Week 01 Questions

Scott Graham 10131323

September 15, 2017

1.1

a.

The explanatory variables are Gender (male, female), and Mother's education (high school, college). The response variable is Attitude towards gun control (favour, oppose).

b.

The explanatory variables are Blood pressure, and Cholesterol level. The response variable is Heart Disease (yes, no).

c.

The explanatory variables are Race (white, nonwhite), Religion (Catholic, Jewish, Protestant), and Annual income. The response variable is Vote for president (Democrat, Republican, Other).

d.

The explanatory variable is Marital status (married, single, divorced, widowed). The response variable is Quality of life (excellent, good, fair, poor).

1.3

a.

Let $X =$ the number of correct answers by the student on the exam. Let $p = \frac{1}{4}$ be the probability of guessing the correct answer for a given question. Let $n = 100$ be the number of questions on the exam. Then:

$$X \sim Binom\left(n = 100, p = \frac{1}{4}\right)$$

b.

Assuming a Type I error rate of 5%:

$$H_0 : p = 0.25$$

$$H_1 : p > 0.25$$

$$\mathrm{P}\left(X \geq 50\right) = \sum_{x=50}^{100} \binom{100}{x} 0.25^x (1 - 0.25)^{100-x}$$

or equivalently in R:

```r
pbinom(50 - 1, 100, 0.25, lower.tail = FALSE)
```

```
## [1] 6.638502e-08
```

which is $< 5\%$. As such we reject the null hypothesis that the student was merely guessing for each question, and assume that their probability of getting the correct answer is > 0.25 based on the sample.

1.5

We need to find $\hat{p}$ s.t. we maximize $\mathrm{P}(Y = 0|n = 2)$ subject to $p \in [0, 1]$

$$\mathrm{P}(Y = 0|n = 2) = \binom{2}{0} p^0 (1-p)^{2-0} = (1-p)^2 \implies$$

$$\frac{d}{dp}(1-p)^2 = -2(1-p) \implies 0 = -2(1-p) \implies p = 1$$

$p = 0$ is also an endpoint, so we can check that as well. For $p = 1$, we have a likelihood of 0, and for $p = 0$, we have a likelihood of 1.

```r
# p = 0
dbinom(x = 0, size = 2, prob = 0)
```

```
## [1] 1
```

```r
# p = 1
dbinom(x = 0, size = 2, prob = 1)
```

```
## [1] 0
```

Since our function is monotonic over this domain, $p = 0$ is our MLE.

This may not be reasonable however, as this would require the coin to have both sides as tails. As well, we have a small sample size of $n = 2$, which makes it hard to draw a reasonable conclusion. If we use the bayesian estimator as given in the text, $p = \frac{1}{4}$, which may be a more reasonable estimator for p based on our limited data.

1.7

a.

Let X be the bullet firing (a somewhat morbid success…), $n = 6$, $p = \frac{1}{6}$:

$$\mathrm{P}\left(X = 0 \middle| n = 6, p = \frac{1}{6}\right) = \binom{6}{0}\left(\frac{1}{6}\right)^0\left(1 - \frac{1}{6}\right)^{6-0} = \left(\frac{5}{6}\right)^6 = 0.3349$$

b.

We have the case where we have $y - 1$ number of non-fires, and then 1 fire. Since the order of the non-fires doesn't matter, we don't need a binomial term in our equation, as it would $= 1$. As well, each spin is independent, which allows us to multiply our probabilities together.

$$\mathrm{P}(X = 0|n = y-1) \times \mathrm{P}(X = 1|n = 1) = (1-p)^{y-1} p \overset{p=\frac{1}{6}}{=} \left(\frac{5}{6}\right)^{y-1}\left(\frac{1}{6}\right)$$

1.11

$$\hat{p} = 0.86, n = 1158$$

$$\text{se}(\hat{p}) = \sqrt{\frac{0.86(1-0.86)}{1158}} = 0.0102$$

95% CI for p:

$$0.86 \mp 1.96(0.0102) = [0.84, 0.88]$$

That is, we can say with 95% confidence the true proportion of american adults that believe in heaven is somewhere between 84% and 88% based on our sample.

1.13

a.

$$\ell = \prod_{i=1}^{25} \text{P}(Y_i = 0) = \prod_{i=1}^{25} \binom{25}{0} p^{y_i}(1-p)^{25-y_i} =$$

Note: We can ignore the binomial term at the start, as it doesn't contain p.

$$\ell = p^{\sum_{i=1}^{25} y_i}(1-p)^{25-\sum_{i=1}^{25} y_i}$$

Since $y_i = 0, \forall 1 = 1, 2, \ldots, 25$:

$$\ell = (1-p)^{25}$$

Under H_0, we assume $p = 0.5$, our likelihood function (ℓ_0) is maximized at $p = 0.5$.

$$\implies \ell_0 = (0.5)^{25}$$

b.

$$\ell_1 = (1-p)^{25} \implies \frac{d}{dp}(1-p)^{25} = -25(1-p)^{24} \implies 0 = -25(1-p)^{24} \implies p = 1$$

We also get $p = 0$ due to it being our other endpoint in the domain $p \in [0, 1]$.

$$\ell_1(0) = (1-0)^{25} = 1$$

$$\ell_1(1) = (1-1)^{25} = 0$$

So our MLE is 0, with a likelihood of 1.

c.

$$\chi^2{}_{calc} = -2\ln\left(\frac{(1-0.5)^{25}}{(1-0)^{25}}\right) = -2\ln\left(0.5^{25}\right) = 34.65736 \approx 34.7$$

```
pchisq(q = -2*log(0.5^25), df = 1, lower.tail = FALSE)
```

```
## [1] 3.931495e-09
```

So we have a p-value of 3.93×10^{-9}, and can then reject our null hypothesis at that level.

d.

$$\chi^2{}_{calc} = -2\ln\left(\frac{(1-0.074)^{25}}{(1-0)^{25}}\right) = 3.844052 \approx 3.84$$

```
pchisq(q = -2*log((1-0.074)^25), df = 1, lower.tail = FALSE)
```

```
## [1] 0.04992273
```

So our p-value is ≈ 0.05 meaning it is the upper bound for the 95% likelihood-ratio confidence interval.