Lab 3

Lily Heidger

2024-02-01

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.5
## v ggplot2   3.4.4     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.0

## -- 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 error
```

```
library(here)
```

```
## here() starts at C:/Users/lily/Documents/210B/210B Lab 3
```

```
library(patchwork)
```

pnorm = find probabilities associated with values in the distribution qnorm = find values associated with probabilities in the distribution

1. Snowfall for a location is found to be normally distributed with mean 96 inches and standard deviation 32 inches.

a) What is the probability that a given year will have more than 120 inches of snow? The probability any given year will have more than 120 inches of snow is 0.227 or ~23%.

```
z_score <- (120-96)/32

probability <- pnorm(z_score)
```

```r
# Print the result
print(1-probability)
```

```
## [1] 0.2266274
```

b) What is the probability that the snowfall will be between 90 and 100 inches? The probability that the snowfall will be between 90 and 100 inches is 0.1241039 or ~12%.

```r
hundred_z_score <- (100-96)/32

hundred_probability <- pnorm(hundred_z_score)

print(hundred_probability)
```

```
## [1] 0.5497382
```

```r
# 0.5497382


ninety_z_score <- (90-96)/32

ninety_probability <- pnorm(ninety_z_score)

print(ninety_probability)
```

```
## [1] 0.4256343
```

```r
# 0.4256343

hundred_probability- ninety_probability
```

```
## [1] 0.1241039
```

c) What level of snowfall will be exceeded only 10% of the time? An amount of snow over 137 inches will only be exceeded 10% of the time.

```r
mean_snowfall <- 96
sd_snowfall <- 32
probability_exceeded <- 0.10

#qnorm calculates quantile based on normal distribution
level_of_snowfall <- qnorm(1 - probability_exceeded, mean_snowfall, sd_snowfall)

level_of_snowfall
```

```
## [1] 137.0097
```

2. Assume that the prices paid for housing within a neighborhood have a normal distribution, with mean \$100,000, and standard deviation \$35,000.

a) What percentage of houses in the neighborhood have prices between \$90,000 and \$130,000? Approximately 42% of houses have prices between \$90,000 and \$130,000.

```r
one_thirty_z <- (130000-100000)/35000

one_thirty_probability <- pnorm(one_thirty_z)
```

```r
print(one_thirty_probability)
```

```
## [1] 0.804317
```

```r
#0.804317

ninety_thou_z <- (90000-100000)/35000

ninety_thou_probability <- pnorm(ninety_thou_z)

print(ninety_thou_probability)
```

```
## [1] 0.3875485
```

```r
#0.3875485

one_thirty_probability- ninety_thou_probability
```

```
## [1] 0.4167685
```

b) What price of housing is such that only 12% of all houses in the neighborhood have lower prices? Only 12% of all houses in the neighborhood cost less than $58,875.46.

```r
mean_price <- 100000
sd_price <- 35000
probability_lower <- 0.12

housing_price <- qnorm(probability_lower, mean_price, sd_price)

