p8130_hw2_zj2357

Zekai Jin(zj2357)

2022-10-05

```
library(tidyverse)
knitr::opts_chunk$set(
  fig.width = 6,
  fig.asp = .6,
  out.width = "90%"
)
```

Problem 1

a)
Denote X to be the number of individuals having one or more dental checkup.
Obviously,

$$X \sim B(56, 0.73)$$

Thus,

$$P(X = 40) = \binom{56}{40} 0.73^{40} \times 0.27^{16} = 0.113$$

b)

$$P(X \geq 40) = \sum_{i=40}^{56} P(X = i)$$

To calculate this, use the following R code

```
sum(dbinom(40:56,56,0.73))
```

```
## [1] 0.6678734
```

Thus, the probability of at least 40 individuals having dental checkup is 0.67.

c)
We can use normal distribution, considering

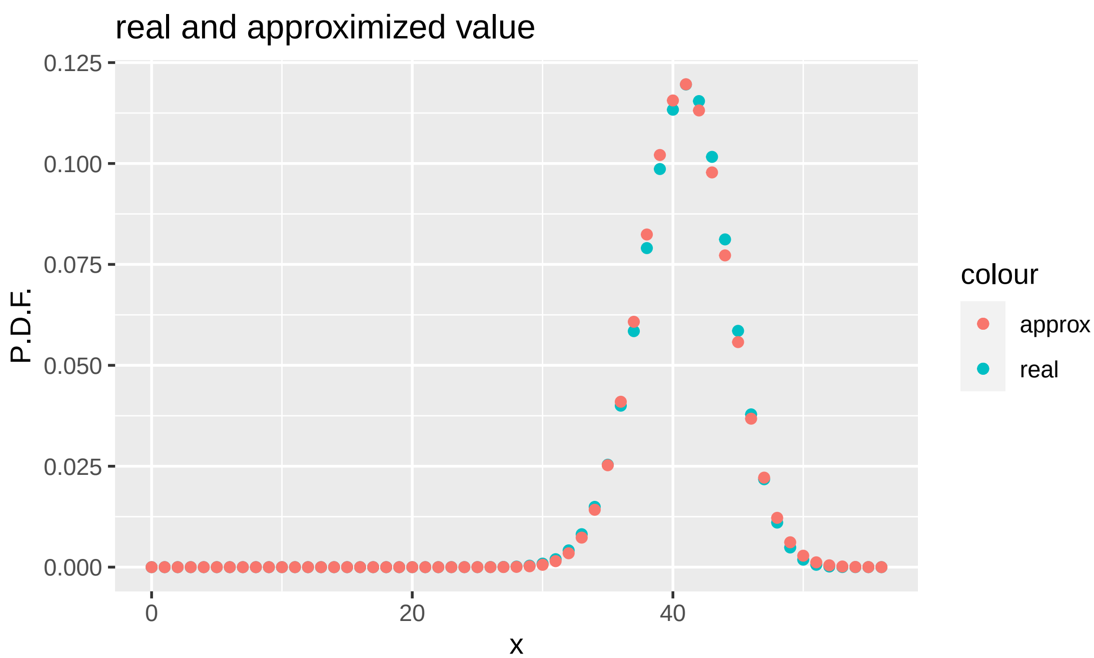
$$np = 40.88 > 10$$

$$nq = 15.72 > 10$$

to compare the differences between real and approximized values, use the following R code:

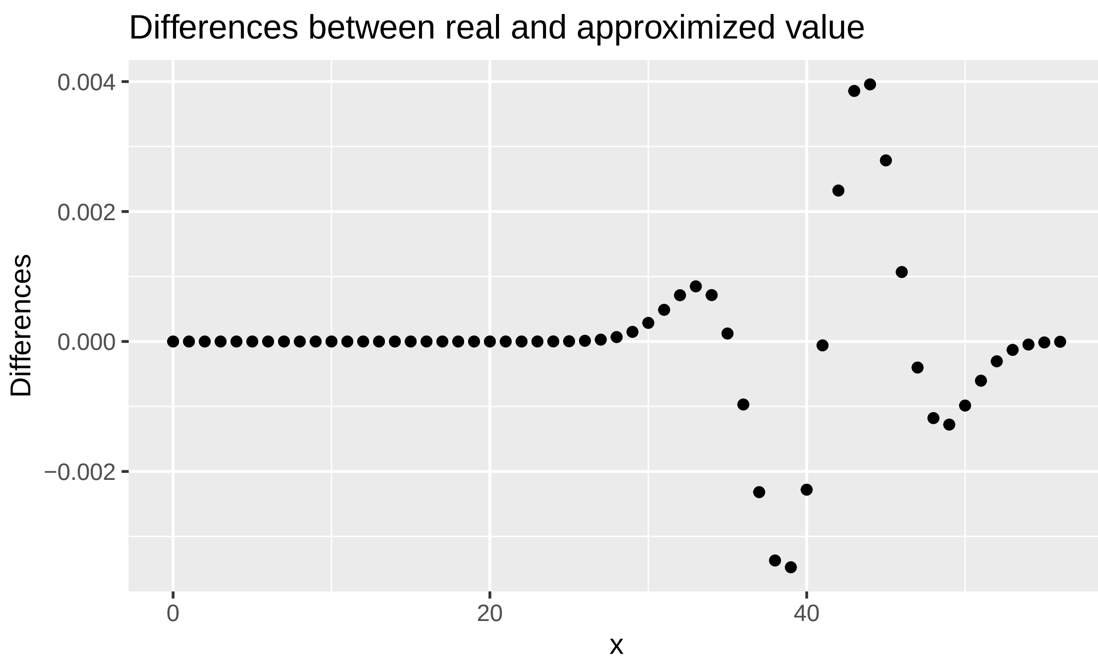
```
tibble(
  x = 0:56,
  real = dbinom(x,56,0.73),
  approximized = pnorm(x+0.5,40.88,3.32)-pnorm(x-0.5,40.88,3.32)
```

```
) %>%
ggplot() +
geom_point(aes(x=x,y=real,color="real")) +
geom_point(aes(x=x,y=approximized,color="approx"))+
labs(
  title = "real and approximized value",
  x = "x",
  y = "P.D.F."
)
```



