You may submit a formula sheet (1 sheet, 2 pages, hand-written), worth up to 5% extra.

Part I: You will be given 2 out of the following 3 types of problems. You must use a graphing calculator.

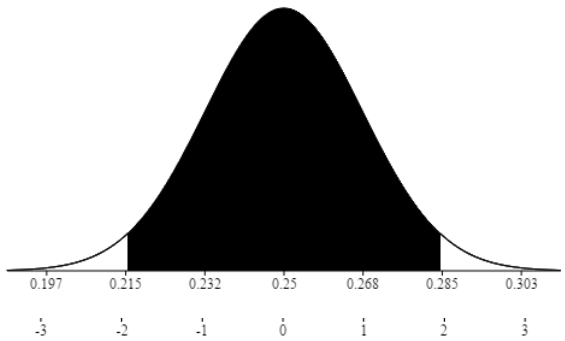
1. (20 pts) Recently married, a young couple plans to continue having children until they have their first girl. Suppose the probability that a child is a girl is 1/2 , the outcome of each birth is an independent event, and the birth at which the first girl appears has a geometric distribution. What is the couple's expected family size? Is the geometric pdf a reasonable model here? Discuss. (Geometric)
   E(Y)=1/p=2 The couple's expected family size is 2. This is basically the flip of a situation where a male heir is needed in a patriarchal society. Let's suppose that there is a matriarchal society. This might be a reasonable model. It has limitations, namely a woman can have a limited number of children within her lifetime, perhaps 20 or 30, assuming a single child is born each time. The world record is 69, but that was from a woman who had multiple children births many times. The model does not assume a limit on the number of births.
2. (20 pts) When a machine is improperly adjusted, it has probability 0.15 of producing a defective item. Each day, the machine is run until three defective items are produced. When this occurs, it is stopped and checked for adjustment. What is the probability that an improperly adjusted machine will produce five or more items before being stopped? What is the average number of items an improperly adjusted machine will produce before being stopped? (Negative binomial)
   The trick is to think complement. The machine must run at least 3 times so the complement of 5 or more is 3 or 4. P(Y>=5)=1-(P(Y=3)+P(Y=4))
   P(Y=3)=2nCr2(.15)^3= 0.003375
   P(Y=4)=3nCr2(.85)(.15)^3=0.00860625
   P(Y>=5)= 0.98801875~98.8%
   E(Y)=r/p=3/.15=20
3. (20 pts) The next generation of space shuttle will include two fuel pumps —one active, the other in reserve. If the primary pump malfunctions, the second is automatically brought on line. Suppose a typical mission is expected to require that fuel be pumped for at most 50 hours. According to the manufacturer's specifications, pumps are expected to fail once every 100 hours. What are the chances that such a fuel pump system would not remain functioning for the full 50 hours? (Gamma)

$$\frac{1}{100^2}\int_0^{50} ye^{-y/100}dy = \frac{1}{100^2}\left(-100ye^{-y/100}\Big|_0^{50} + 100\int_0^{50} e^{-y/100}dy\right)$$

$$= -\frac{e^{-1/2}}{2} - e^{-y/100}\Big|_0^{50} = 1 - \frac{3}{2}e^{-1/2} = 0.09020401\sim 9\%$$

Part II You must use Excel or R. No picture no credit for this section.

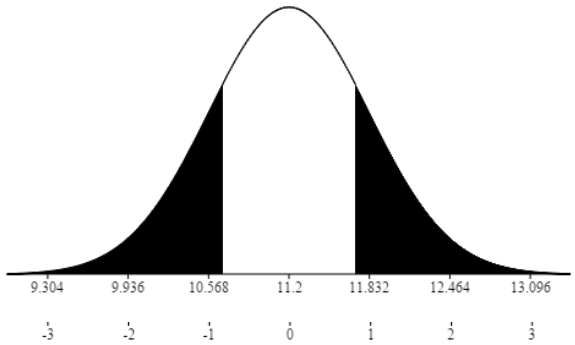
4. (20 pts) A recent poll (Tuesday 11/24/15):
   http://www.quinnipiac.edu/news-and-events/quinnipiac-university-poll/iowa/release-detail?ReleaseID=2305
   has Trump leading the Iowa Caucus polls 25% to Rubio's 23%. However, the lead is within the claimed margin of error of ± 4%. Focus just on Trump's numbers.
   a. Find the standard deviation for whether one voter is a supporter of Trump or not.
      This is a binomial distribution (coin flipping) with just one trial (flip).
      σ= √(pq)= √(.25*.75q)=.43
   b. Given that the sample size was 600, find the standard error (the standard deviation for the sample mean). **Std err=σ/√n=.43/√600=.0177** Draw the appropriate normal curve and label the horizontal axis with both Xbar and Z labels.



c. The margin of error is approximately ± twice the standard error. Find it. Is it roughly what the polltakers claim? **Margin of error = ±2(.0177)= ±3.5% which if we round up is the 4% provided by the polltakers.**

d. By what factor will the sample size need to be increased (assuming that results will stay the same) for us to conclude that Trump is definitely leading. The difference between the candidates is 2%. **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.**

5. (20 pts) A previous sample of fish in Lake Michigan indicated that the mean polychlorinated biphenyl (PCB) concentration per fish 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, and test the hypothesis that the mean PCB concentration has also remained unchanged at 11.2 parts per million. Use the 5 percent level of significance. NOTE: this is a 2-tailed test.
H_0 = 11.2, $H_1 \neq$ 11.2 Sample mean is 11.72. Std err=$\sigma/\sqrt{n}$=2/√10=.632
So the z value for the sample mean is z=(11.72-11.2)/.632=.822
and the pvalue=2*(1-norm.s.dist(.822,true))=.411~41%
This is way above 5%, so we do not reject H_0. We do not have enough evidence to say that the concentration has changed.



6. (20 pts) A farmer claims to be able to produce larger tomatoes. To test this claim, a tomato variety that has a mean diameter size of 8.2 centimeters with a standard deviation of 2.4 centimeters is used. If a sample of 36 tomatoes yielded a sample mean of 9.1 centimeters, does this prove that the mean size is indeed larger? Assume that the population standard deviation remains equal to 2.4, and use the 5 percent level of significance. NOTE: this is a 1-tailed test.
H_0 = 8.2, H_1 >8.2 Sample mean is 9.1. Std err=$\sigma/\sqrt{n}$=2.4/√36=.4
So the z value for the sample mean is z=(9.1-8.2)/.4=.9/.4=9/4=2.25
and the pvalue=1-norm.s.dist(2.25,true)=.0122~1.2%

This is below 5%, so we do reject H_0. We do have enough evidence to say that the farmer is producing larger tomatoes.