stats100b_hw5

Takao

5/10/2022

1

The null and alternative hypotheses are where the alternative hypothesis is a two sided test

$$H_0 : \bar{d}_{placebo} = 0$$

$$H_1 : \bar{d}_{placebo} \neq 0$$

The sample size is 5

```
placebo_before <- c(1.9, 1.5, 1.7, 2.4, 1.5)
placebo_after <- c(1.91, 1.45, 1.54, 2.54, 1.54)
difference <- placebo_after - placebo_before
e.s.e <- sd(difference) / sqrt(5)
t_stats <- (mean(difference) - 0) / e.s.e
cat("P-value = ", 2*pt(t_stats, 5-1))
```

```
## P-value =  0.9396693
```

The P-Value is greater than 0.05. Thus, we do not have significant evidence to reject the null hypothesis that there is no placebo effect

```
# For verification, we have
t.test(difference)
```

```
##
##  One Sample t-test
##
## data:  difference
## t = -0.08055, df = 4, p-value = 0.9397
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.141875  0.133875
## sample estimates:
## mean of x
##    -0.004
```

2

$$H_0 : \mu_X - \mu_Y = 0$$

$$H_1 : \mu_X - \mu_Y \neq 0$$

```r
x <- c(1.16, 0.63, 0.075, 0.35)
y <- c(0.3, 2.70, 1.06, 2.79, 1.26)
Sp <- sqrt( ((4-1)*var(x) + (5-1)*var(y)) / (4+5-2) )
e.s.e <- Sp * sqrt(1/4 + 1/5)
t_stats <- (mean(x) - mean(y)) / e.s.e
cat("p-value =", 2*pt(t_stats, 4+5-2, lower.tail = TRUE))
```

```
## p-value = 0.111732
```

The p-value is greater than 0.05, so we fail to reject the null hypothesis that the means used in generating these two sets of numbers are the same.

```r
t.test(x,y,var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  x and y
## t = -1.819, df = 7, p-value = 0.1117
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.4569299  0.3204299
## sample estimates:
## mean of x mean of y
##   0.55375   1.62200
```

The p-value and the conclusion aligns.

3

The hypotheses are given by

$$H_0 : p = 0.5$$
$$H_a : p \neq 0.5$$

```r
p_hat <- 9207/17950
e.s.e <-sqrt(p_hat*(1-p_hat) / 17950)
z_stats <- (p_hat - 0.5) / e.s.e
cat("p-value = ", 2*pnorm(z_stats, lower.tail = FALSE))
```

```
## p-value =  0.0005313704
```

The p-value is less than 0.05, thus we have significant evidence to reject the null hypothesis that the coin is fair

4

a

We want to conduct a Pearson's Chi-square test

$$H_0 : P(\#ofheads = k) = \binom{5}{k} 0.5^k (1 - 0.5)^{5-k} for\ k = 0, 1, 2, 3, 4, 5$$

```
obs <- c(100, 524, 1080, 1126, 655, 105)
exp <- ratio <- numeric(length(obs))
n <- sum(obs)

for (i in 1:length(obs)){
  count <- i - 1
  exp[i] <- n*choose(5,count)*0.5^count*(1-0.5)^(5-count)
  ratio[i] <- (obs[i] - exp[i])^2/exp[i]

}

cat("p-value = ", pchisq(sum(ratio), 6-1, lower.tail = FALSE))
```

```
## p-value =  0.0006323943
```

The p-value is less than the significance level of 0.05 so we reject the null hypothesis that all coins are fair.

b

$$H_0 : P(\#ofheads = k) = \binom{5}{k} p^k (1 - p)^{5-k}\ for\ k = 0, 1, 2, 3, 4, 5$$

Since we do not know what p is, we will have to estimate for p so we will use MLE, for the binomial distribution

$$X \sim \binom{m}{x} p^x (1 - p)^{m-x}$$

$$\hat{p}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n \times m} = \frac{0 \times 100 + 1 \times 524 + 2 \times 1080 + 3 \times 1126 + 4 \times 655 + 5 \times 105}{(100 + 524 + 1080 + 1126 + 655 + 105) \times 5}$$

Notice the difference from the prior question

```
p_MLE <- sum(obs * c(0:5)) / (n*5)
exp <- ratio <- numeric(length(obs))

for (i in 1:length(obs )){
  count <- i - 1
  exp[i] <- n*choose(5, count)*p_MLE^count*(1-p_MLE)^(5-count)
  ratio[i] <- (obs[i] - exp[i])^2 / exp[i]
}
cat("p-value =", pchisq(sum(ratio), 6-1-1, lower.tail = FALSE))
```

```
## p-value = 0.06783489
```

Since the p-value is greater that 0.05, we do not have sufficient evidence to reject the null hypothesis that five coins have the same probability of heads.

