week3script

Summer Heschong

2025-01-29

#load packages

```
library(here)
```

```
## here() starts at /Users/summerheschong/Documents/GitHub/stats_spring25
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.2
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(moments)
```

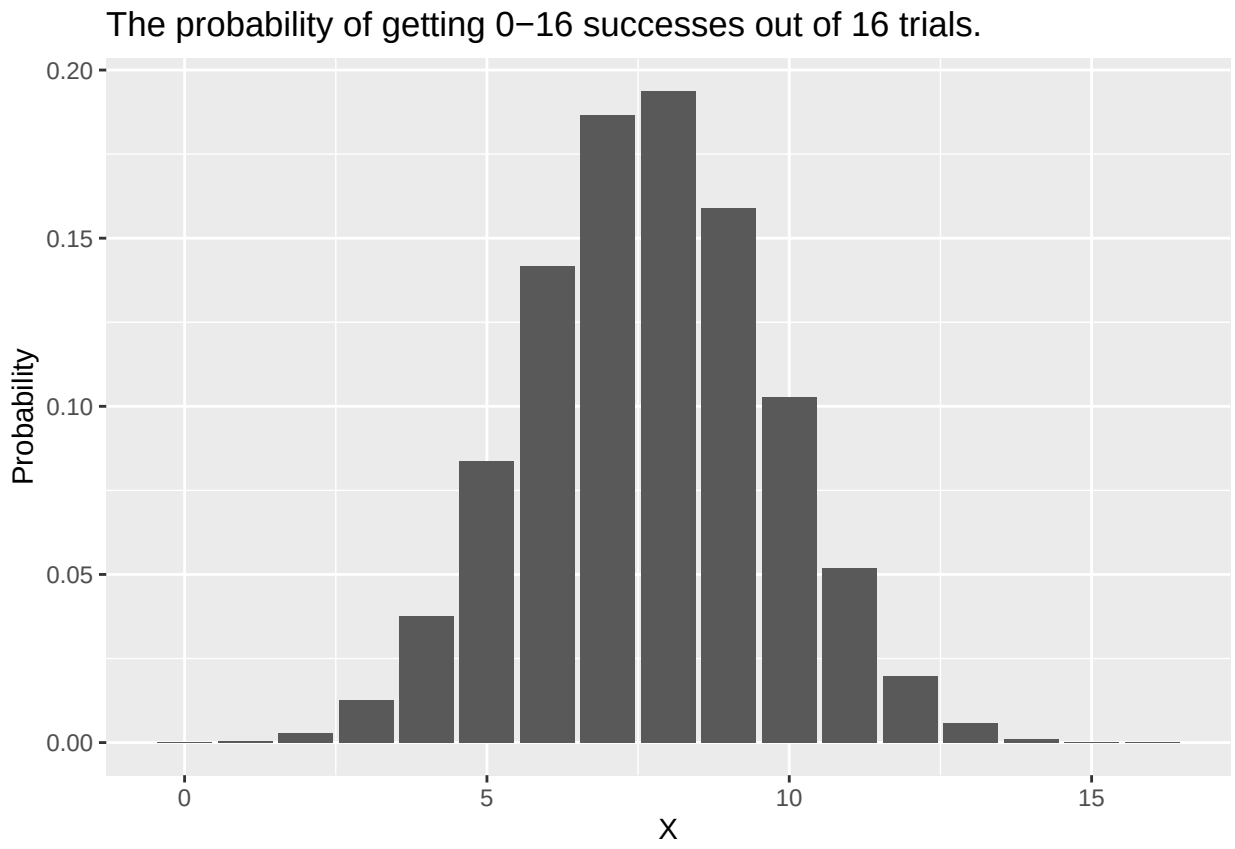
#Discrete Distributions (Bernoulli Trial and Poisson Distribution) ###Bernoulli Trial

```
#probability of three successes
dbinom(x = 3, size = 16, prob = 0.48)
```

```
## [1] 0.01258796
```

```
#probability of 0-16 successes
prob <- dbinom(x = 0:16, size = 16, prob = 0.48)

#create a list of numbers of successes
success <- 0:16

#bind two columns together to create a dataframe
binom_df <- cbind(prob, success)
```

```r
#create a fugure
ggplot(binom_df, aes(x = success,
                     y = prob)) +
  geom_bar(stat = "identity") +
  labs(x = 'X',
       y = 'Probability',
       title = 'The probability of getting 0-16 successes out of 16 trials.')
```




```r
#Cumulative probability using dbinom
sum(dbinom(x = 0:3, size = 16, prob = 0.48))
```

```
## [1] 0.01596084
```

```r
#cumulative probability using generated data
#adding together rows 1 through 4
sum(prob[1:4])
```

```
## [1] 0.01596084
```

```r
#Probability of three of fewer successes
pbinom(q = 3, size = 16, prob = 0.48)
```

```
## [1] 0.01596084
```

