Workshop 06: More Plots & Normal

Wesley Burr

11/10/2018

Making Plots for Normal Distributions

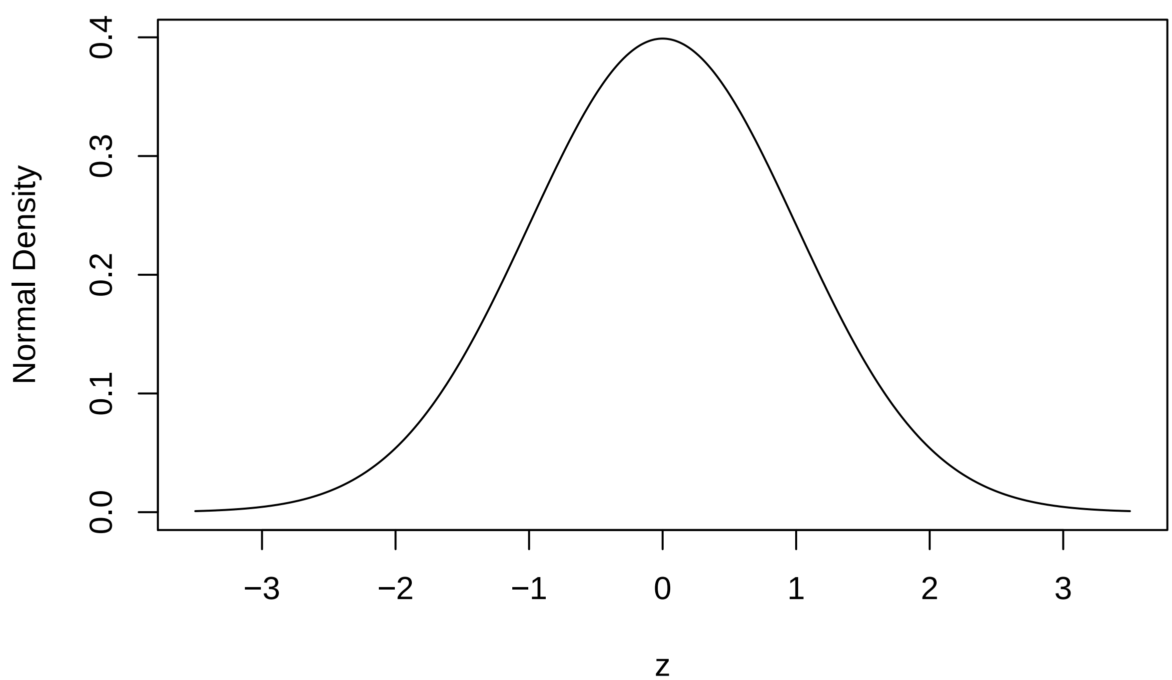
Making pictures for the normal problems can help a lot when problem solving. They're also examples of **line graphs** (scatterplots with only one observation per x value), which a few people have been asking about examples for. So let's make one.

