Stat 571 - HW3 - Nathanael Fillmore

1. We use point estimate $p' = (x+2)/(n+4)$ and confidence interval $(p' - 1.96\sqrt{p'(1-p')/n'}, p' + 1.96\sqrt{p'(1-p')/n'})$ where $n' = n+4$.

```
> input <- data.frame(n = c(10, 50, 1000), x = c(3, 24, 635))
> output <- adply(input, 1, function(row) {
+     x2 <- row$x + 2
+     n4 <- row$n + 4
+     p <- x2/n4
+     ci <- 1.96 * sqrt(p * (1 - p)/n4)
+     c(`point estimate` = p, `ci lower` = p - ci, `ci upper` = p +
+         ci)
+ })
> rownames(output) <- c("(a)", "(b)", "(c)")
> print(xtable(output))
```

	n	x	point estimate	ci lower	ci upper
(a)	10.00	3.00	0.36	0.11	0.61
(b)	50.00	24.00	0.48	0.35	0.61
(c)	1000.00	635.00	0.63	0.60	0.66

2. (a) For $p = .5, n = 10$, note that $\{0,\dots,4\} \cup \{6,\dots,10\}$ are at least as far from the mean as $x = 4$. So

```
> (p.value <- sum(dbinom(c(0:4, 6:10), size = 10, prob = 0.5)))
[1] 0.7539063
```

(b) For $n = 50, p = .35, H_A : p > .35$, doublecheck that $21/50 > .35$ and note that on this side of the mean $\{21,22,\dots,50\}$ are at least as far from the mean as $x = 21$. So

```
> (p.value <- sum(dbinom(c(21:50), size = 50, prob = 0.35)))
[1] 0.1860549
```

(c) For $n = 37, p = .45, H_A : p < .45$, doublecheck that $16/37 < .45$ and note that on this side of the mean $\{0,1,\dots,16\}$ are at least as far from the mean as $x = 16$. So

```
> (p.value <- sum(dbinom(c(0:16), size = 37, prob = 0.45)))
[1] 0.4824959
```

3. Let $X \sim \text{Binomial}(n = 1000, p = 0.4)$.

(a) Want $P(X = 380)$.

```
> dbinom(380, size = 1000, prob = 0.4)
[1] 0.01123732
```

(b) Want $P(370 \leq X \leq 390)$

```
> sum(dbinom(370:390, size = 1000, prob = 0.4))
[1] 0.2462151
```

(c) Want $P(X \leq 380)$.

```
> sum(dbinom(0:380, size = 1000, prob = 0.4))
[1] 0.1038245
```

(d) Let Y be $\text{Binomial}(n = 1000, p = 0.37)$. Want $P(Y \leq 380)$.

```
> sum(dbinom(0:380, size = 1000, prob = 0.37))
[1] 0.7546507
```

4. Since $P(\Omega) = 1$, we have $P(X = 5) = 1 - 0.2 - 0.3 - 0.2 - 0.1$

```
> (prob.X.equals.5 <- 1 - 0.2 - 0.3 - 0.2 - 0.1)

[1] 0.2
```

Let k_0 be the missing value of k. Since $EX = -.3$, we have $-.3 = 0.2*(-10) + 0.3*k_0 + 0.2*3 + 0.2*5 + 0.1*7$ so

```
> (missing.k <- (0.3 + 0.2 * (-10) + 0.2 * 3 + 0.2 * 5 + 0.1 *
+     7)/(-0.3))

[1] -2
```

5. Binomial assumptions apply. We have $n = 169$, $x = 27$. As in question 1, we use point estimate $p' = (x+2)/(n+4)$ and confidence interval $p' \pm 1.96\sqrt{p'(1-p')/(n+4)}$:

```
> (function() {
+     x <- 27
+     n <- 169
+     p <- (x + 2)/(n + 4)
+     ci <- 1.96 * sqrt(p * (1 - p)/(n + 4))
+     print(xtable(data.frame(`point estimate` = p, `ci lower` = p -
+         ci, `ci upper` = p + ci)))
+ })()
```

	point.estimate	ci.lower	ci.upper
1	0.17	0.11	0.22

We are 95% confident that, out of women who have a family history of breast cancer, the proportion with a BRCA1 mutation is between 0.11 and 0.22.

6. The number of crickets is $n = 141$, and the number of spiders who choose dead crickets is $x = 98$. The null hypothesis that the spiders have no preference for dead crickets is $H_0 : p = 0.5$. The alternative that they prefer dead crickets is $H_A : p > 0.5$. The test statistic is the number of spiders who choose dead crickets. The null distribution of the test statistic is Binomial$(141, 0.5)$. The p-value is $P(X > 98)$ where X has the null distribution.

```
> (p.value <- sum(dbinom(98:141, size = 141, prob = 0.5)))

[1] 2.076751e-06
```

The proportion of spiders who prefer dead crickets, 98/141, is significantly larger than one half (binomial test, $P = 2.08 \times 10^{-6}$).

7. $n = 12, x = 5$.
   (a) We use the point estimate $p' = (x+2)/(n+4)$:

   ```
   > (point.estimate <- (5 + 2)/(12 + 4))

   [1] 0.4375
   ```

   (b) We use $p' \pm 1.96\sqrt{p'(1-p')/(n+4)}$:

   ```
   > (function() {
   +     p <- (5 + 2)/(12 + 4)
   +     ci <- 1.96 * sqrt(p * (1 - p)/(12 + 4))
   +     cat(sprintf("ci.lower is %f\n", p - ci))
   +     cat(sprintf("ci.upper is %f\n", p + ci))
   + })()

   ci.lower is 0.194422
   ci.upper is 0.680578
   ```

(c) Let $y = n - x$ be the number of students who failed to notice the woman. We use the point estimate $p' = (y+2)/(n+4)$:

```
> (point.estimate <- (12 - 5 + 2)/(12 + 4))
[1] 0.5625
```

Note that this answer differs from one minus the answer to (a), basically because we are using different priors.

8. (a) $n = 200$, $x = 109$. Our point estimate and confidence interval for the proportion of shoppers who have injured themselves are:

```
> (function() {
+     p <- (109 + 2)/(200 + 4)
+     ci <- 1.96 * sqrt(p * (1 - p)/(200 + 4))
+     print(xtable(data.frame(`point estimate` = p, `ci lower` = p -
+         ci, `ci upper` = p + ci)))
+ })()
```

	point.estimate	ci.lower	ci.upper
1	0.54	0.48	0.61

(b) It's not exactly a simple random sample of all UK consumers. For example, people who do not cook frequently are less likely to be included.