###Poisson distribution models counts of outcomes rather than modeling successes and failures

```
#probability of 10 occurrences
dpois(x = 10, lambda = 3)
```

```
## [1] 0.0008101512
```

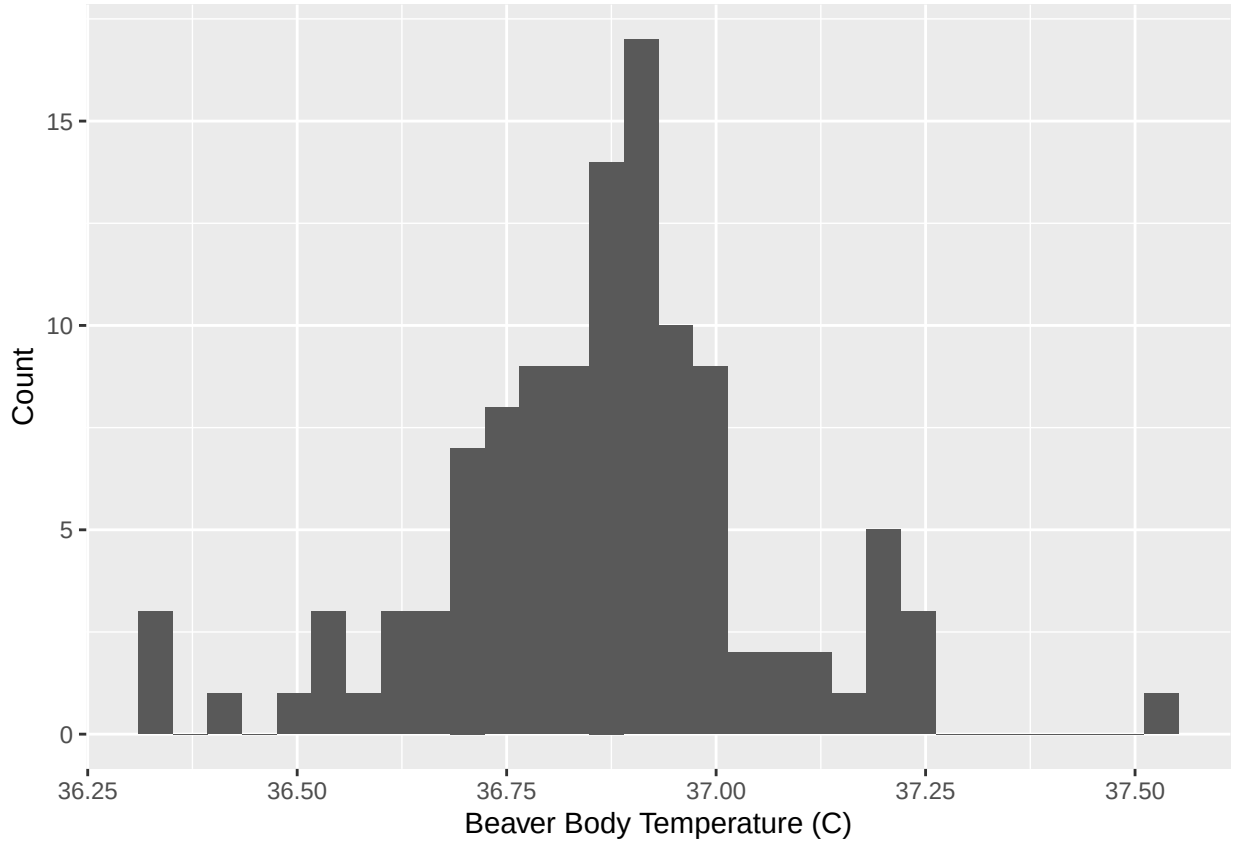
```
#Probability of 3 or fewer occurrences
ppois(q = 3, lambda = 3)
```

```
## [1] 0.6472319
```