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

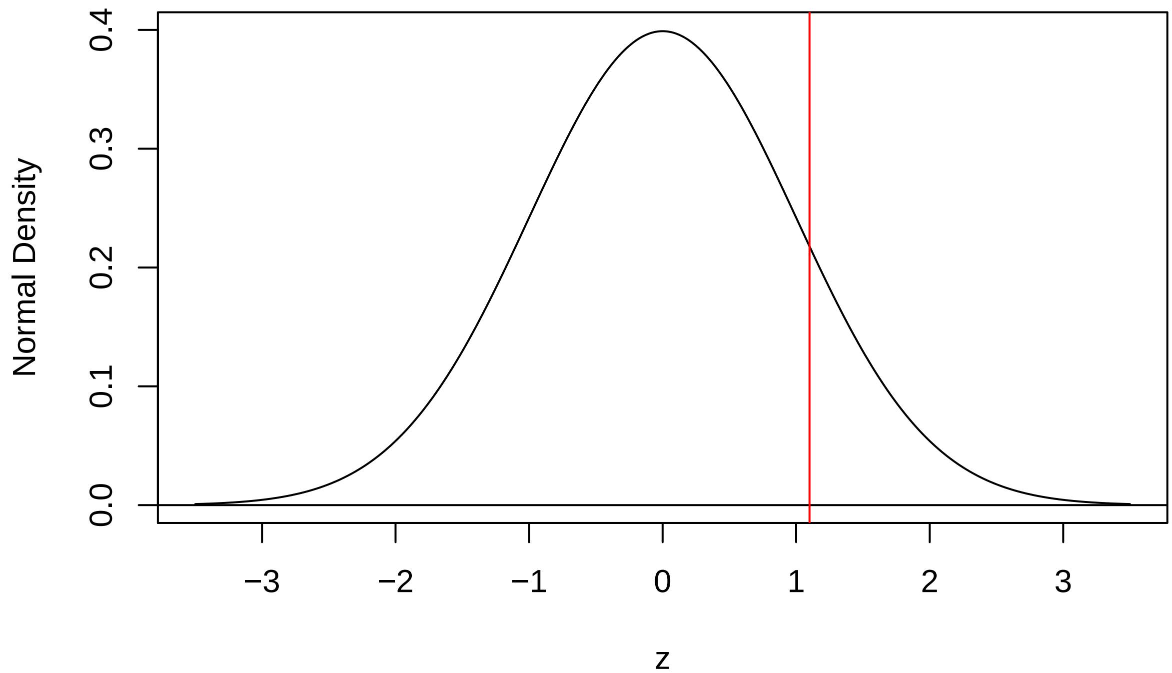
Let's start by creating a simple plot of the normal density. To make any plot, we need an x and a y. All simple line plots are y versus x. What should the x be for this plot . . . ?

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



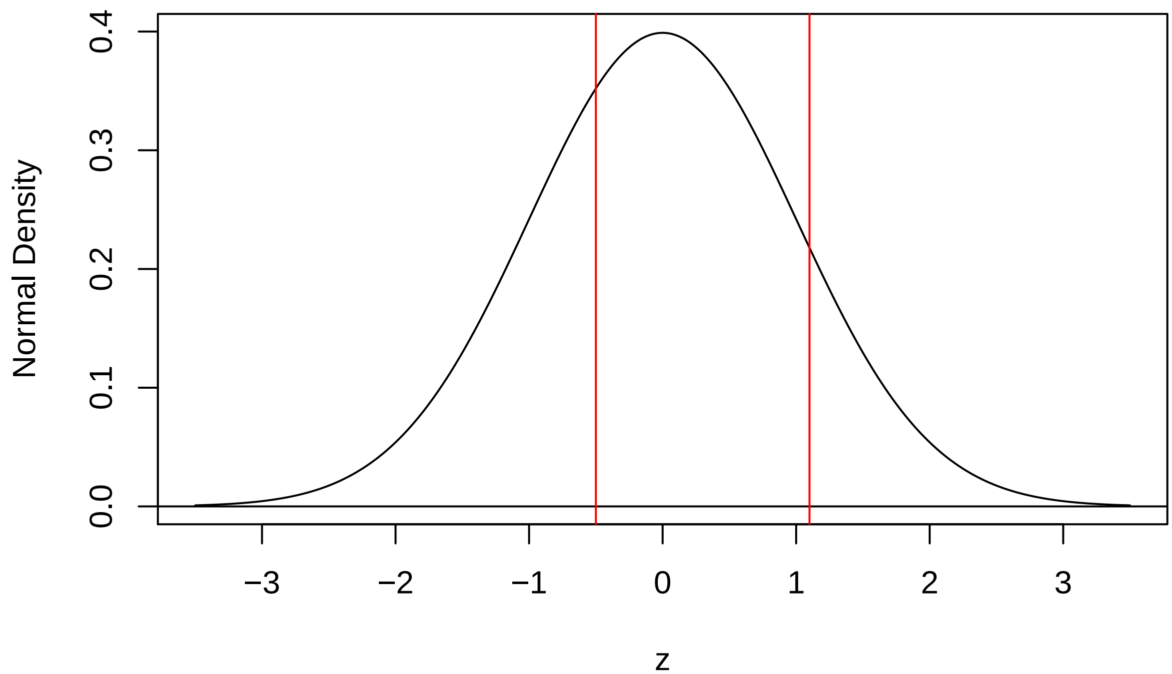
We can add vertical lines to this:

```
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. We can add more than one line at once by using a vector:

```
plot(x = x, y = dnorm(x), type = "l",
     xlab = "z", ylab = "Normal Density")
abline(h = 0)
abline(v = c(-0.5, 1.1), col = "red")
```



Solving Problems using dnorm() and pnorm()

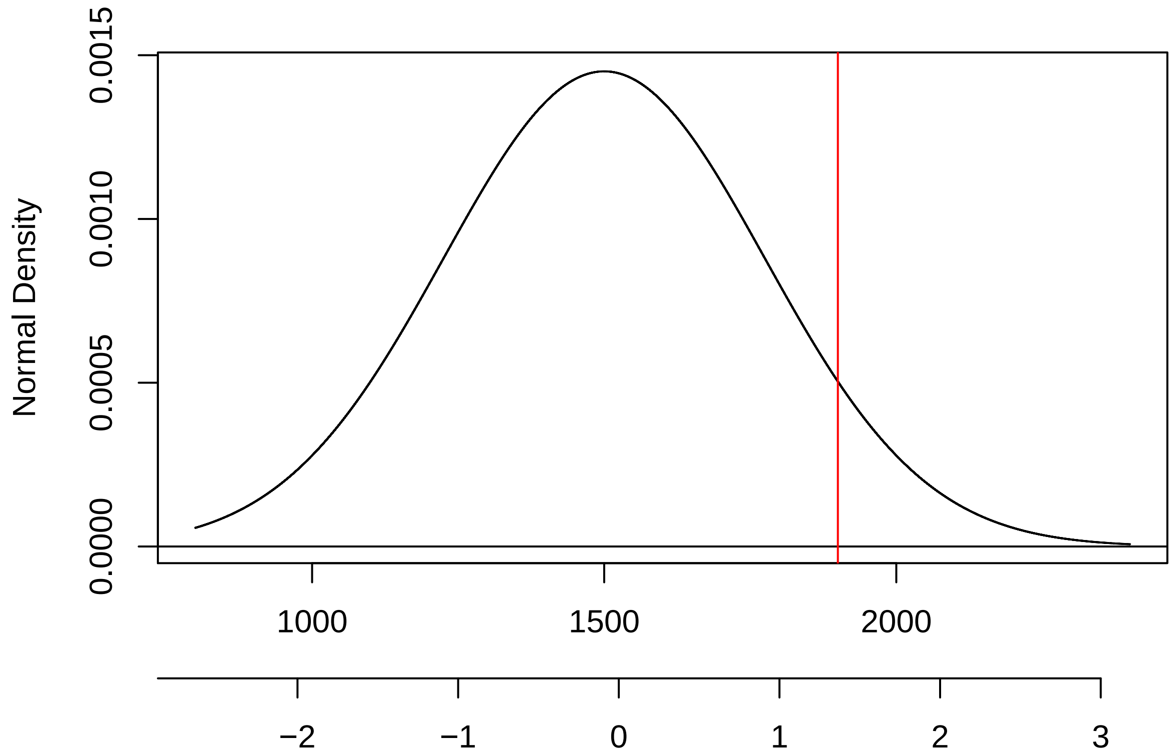
Let's solve some problems involving the standard normal. To do this, **dnorm()** is useful for only the "draw a picture" part of the problem. We have to use **pnorm()**.

Example 1

If the grades on the SAT are normally distributed, with mean 1500 and standard deviation 275, what is the probability that a randomly selected student will get a score above 1900?

Solution: this is asking P[X $\geq$ 1900] for an X which has $\mu = 1500$ and $\sigma = 275$. We can solve this using the **pnorm()** function. First, let's draw a picture for intuition.

