Lecture 3 practice

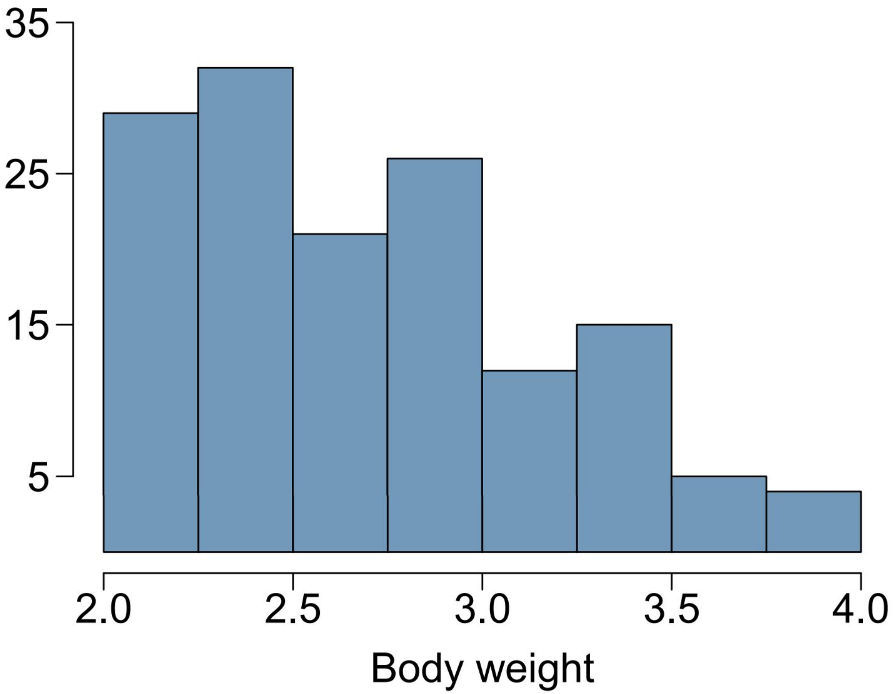
August 9, 2022

Cat weights

The histogram shown below represents the weights (in kg) of 47 female and 97 male cats.

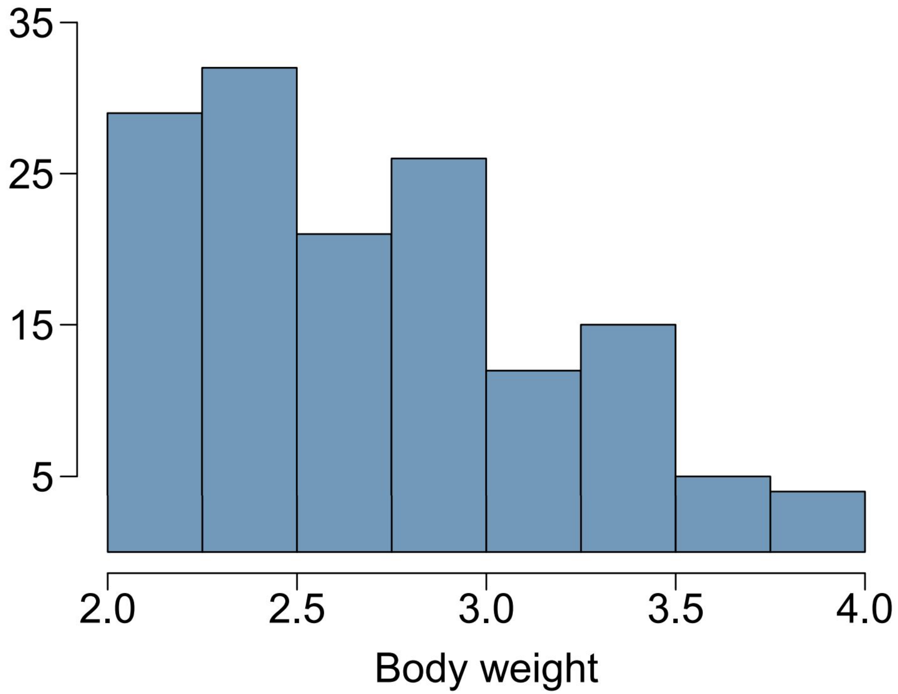
(a) What fraction of these cats weigh less than 2.5 kg?