housing_price
```

```
## [1] 58875.46
```

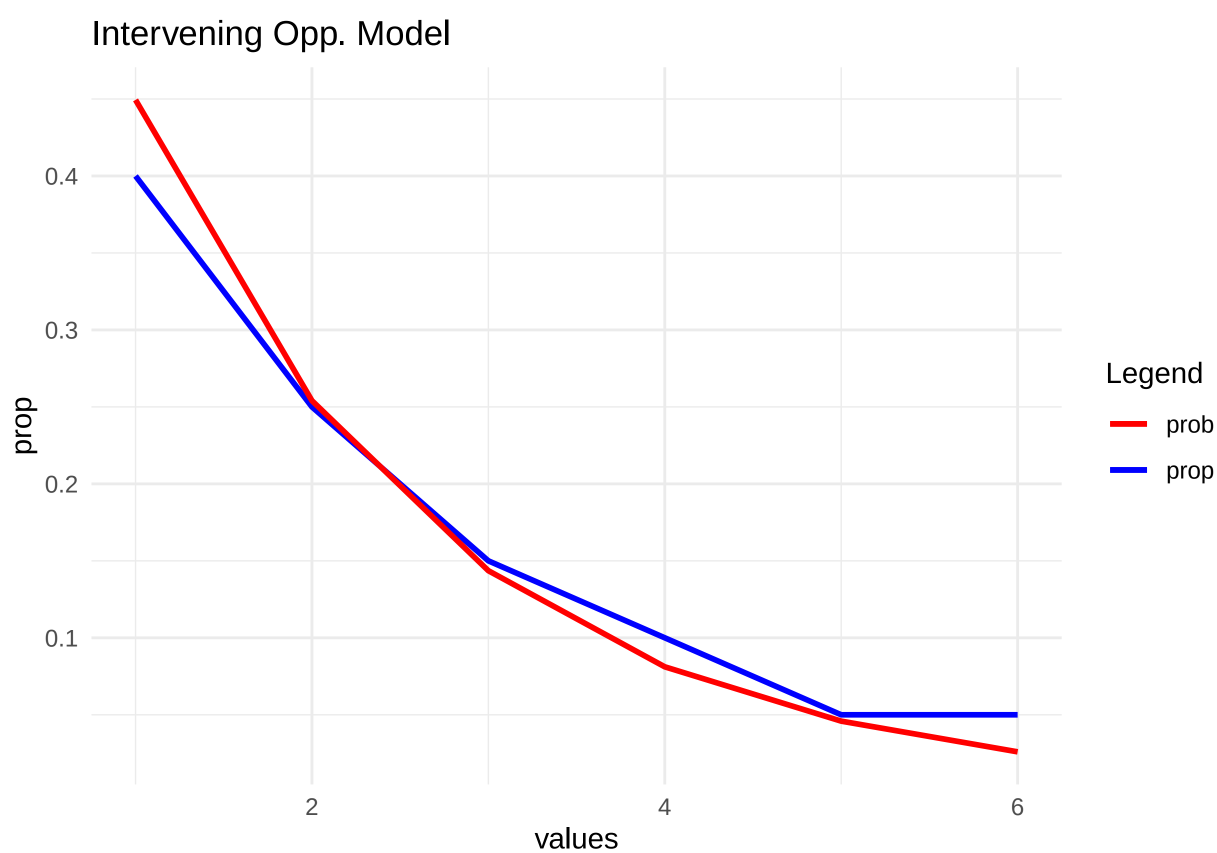
```r

mean <- 2.2
L <- 1/2.2

df1<- data.frame(values = 1:6,
prop = c(0.4, 0.25, 0.15, 0.1, 0.05, 0.05),
prob = NA)

prob_function <- function(L, n) {
numerator <- (1 - L)^(1:n - 1) * L
denominator <- sum((1 - L)^(1:n - 1) * L)
result <- numerator / denominator
return(result)
}

L <- 1 / sum(df1$values * df1$prop)
n <- 6
result <- prob_function(L, n)
df1<- df1 %>%
mutate(prob = (result))

df1
```

3. Residents in a community have a choice of six different grocery stores. The proportions of residents observed to patronize each are p(1) = 0.4, p(2) = 0.25, p(3) = 0.15, p(4) = 0.1, p(5) = 0.05, and p(6) = 0.05, where the stores are arranged in terms of increasing distance from the residential community. Fit an intervening opportunities model to these data by estimating the parameter L.

```
##   values prop       prob
## 1      1 0.40 0.44943680
## 2      2 0.25 0.25402949
## 3      3 0.15 0.14358189
## 4      4 0.10 0.08115498
## 5      5 0.05 0.04587021
## 6      6 0.05 0.02592664
```

```r
ggplot(df1, aes(x = values)) +
  geom_line(aes(y = prop, color = "prop"), size = 1) +
  geom_line(aes(y = prob, color = "prob"), size = 1) +
  labs(title = "Intervening Opp. Model",
       color = "Legend") +
  scale_color_manual(values = c("prop" = "blue", "prob" = "red")) +
  theme_minimal()
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



4. The annual probability that suburban residents move to the central city is 0.08, while the annual probability that the city residents move to the suburbs is 0.11. Starting with respective populations of 30,000 and 20,000 in the central city and suburbs, respectively, forecast the population redistribution that will occur over the next three years. Use the Markov model assumption that the probabilities of movement will remain constant. Also find the long-run, equilibrium populations. The population forecasts for the next three years are as follows: Year 1: City: 28300 Suburbs: 21700

Year 2: City: 21577 Suburbs: 28423

Year 3: City: 21477 Suburbs: 28523

City Equilibrium = 21053 Suburbs Equilibrium = 28945

```
total_pop <- 50000
Pcity1 <- (0.89*30000)+(0.08*20000)
Psub1 <- (0.92*20000)+(0.11*30000)

50000-Psub1
```

```
## [1] 28300
```

```
#year 2
Psub2 <-  (0.92*28300)+(0.11*21700)

50000- Psub2
```

```
## [1] 21577
```

```
#year 3
Psub3 <- (0.92 *28423) + (0.11*21577)
Psub3
```

```
## [1] 28522.63
```

```
50000-Psub3
```

```
## [1] 21477.37
```

```
#PcityEquil <- (0.89*x) + (0.08 * y)
#PcityEquil

# Pcity = 0.89(Pcity) +0.08(50,000 - Pcity)

PcityEquil <- 21052.63
50000- PcityEquil
```

```
## [1] 28947.37
```

5. The magnitude (Richter scale) of earthquakes along a Californian fault is exponentially distributed, with l = (1/2.35).

