hw_3

Nicholas Wendt

2025-02-27

Questions: 1.14, 2.2, 2.3

Question 1.14

Part a $H_0 : \pi = 0.5$

(i) $H_a : \pi > 0.5$

```r
p.val.greater <- sum(dbinom(8:10, size = 10, prob = 0.5))
print(p.val.greater)
```

```
## [1] 0.0546875
```

(ii) $H_a : \pi < 0.5$

```r
p.val.less <- sum(dbinom(0:8, size = 10, prob = 0.5))
print(p.val.less)
```

```
## [1] 0.9892578
```

```r
#(i)
mid.p.val.greater <- sum(dbinom(9:10, size = 10, prob = 0.5)) + (0.5*dbinom(8,size = 10, prob = 0.5))
cat("Mid p-val for Ha > 0.5",mid.p.val.greater)
```

Part b

```
## Mid p-val for Ha > 0.5 0.03271484
```

```r
mid.p.val.less <- sum(dbinom(0:7, size = 10, prob = 0.5)) + (0.5*dbinom(8,size = 10, prob = 0.5))
cat("Mid p-val for Ha < 0.5", mid.p.val.less)
```

```
## Mid p-val for Ha < 0.5 0.9672852
```

Part c In part a, the reason that the sum of the p-values is greater than 1 is because P(Y=8) is accounted for in both the greater and lesser calculations. In part b, the probability of P(Y=8) is split evenly between the greater and lesser calculations.

```r
library(binom)
```

Part d

```
## Warning: package 'binom' was built under R version 4.3.3
```

```r
ci <- binom.exact(7, 10, alternative = "two.sided", midp = TRUE)
cat("Lower bound:", ci$lower, "\n")
```

```
## Lower bound: 0.3475471
```

```r
cat("Upper bound:", ci$upper)
```

```
## Upper bound: 0.9332605
```

Question 2.2

	Positive test	Negative test	
Disease	n11	n12	Lambda
No Disease	n21	n22	

Part a

$$P(X = 1|Y = 1) = \frac{P(Y = 1|X = 1)P(X = 1)}{P(Y = 1)}$$

where

$$P(Y = 1) = P(Y = 1|X = 1) * P(X = 1) + P(Y = 1|X = 2) * P(X = 2)$$

then

$$P(Y = 1) = \pi_1\gamma + \pi_2(1 - \gamma)$$

then

$$P(X = 1|Y = 1) = \frac{\pi_1\gamma}{\pi_1\gamma + \pi_2(1 - \gamma)}$$

Part b $\pi_1 = 0.86, \pi_2 = 0.12, \gamma = 0.01$

$P(Y = 1|X = 1) = \frac{0.86\times0.01}{(0.86\times0.01)+(0.12\times0.99)} = \frac{0.0086}{0.1274} \approx 0.0675$

The positive predicted values is about 6.75%

Part c

1. $P(X = 1, Y = 1) = \pi_1 \gamma = 0.86 * 0.01 = 0.0086$
2. $P(X = 1, Y = 2) = (1 - \pi_1)\gamma = 0.14 * 0.01 = 0.0014$
3. $P(X = 2, Y = 1) = \pi_2(1 - \gamma) = 0.12 * 0.99 = 0.1188$
4. $P(X = 2, Y = 2) = (1 - \pi_2)(1 - \gamma) = 0.88 * 0.99 = 0.8712$

	Y=1 Positive Test	Y=2 Negative Test
X=1 Disease	0.0086	0.0014
X=2 No Disease	0.1188	0.8712

False positives are much more common than true positives due to the low prevalence of the disease (0.01). Even though the sensitive (0.86), it's not sensitive enough to account for the low prevalence. This is probably good in this case since type 2 error in a test like this would be bad.

Question 2.3

$P_{us} = \frac{62.4}{10^6} = 0.0000624$

$P_{britian} = \frac{1.3}{10^6} = 0.0000013$

Difference in proportion:

Difference of proportions $= P_{us} - P_{britian} = 0.0000624 - 0.0000013 = 0.0000611$

This means the homicide rate in the us is 0.0000611 higher per resident when compared to Britain

Relative Risk:

Relative Risk $= \frac{P_{us}}{P_{britian}} = \frac{0.0000624}{0.0000013} \approx 48$

This means that the relative risk of being killed by a gun in the us is 48 times more likely than Britain

Relative risk is more useful for describing the strength of this association since it shows the relative difference between 2 statistics rather than the absolute difference. It is more easy to understand that something is 48 times more likely to happen rather than a 0.0000611 difference in probability.