Lab 4 - Problems

Aaron Roche

Load the mosaic and tidyverse (installed last week) packages:

```
library(mosaic)
library(tidyverse)
```

Problem 1

Say you have a machine that has 5 components, and the probability each one fails is 0.05. If component failures are independent of each other, you can model the number of failures as binomial. Use the R functions `dbinom(x,size,prob)`, `pbinom(x,size,prob)`, and `rbinom(n,size,prob)` that are analogous to the Poisson functions above.

a) What are the probabilities that 0, 1, 2, 3, 4, or 5 components fail?

```
# The probability that 0 components fail is dbinom(0, 5, 0.05) = 0.7737809.
# The probability that 1 components fail is dbinom(1, 5, 0.05) = 0.2036266.
# The probability that 2 components fail is dbinom(2, 5, 0.05) = 0.02143438.
# The probability that 3 components fail is dbinom(3, 5, 0.05) = 0.001128125.
# The probability that 4 components fail is dbinom(4, 5, 0.05) = 2.96875e-05.
# The probability that 5 components fail is dbinom(5, 5, 0.05) = 3.125e-07.
```

b) The machine jams if three or more components fail. What's the probability two or fewer components fail?

```
# The probability that two or fewer components fail is pbinom(2, 5, 0.05) =
0.9988419.
```

c) What is the population mean (expected value) of the number of component failures? (You don't really need R for this–it's a formula from class–but you can use R as a calculator:

```
# The population mean (expected value) of the number of component failures is
n * p or 5 * 0.05 = 0.25.
```

d) Generate a random sample of 10,000 draws from the $Bin(5,0.05)$ distribution. What are the sample proportions for 0-5 failures (i.e. what proportion of draws had 0 failures, etc.)? What is the (sample) mean number of failures in the 10,000 draws? How do these compare with your answers to (a) and (c)?

```
# randBinom <- rbinom(10000, 5, 0.05)
# tally(~randBinom,format="proportion")

# The sample proportion for 0 failures is 0.7688.
# The sample proportion for 1 failures is 0.2066.
# The sample proportion for 2 failures is 0.0236.
# The sample proportion for 3 failures is 0.0010.
# The sample proportion for 4 failures is 0.0000 (very close to 0).
```

```
# The sample proportion for 5 failures is 0.0000 (very close to 0).

# The sample mean number of failures in the 10,000 draws is mean(randBinom) =
0.2568.

# The answers from (a) are very close to the sample proportions in (d). The
difference between each value is than 0.01 (ex. 0.7737809 - 0.7688 =
0.0049809). Also, the answer from (c) is very close to (d) because the
difference between 0.25 and 0.2568 is only 0.0068.
```

Problem 2

If $X \sim N(-4,0.5)$ what is: a) $Pr(X < -5)$?

```
# pnorm(-5, mean = -4, sd = 0.5) = 0.02275013
```

b) $Pr(-5 < X < -4.5)$?

```
# pnorm(-4.5, mean = -4, sd = 0.5) - pnorm(-5, mean = -4, sd = 0.5) =
0.1359051
```

c) What is the number x such that $Pr(X < x) = 0.2$?

```
# The value of x that satisfies Pr(X < x) = 0.2 is qnorm(0.2, mean = -4, sd =
0.5) = -4.420811
```

If $Y \sim U(5,10)$ what is: d) $Pr(Y < 8)$?

```
# punif(8, min = 5, max = 10)
```

e) $Pr(7 < Y < 8)$?

```
# punif(8, min = 5, max = 10) - punif(7, min = 5, max = 10) = 0.2
```

f) The following code draws a random sample of size 100 from the $U(5,10)$ distribution called `randUnif`. Make a normal QQ Plot of `randUnif`. Does the normal approximation look appropriate?

```
randUnif <- runif(100,5,10)
# gf_qq(~randUnif)%>%gf_qqline()

# The normal approximation look appropiate because the line sits on top of
most of the data points on the graph.
```