5

The null hypothesis is given by

$$H_0 : P(starchy\ green) = 0.25(2 + \theta)$$
$$P(starchy\ white) = 0.25(1 - \theta)$$

$$P(sugary\ green) = 0.25(1 - \theta)$$

$$P(sugary\ white) = 0.25\theta$$

Since we do not know what the value of theta is, we will utilize MLE

```
t_1 <- 1997; t_2 <- 906; t_3 <- 904; t_4 <- 32; n <- t_1 + t_2 + t_3 + t_4
part <- t_1 - 2*t_2 - 2*t_3 - t_4; theta_MLE <- (part + sqrt(part^2 + 8*n*t_4)) / (2*n)

obs <- c(t_1, t_2, t_3, t_4)
exp <- ratio <- numeric(length(obs))
prob <- 0.25 * c(2 + theta_MLE , 1- theta_MLE, 1 - theta_MLE, theta_MLE)

for(i in 1:length(obs)){
  exp[i] <- n*prob[i]
  ratio[i] <- (obs[i] - exp[i])^2 / exp[i]
}

cat("P-value = ", pchisq(sum(ratio), 4-1-1, lower.tail = FALSE))
```

```
## P-value =  0.3650511
```

Since the p-value is greater than 0.05, we fail to reject the null hypothesis

6

The hypotheses are given through

$$H_0 : p = 0.5$$
$$H_a : p \neq 0.5$$

The level of significance is the probability of rejecting the null when the null is true which is the type 1 error

```
choose(10, 0)*0.5^0*(1-0.5)^10 + choose(10,10)*0.5^10*(1-0.5)^0
```

```
## [1] 0.001953125
```

The power 1 - beta is the probability of rejecting the null hypothesis when a specific alternative hypothesis alternative hypothesis p = 0.1 is true

```
choose(10,0)*0.1^0*(1-0.1)^10 + choose(10,10)*0.1^10*(1-0.1)^0
```

```
## [1] 0.3486784
```

7

a

$$H_0 : \mu = 0$$
$$H_a : \mu = -1$$

$$X \sim N(\mu, 4)$$
$$\bar{X} \sim N(\mu, \frac{1}{4})$$

The level of significance is the probability of rejecting the null hypothesis when the null is true or the Type 1 error

$$P(\bar{X} < c|\mu = 0) = 0.01$$

```
c <- qnorm(0.01, mean = 0, sd = sqrt(0.25))
cat("C = ", c)
```

```
## C =  -1.163174
```

b

The power 1 - beta is the probability of rejecting the null hypothesis when a specific alternative hypothesis mu = -1 is true

$$P(\bar{X} < c|\mu = -1) = 1 - \beta$$

```
cat("Power = ", pnorm(c, mean = -1, sd = sqrt(0.25), lower.tail = TRUE))
```

```
## Power =  0.3720806
```

8

We want to know how large n has to be

$$X \sim N(\mu_X, 100)$$
$$Y \sim N(\mu_Y, 100)$$

$$\bar{X} - \bar{Y} \sim N(\mu_X - \mu_Y, \frac{2\sigma^2}{n} = \frac{200}{n}$$

We reject the null hypothesis if

$$\bar{X} - \bar{Y} > C$$

Under the null hypothesis, we have that

$$P(\bar{X} - \bar{Y} > C | \mu_X - \mu_Y = 0) = \alpha$$
$$P(\frac{\bar{X} - \bar{Y} - 0}{\sqrt{200/n}} > \frac{C}{\sqrt{200/n}}) = \alpha$$

$$P(Z > \frac{C}{\sqrt{200/n}}) = 0.1$$

$$\frac{C}{\sqrt{200/n}} = Z_{1-0.1}$$

Additionally, under the alternative hypothesis, we have that

$$P(\bar{X} - \bar{Y} > C | \mu_X - \mu_Y = 2) = 1 - \beta$$

$$P(\frac{\bar{X} - \bar{Y} - 2}{\sqrt{200/n}} > \frac{C - 2}{\sqrt{200/n}}) = 1 - \beta$$

Since power = 0.5

$$P(Z > \frac{C - 2}{\sqrt{200/n}}) = 0.5$$

$$\frac{C - 2}{\sqrt{200/n}} = 0$$

Given that C = 2, we have

$$\frac{2}{\sqrt{200/n}} = Z_{1-0.1}$$

$$n = \frac{Z_{1-0.1}^2 \times 200}{4}$$

```r
cat("n = ", ceiling(qnorm(0.9, 0.1, lower.tail = FALSE)^2 * 200 /4))
```

```
## n =  70
```

9

a

FALSE

b

FALSE

c

FALSE

d

TRUE

e

FALSE

10

a

FALSE

b

TRUE

c

FALSE

d

FALSE

```r
pchisq(8.5, 4, lower.tail = FALSE)
```

```
## [1] 0.07488723
```

e

FALSE