```
x <- seq(from = 800, to = 2400, by = 0.01)
plot(x = x, y = dnorm(x, mean = 1500, sd = 275), type = "l",
     xlab = "", ylab = "Normal Density")
axis(side = 1, line = 3, at = seq(700, 2400, 275),
     labels = round((seq(700, 2400, 275) - 1500) / 275), 4)
abline(h = 0)
abline(v = 1900, col = "red")
```



How big do we think this is?

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

```
pnorm(q = 1900, mean = 1500, sd = 275, lower.tail = FALSE)
```

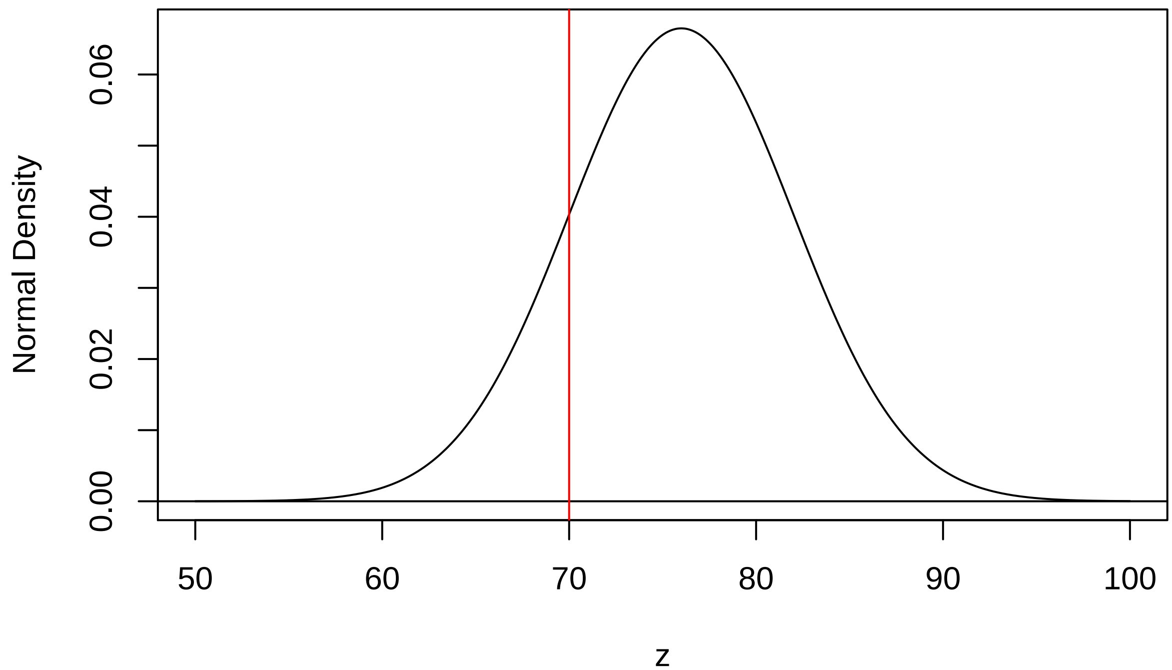
```
## [1] 0.07289757
```

Example 2

Suppose that student grades in MATH 1051 are normally distributed with mean 76% and standard deviation 6%. What is the probability that a randomly selected student from the class has a final grade lower than a B (that is, below 70%)?

Solution: this is asking P[X < 70] for an X with $\mu = 76$ and $\sigma = 6$. Start with a picture:

```
z <- seq(from = 50.0, to = 100.0, by = 0.1)
plot(x = z, y = dnorm(z, mean = 76, sd = 6), type = "l",
     xlab = "z", ylab = "Normal Density")
abline(h = 0)
abline(v = c(70), col = "red")
```



This area looks bigger than the last one. Let's guess 0.25. Now, how do we compute it?

