1. Ten students got the following scores on an exam:

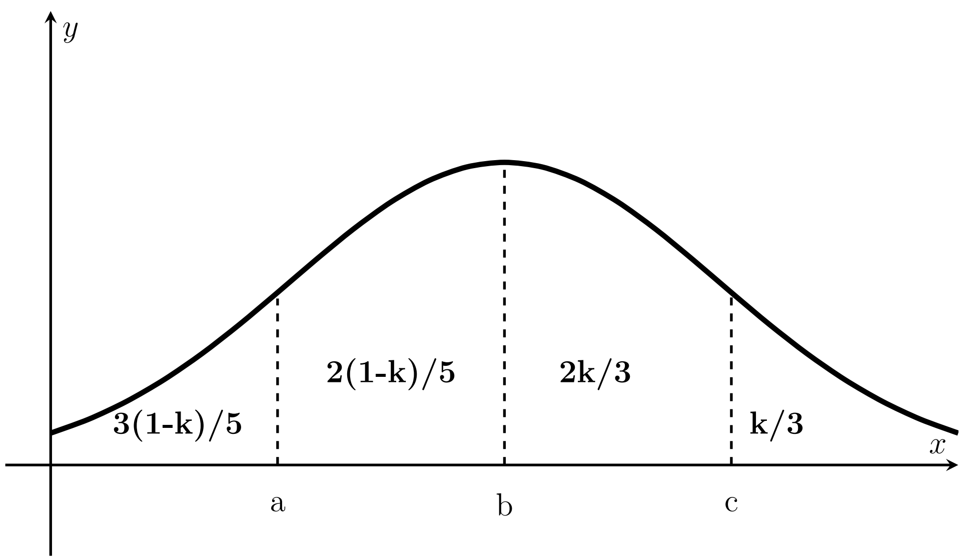
$$77,\ 77,\ 77,\ 82,\ 82,\ 82,\ 82,\ 89,\ 89,\ 94.$$

   (a) Find the mean of these scores. 83.1

   (b) What is the standard deviation of these scores? 5.82

   (c) How many scores fall within one standard deviation of the mean? 4

2. The following graph corresponds to the probability density function (pdf) for a random variable X. The regions separated by dashed lines have areas as shown. Here, k is a constant that is less than 1.



   (a) Find $P(x < b)$ (answer in terms of k). $1 - k$

   (b) Find $P(x > a)$ (answer in terms of k). $k + 2(1 - k)/5$

   (c) Find r such that $P(x > r) = k$ (answer in terms of a, b, c). $r = b$

   (d) Looking at the above graph, about what do you think k is equal to? (Your answer should be an explicit real number.) Please explain. It looks like k is about 0.5, because it looks like about half the area under the curve (probability) is to the left of b, and half is to the right. So by part (c) above, k should equal 0.5.

3. The mean of a set of 30 exam scores is computed and found to equal 82. (Not all scores are the same.) Later, a 31st exam is found. (It's a long story. It turns out the dog *hadn't* eaten that exam.) The exam is graded, and the score comes out also to be 82.

   (a) Is the *mean* of the new set of 31 scores less than, equal to, or greater than the mean of the original set of 30 scores? Please explain.

   The means for the two sets are the same. Since the new score equals the original mean, it can't pull this mean up or down.

   (b) Is the *standard deviation* of the new set of 31 scores less than, equal to, or greater than the standard deviation of the original original set of 30 scores? Please explain.

   The standard deviation for the initial set is larger. Adding the one score of 82 does not increase the numerator in the definition of standard deviation, but it increases the denominator, resulting in a smaller s.

4. Given below are four intervals of real numbers.

$$(3.878, 3.922), \quad (3.874, 3.926), \quad (3.876, 3.920), \quad (3.871, 3.929)$$

   Three of them are confidence intervals for a population mean computed from the same sample, from some population X. One of these three is a 95% confidence interval, one a 98% confidence interval, and one a 99% confidence interval. The fourth interval is *not* a confidence interval computed from this sample. Which of the intervals is which? Please explain.