Problem 3

a) Use the `rpois` function to generate 10,000 draws from the Poisson distribution for λ values $\lambda = 5,50,500$. Make histograms and Normal QQ plots for each value of λ. When do you think a normal approximation might be valid? Example:

```
draws <- rpois(10000,lambda = 1)
gf_histogram(~draws,title=expression(paste(lambda,"=1")))
gf_qq(~draws,title=expression(paste(lambda,"=1")))%>%gf_qqline()

# lambda = 5
# draws <- rpois(10000, lambda = 5)
```

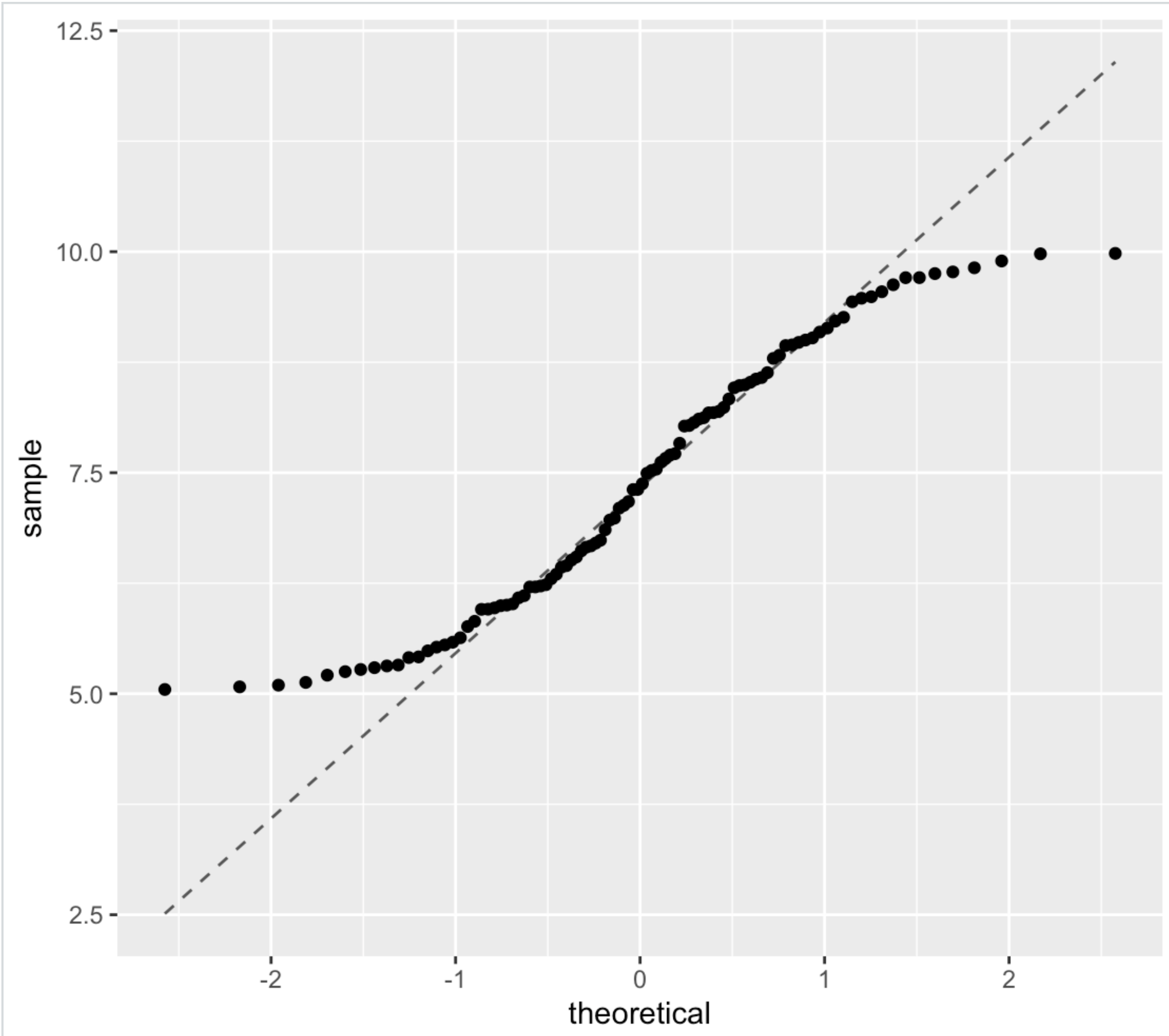
```
# gf_histogram(~draws,title=expression(paste(lambda,"= 5")))
# gf_qq(~draws,title=expression(paste(lambda,"= 5")))%>%gf_qqline()

# lambda = 50
# draws <- rpois(10000, lambda = 50)
# gf_histogram(~draws,title=expression(paste(lambda,"= 50")))
# gf_qq(~draws,title=expression(paste(lambda,"= 50")))%>%gf_qqline()

# lambda = 500
# draws <- rpois(10000, lambda = 500)
# gf_histogram(~draws,title=expression(paste(lambda,"= 500")))
# gf_qq(~draws,title=expression(paste(lambda,"= 500")))%>%gf_qqline()

# A normal approximation becomes more valid as the value of lambda increaes.
For example, 50 and 500 are both high enough that the normal approximation
will be accurate. It is best to have a value of lambda greater than 10.
```