(29 + 32) / (47 + 97) = 0.42



Cat weights

The histogram shown below represents the weights (in kg) of 47 female and 97 male cats.



(b) What fraction of these cats weigh between 2.5 and 2.75 kg?

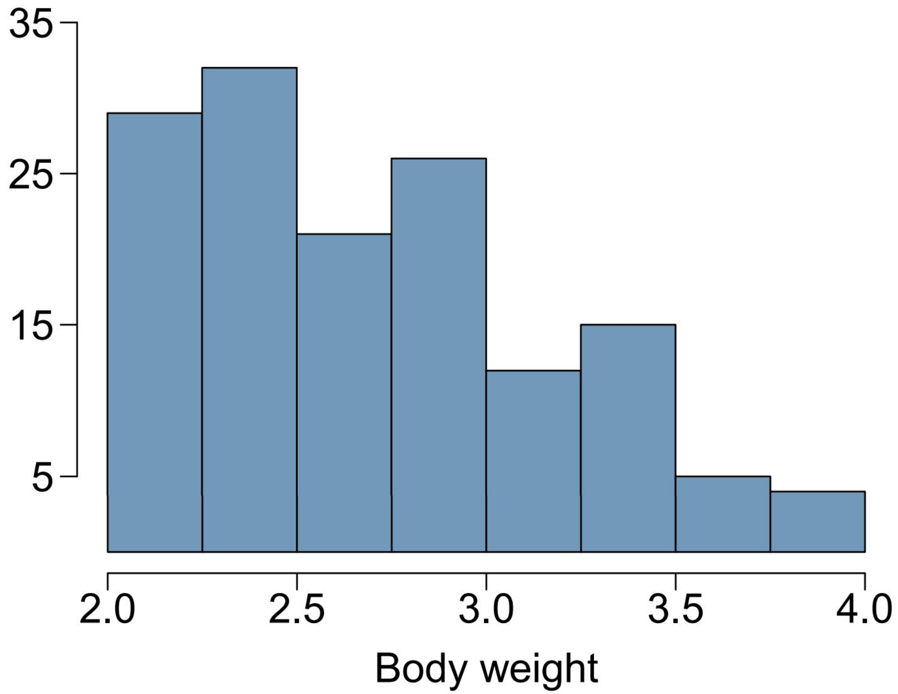
21 / (47 + 97) = 0.15

Cat weights

The histogram shown below represents the weights (in kg) of 47 female and 97 male cats.

(c) What fraction of these cats weigh between 2.75 and 3.5 kg?

(26 + 12 + 15) / (47 + 97) = 0.37



Standard normal distribution

Recall that a Z-score of normal data has a normal distribution with mean 0 and standard deviation 1, i.e.

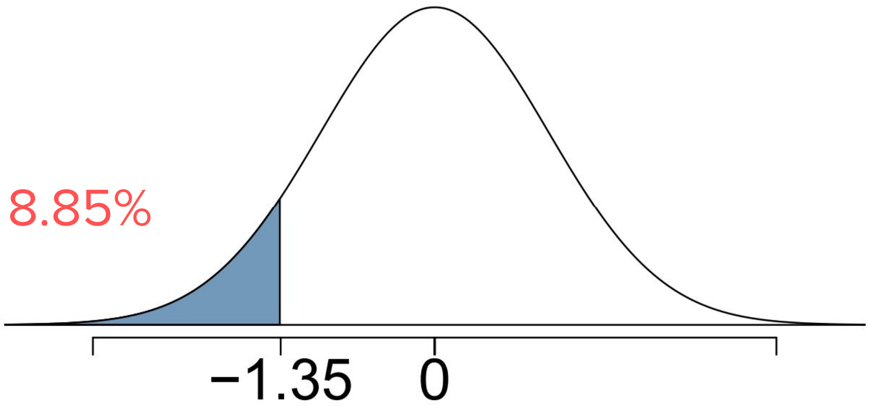
$$Z \sim N(\mu = 0, \sigma = 1)$$

The normal distribution with $\mu = 0$ and $\sigma = 1$ is a special case called the standard normal distribution

Area under the curve

What percent of a standard normal distribution N(μ = 0, σ = 1) is found in each region? Be sure to draw a graph.

