Lecture 09 - Normal Demo

Wesley Burr

05/10/2018

Making Plots for Normal Distributions

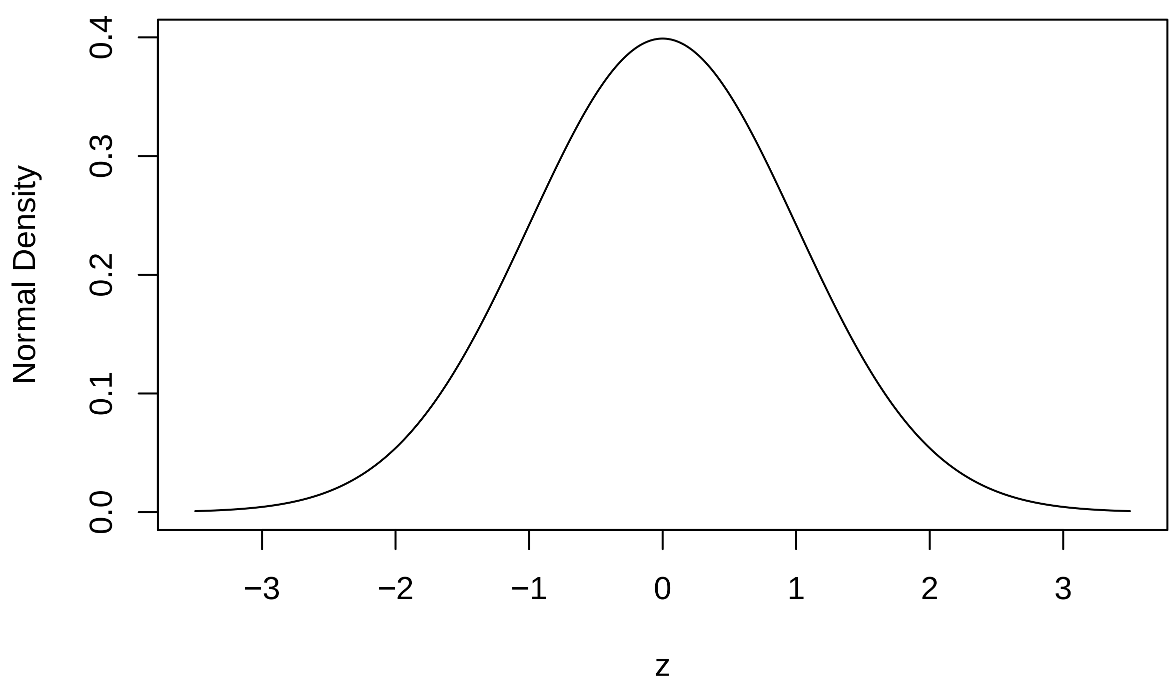
Making pictures for the normal problems can help a lot when problem solving. They're also examples of **line graphs** (scatterplots with only one observation per x value), which a few people have been asking about examples for. So let's make one.

Recall from the recent lectures that we have continuous functions which describe probabilities for **continuous random variables**. When working with the Binomial, we used **dbinom()** and **pbinom()**. We're going to do that now, but for the **normal** distribution. The R developers were really careful, and kept most of the syntax the same. The two functions are:

- **pnorm()**: corresponds to $P[X \leq q]$
- **dnorm()**: the density function - like the $P[X = x]$ from the binomial, but continuous

Let's start by creating a simple plot of the normal density. To make any plot, we need an x and a y. All simple line plots are y versus x. What should the x be for this plot . . . ?

```
x <- seq(from = -3.5, to = 3.5, by = 0.01)
plot(x = x, y = dnorm(x), type = "l",
     xlab = "z", ylab = "Normal Density")
```