b) Use the `rbinom` function to generate 10,000 draws from the Binomial distribution for $n=$5, 50, & 500 and $\pi = 0.5$. Make histograms and Normal QQ plots for each value of n. When do you think a normal approximation might be valid? Example:

```
draws <- rbinom(10000,size=10,prob=0.5)
gf_histogram(~draws,title="n=10")
gf_qq(~draws,title="n=10")%>%gf_qqline()
```

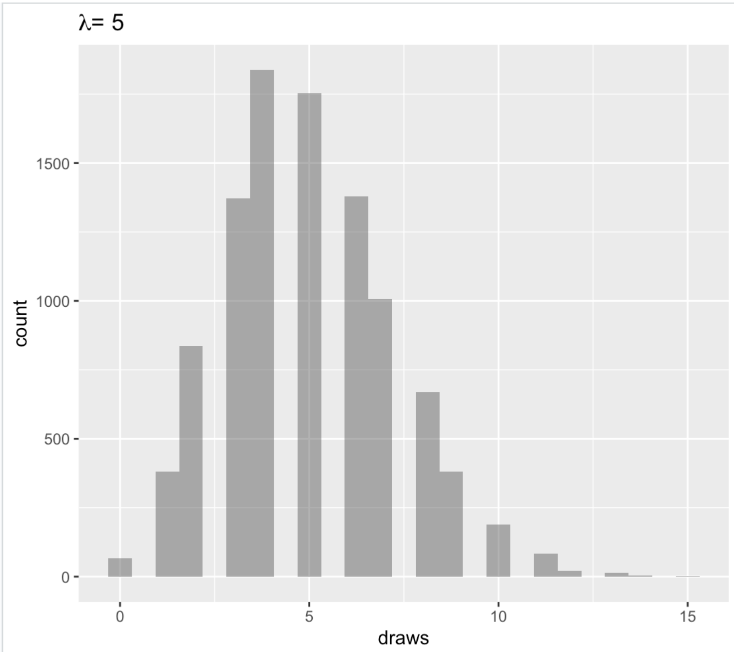
```
# n = 5
# draws <- rbinom(10000,size=5,prob=0.5)
# gf_histogram(~draws,title="n = 5")
# gf_qq(~draws,title="n = 5")%>%gf_qqline()

# n = 50
# draws <- rbinom(10000,size=50,prob=0.5)
# gf_histogram(~draws,title="n = 50")
# gf_qq(~draws,title="n = 50")%>%gf_qqline()

# n = 500
# draws <- rbinom(10000,size=500,prob=0.5)
# gf_histogram(~draws,title="n = 500")
# gf_qq(~draws,title="n = 500")%>%gf_qqline()

# When the value of n is increasing, the normal approximation becomes more
valid. This is because the number of data points that sits on the line in the
QQ plot increase as the the value of n increases. Also, the graph of the
histogram is closer to a bell-curve shape. It is ideal to have a value of n
that is greater than or equal to 10.
```