(a) Z < -1.35

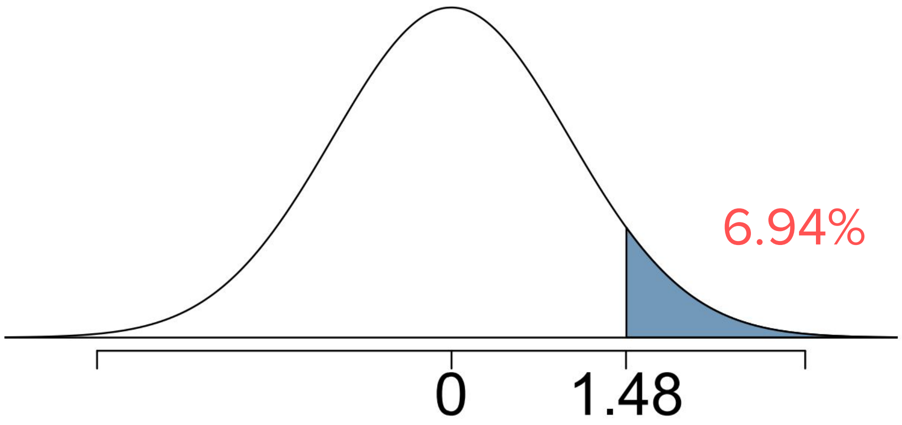



```
> pnorm(-1.35, mean = 0, sd = 1)
[1] 0.08850799
```

Area under the curve

What percent of a standard normal distribution N(μ = 0, σ = 1) is found in each region? Be sure to draw a graph.

(b) Z > 1.48

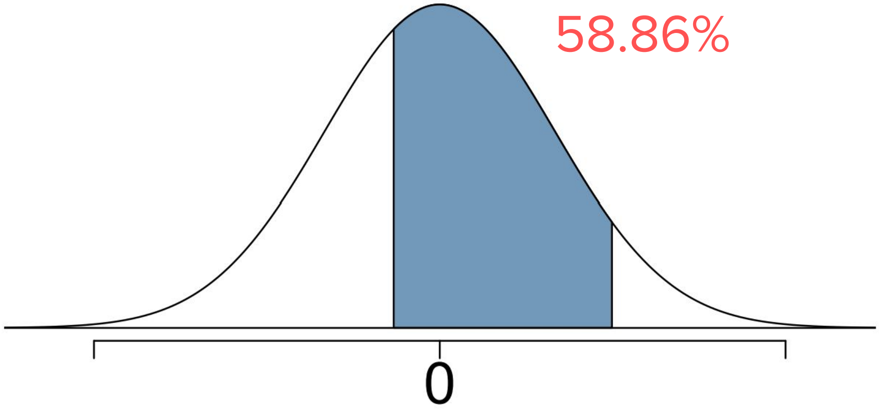



```
> 1 - pnorm(1.48, mean = 0, sd = 1)
[1] 0.06943662
```

Area under the curve

What percent of a standard normal distribution N(μ = 0, σ = 1) is found in each region? Be sure to draw a graph.

(c) -0.4 < Z < 1.5

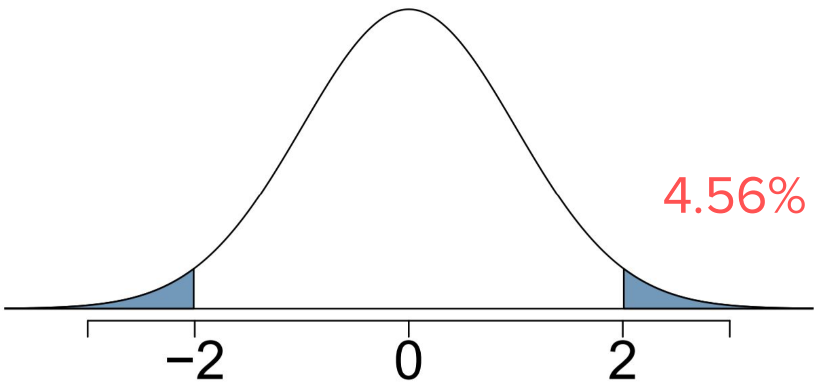



```
> pnorm(1.5, mean = 0, sd = 1) - pnorm(-0.4, mean = 0, sd = 1)
[1] 0.5886145
```

Area under the curve

What percent of a standard normal distribution N(μ = 0, σ = 1) is found in each region? Be sure to draw a graph.

