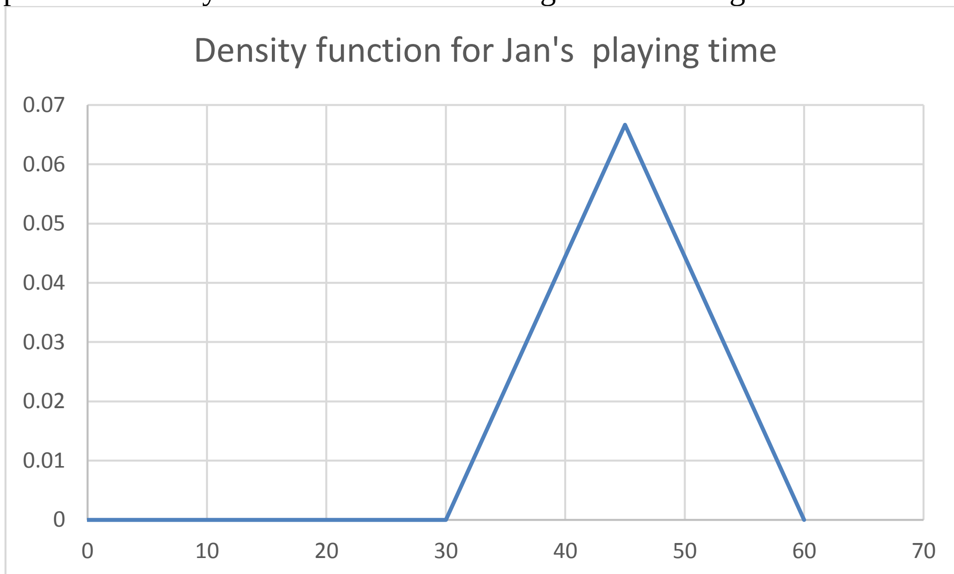
- Excel or R will be allowed and you may use a formula sheet (1 sheet, 2 pages, hand-written).

1. (25 pts) The playing time for high school basketball player Jan is a continuous random variable Y defined by a density function which is 0 except between 30 and 60 minutes, where it has the form of an isosceles triangle (symmetric with peak at 45).

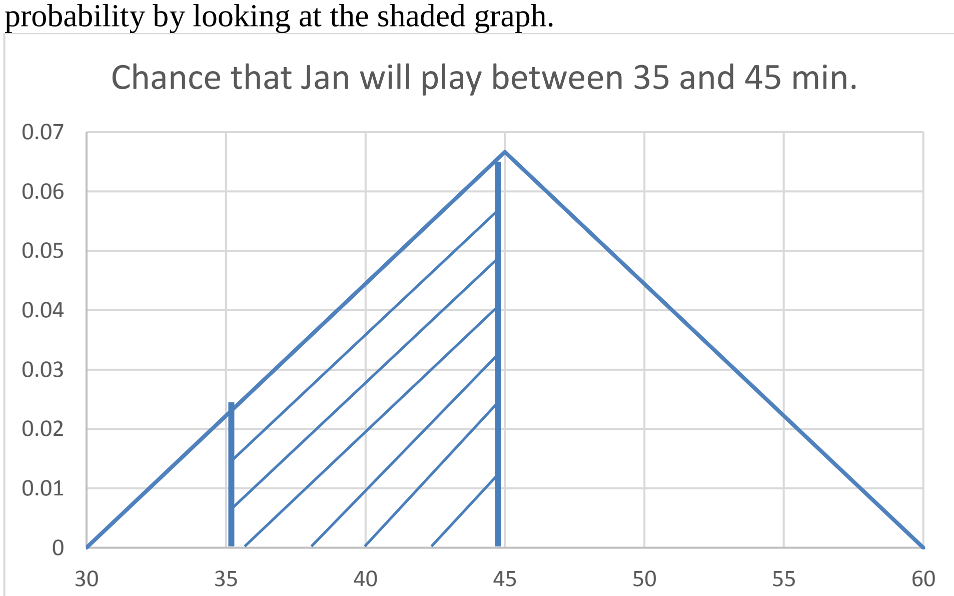
   a. Draw a graph of the density function and find the height of the triangle



Let height be h. A=1/2bh $= \frac{1}{2}(60-30)h = 1$ **so** $15h = 1$ **or h=1/15**

   b. What is the chance that Jan will play between 35 and 45 minutes?

      i. Shade and label the appropriate region of the graph for P(35<Y<45) and guess the probability by looking at the shaded graph.



