0 < 0.05, reject Ho.

5. We reject Hoand conclude that there is significant evidence of a difference in the two treatments

relative to the proportions of patients who experienced a significant reduction in pain.

**c)**

The 95% Agresti and Caffo C.I. is (-0.162, -0.078).

(Same confidence interval as part b)

**10.21**

a)

For new treatment

The 95% Agresti-Coull C.I. is (0.039, 0.218). The 95% Wilson C.I. is also (0.043, 0.214).

> # a

> # New

> y1 <- 5

> n1 <- 50

> alpha <- 0.05

> binom.confint(x = y1, n = n1, conf.level = 1-alpha, methods= "agresti-

 coull")

 method x n mean lower upper

1 agresti-coull 5 50 0.1 0.03914035 0.2179377

> binom.confint(x = y1, n = n1, conf.level = 1-alpha, methods = "wilson")

 method x n mean lower upper

1 wilson 5 50 0.1 0.04347576 0.2136023

For old treatment

The 95% Agresti-Coull C.I. is (0.095, 0.310). The 95% Wilson C.I. is also (0.098, 0.308).

> # old

> y2 <- 9

> n2 <- 50

> alpha <- 0.05

> binom.confint(x = y2, n = n2, conf.level = 1-alpha, methods= "agresti-

 coull")

 method x n mean lower upper

1 agresti-coull 9 50 0.18 0.09542432 0.3102381

> binom.confint(x = y2, n = n2, conf.level = 1-alpha, methods = "wilson")

 method x n mean lower upper

1 wilson 9 50 0.18 0.09770193 0.3079605

b)

The level of significance is α = 0.05.

Confidence interval method:

1.

Ho: πN – πo $\geq $ 0
Ha: πN – πo < 0

2.

The one-side 95% Agresti-Coull C.I. for πN – πo is (-1,0.039).

> y1<-5

> n1<-50

> y2<-9

> n2<-50

> alpha <- 0.05

> pi.tilde1<-(y1+1)/(n1+2)

> pi.tilde2<-(y2+1)/(n2+2)

> var.AC<-pi.tilde1\*(1-pi.tilde1)/(n1+2)+pi.tilde2\*(1-pi.tilde2)/(n2+2)

> upper.AC<-pi.tilde1 - pi.tilde2 + qnorm(p = 1-alpha)\*sqrt(var.AC)

> upper.AC

[1] 0.03880184

3. Because 0 is not in this interval, reject Ho.

4. We fail to reject H0 and conclude that there is not significant evidence that the new treatment

would have a lower proportion of plants having a toxic level of nickel.

Test statistic method:

1.

Ho: πN – πo $\geq $ 0
Ha: πN – πo < 0

2. z=-1.152

> pi.hat1 <- y1/n1

> pi.hat2 <- y2/n2

> pi.hat.p<-(y1 + y2)/(n1 + n2)

> z.stat<-(pi.hat1 - pi.hat2)/sqrt(pi.hat.p\*(1-pi.hat.p)\*(1/n1 + 1/n2))

> z.crit<-qnorm(p = 1-alpha, mean = 0, sd = 1)

> pvalue<-pnorm(q = z.stat, mean = 0, sd = 1)

> data.frame(pi.hat1, pi.hat2, z.stat, z.stat^2, z.crit, pvalue)

 pi.hat1 pi.hat2 z.stat z.stat.2 z.crit pvalue

1 0.1 0.18 -1.152781 1.328904 1.644854 0.1245002

> prop.test(x = c(y1, y2), n = c(n1, n2), conf.level = 0.95, correct = FALSE,

 alternative = "less")

 2-sample test for equality of proportions without continuity correction

data: c(y1, y2) out of c(n1, n2)

X-squared = 1.3289, df = 1, p-value = 0.1245

alternative hypothesis: less

95 percent confidence interval:

 -1.00000000 0.03338758

sample estimates:

prop 1 prop 2

 0.10 0.18

4. Because -1.152 < -1.96, do not reject Ho.

5. We fail to reject Hoand conclude that there is not significant evidence that the new treatment

would have a lower proportion of plants having a toxic level of nickel.

3. P-value method

1.

Ho: πN – πo $\geq $ 0
Ha: πN – πo < 0

2. z=-1.152. The p-value is 1 - P(Z>-1.152) = 0.1245.

(using the same R code as above)

3. α = 0.05

4. Since 0.1245> 0.05, do not reject Ho.

5. We fail to reject Hoand conclude that there is not significant evidence that the new treatment

would have a lower proportion of plants having a toxic level of nickel.

d)

The Agresti-Coull C.I. is (-0.215,0.061)

> y1<-5

> n1<-50

> y2<-9

> n2<-50

> alpha <- 0.05

> pi.tilde1<-(y1+1)/(n1+2)

> pi.tilde2<-(y2+1)/(n2+2)

> var.AC<-pi.tilde1\*(1-pi.tilde1)/(n1+2)+pi.tilde2\*(1-pi.tilde2)/(n2+2)

> lower.AC<-pi.tilde1 - pi.tilde2 - qnorm(p = 1-alpha/2)\*sqrt(var.AC)

> upper.AC<-pi.tilde1 - pi.tilde2 + qnorm(p = 1-alpha/2)\*sqrt(var.AC)

> data.frame(lower.AC, upper.AC)

 lower.AC upper.AC

1 -0.2148178 0.06097166

or

> prop.test(x = c(y1, y2), n = c(n1, n2), conf.level = 0.95, correct = FALSE,

 alternative = "two.sided")

 2-sample test for equality of proportions without continuity correction

data: c(y1, y2) out of c(n1, n2)

X-squared = 1.3289, df = 1, p-value = 0.249

alternative hypothesis: two.sided

95 percent confidence interval:

 -0.21510963 0.05510963

sample estimates:

prop 1 prop 2

 0.10 0.18