(d) |Z| > 2




```
> pnorm(-2, mean = 0, sd = 1) + 1 - pnorm(2, mean = 0, sd = 1)
[1] 0.04550026
> 2 * pnorm(-2, mean = 0, sd = 1)
[1] 0.04550026
```

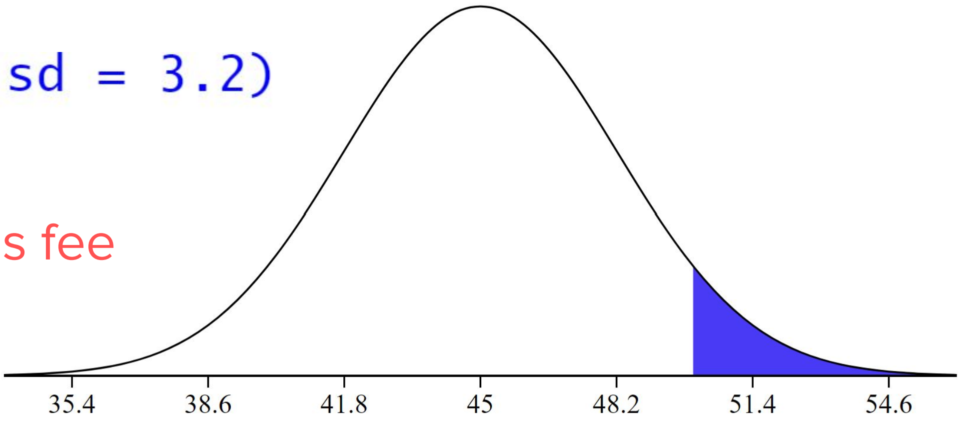
Overweight baggage

Suppose weights of the checked baggage of airline passengers follow a nearly normal distribution with mean 45 pounds and standard deviation 3.2 pounds. Most airlines charge a fee for baggage that weigh in excess of 50 pounds. Determine what percent of airline passengers incur this fee.

We want to calculate the area past 50 under the normal distribution curve.

```
> 1 - pnorm(50, mean = 45, sd = 3.2)
[1] 0.05908512
```

5.9% of airline passengers incur this fee



GRE scores

Sophia who took the Graduate Record Examination (GRE) scored 160 on the Verbal Reasoning section and 157 on the Quantitative Reasoning section. The mean score for Verbal Reasoning section for all test takers was 151 with a standard deviation of 7, and the mean score for the Quantitative Reasoning was 153 with a standard deviation of 7.67. Suppose that both distributions are nearly normal.

(a) Write down the short-hand for these two normal distributions.

Verbal: $N(\mu = 151, \sigma = 7)$

Quantitative: $N(\mu = 153, \sigma = 7.67)$

GRE scores

Verbal Reasoning section score: 160
Quantitative Reasoning section: 157
Verbal: N(μ = 151, σ = 7)
Quantitative: N(μ = 153, σ = 7.67)

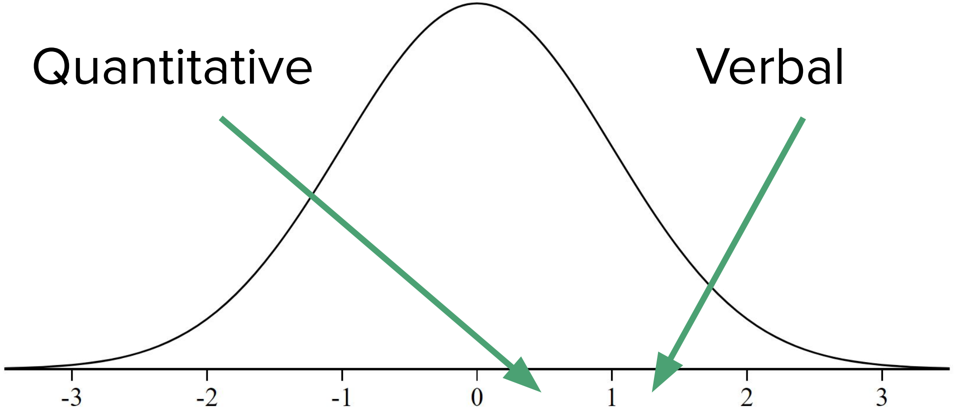
(b) What is Sophia's Z-score on the Verbal Reasoning section? On the Quantitative Reasoning section? Draw a standard normal distribution curve and mark these two Z-scores.