   Confidence intervals for a population mean, taken from the same sample, will all be centered around the sample mean $\overline{x}$. Note that the first, second, and fourth intervals are all centered about the same number – namely, 3.900 – while the third interval is not. So the third interval must be the one that's *not* a confidence interval taken from the same sample.

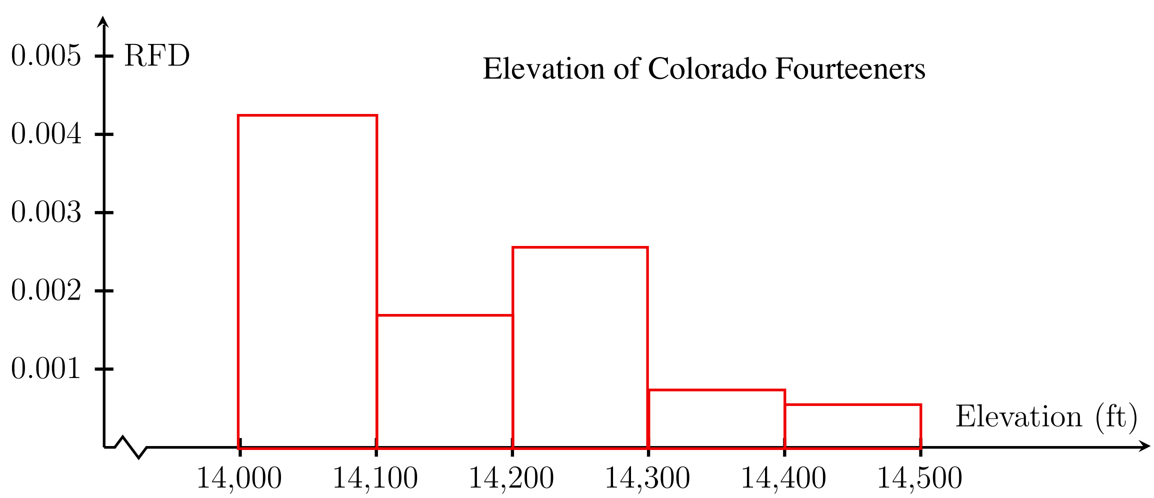
   To distinguish the other three intervals, we note that the wider the interval, the higher the confidence level. So the first interval must be at the 95% level, the second 98%, and the fourth 99%.

5. Colorado has 53 "fourteeners," or mountains over 14,000 feet elevation (with at least 300 feet of prominence). Below is a frequency table for the elevations of these 53 mountains.

Elevation range (ft)	**Frequency**	**Relative Frequency Density (RFD)**
$[14000, 14100)$	23	$23/(100 \cdot 53) = 0.0043$
$[14100, 14200)$	9	$9/(100 \cdot 53) = 0.0017$
$[14200, 14300)$	14	$14/(100 \cdot 53) = 0.0026$
$[14300, 14400)$	4	$4/(100 \cdot 53) = 0.0008$
$[14400, 14500)$	3	$3/(100 \cdot 53) = 0.0006$

(a) Fill in the table above with RFD values. Include four decimal places.

(b) On the axes below, draw an RFD histogram for this data.



(c) Given a randomly chosen fourteener, what is the probability that its elevation is at least 14,200 but less than 14,400 feet? 33.96%

(d) If one of the five elevation ranges given in the above table is selected at random, what's the probability that fewer than 10 Colorado fourteeners have elevation within that range? 60%

(e) Consider the following events:

E: A Colorado fourteener, chosen at random, has elevation in the range $[14,000, 14,300)$.

F: A Colorado fourteener, chosen at random, has elevation in the range $[14,200, 14,500)$.