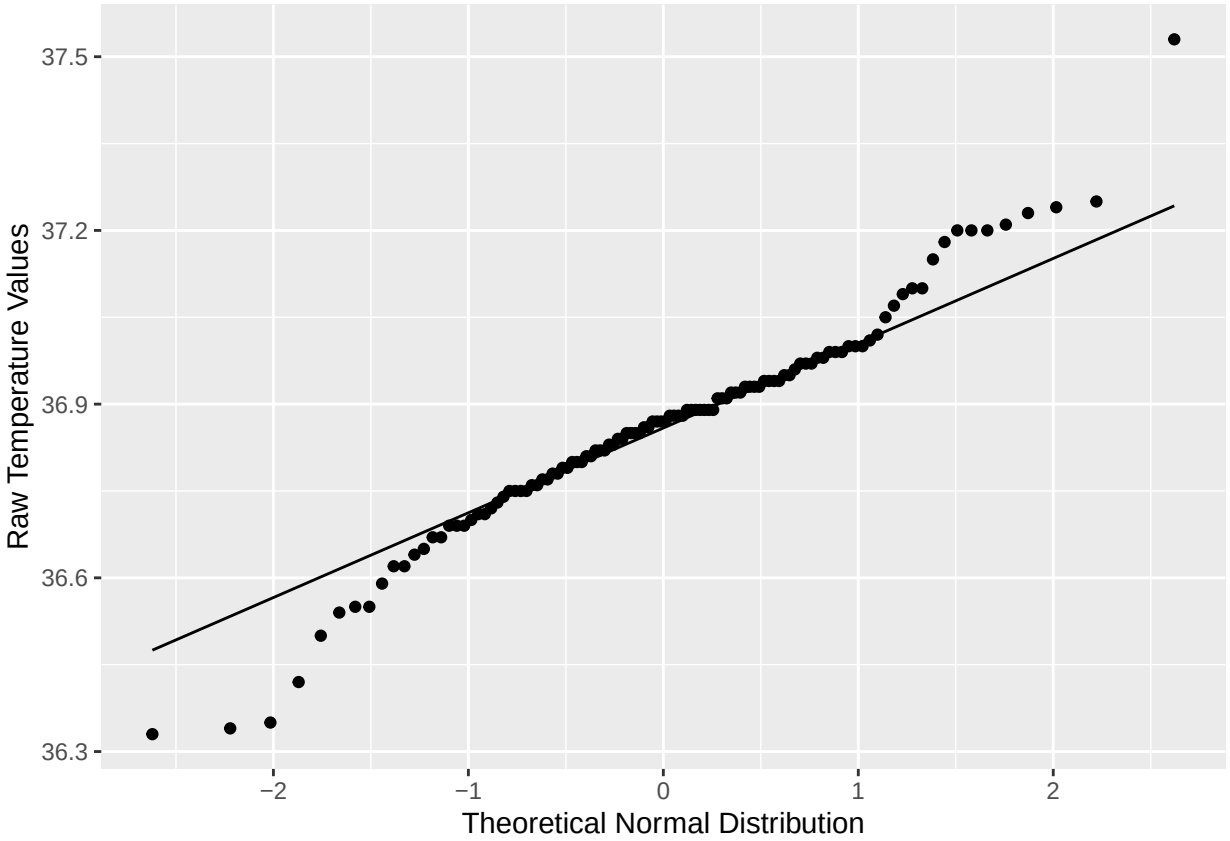
#Assessing Normality

```
#consider information on how this data was collected
#?beaver1()

#visually inspect data
#View(beaver1)

#create a histogram of beaver body temperature
ggplot(beaver1, aes(x = temp)) +
  geom_histogram() +
  labs(x = 'Beaver Body Temperature (C)',
       y = 'Count')
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
#create a Q-Q plot of beaver body temperature
ggplot(beaver1, aes(sample = temp)) +
  geom_qq() +
  geom_qq_line() +
  labs(x = 'Theoretical Normal Distribution',
       y = 'Raw Temperature Values')
```




```
#calculate skewness of beaver body temperature
skewness(beaver1$temp)
```

```
## [1] -0.02782567
```

```
#calculate kurtosis of beaver body temperature
kurtosis(beaver1$temp)
```

```
## [1] 4.351118
```

the absolute value of skewness for beaver body temperature is > 0 and < 0.5 therefore it is approximately symmetrical

kurtosis for beaver body temperature is >3 therefore it is leptokurtic

#Continuous Distributions pnorm() calculates area under the probability density function of a normal distribution to the left of a specified value of x

