Week 4 Lab

AUTHOR
Jessica Tran

PUBLISHED
February 16, 2024

Problem 1

Suppose a genetic condition is in 2% of the population. If you have the genetic condition, you are 100 times more likely to develop a specific type of cancer. Suppose you have a group of 4 patients who have cancer. Let `X` equal the number of cancer patients that have the genetic condition. Determine if this is a binomial distribution or a poisson distribution and find the corresponding PDF and a CDF in table form (that is, use `data.frame`).

```
data.frame(
  "X" = 0:4,
  "PDF" = dbinom(c(0:4), 4, p = 0.02),
  "CDF" = pbinom(c(0:4), 4, p = 0.02)
)
```

```
  X        PDF       CDF
1 0 0.92236816 0.9223682
2 1 0.07529536 0.9976635
3 2 0.00230496 0.9999685
4 3 0.00003136 0.9999998
5 4 0.00000016 1.0000000
```

Find the probability less than two patients have the genetic condition using either the PDF or the CDF.

```
pbinom(1, 4, p = 0.02)
```

```
[1] 0.9976635
```

Problem 2

The average exam score is normally distributed with mean 70 with a standard deviation of 4. Find the probability someone gets below a 60.

```
pnorm(60, mean = 70, sd = 4)
```

```
[1] 0.006209665
```

Problem 3

Suppose the price for generic ibuprofen is normally distributed with mean of $5.00 and standard deviation of $0.50. If the price you pay is the 90th percentile, how much are you paying for the ibuprofen?

```
Z <- qnorm(.90)
5 + Z*0.5
```

```
[1] 5.640776
```

Problem 4

Find a 90% confidence interval for a random sample of twenty students who took a driving test where 65% of the student's passed. Make sure to check any assumptions. Comment out each line of code to explain what you are doing.

```
phat <- 13/20        # sample proportion
n <- 20             # sample size
n*phat
```

```
[1] 13
```

```
n*(1-phat)
```

```
[1] 7
```

```
Z <- qnorm(1 - 0.1/2)               # z score
L1 <- phat - Z*sqrt(phat*(1-phat)/n) # lower endpt
L2 <- phat + Z*sqrt(phat*(1-phat)/n) # upper endpt
c(L1, L2)
```

```
[1] 0.4745704 0.8254296
```

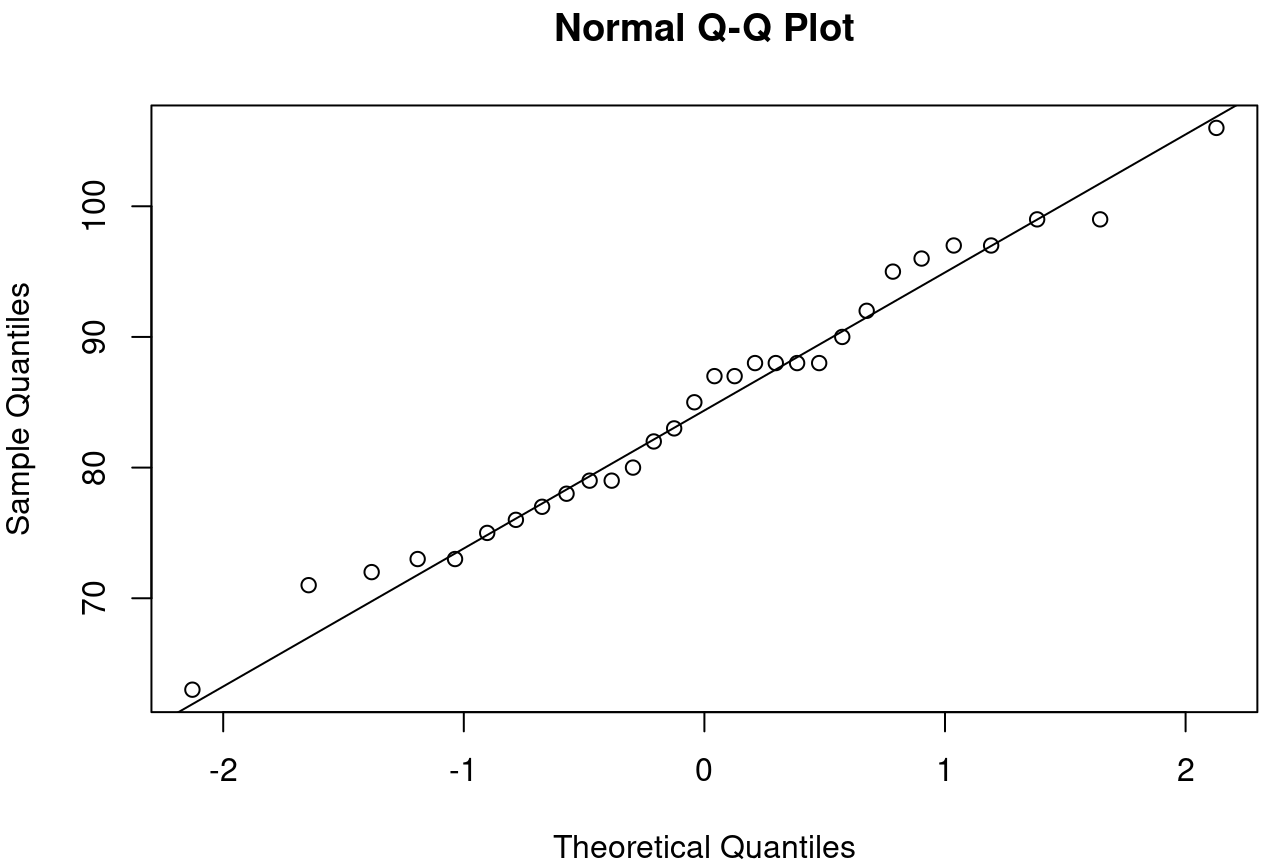
Problem 5

Find a 95% confidence interval for a random sample of thirty highway speeds of cars on I-24 was taken. The data is below. Make sure to check any assumptions. Comment out each line of code to explain what you are doing.

85	73	82
106	83	87
75	96	97

79	88	88
99	88	80
99	97	72
92	77	95
63	78	87
79	73	88
90	76	71

```
speed <- c(85, 73, 82, 106, 83, 87, 75, 96, 97, 79, 88, 88, 99, 88, 80, 99, 97, 72, 92, 7
qqnorm(speed)
qqline(speed)
```




```
xbar <- mean(speed)              # sample mean
t <- qt(0.95, df = 30 - 1)       # t score (1 - .05/2)
s <- sd(speed)                   # sample sd
n <- 30                           # sample size
```

```
L1 <- xbar - t*s/sqrt(n)          # lower endpt
L2 <- xbar + t*s/sqrt(n)          # upper endpt
c(L1, L2)
```

```
[1] 81.61180 87.92153
```

Submitting

Submit the following to Canvas:

- Your rendered PDF titled Lastname_4R. Make sure your name is at the top of the document.
- Your .qmd file