Problem 2f

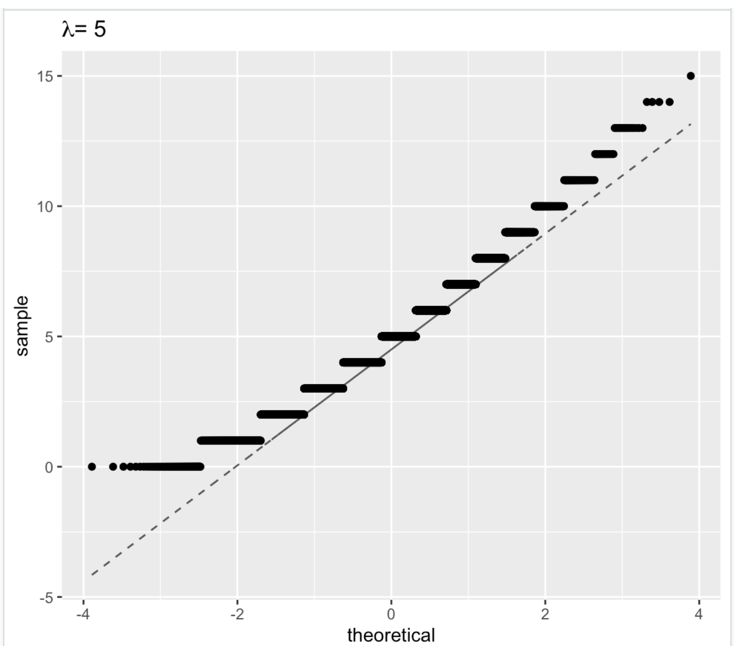


Problem 3a (lambda = 5)

