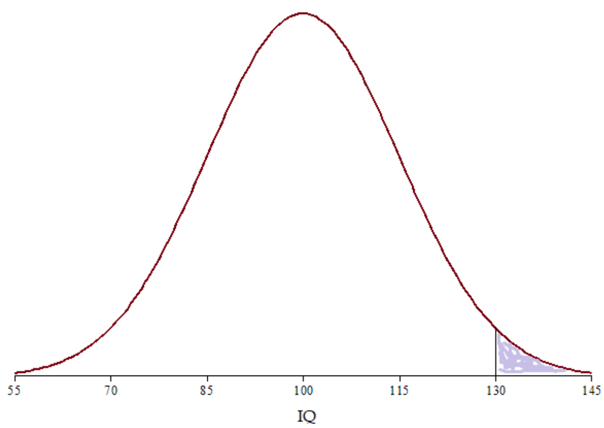
Lab 3: The Normal Distribution (Section 1.4)

Insert the answers to all the questions below into this document and then submit through Blackboard.

1. The Wechsler Adult Intelligence Scale (WAIS) is the most common IQ test. The scale of the scores is set separately for each age group, and the scores are approximately Normal with mean 100 and standard deviation 15.

Use the 69-95-99.7% rule to answer the following questions.

a. The organization MENSA, which calls itself "the high-IQ society, " requires a WAIS score of 130 or more for membership. What percent of adults would qualify for membership? (The proportion is represented by the shaded area under the curve below.)
b. 2.35%



b. Find the percentage of adults with IQ scores between 70 and 100.

47.5%

c. Find the percentage of adults with IQ scores below 55.

.3%

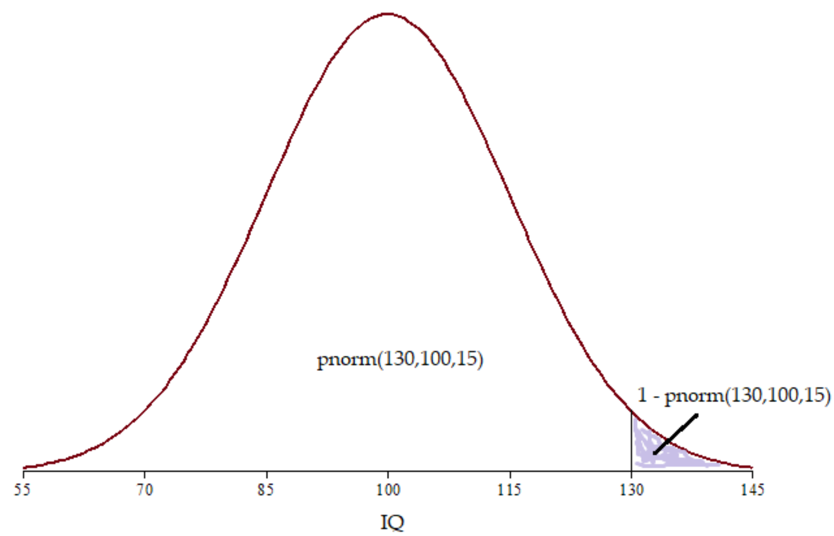
d. Find the percentage of adults with IQ scores between 85 and 130.

81.5%

The **68-95-99.7% rule** gives a good approximation of the percentages for Normal distributions. But we can get more accurate results using R. Moreover, unlike the instruction given in the textbook, we won't have to first change to z-scores and use the standard Normal tables in the back of the book. (Recall: You calculate the z-score as follows: $z = \frac{x-\mu}{\sigma}$.)

2. Return the distribution of IQ scores from question 1. You will repeat the same questions but this time you will get more accurate results using R. R will give the areas in terms of the proportions that fall in a certain category; you will just need to multiply the results by 100 to turn proportions (or probabilities) into percentages.

a. The organization MENSA, which calls itself "the high-IQ society, " requires a WAIS score of 130 or more for membership. What percent of adults would qualify for membership?



To answer this question we use the pnorm command: pnorm(x-value, mean, standard deviation). This will give the area to the left of the x-value, which is the white area under the normal curve (see above). So, pnorm(130, 100, 15) gives the white area or the proportion of adults whose IQ scores were below 130. For this situation, we need the purple area. Since the area under the entire normal curve is 1, the area shaded in purple must be:

1 – pnorm(130,100,15)

The default of pnorm is the area to the left of the x-value. To find the area to the right of the x-value, you could also change the default by adding lower.tail = FALSE as shown below.

pnorm(130, 100, 15, lower.tail = FALSE)

```
> pnorm(130, 100, 15, lower.tail = FALSE)
[1] 0.02275013  =  2.28%
```

Try both commands and compare your result with 1(a). How good was the 68-95-99.7% rule?

b. Calculate the percentages for (b) – (d) using R.

b. [1] 0.4772499 = 47.72%

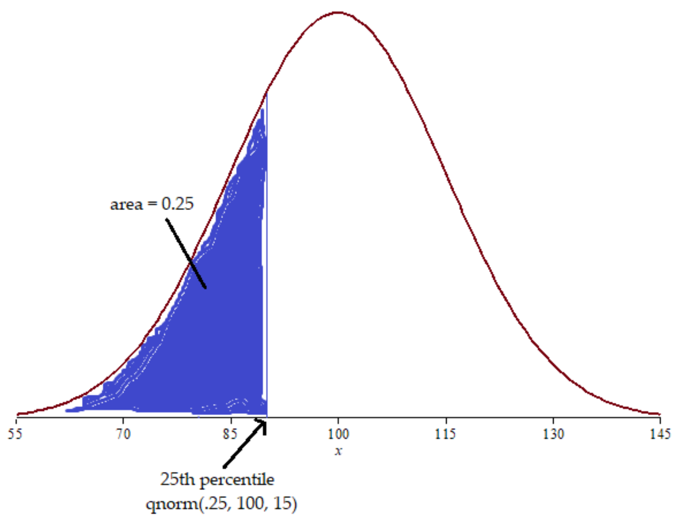
c. [1] 0.001349898 = .13%

d. [1] 0.8185946 = 81.86%

It was pretty accurate for b and d. Bit off for c but nonetheless it was a good tool to use.