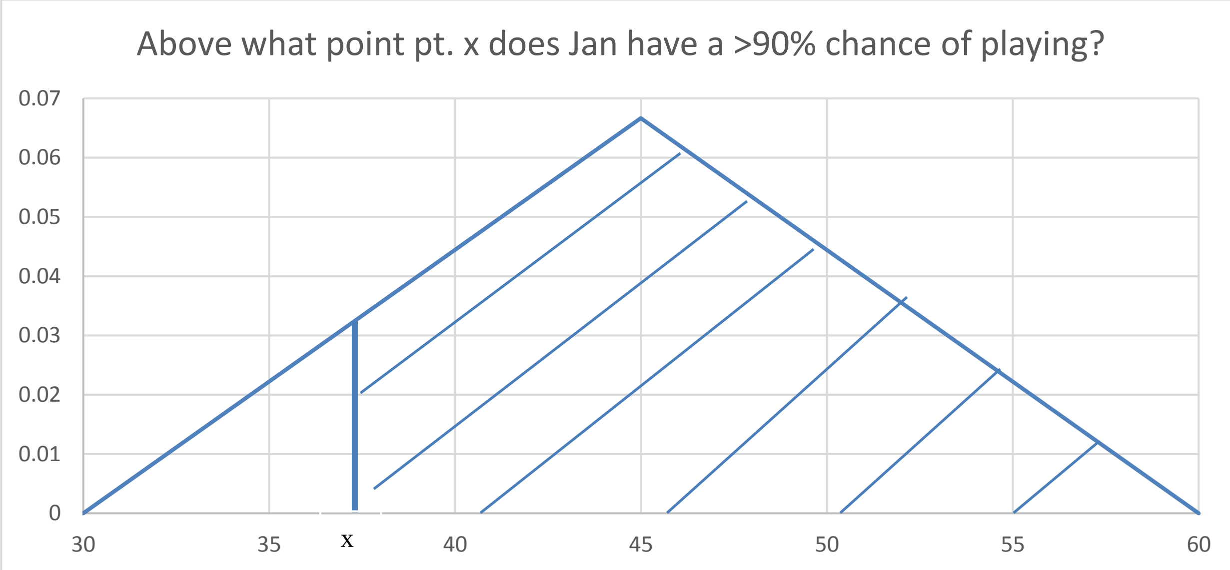
Area is less than 50% (due to symmetry), perhaps 40 or 45%.

      ii. Find the probability. Does your answer more or less match what you guessed?

This is a trapezoid (on its side). "Height" is (45-35)=10 and one base is 1/15. To find other base, note that the height grows linearly as we go from left to right. 35 is 1/3 of the way, so the height will be 1/3 of 1/15 = 1/45
Area = ½ ("height")("base1"+ "base2") = ½ (10)(1/15+1/45)=5(4/45)=4/9 or 44%
Which is in agreement with our visual estimation.
Alternatively, we could have found area of the small triangle [12/bh=1/2(1/45)5=1/18] on the left and subtracted from 50% [1/2-1/18=(9-1)/18=8/18=4/9].

   c. Jan has a 90% chance of playing above how many minutes?

i. Make a new drawing of the distribution and shade and label the appropriate region of the graph and guess the x-value.



From our work in part b, at 35, the area above is 95% so it will be perhaps x = 37 or 38 when the area to the right is 90%.

ii. Find the x-value. Does your answer more or less match what you guessed?

Clever solution: **As we move from the left the to right, the area in the excluded triangle will increase with the square of the linear dimension $(x-30)$, i.e.,**

$$A = c(x-30)^2$$

To find the constant of proportionality c, we note that at x=45 the area would be at 50%, i.e., $c(45-30)^2 = \frac{1}{2}$. So, solving for c, we get $c = \frac{1}{450}$.