Find the following values (it's OK to leave your answers as fractions):

- $P(E) = \dfrac{46}{53}$
- $P(F) = \dfrac{21}{53}$
- $P(F^c)$ (recall that F^c means the complement of F). $= \dfrac{32}{53}$
- $P(E|F) = \dfrac{14}{21}$
- $P(E|F^c) = 1$

(f) Verify that Bayes' Law holds here for $P(E)$.

$$P(E) = \frac{46}{53}$$

$$P(F)P(E|F) + P(F^c)P(E|F^c) = \frac{21}{53} \cdot \frac{14}{21} + \frac{32}{53} \cdot 1 = \frac{14}{53} + \frac{32}{53} = \frac{46}{53} = P(E).$$

6. Suppose the radii X of ball bearings produced by a certain machine are normally distributed, with mean $\mu = 10$ mm and standard deviation $\sigma = 0.1$ mm.

(a) Find the proportion of ball bearings produced by this machine that have radius between 9.8 mm and 10.2 mm. 95.5%

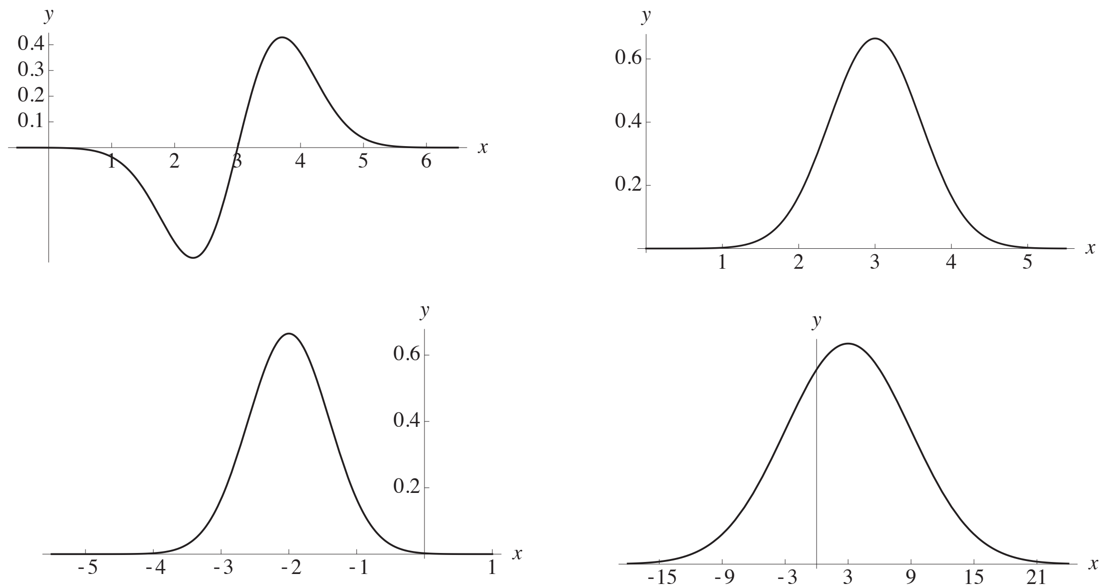
(b) Find the proportion that have radius greater than or equal to 10.2 mm. Hint: think about the *symmetry* of the standard normal curve about the mean. 2.25%

(c) Find the number L such that 95% of the bearings have radius less than L standard deviations from the mean radius. $L = 1.96$

(d) If a random sample, of size $n = 100$, of ball bearings from this machine is chosen, what is the probability that the mean radius $\overline{x}$ of this sample will be in the interval $(9.9767, 10.0233)$?

$$\begin{aligned} P\Big(9.9767 < \overline{x} < 10.0233\Big) &= P\left(\frac{9.9767 - 10}{0.1/\sqrt{100}} < \frac{\overline{x} - 10}{0.1/\sqrt{100}} < \frac{10.0233 - 10}{0.1/\sqrt{100}}\right) \\ &= P\left(-2.33 < \frac{\overline{x} - 10}{0.1/\sqrt{100}} < 2.33\right) = 98\%. \end{aligned}$$

7. Suppose X is a random variable of mean $\mu = 3$ and standard deviation $\sigma = 6$. Which of the four pdfs below could possibly be a pdf for the random variable $\overline{X}$ of all sample means of X, of sample size $n = 100$? Please explain. (Describe completely how you are eliminating the incorrect choices.)