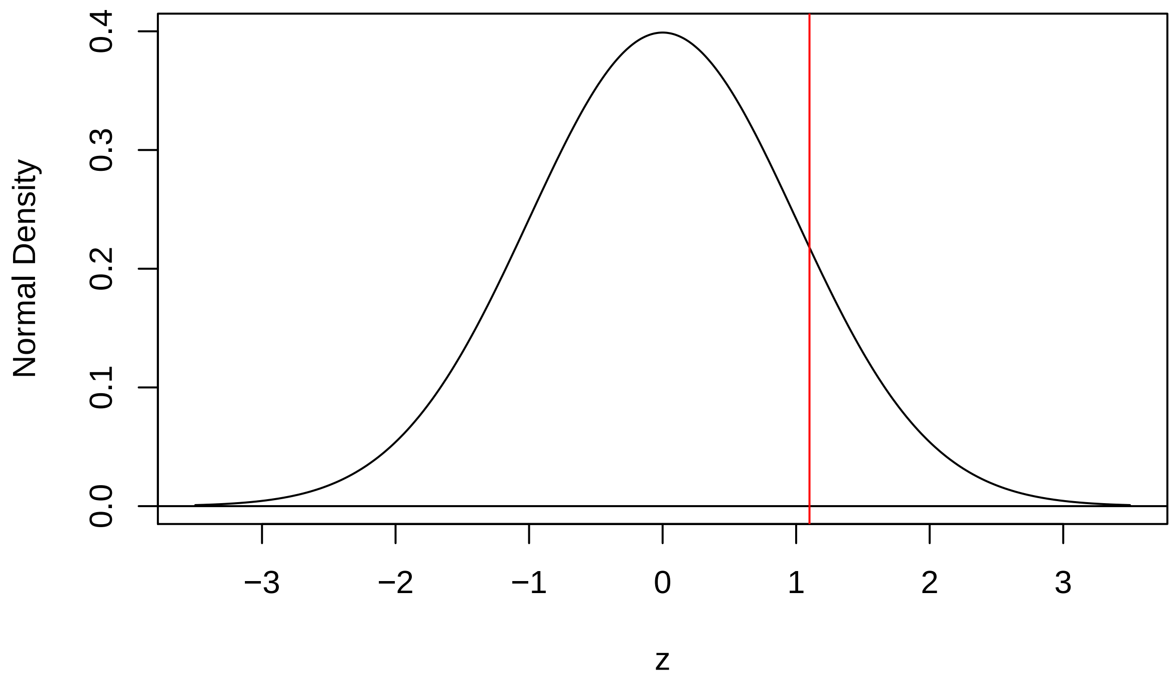


This is the famous “bell curve”: the **standard normal distribution**. For a lot of problems we’ll solve with these normal distributions, we want to consider “greater than” or “less than”, so it’s handy to put a vertical line on the curve to indicate our cutoff. Let’s say we had a question that asked:

What is P[Z $\leq$ 1.1]?

The graph that would go along with this would be:

```
plot(x = x, y = dnorm(x), type = "l",
     xlab = "z", ylab = "Normal Density")
abline(h = 0)
abline(v = 1.1, col = "red")
```



The **abline()** command puts **h** (horizontal) or **v** (vertical) lines onto plots: that’s it.

Solving Problems using dnorm() and pnorm()

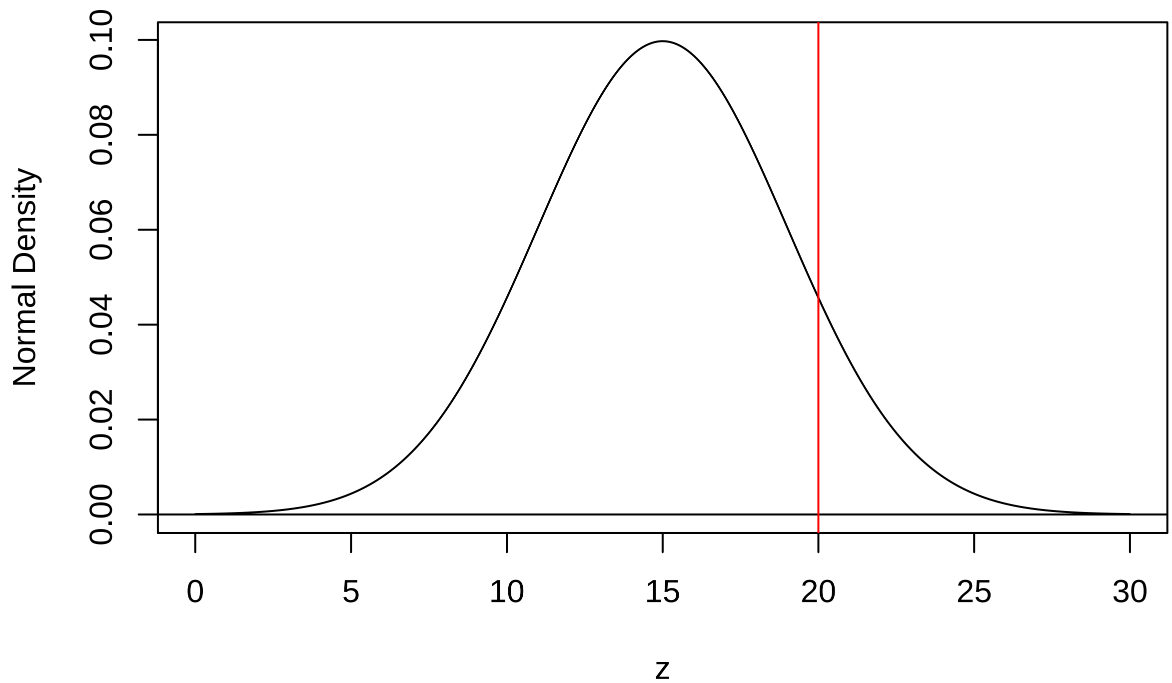
Let’s solve some problems involving the standard normal. To do this, **dnorm()** is basically useless: it just lets us draw pretty pictures. So we can use it for a picture, then ignore it.