We want the excluded triangle to have area 10% or 1/10, so solve the equation

$$A = \frac{1}{450}(x-30)^2 = \frac{1}{10}$$

Multiply both sides by 450 and we get: $(x-30)^2 = 45$. Taking the square root of both sides and solving for x we get $x = 30 \pm 3\sqrt{5}$.

x must be >30 so we take just the positive square root and approximate to get 36.7.

As you can see, we overestimated a bit.

***More straightforward solution*: Use the point-slope form of the equation to find the equation of the line for the left side of the isosceles triangle:**

$$y - 0 = \frac{\frac{1}{15}-0}{45-30}(x-30) \text{ or } y = \frac{1}{225}(x-30)$$

Now use the formula for the area of a triangle to get

$$A = \frac{1}{2}bh = \frac{1}{2}(x-30)\left[\frac{1}{225}(x-30)\right] = \frac{1}{10}$$

which when simplified gives us the same equation in the "clever" solution.

Conclude with a sentence:

Jan has a 90% chance of playing more than 36.7 minutes (36 minutes and 42 seconds).

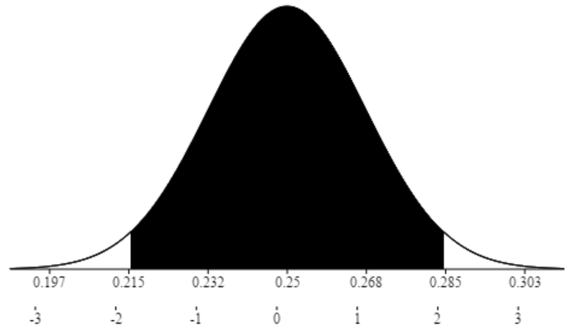
2. (25 pts) A poll (Tuesday 11/24/15): http://www.quinnipiac.edu/news-and-events/quinnipiac-university-poll/iowa/release-detail?ReleaseID=2305 had Trump leading the Iowa Caucus 25% to Rubio's 23%. However, the lead was within the claimed margin of error of $\pm$ 4%. Focus just on Trump's numbers.

a. Find the standard deviation for whether one Iowa caucus voter is a supporter of Trump or not.

This is binomial distribution (coin flipping) with just one trial (flip). $\sigma = \sqrt{(pq)} = \sqrt{(.25*.75)} = 0.43$

b. Given that the sample size was 600, find the standard error (the standard deviation for the sample mean). **Std err=$\sigma/\sqrt{n} = 0.43/\sqrt{600} = 0.0177$**

Draw the appropriate normal curve and label the horizontal axis with both $\bar{X}$ and Z labels.



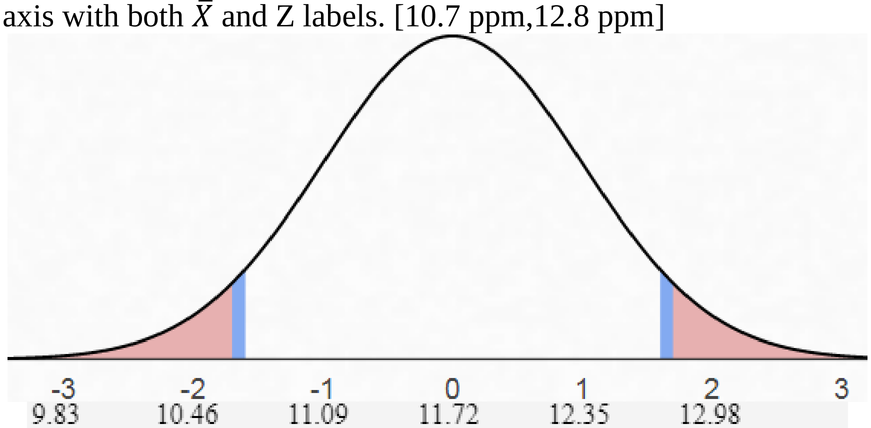
c. The margin of error is approximately ± twice the standard error. Find it. Was it roughly what the polltakers claim? **Margin of error = ±2(.0177)= ±0.035% which if we round up is the 4% provided by the polltakers.**