```
pnorm(q = 70, mean = 76, sd = 6, lower.tail = TRUE)
```

```
## [1] 0.1586553
```

So the probability is 0.1586, so the chance of a random student getting less than a B in 1051H, if the grades are actually distributed normally, is 15.9%.

Challenge Problems

1. A radar unit is used to measure speeds of cars on a motorway. The speeds are normally distributed with a mean of 90 km/hr and a standard deviation of 10 km/hr. What is the probability that a car picked at random is travelling at more than 100 km/hr?
2. The length of similar components produced by a company are approximated by a normal distribution model with a mean of 5 cm and a standard deviation of 0.02 cm. If a component is chosen at random
    - What is the probability that the length of this component is between 4.98 and 5.02 cm?
    - What is the probability that the length of this component is between 4.96 and 5.04 cm?
3. The time taken to assemble a car in a certain plant is a random variable having a normal distribution of 20 hours and a standard deviation of 2 hours. What is the probability that a car can be assembled at this plant in a period of time
    - Less than 19.5 hours?
    - Between 20 and 22 hours?

Inverse Probability Problems

The problems we just looked at are **forward** probability problems: we're given a situation, and asked to find the probability. What happens if we have a **backward** probability problem: given a probability, and asked what situation it came from? These are called **inverse** problems.

Solving an Inverse Problem

The **qnorm()** function does the inverse of the **pnorm()** function. Check out the help system for them now!

```
?qnorm
```

From this, we see that the syntax for qnorm() is very similar to that of pnorm().

```
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
```

So we specify a **p**, the mean and the sd, and then specify which tail we want, and it inverts the problem to get our original X that we're interested in.

Example 3

The annual salaries of employees in a large company are approximately normally distributed with a mean of $50,000 and a standard deviation of $20,000. If the probability of a randomly selected employee making more than a certain amount X per year is 0.21, what is the dollar figure?

Solution: This is saying $P[X ≥ x_0] = 0.21$, and we're trying to solve for x_0.