Example 1

If the length of time you spend in a dentist’s chair is normally distributed with mean $\mu = 15$ minutes and standard deviation $\sigma = 4$ minutes, what is the probability your time in the chair exceeds 20 minutes?

Solution: this is asking P[Z $\geq$ 20] for a Z which has $\mu = 15$, $\sigma = 4$. We can solve this using the **pnorm()** function. First, let’s draw a picture for intuition.

```r
x <- seq(from = 0, to = 30, by = 0.01)
plot(x = x, y = dnorm(x, mean = 15, sd = 4), type = "l",
     xlab = "z", ylab = "Normal Density")
abline(h = 0)
abline(v = 20, col = "red")
```



How big do we think this is? Looks like no more than 1/5, probably less than 15%, so let's guess 0.15.

Now, solve it using **pnorm()**:

```r
pnorm(q = 20, mean = 15, sd = 4, lower.tail = FALSE)
```

```
## [1] 0.1056498
```

10.56% is our answer, so our "guess" of 15% isn't that bad! So the probability of you sitting in your dentist's chair for 20 or more minutes is 0.106.

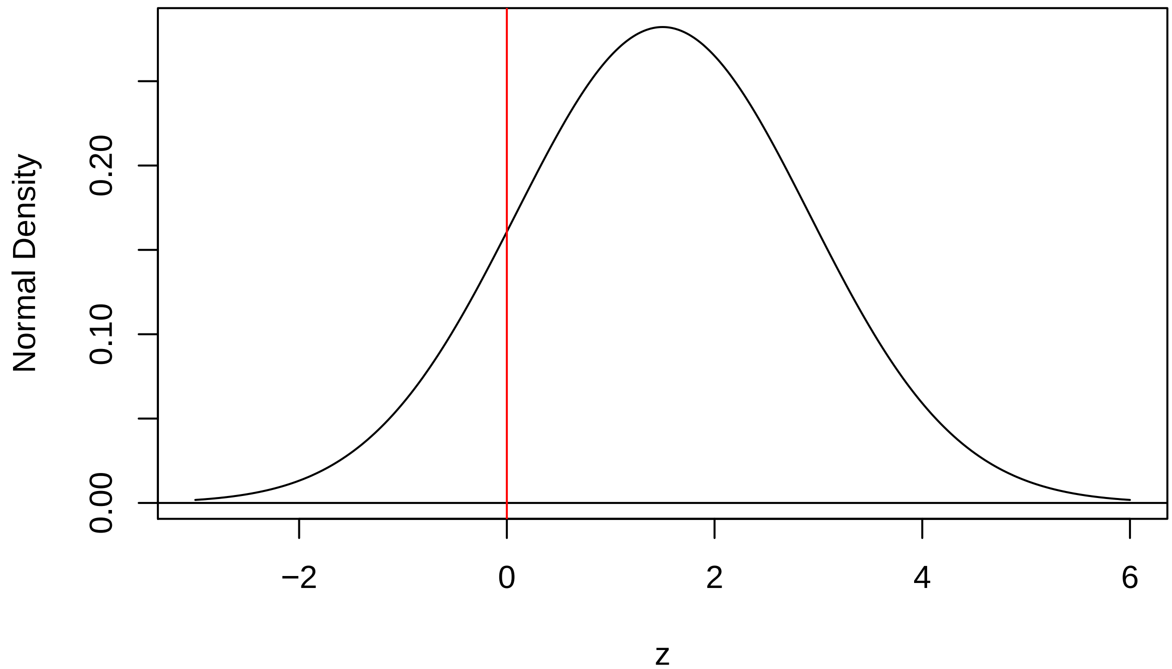
Example 2

If the amount of money won in a game is thought to be a normal random variable, with expectation 1.5 and variance 2, what is the probability that if you played the game, you would lose money?