Verbal:

Z-score = (160 - 151) / 7 = 1.29

Quantitative:

Z-score = (157 - 153) / 7.67 = 0.52



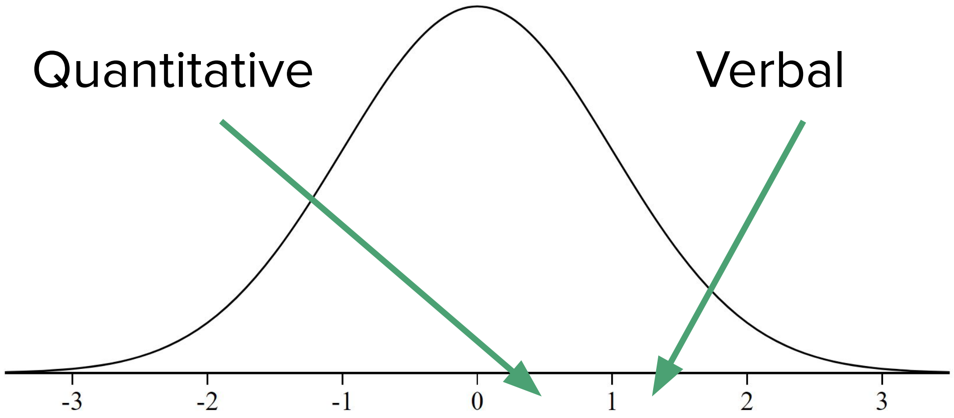
GRE scores

Verbal Reasoning section score: 160
Quantitative Reasoning section: 157
Verbal: $N(\mu = 151, \sigma = 7)$
Quantitative: $N(\mu = 153, \sigma = 7.67)$

(c) What do these Z-scores tell you?

The Z-score is the number of standard deviations from the mean the value is.

Verbal:
Z-score = (160 - 151) / 7 = 1.29
Quantitative:
Z-score = (157 - 153) / 7.67 = 0.52



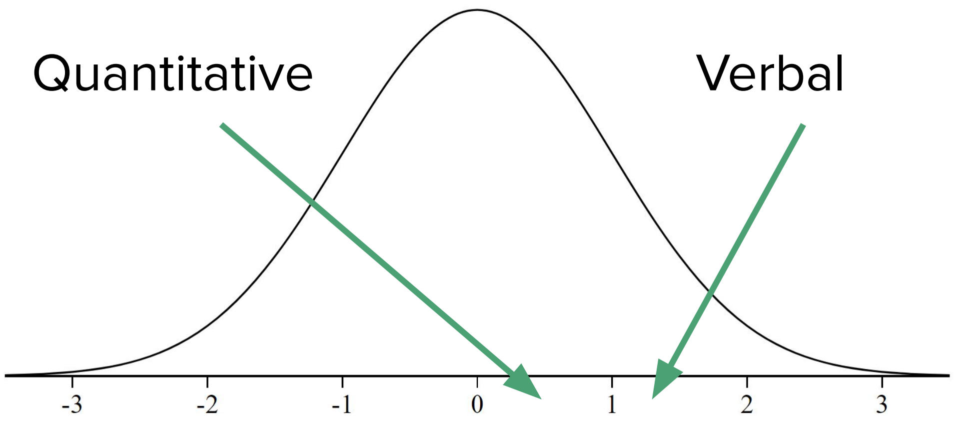
GRE scores

Verbal Reasoning section score: 160
Quantitative Reasoning section: 157
Verbal: N(μ = 151, σ = 7)
Quantitative: N(μ = 153, σ = 7.67)

Verbal:
Z-score = (160 - 151) / 7 = 1.29
Quantitative:
Z-score = (157 - 153) / 7.67 = 0.52

(d) Relative to others, which section did she do better on?