```
qnorm(p = 0.21, mean = 50000, sd = 20000, lower.tail = FALSE)
```

```
## [1] 66128.42
```

Notice how we set **lower.tail = FALSE**, because this is a $\geq$ (greater-than or equal to) problem.

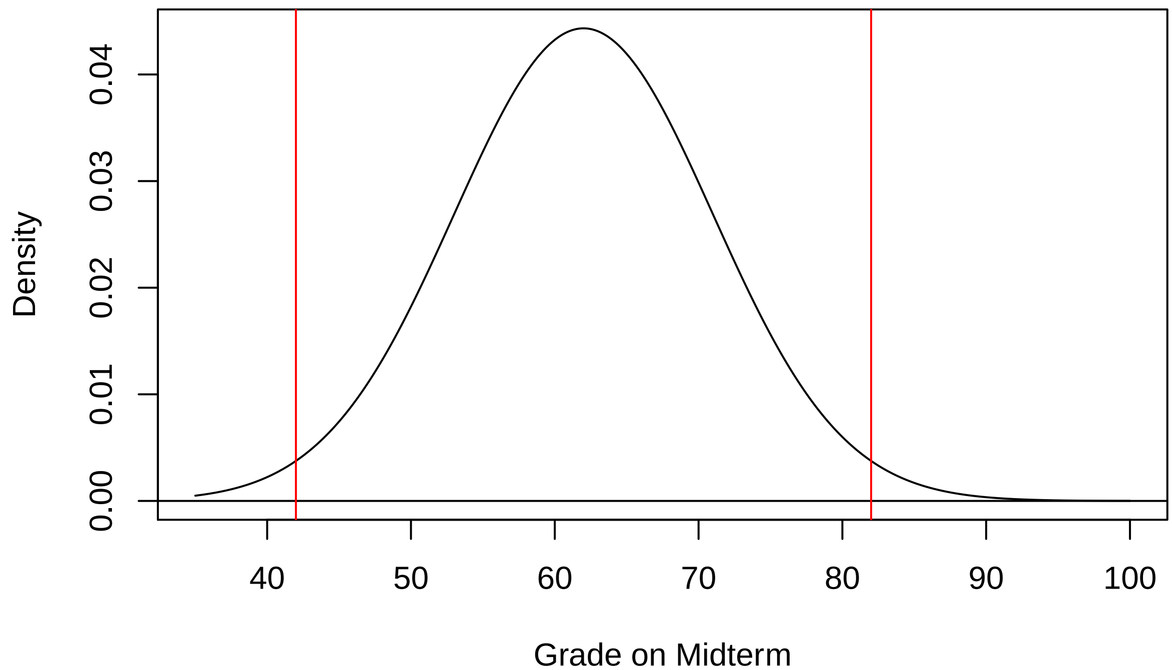
So our solution is $66,128.42 - so 21% of the company makes more than that amount.

Example 4

If I tell you that the midterm exam for MATH 1051H had normally distributed scores, with mean 62% and standard deviation 9%, and then ask you to find the two scores which contain the middle approximately 90% of the class, what are they?

Solution: This is a more complicated problem. Let's slow down, draw a picture, and think a moment.

```
x <- seq(35, 100, 0.01)
plot(x, dnorm(x, mean = 62, sd = 9), type = "l", xlab = "Grade on Midterm", ylab = "Density")
abline(h = 0)
abline(v = c(62-20, 62+20), col = "red")
```



Now, I just made up those numbers (42 and 82) to get lines which cover "about" 90% of the curve. This is to help us think through the problem. We know that the 90% is supposed to be the "middle", which means that since the normal curve is symmetric, we should have 5% area to the left of the left red line, and 5% area to the right of the right red line, and 90% of the area between them. Write this as a probability statement:

$$P[X_1 \leq X \leq X_2] = 0.90$$

and then think about it in terms of Z-scores:

$$Z_1 = \frac{X_1 - 62}{9} \qquad Z_2 = \frac{X_2 - 62}{9}$$

But from the symmetry, we know that X_1 is the same distance to the left of the mean, $\mu = 62$, as X_2 is to the right. So that means the Z scores, which are centered on 0, need to be the same distance as well. In other words,

$$P[X_1 \leq X \leq X_2] = 0.90 = P[-Z_0 \leq Z \leq Z_0].$$

And this is the form we need to solve this problem! Because it says that the area to the left of $-Z_0$ is 0.05, just like the area to the right of $+Z_0$.

```r
qnorm(p = 0.05, mean = 62, sd = 9, lower.tail = TRUE)
```

```
## [1] 47.19632
```

```r
qnorm(p = 0.05, mean = 62, sd = 9, lower.tail = FALSE)
```

```
## [1] 76.80368
```

So these results tell us that 90% of the class (approximately!) had grades on the midterm of between 47.2% and 76.8%.

These are the hardest problems in the first half of the course, so make sure you think about them carefully! They will absolutely be on your midterm (and your final).

Challenge Problems

1. Calculate the value of k for which $P[0 < Z < k] = 0.4370$.
2. A manufacturing process produces bolts with mean weight 75g and standard deviation 4g. 10% of the bolts are rejected because they are too heavy. Calculate this weight to the nearest g.
3. Given a normal distribution of values for which the mean is 70 and the standard deviation is 4.5. Find:

- the probability that a value is between 65 and 80, inclusive.
- the 90th percentile for this distribution.