We eliminate the upper left graph because a pdf can't be negative. We eliminate the lower left because $\overline{X}$ is $N(3, 6/\sqrt{100}) = N(3, .6)$, and the mean in the lower left curve not 3. We eliminate the lower right because, again, $\overline{X}$ is $N(3, 6/\sqrt{100}) = N(3, .6)$, so 99.7% of the data must fall withing 3(.6)=1.8 units of 3, and this is clearly not the case for the lower right curve. So the upper right curve must be the correct one.

8. For healthy adults, the ideal red blood cell volume in ml per kg of body weight is 28 ml/kg. Deviation from this level indicates a serious condition. A patient has his blood tested 36 times. The mean of these tests is 32.7 ml/kg and the standard deviation is 2.18 ml/kg. Test at the 98% level the hypothesis that the patient has abnormal red blood cell volume by looking at the test statistic, z.

We compute that

$$z = \frac{\overline{x} - \mu_0}{(s/\sqrt{n})} = \frac{32.7 - 28}{(2.18/\sqrt{36})} = 12.94.$$

Since $|z| = 12.94 > 2.33$, we reject the null hypothesis $H_0\colon \mu = 28$, and accept the alternative hypothesis $H_A\colon \mu \neq 28$.

9. Biologists are studying Allen's hummingbird, which lives in AZ. They capture a sample of 40 hummingbirds and find a mean mass of 3.15 grams and a standard deviation of 0.33 grams. Find a 98% confidence interval for the mean mass of these hummingbirds.

The 98% confidence interval is

$$\left(\overline{x} - 2.33\frac{s}{\sqrt{n}}, \overline{x} + 2.33\frac{s}{\sqrt{n}}\right) = \left(3.15 - 2.33\frac{0.33}{\sqrt{40}}, 3.15 + 2.33\frac{0.33}{\sqrt{40}}\right) = (3.03, 3.27).$$

10. (a) Find the maximum value of the function $f(\widehat{p}) = \widehat{p}(1-\widehat{p})$ on the interval $[0.6, 0.8]$.

$f'(\widehat{p}) = 1 - 2\widehat{p}$, which equals zero only at $\widehat{p} = 1/2$, which is out side the interval in question. So we check the endpoints: $f(0.6) = 0.6(1-0.6) = 0.24$, while $f(0.8) = 0.8(1-0.8) = 0.16$. So the maximum value of 0.24, on the interval $[0.6, 0.8]$, is attained at $\widehat{p} = 0.6$.

(b) Pollsters conducting a public opinion poll on the proportion of the population that hates zucchini are fairly certain that their sample proportion will be somewhere between 60% and 80%. If this is correct, how large a sample should they choose to assure a margin of error, at the 99% level, of no more than 3%?

Again, on the interval $[0.6, 0.8]$, $\widehat{p}(1-\widehat{p}) \le 0.24$, so

$$2.58\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \le 2.58\sqrt{\frac{0.24}{n}} = \frac{2.58\sqrt{0.24}}{\sqrt{n}} = \frac{1.26394}{\sqrt{n}}.$$

We solve for n:

$$\begin{aligned} \frac{1.26394}{\sqrt{n}} &\le 0.03 \\ 1.26394 &\le 0.03\sqrt{n} \\ n &\ge \left(\frac{1.26394}{0.03}\right)^2 = 1775.04. \end{aligned}$$

(c) Is the margin of error under the above assumption larger than, smaller than, or the same as what you would get if you made no prior assumptions about the sample proportion?

Somewhat smaller. The maximum value of $\widehat{p}(1-\widehat{p})$ overall is $1/4 = 0.25$, which is slightly larger than 0.24.

(d) Suppose these pollsters conduct a poll with a sample of size $n = 2000$. Suppose they find that the proportion of that sample that hates zucchini is 74%. Find a 99% confidence interval for the proportion of the population that hates zucchini. Write your answer to three decimal places. What is the margin or error for this interval? The margin of error is

$$2.58\sqrt{\frac{0.74(1-0.74)}{2000}} = 0.025.$$

The confidence interval is

$$(0.74 - 0.025, 0.74 + 0.025) = (0.715, 0.765).$$