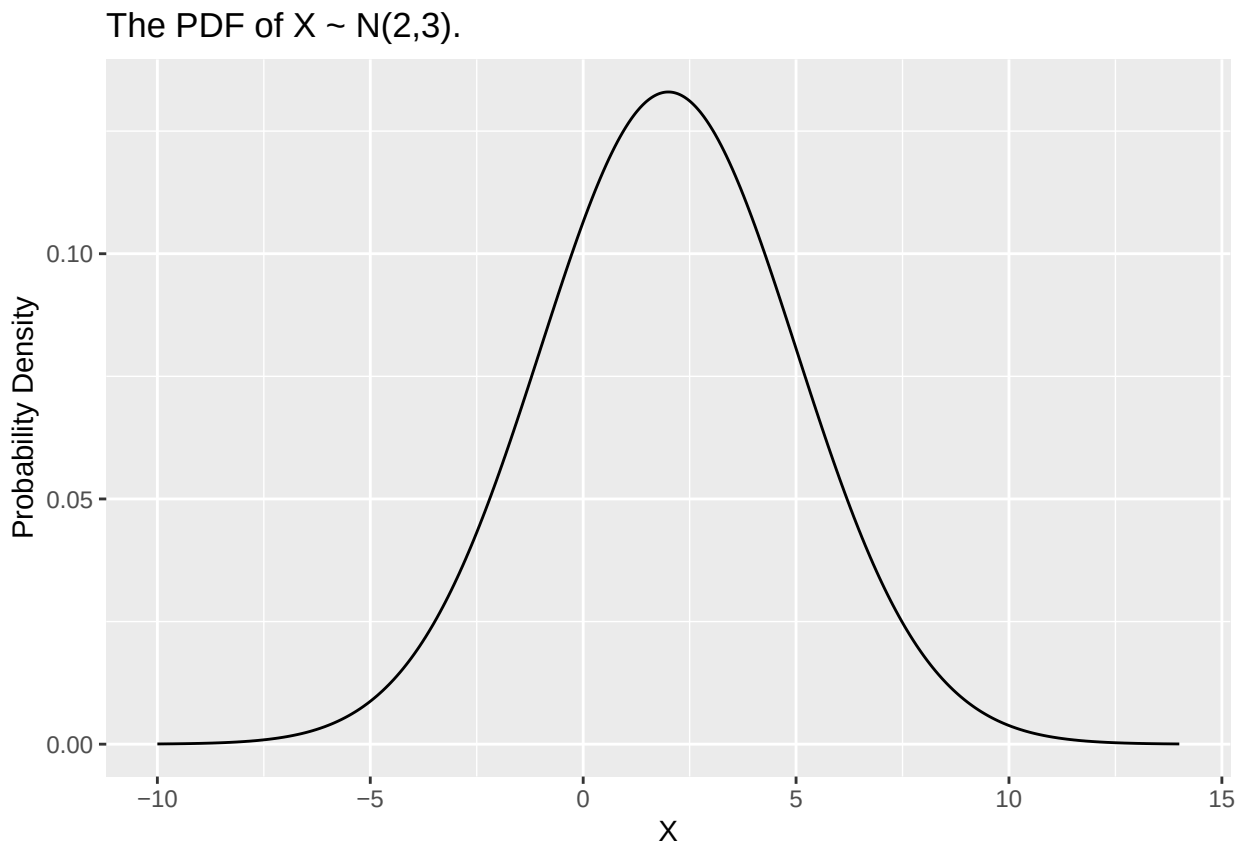
```
#probability of value < 1
pnorm(q = 1, mean = 2, sd = 3)
```

```
## [1] 0.3694413
```

```r
#prob of value < 1
pnorm(q = 1, mean = 2, sd = 3)
```

```
## [1] 0.3694413
```

```r
#create a list of values with lots of values
#in between whole numbers (so our curve looks nice and smooth)
values <- seq(-10, 14, length = 1000)

#probability densities for values -10 to 14
prob_norm <- dnorm(x = values, mean = 2, sd = 3)

#bind two columns together to create a dataframe
norm_df <- cbind(prob_norm, values)

#create a figure
ggplot(norm_df, aes(x = values,
                    y = prob_norm)) +
  geom_line() +
  labs (x= "X",
        y = 'Probability Density',
        title = 'The PDF of X ~ N(2,3).')
```



qnorm() returns the value of x that is the upper bound of a particular probabiltiy

```r
#value of 'x' for a given probability
qnorm(p = 0.80, mean = 2, sd = 3)
```

```
## [1] 4.524864
```