Histogram

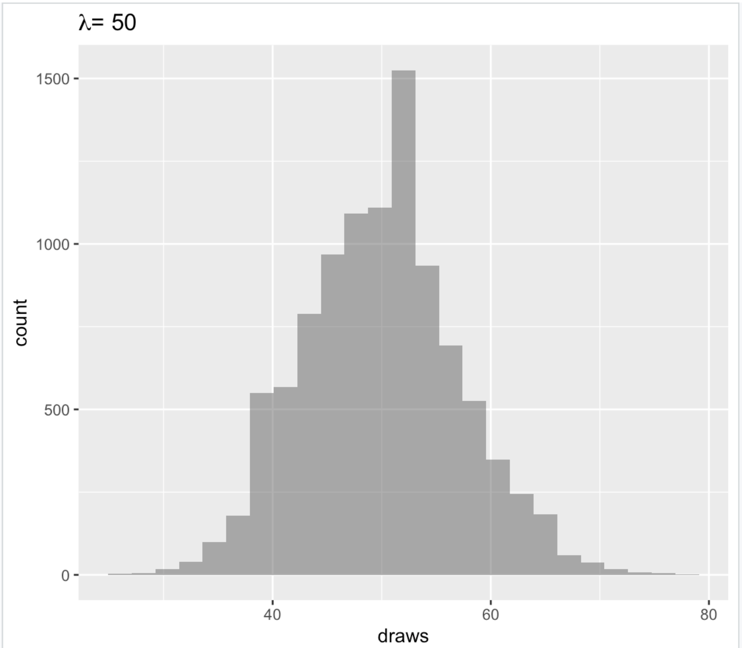


QQ Plot

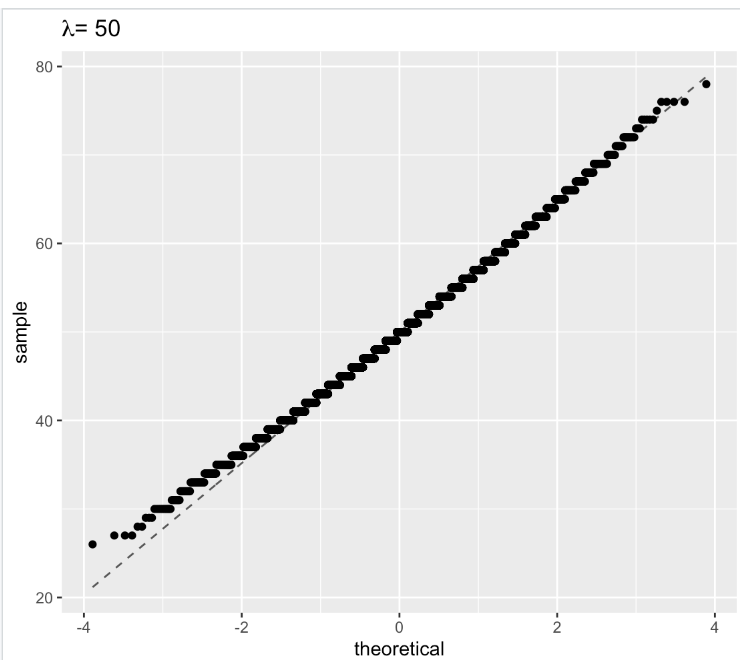


Problem 3a (lambda = 50)

Histogram

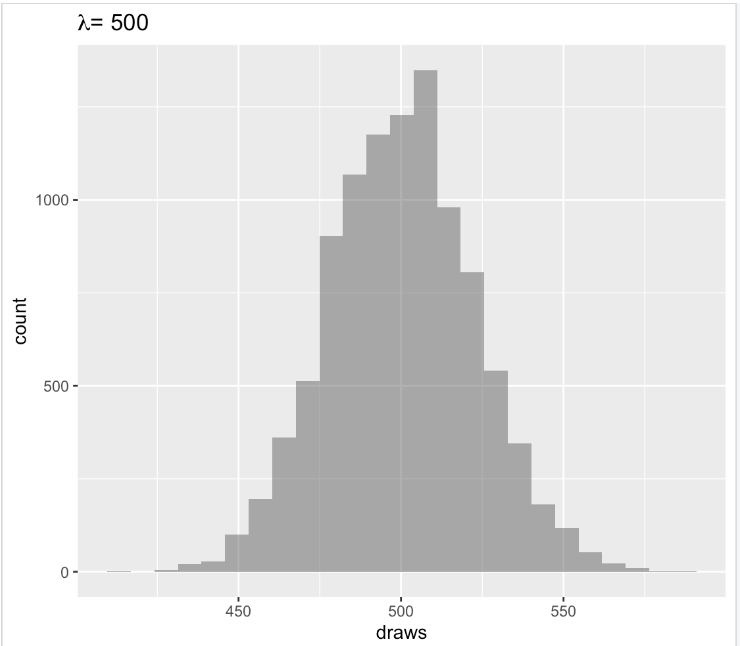


QQ Plot

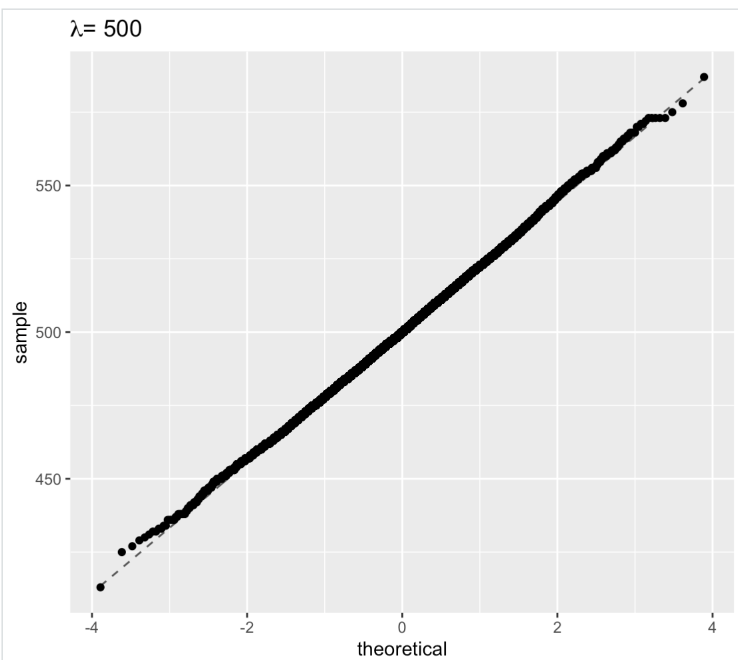


Problem 3a (lambda = 500)