The difference is plotted below

```
tibble(
  x = 0:56,
  real = dbinom(x,56,0.73),
  approximized = pnorm(x+0.5,40.88,3.32)-pnorm(x-0.5,40.88,3.32)
  ) %>%
  ggplot() +
  geom_point(aes(x=x,y=real-approximized)) +
  labs(
    title = "Differences between real and approximized value",
    x = "x",
    y = "Differences"
  )
```



d)

$$\bar{X} = np = 40.88$$

e)

$$sd = \sqrt{npq} = \sqrt{56 \times 0.73 \times 0.27} = 3.32$$

Problem 2

Denote X to be the number of tornadoes in the next year.
Obviously,

$$X \sim Pois(6)$$

a)

$$P(X < 3) = \sum_{i=0,1,2} P(X = i)$$

To calculate this, use the following R code

```
sum(dpois(0:2,6))
```

```
## [1] 0.0619688
```

Thus, the probability of having less than 3 tornadoes is 0.0619688.

b)

$$P(X = 3) = \frac{\lambda^3}{3!} e^{-6} = 0.089$$

c)

$$\begin{aligned} &P(X > 3) \\ &= 1 - P(x < 3) - P(X = 3) \\ &= 1 - 0.062 - 0.089 \\ &= 0.85 \end{aligned}$$

Problem 3

a)
Denote X to be the blood pressure of a randomly selected male and $Z = \frac{X-u}{\sigma}$.
Obviously,

$$X \sim N(128.0, 10.2)$$
$$Z \sim N(0, 1)$$

Thus,

$$P(X > 137) = P(Z > 0.882) = 1 - \Phi(0.882) = 0.19$$

b)
Denote Y to be the mean blood pressure of 50 randomly selected male and $Z = \frac{Y-u}{\sigma\sqrt{n}}$. Obviously,

$$Y \sim N(128.0, 1.44)$$
$$Z \sim N(0, 1)$$

Thus,

$$P(Y < 125) = P(Z < -2.083) = 1 - \Phi(2.083) = 0.019$$

c)
Denote X to be the mean blood pressure of 40 randomly selected male and $Z = \frac{X-u}{\sigma/\sqrt{n}}$. Obviously,

$$X \sim N(128.0, 1.61)$$
$$Z \sim N(0, 1)$$

Because

$$\Phi(1.28) = 0.9$$

The 90th percentile of Z is 1.28
Thus, the 90th percentile of the sample mean

$$X = u + 1.28\frac{\sigma}{\sqrt{n}} = 130.1$$

Problem 4

Denote X to be the pulse of a randomly selected women with fibromyalgia.
Accoringly, we have

$$n = 40$$
$$\hat{u} = \bar{X} = 80$$
$$\hat{\sigma} = 10$$

a)
Because X follows t-distribution with $df = 39$, the 95% confidence interval is

$$[\bar{X} - 2.02\frac{\hat{\sigma}}{\sqrt{n}}, \bar{X} + 2.02\frac{\hat{\sigma}}{\sqrt{n}}] = [76.8, 83.2]$$

b)
The 95% confidence interval means that if we repeat sampling, using the same method, 95% of the CIs will cover the true population mean.
c)

$$H_0 : u = 70$$

$$H_1 : u \neq 70$$

Under H_1, we have

$$t = \frac{\bar{X} - u_0}{\hat{\sigma}/\sqrt{n}} = 6.32$$

Because $t_a = 2.704 < z$, we can reject H_0. The mean pulse of young women suffering from fibromyalgia is not equal to 70