d. By what factor did the sample size need to be increased (assuming that results would stay the same) for us to conclude that Trump was definitely leading. **The difference between the candidates is 2%. Roughly, to definitively say that Trump is in the lead, we need to halve the margin of error which corresponds to halving the standard error. To halve the standard error, we need to increase the sampling by a factor of 4. More precisely, we solve the equation:**

$$\frac{\mathbf{0.43}}{\sqrt{\boldsymbol{n}}}\mathbf{1.96} = \mathbf{0.02}$$

[1.96=norm.s.inv(0.975)] or $\boldsymbol{n} = (\mathbf{0.43} * \mathbf{1.96} / \mathbf{0.02})\text{^}\mathbf{2} = \mathbf{1776} \approx \mathbf{1800}$ in other words by a factor of 3 rather than the 4 that resulted from the rough calculation.

3. (25 pts) A previous sample of fish in Lake Michigan indicated that the polychlorinated biphenyl (PCB) concentration per fish was distributed normally and that its mean was 11.2 parts per million with a standard deviation of 2 parts per million. Suppose a new random sample of 10 fish has the following concentrations:
   11.5, 12.0, 11.6, 11.8, 10.4, 10.8, 12.2, 11.9, 12.4, 12.6
   Assume that the standard deviation has remained equal to 2 parts per million.
   a. Why do we need to know that the distribution X is itself normal?
   Since the sample size is <30, then the distribution for X must normal or roughly normal for us to be able to say that the sample mean distribution is normal.
   b. Find the standard error for the sample mean. **2/√10 = 0.63 ppm**
   c. Find and draw the curve for the 90th confidence interval for the sample mean. Be sure to label the horizontal axis with both $\bar{X}$ and Z labels. [10.7 ppm,12.8 ppm]



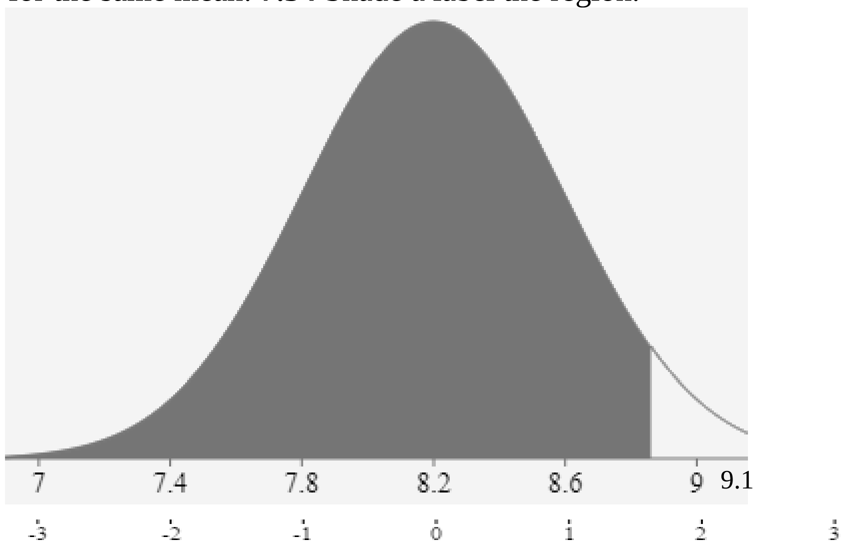
   d. Is the old concentration within this confidence interval? **11.2 ppm is within this interval.**
   e. Do you think that the concentration has changed? **Although the concentration seems to have increased by ~5%, we do not have enough evidence to say that the concentration has changed. The present mean could very believably be the old mean.**
4. (25 pts) Orelia claims that she is able to produce larger tomatoes than average. She plants a tomato variety that results in tomatoes with mean diameter of 8.2 centimeters and standard deviation 2.4 centimeters. A sample of 36 of her tomatoes yields a sample mean of 9.1 centimeters.