Histogram

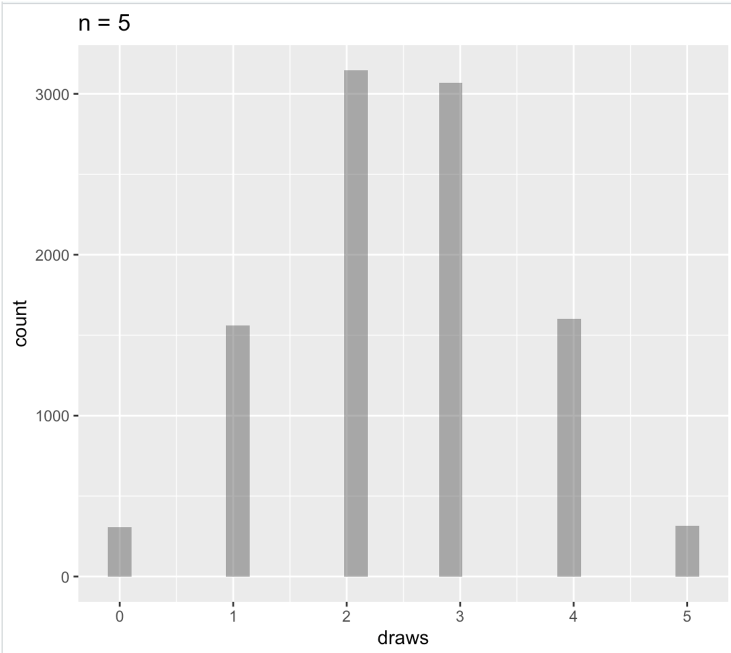


QQ Plot

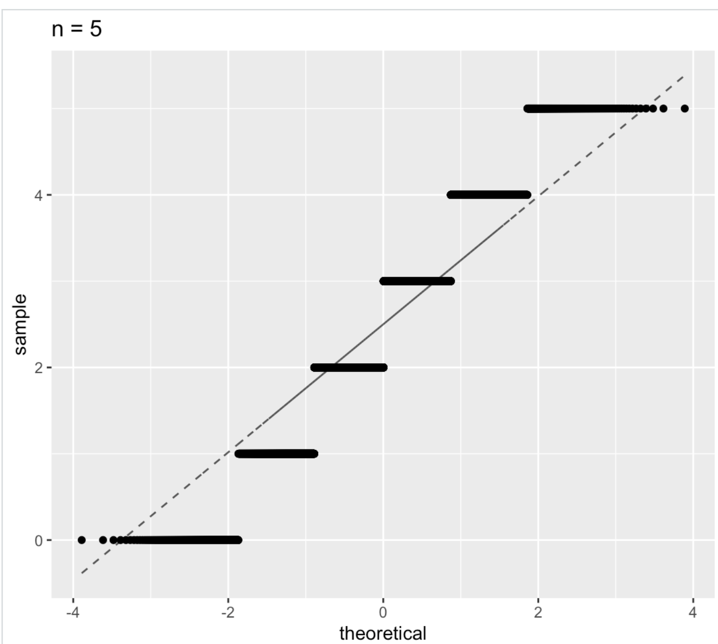


Problem 3b (n = 5)