Solution: this is asking P[Z < 0] for a Z with $\mu = 1.5$ and $\sigma = \sqrt{2}$. Let's draw a picture again:

```r
z <- seq(from = -3.0, to = 6.0, by = 0.01)
plot(x = z, y = dnorm(z, mean = 1.5, sd = sqrt(2)), type = "l",
     xlab = "z", ylab = "Normal Density")
```

```r
abline(h = 0)
abline(v = 0, col = "red")
```



This looks a little bigger than last time, so let's guess 0.15 again.

```r
pnorm(q = 0, mean = 1.5, sd = sqrt(2), lower.tail = TRUE)
```

```
## [1] 0.1444222
```

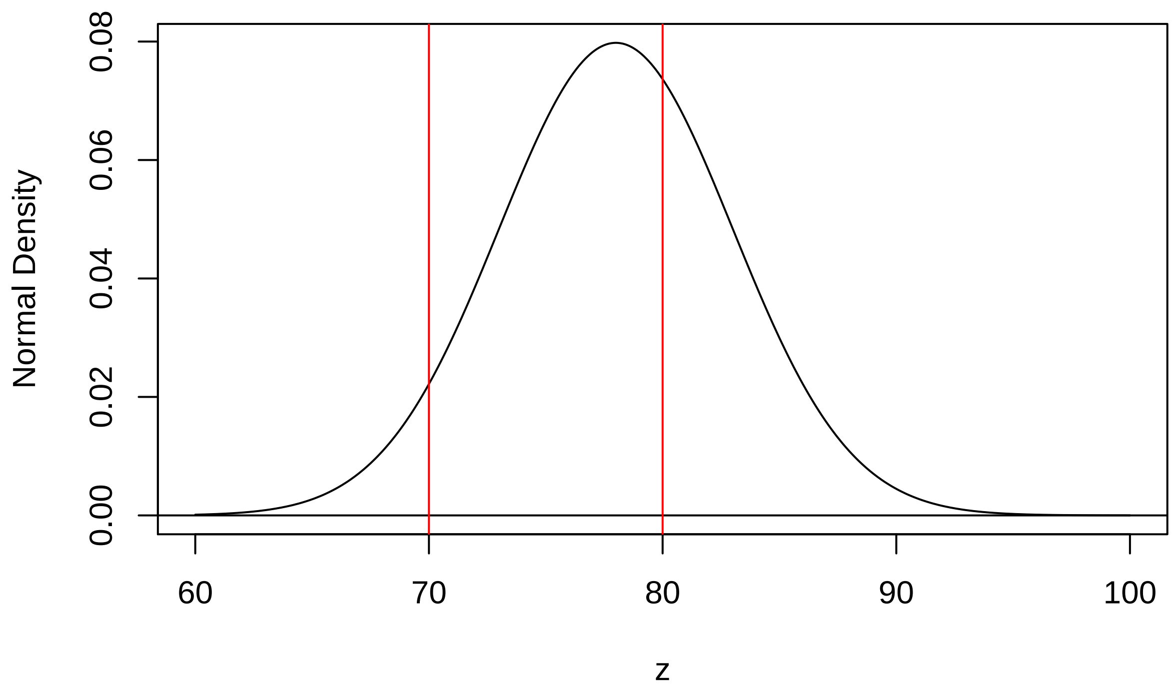
So in this case, the probability of you losing money by playing the game is 0.144.

Example 3

Suppose that student grades in MATH 1051 are normally distributed with mean 78% and standard deviation 5%. What is the probability that a randomly selected student from the class has a final grade of a B (that is, between a 70% on the bottom and an 80% on the top, but not including 80%)?

Solution: this is asking P[$70 \leq$ Z ≤ 79] for a Z with $\mu = 78$ and $\sigma = 5$. But there's a subtle point here: is 79% the cap? Or 80%? What about a 79.3%? So actually, what we want is P[$70 \leq$ Z < 80]. Let's draw a picture:

```r
z <- seq(from = 60.0, to = 100.0, by = 0.1)
plot(x = z, y = dnorm(z, mean = 78, sd = 5), type = "l",
     xlab = "z", ylab = "Normal Density")
abline(h = 0)
abline(v = c(70, 80), col = "red")
```



This area looks much bigger than the last two, maybe half of the total? Let's guess 0.5. Now, how do we compute it? All we have is a single function that can go from **one** line to the end: either from a line up to the very top, or from a line down to the very bottom. How do we go **between** two lines?

```
pnorm(q = 80, mean = 78, sd = 5, lower.tail = TRUE)
```

```
## [1] 0.6554217
```

```
pnorm(q = 70, mean = 78, sd = 5, lower.tail = TRUE)
```

```
## [1] 0.05479929
```

So by computing both, we see that the area from 80 down to 0 is 0.655 probability, and the area from 70 down to 0 is 0.055. So what's the area in between? **Their difference!**

```
pnorm(q = 80, mean = 78, sd = 5, lower.tail = TRUE) -
      pnorm(q = 70, mean = 78, sd = 5, lower.tail = TRUE)
```

```
## [1] 0.6006224
```

So the probability is 0.6006, so the chance of a random student getting a B in 1051H, if the grades are actually distributed normally, is 60%.