a. Why do we NOT need to know that the distribution X is normal or not? **The sampling size of 36 is above threshold size of 30, below which the original distribution must be normal.**

b. Find the standard error for the sample mean: **std err = 2.4/√36 = 0.4**.

c. Draw the curve for the sample mean. Label the horizontal axis with both $\bar{X}$ and Z labels.

d. Find 95th percentile for the same mean: **7.54** Shade a label the region.



e. Mark our sample mean of 9.1. Does it lie in the tail outside the shaded region?
9.1 does lie in the tail outside the shaded region.

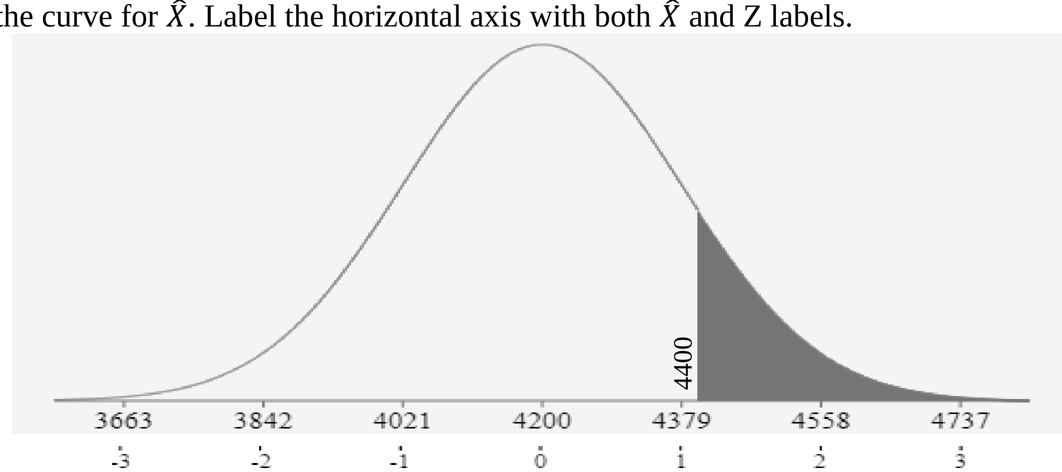
f. Do you think that Orelia's claim is true? **Yes, it is quite believable that Orelia's claim is true.**

5. (25 pts) Airplanes have a harder time lifting off in the thinner air of high-elevation airports. Weights of passengers including their luggage are normally distributed with an average of 210 pounds and standard deviation of 40. A 20-passenger plane can safely take off from Denver with at most 4400 lbs of passengers and cargo. Our plane has the maximum number of passengers allowed.

a. Why do we need to know that the distribution X is normal? **Since the sample size is <30, then the distribution for X must normal or roughly normal for us to be able to say that the sample mean distribution is normal.**

b. Find the standard error for $\hat{X}$ **Std err=σ*√n=40*√20=179**

c. Draw the curve for $\hat{X}$. Label the horizontal axis with both $\hat{X}$ and Z labels.



d. What is the chance that the plane will be over the safe limit for takeoff?
P($\hat{X}$ >4400)=P(Z>1.12)=13%

e. Suppose that you want to reduce the risk of being over the limit to 1%, what should the maximum allowed number of passengers be for this flight?
We note that the mean is now 210*n* and the std err = 40√*n*. Standardizing the inequality

$$\hat{X} > 4400 \Rightarrow \hat{X} - 210\,n > 4400 - 210\,n \Rightarrow (\hat{X} - 210\,n)/\,40\sqrt{n} > (4400 - 210\,n)\,/\,40\sqrt{n}$$

$$\Rightarrow Z > (4400 - 210\,n)\,/\,40\sqrt{n} = z(n)$$

We want P(Z>z(n)) < 1%. Solution is arrived at by substituting in values for n, starting at n=20 & reducing by 1 each time until probability is < 1%. For n=19, we get 0.9% which is less than 1%, so for airplane to be safe, there should be at most 19 passengers.