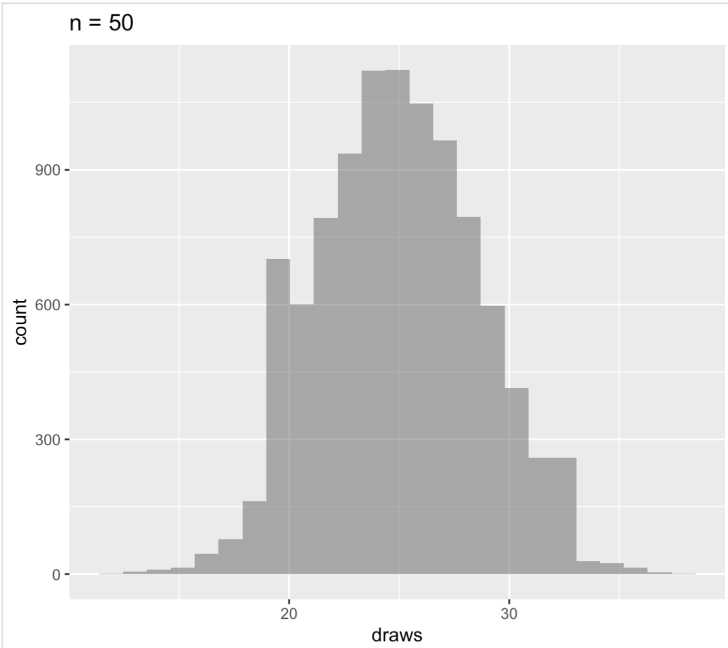
Histogram



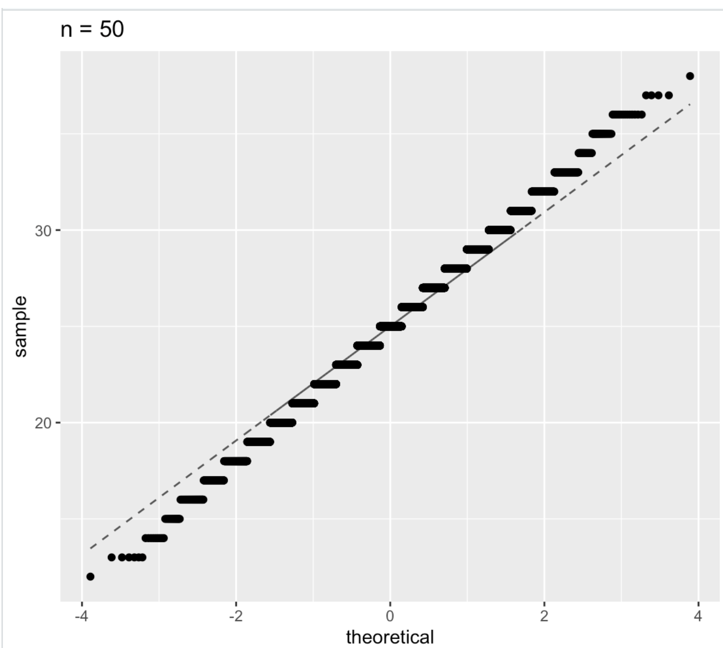
QQ Plot



<u>**Problem 3b (n = 50)**</u>
Histogram

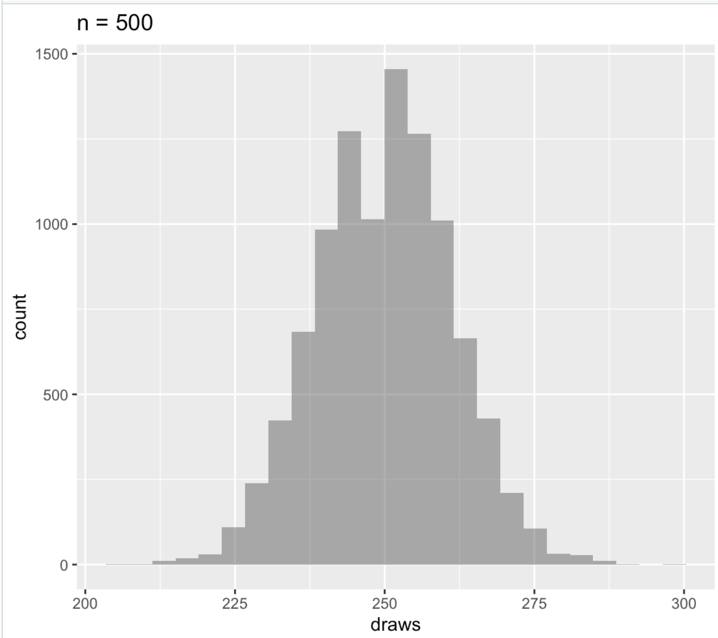


QQ Plot



<u>Problem 3b (n = 500)</u>

Histogram



QQ Plot