3.a. Sometimes, we want to find percentiles. For example, the 25th percentile, would be the IQ score such that 25% of adults have scores at or below this score. Find the 25th percentile for the IQ scores using the command qnorm(.25, 100, 15).

> qnorm(.25, 100, 15)
[1] 89.88265



b. What is the 75th percentile?

110.11173

c. To find the upper 25th percentile, we need to use the lower.tail = FALSE command:

qnorm(.25, 100, 15, lower.tail = FALSE)

Run the command above.

d. Compare your answers to (b) and (c). What does this tell you about the 75th percentile and the upper 25th percentile?

The upper 25th percentile is the 75th percentile.

e. Find the 40th percentile.

[1] 96.19979

f. Find the upper 40th percentile.

[1] 103.8002

4. The deciles of any distribution are the 10th, 20th, . . . , 90th percentiles. The first and last deciles are the 10th and 90th percentiles, respectively.

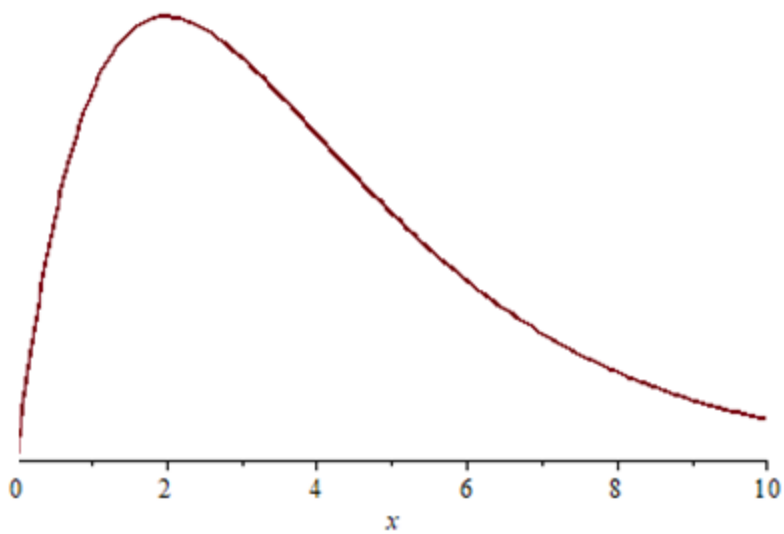
a. What are the first and last deciles of a standard normal distribution? (The standard normal distribution has mean 0 and standard deviation 1; we can specify this as N(0,1).)

```
> qnorm(.1,0,1)
[1] -1.281552

 qnorm(.9,0,1)
[1] 1.281552
```

b. The weights of 9-ounce potato chip bags are approximately Normal with mean 9.12 ounces and standard deviation 0.15 ounce. What are the first and last deciles of this distribution?

```
> qnorm(.1,9.12,.15)
[1] 8.927767

> qnorm(.9,9.12,.15)
[1] 9.312233
```

5. The Normal density function is a good model when data are roughly mound-shaped and symmetric. But some datasets are skewed and thus, there is a need for density curves that have different shapes. One example is the chi-square distribution. In Lab 4 you will be introduced to the formula for this distribution. While the Normal density function is completely determined by the mean and standard deviation, the chi-square distribution is completely determined by its degrees of freedom. For example, here is the density curve for a chi-square with 4 degrees of freedom.

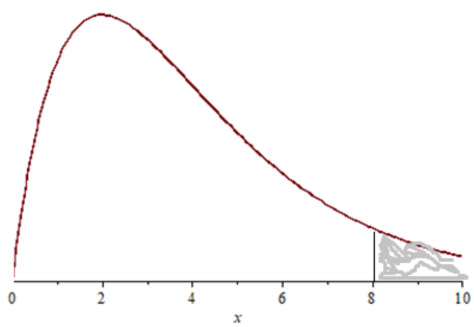


What percentage of data from this distribution is above 8? In other words, find the area of the shaded region shown below. You will need a new command pchisq(x, df, lower.tail = FALSE) – enter in the x-value followed by the degrees of freedom. So, in this case, pchisq(8, 4, lower.tail = FALSE) and then multiply by 100 to get the percentage.

```
> pchisq(8, 4, lower.tail = FALSE)
```

[1] 0.09157819   = 9.16%

.



b. What percentage of data from this distribution is greater than 10? Before you make this calculation in R, would you expect your answer to be larger or smaller than your answer to (b)?

Smaller.

> pchisq(10, 4, lower.tail = FALSE)
[1] 0.04042768 = 4.04%

c. Suppose I want to know the cut-off for the upper 5th percentile. (The area to the right of this x-value will be 0.05.) All that is needed is to change the p to q: qchisq(0.05, 4, lower.tail = FALSE).

> qchisq(.05,4,lower.tail = FALSE)
[1] 9.487729

d. What is the upper 10th percentile?
> qchisq(.9,4)

[1] 7.77944

> qchisq(.1,4,lower.tail = FALSE)

[1] 7.77944