Sophia's verbal score is more impressive than her quantitative score, relative to their respective distributions.



GRE scores

Verbal Reasoning section score: 160
Quantitative Reasoning section: 157
Verbal: N(μ = 151, σ = 7)
Quantitative: N(μ = 153, σ = 7.67)

Verbal: 90th percentile
Quantitative: 70th percentile

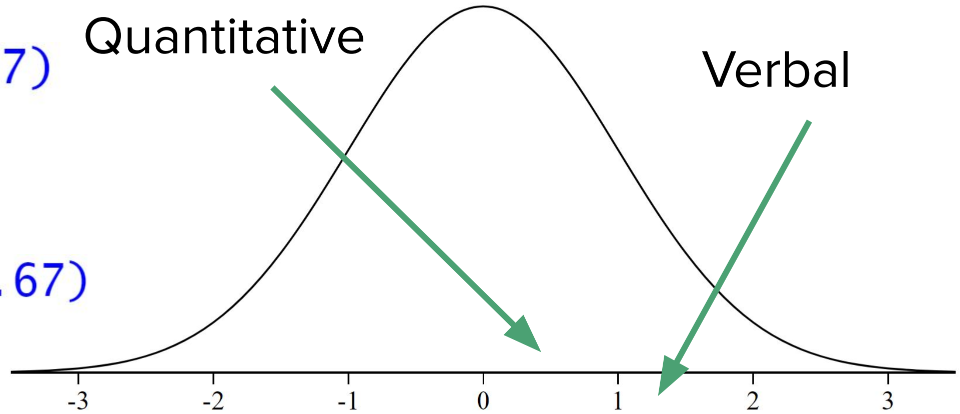
(e) Find her percentile scores for the two exams.

Verbal:

```
> pnorm(160, mean = 151, sd = 7)
[1] 0.9007286
```

Quantitative:

```
> pnorm(157, mean = 153, sd = 7.67)
[1] 0.6989951
```



GRE scores

Verbal Reasoning section score: 160
Quantitative Reasoning section: 157
Verbal: N(μ = 151, σ = 7)
Quantitative: N(μ = 153, σ = 7.67)

Verbal: 90th percentile
Quantitative: 70th percentile

(f) What percent of the test takers did better than her on the Verbal Reasoning section? On the Quantitative Reasoning section?

Only 10% of test takers did better than Sophia on the verbal reasoning section.
30% of test takers did better than Sophia on the quantitative reasoning section.

GRE scores

(g) Explain why simply comparing raw scores from the two sections could lead to an incorrect conclusion as to which section a student did better on.

We cannot compare the raw scores since they are on different scales. Comparing her percentile scores is more appropriate when comparing her performance to others.

GRE scores

(h) If the distributions of the scores on these exams are not nearly normal, would your answers to parts (b) - (f) change? Explain your reasoning.

Answer to part (b) would not change as Z-scores can be calculated for distributions that are not normal. However, we could not answer parts (d)-(f) since we cannot use the normal probability table to calculate probabilities and percentiles without a normal model

GRE scores continued

Verbal: N(μ = 151, σ = 7)
Quantitative: N(μ = 153, σ = 7.67)
Using the two distributions for GRE scores, find the following:

(a) The score of a student who scored in the 80th percentile on the Quantitative Reasoning section.

Here we know the area and we want to compute the corresponding value

```
> qnorm(0.8, mean = 153, sd = 7.67)
[1] 159.4552
```

So the score of someone who scored in the 80th percentile on the Quantitative Reasoning section is 159

GRE scores continued

Verbal: N(μ = 151, σ = 7)
Quantitative: N(μ = 153, σ = 7.67)
Using the two distributions for GRE scores, find the following:

(b) The score of a student who scored worse than 70% of the test takers in the Verbal Reasoning section.