What is the probability of an earthquake exceeding magnitude 6.3? The probability of an earthquake exceeding magnitude 6.3 is 0.069.

```
1- (1- exp(-6.3 / 2.35))
```

```
## [1] 0.06850483
```

What is the probability of an earthquake during the year that exceeds magnitude 7.7? The probability that an earthquake during the year exceeds magnitude 7.7 is 0.038.

```
1- (1 - exp(-7.7 / 2.35))
```

```
## [1] 0.03775657
```

6. A variable, X, is uniformly distributed between 10 and 24.

(a) What is P(16 x 20)? The probability that x is greater than or equal to 16 and less than or equal to 20 is 0.286.

```
(20 - 16)/(24 - 10)
```

```
## [1] 0.2857143
```

(b) What is the mean and variance of X? The mean is 17, and the variance is 16.33.

```
mean <- (10 + 24)/2
var <- ((24-10)^2)/12

mean
```

```
## [1] 17
```

```
var
```

```
## [1] 16.33333
```

```

mean_duration <- 4.76
```

7. The duration of residences in households is found to be exponentially distributed with mean 4.76 years.

What is the probability that a family is in their house for more than 8 years? The probability that a family is in their house for more than 8 years is 0.18.

```
years <- 8

prob_more_than_8 <- 1 - pexp(years, rate = 1 / mean_duration)

prob_more_than_8
```

```
## [1] 0.1862487
```

Between 5 and 8 years? The probability that a family is in their house for more between 5 and 8 years is 0.162.

```
lower_bound <- 5
upper_bound <- 8

prob_lower <- pexp(lower_bound, rate = 1/ mean_duration)
prob_upper <- pexp(upper_bound, rate = 1/mean_duration)

prob_upper - prob_lower
```

```
## [1] 0.163542
```

```r
mean_imports <- 30
sd_imports <- 16
```

8. The mean value of annual imports for a country is normally distributed with mean $30 million and standard deviation $16 million.

What dollar value of imports is exceeded only 5% of the time? A dollar value of $56.32 Million is only exceeded 5% of the time.

```r
prob_5_percent <- 0.05

imports <- qnorm(1 - prob_5_percent, mean_imports, sd_imports)

imports
```

```
## [1] 56.31766
```

What fraction of years have import values between 29 and 45 million? 0.351 of years have import values between 29 and 45 million.

```r
lower_bound <- 29
upper_bound <- 45

probability_lower <- pnorm(lower_bound, mean_imports, sd_imports)
probability_upper <- pnorm(upper_bound, mean_imports, sd_imports)

fraction <- probability_upper - probability_lower

fraction
```

```
## [1] 0.350667
```

```r
mean_customers <- 250
sd_customers <- 110
```

9. The number of customers at a bank each day is found to be normally distributed with mean 250 and standard deviation of 110.

What fraction of days will have less than 100 customers? 0.0863 or ~8.6% of days will have less than 100 customers.

```r
less_than_100 <- 100

probability_less_than_100 <- pnorm(less_than_100, mean_customers, sd_customers)


probability_less_than_100
```

```
## [1] 0.08634102
```

More than 320? 0.2622 or ~26% of days will have more than 320 customers.

```r
more_than_320 <- 320
```

```r
probability_more_than_320 <- pnorm(more_than_320, mean_customers, sd_customers, lower.tail = FALSE)

probability_more_than_320
```

```
## [1] 0.2622697
```

What number of customers will be exceeded 10% of the time? 10% of the time, the number of customers will exceed 109.

```r
prob_10_percent <- 0.10

customers_exceeded_10_percent <- qnorm(prob_10_percent, mean_customers, sd_customers)

customers_exceeded_10_percent
```

```
## [1] 109.0293
```

10. Incomes are exponentially distributed with a mean of $10,000. What fraction of the population has income:

a) Less than or equal to $8000? About 0.550671 or 55% of the population has income less than or equal to $8000.

```r
mean_income <- 10000
rate <- 1/mean_income

desired_income <- 8000

#cdf function for exp model
probability_less_than_8000 <- pexp(desired_income, rate)

probability_less_than_8000
```

```
## [1] 0.550671
```

b) Greater than $12,000? About 0.3011 or 30% of the population has income greater than $12,000.

```r
income_greater_than_12000 <- 12000

# cdf for exp but only upper tail
probability_greater_than_12000 <- pexp(income_greater_than_12000, rate, lower.tail = FALSE)

probability_greater_than_12000
```

```
## [1] 0.3011942
```

c) Between $9,000 and $12,000? About 0.1053754 or 11% of the population has income between $9,000 and $12,000.

```r
lower_bound <- 9000
upper_bound <- 12000

probability_lower <- pexp(lower_bound, rate)
probability_upper <- pexp(upper_bound, rate)
```

```r
probability_between_9000_and_12000 <- probability_upper - probability_lower

probability_between_9000_and_12000
```

```
## [1] 0.1053754
```