

Module 4 Notes

Jeremiah Lowhorn

2023-09-11

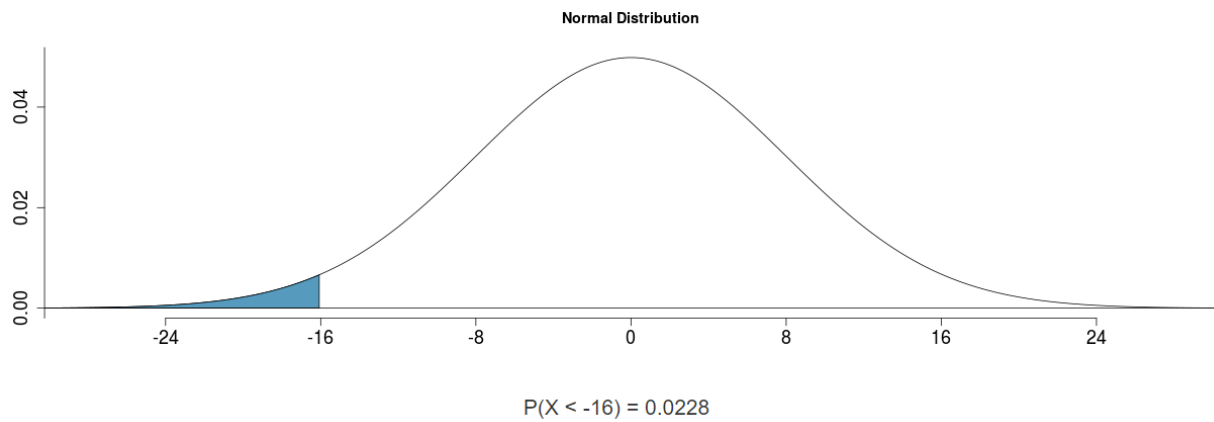


Figure 1: Caption for the picture.

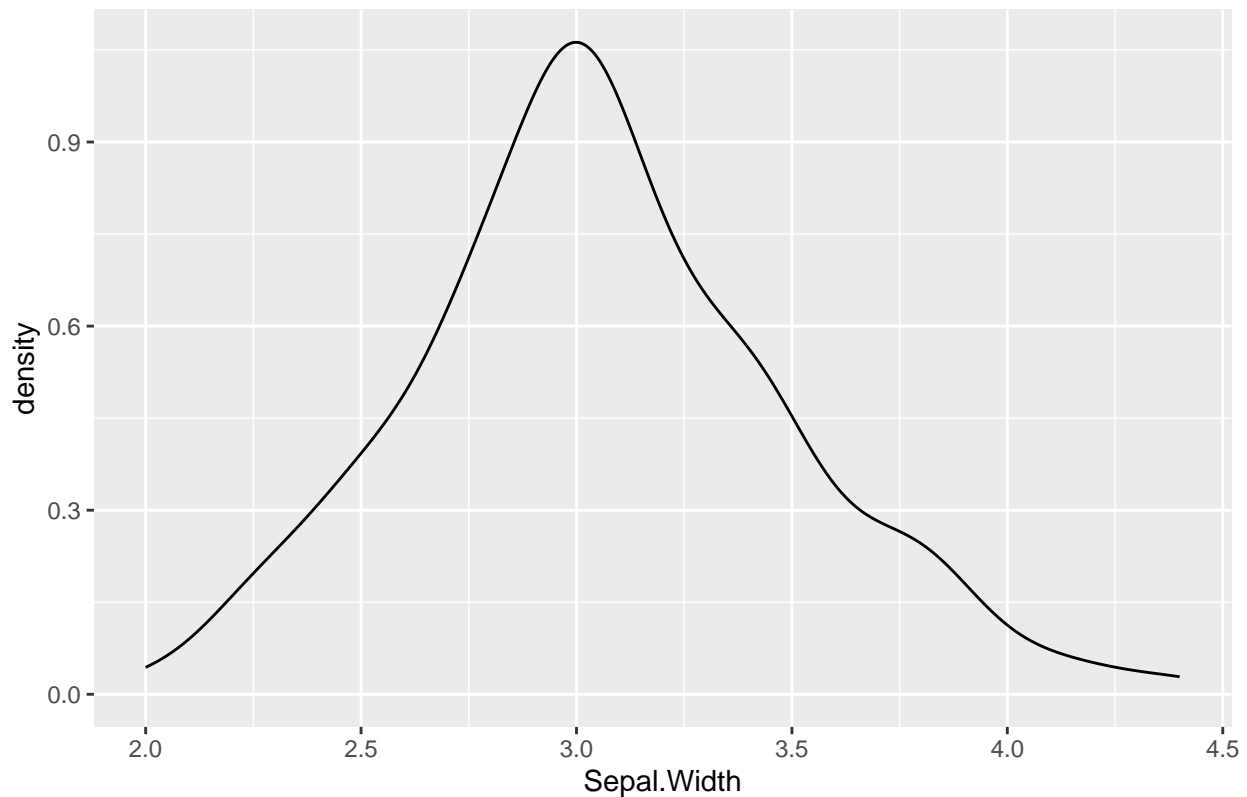
Chapter 4:

Normal Distributions

```
set.seed(1233)

data <- iris
ggplot(iris) +
  geom_density(aes(Sepal.Width)) +
  labs(title='KDE of Sepal Width')
```

KDE of Sepal Width



```
mean_s_width = mean(data$Sepal.Width)
sd_s_width = sd(data$Sepal.Width)
```

How many standard deviations above the mean is a sepal width of 4?

```
x = 4
(x - mean_s_width)/sd_s_width
```

```
## [1] 2.162743
```

What about a sepal width of 2?

```
x = 2
(x - mean_s_width)/sd_s_width
```

```
## [1] -2.42582
```

These are called Z-Scores and are very important as they can be used to calculate percentiles, that is what is the percentage of the population that falls above or below a specific number?

- Anything more than ± 2 standard deviations from the mean are considered unusual (anomalous)

What is the percentage of observations that fall below a sepal width of 4?

```
pnorm(4,mean=mean_s_width,sd=sd_s_width)
```

```
## [1] 0.9847195
```

What about 2?

```
pnorm(2,mean=mean_s_width,sd=sd_s_width)
```

```
## [1] 0.007636913
```

What is the percentage of population that falls within 2-4?

```
pnorm(4,mean=mean_s_width,sd=sd_s_width) - pnorm(2,mean=mean_s_width,sd=sd_s_width)
```

```
## [1] 0.9770826
```

What about above 2?

```
1 - pnorm(2,mean=mean_s_width,sd=sd_s_width)
```

```
## [1] 0.9923631
```

Geometric Distributions

Bernoulli Random Variables

Geometric distribution describes the waiting time until a success for independent and identically distributed (iid) Bernoulli random variables. * independence: outcomes of trials don't affect each other * identical: the probability of success is the same for each trial

```
bern_trial <- function(p,n_trials){  
  prob_not = 1 - p  
  prob_not^(n_trials) * p  
}
```

```
bern_trial(1/6,5)
```

```
## [1] 0.0669796
```

Dr Smith's Experiment Dr. Smith wants to repeat Milgram's experiments but she only wants to sample people until she finds someone who will not inflict a severe shock. What is the probability that she stops after the first person?

```
p_true = 0.35  
p_not = 1 - p_true  
  
#prob after the first person?  
bern_trial(p_true,2)
```

```
## [1] 0.147875
```

Expected value or the mean of a geometric distribution is simply $\frac{1}{p}$. The standard deviation is $\sqrt{\frac{(1-p)}{p^2}}$ How many people does Dr. Smith have to test before finding the first one that refuses to administer the shock?

```
1/p_true
```

```
## [1] 2.857143
```

```
s_dev_bern <- function(p){  
  sqrt( ((1-p)/(p^2)) )  
}
```

```
s_dev_bern(p_true)
```

```
## [1] 2.303502
```

Dr. Smith expects to test 2.86 people give or take 2.3...

Binomial Distribution

The Binomial distribution describes the probability of having exactly k successes in n independent Bernoulli trials with probability of success p .

In R you can use the `choose()` function to calculate the number of ways to choose k successes in n trials.

```
choose(4,1)
```

```
## [1] 4
```

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

```
dbinom(x=8,size=10,prob=0.262)
```

```
## [1] 0.0005441712
```

What is the probability that 2 randomly chosen people share a birthday?

```
dbinom(x=2,size=121,prob=1/365)
```

```
## [1] 0.03931558
```

Negative Binomial Distribution

A college student working at a psychology lab is asked to recruit 10 couples to participate in a study. She decides to stand outside the student center and ask every 5th person leaving the building whether they are in a relationship and, if so, whether they would like to participate in the study with their significant other. Suppose the probability of finding such a person is 10%. What is the probability that she will need to ask 30 people before she hits her goal?

Given: $p = 0.10$, $k = 10$, $n = 30$. We are asked to find the probability of 10th success on the 30th trial, therefore we use the negative binomial distribution.

How is the negative binomial distribution different from the binomial distribution? * In the binomial case, we typically have a fixed number of trials and instead consider the number of successes. * In the negative binomial case, we examine how many trials it takes to observe a fixed number of successes and require that the last observation be a success.

```
dnbinom(x=10,size=30,prob=0.1)
```

```
## [1] 2.216707e-22
```

Poisson

- The Poisson distribution is often useful for estimating the number of rare events in a large population over a short unit of time for a fixed population if the individuals within the population are independent.
- The rate for a Poisson distribution is the average number of occurrences in a mostly-fixed population per unit of time, and is typically denoted by λ .
- Using the rate, we can describe the probability of observing exactly k rare events in a single unit of time.

Suppose that in a rural region of a developing country electricity power failures occur following a Poisson distribution with an average of 2 failures every week. Calculate the probability that in a given week the electricity fails only once.

```
r_pois <- function(lambda,n){  
  (lambda^n* exp(-lambda))/factorial(n)  
}
```

```
r_pois(2,1)
```

```
## [1] 0.2706706
```