Here we know the area and we want to compute the corresponding value.
Worse than 70% of test takers is better than 30% of test takers

```
> qnorm(0.3, mean = 151, sd = 7)
[1] 147.3292
```

The score of someone who scored less than 70% of test takers is 147

LA weather

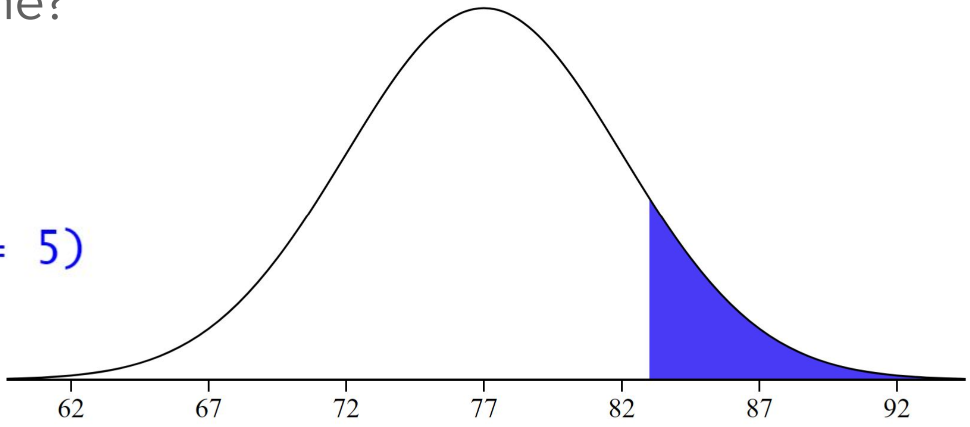
The average daily high temperature in June in LA is 77◦F with a standard deviation of 5◦F. Suppose that the temperatures in June closely follow a normal distribution.

(a) What is the probability of observing an 83◦F temperature or higher in LA during a randomly chosen day in June?

We want to know $P(X > 83)$ with $X \sim N(\mu = 77, \sigma = 5)$

```
> 1 - pnorm(83, mean = 77, sd = 5)
[1] 0.1150697
```

The probability is 0.1151



LA weather

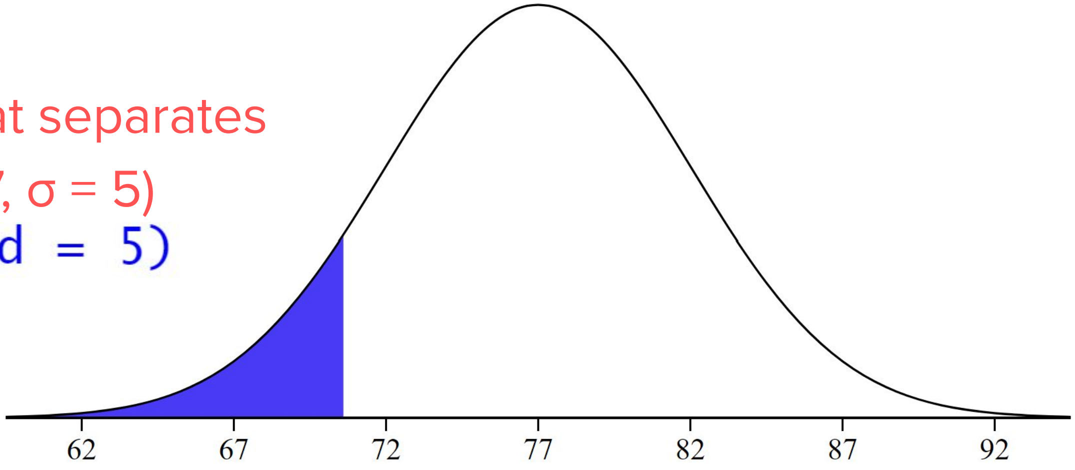
The average daily high temperature in June in LA is 77◦F with a standard deviation of 5◦F. Suppose that the temperatures in June closely follow a normal distribution.

(b) How cool are the coldest 10% of the days (days with lowest high temperature) during June in LA?

We want to know the X value that separates the bottom 10% with X ~ N(μ = 77, σ = 5)

```
> qnorm(0.1, mean = 77, sd = 5)
[1] 70.59224
```

The coolest 10% of days are 70.6◦F or cooler



Credits