HW2

2022-09-08

Group 3

Mahibul, Amira, Minghao

Part 1. Write-up for experiment - roll the dice.

If a dice is fair, it should roll each number approximately the same number of times if I have a large enough try (say 10000). It is obviously impossible to do so unless the payoff of doing such an experiment is high. To accomodate time constraits and to determine the accuracy of rolling a dice, I decided to roll a dice only 35 times and have the following rules set up:

1. A fair dice would have an expected value of 3.5, because
$\frac{1}{6} * 1 + \frac{1}{6} * 2 + \frac{1}{6} * 3 + \frac{1}{6} * 4 + \frac{1}{6} * 5 + \frac{1}{6} * 6 = 3.5.$

2. A fair dice's population variance should follow a discrete probability distribution

$$Var(X) = (x - \mu)^2 P(X = x),$$

Knowing the probability density function,

$$Var(X) = \frac{1}{6}[(1 - 3.5)^2 + (2 - 3.5)^2 + (3 - 3.5)^2 + (4 - 3.5)^2 + (5 - 3.5)^2 + (6 - 3.5)^2] = 2.9167$$

3. Given Variance and population average, I can set up a simple hypothesis testing as follows.

Step 1. Setting up my hypothsis
$H_o : \bar{X} = \mu$

$H_1 : \bar{X} \neq \mu$

Step 2. Testing to see if my reuslts are within normal range based on stats given above. Using $a = 0.1$ because it doesn't have to be that accurate (getting a type I error doesn't cost me anything since I won't go to Las Vegas with this dice).

```
qnorm(0.05) #For two-tailed test
```

```
## [1] -1.644854
```

Step 3. Conclusion This means if my result is $\mu - 1.64\frac{Sample\ SD}{\sqrt{(35)}} < \bar{X} < \mu + 1.64\frac{Sample\ SD}{\sqrt{(35)}}$, I will conclude that my result isn't statistically significant and I would fail to reject the null hypothesis. In other words, I have a fair dice and vice versa.

Part 2. Interesting results from data

Initiate the data

```
load('Household_Pulse_data.RData')
attach(Household_Pulse_data)
```

I want to see if there's a trend associated with higher education vs vaccination status

```
all_doses <- data.frame(matrix(ncol=1+length(summary(EEDUC)),nrow=0))
colnames(all_doses) <- levels(unique(EEDUC))
for (i in 1:length(summary(EEDUC))){
  all_doses[1,i] <- summary(EEDUC[DOSESRV=='yes got all doses' | DOSESRV == 'yes plan to get
all doses'])[i] / summary(EEDUC)[i]
}
all_doses
```

```
##   less than hs   some hs HS diploma some coll assoc deg  bach deg   adv deg NA
## 1    0.6885645 0.6826923  0.7650503 0.8144697 0.8251199 0.9030635 0.9367774 NA
```

I want to also include people who "claim" they are going to get vaccinated.
Some people might say vaccination is a good thing but never do it. Let's see what percentage of people actually got vaccinated.

```
for (i in 1:length(summary(EEDUC))){
  all_doses[2,i] <- summary(EEDUC[DOSESRV=='yes got all doses'])[i] / summary(EEDUC)[i]
}
all_doses
```

```
##   less than hs   some hs HS diploma some coll assoc deg  bach deg   adv deg NA
## 1    0.6885645 0.6826923  0.7650503 0.8144697 0.8251199 0.9030635 0.9367774 NA
## 2    0.6131387 0.6111111  0.7146494 0.7760345 0.7904901 0.8812951 0.9231854 NA
```

```
all_doses[3,] <- all_doses[1,] - all_doses[2,] #for percentage difference
```

Some cleaning -

```
all_doses[8] <- NULL
all_doses_t <- t(all_doses) #For better viewing
colnames(all_doses_t) <- c('Included','Not Included','percentage_difference')
all_doses_t
```

```
##               Included Not Included percentage_difference
## less than hs 0.6885645    0.6131387            0.07542579
## some hs      0.6826923    0.6111111            0.07158120
## HS diploma   0.7650503    0.7146494            0.05040092
## some coll    0.8144697    0.7760345            0.03843519
## assoc deg    0.8251199    0.7904901            0.03462973
## bach deg     0.9030635    0.8812951            0.02176837
## adv deg      0.9367774    0.9231854            0.01359201
```