1. A very large number of cans of cat food are stored in a warehouse. The cans are manufactured at two plants: Plant A and Plant B. Past inspections have indicated that 30% of the cans manufactured in Plant A have melamine contamination, whereas only 2% of the cans manufactured in Plant B have melamine contamination. One-third of the cans come from Plant A and the rest from Plant B. A can is taken from the warehouse at random and tested for melamine. If the can tests positive for melamine, what is the probability that it came from Plant A?

   $P(A) = \frac{1}{3}$, $P(B) = \frac{2}{3}$, $P(M|A) = 0.3$, $P(M|B) = 0.02$ Use Bayes Theorem in some form, get the result $P(A|M) = \frac{15}{17} \approx 0.882$

2. A survey showed that half of all 1,000,000 adults in a city were tobacco smokers. 200,000 of the smokers had some form of lung cancer. 25,000 of the nonsmokers had some form of lung cancer. One adult person is selected at random from this city. Let T= "the person is a tobacco smoker" C= "the person has some form of lung cancer" Translate each of the following into symbols and then find the probabilities:

   a) What is the probability that the person is a tobacco smoker?
   $P(T) = 0.5$ (half of all the adults were tobacco smokers)
   b) What is the probability that the person is a nonsmoker who has lung cancer?
   $P(T' \cap C) = \frac{25,000}{1,000,000} = 0.025$
   c) What is the probability that the person has lung cancer, given that the person is a smoker?
   $P(C|T) = \frac{200,000}{500.000} = 0.4$
   d) What is the probability that the person has lung cancer?
   $P(C) = \frac{225,000}{1,000,000} = 0.225$
   e) If the person has lung cancer, what is the probability the person is a smoker?
   $P(T|C) = \frac{200,000}{225,000} = \frac{8}{9}$

3. Suppose that the drained weight of peaches in an 18-oz can of peaches is known to be normally distributed with standard deviation 0.75 oz.

   (a) What is the probability that a can of peaches, selected at random, has drained weight between 17.0 and 19.0 oz., if the true mean of the population is 18 oz.?
   $P(17.0 < X < 19.0) = \text{normalcdf}(17, 19, \mu = 18, \sigma = 0.75) \approx 0.8164$
   (b) What is the probability that a random sample of 25 cans of peaches, selected from this population, will have mean drained weight between 17.0 and 19.0 ounces?
   $P(17.0 < \bar{X} < 19.0) = \text{normalcdf}(17, 19\mu = 18, \sigma = \frac{0.75}{\sqrt{25}}) =$ very close to 1.00

4. In a random sample of 8 eggs, taken from a normally distributed population with $\sigma = 0.8$, the mean protein content was 5.2. If this mean is used to estimate the mean protein content of all the eggs, what can be said with 98% confidence about the size of the maximum error of the estimate?
   The error term is $Z_{0.01} \cdot \frac{0.8}{\sqrt{8}}$
   With 98% confidence the error is not more than 0.66.

5. If we suppose that our data comes from a normally distributed population with standard deviation 2.6, what sample size is needed to ensure that with probability 99%, the mean of the sample will be in error by at most 0.25?
   $n \geq \left(\frac{Z_{0.005} \cdot 2.6}{0.25}\right)^2$
   n must be at least 718.

6. A randomly chosen sample of NYC college students were asked if they take public transportation to get to classes. 175 out of 200 people in the sample indicated that they do take public transportation. Construct a 95% confidence interval for the true proportion of NYC college students who take public transportation to get to classes.
   0.875 ± 0.046, or we are 95% confident that the true proportion is between 0.829 and 0.920 [Use the confidence interval for a population proportion.]

7. The variance of the amount of drink contained in a random sample of 16 bottles of a certain sports drink was 2.5 ounces. Find a 95% confidence interval for this variance for all bottles of this sports drink, assuming that the amount of drink contained in the bottles is normally distributed.
   The confidence interval is $\dfrac{(16-1)(2.5)}{\chi^2_{0.025}} < \sigma^2 < \dfrac{(16-1)(2.5)}{\chi^2_{0.975}}$ with 16-1 = 15 degrees of freedom.
   With 95% confidence, $1.4 < \sigma^2 < 6.0$

8. Suppose that the average number of calls that come in to a certain switchboard each minute is 0.5.

   (a) What is the probability that the waiting time until the next call comes in will be more than 3 minutes?
   This is an exponential (waiting time for Poisson) distribution with $f(x) = \frac{1}{2}e^{-x/2}$, and the probability that X is greater than 3 is
   $\displaystyle\int_3^\infty \frac{1}{2}e^{-x/2}\mathrm{d}x = e^{-3/2} \approx 0.2231$

   (b) We make 36 random observations of the waiting times between calls at this switchboard.What is the probability that the mean waiting time for this sample will be more than 3 minutes?
   (Use CLT: The means are nearly normally distributed with $\mu_{\bar{X}} = \frac{1}{\lambda} = 2$ and $\sigma_{\bar{X}} = \frac{1}{\lambda\sqrt{36}} = \frac{1}{3}$)
   $P(\bar{X} > 3) = \text{normalcdf}(3, \infty, \mu = 2, \sigma = \frac{1}{3}) \approx 0.0013$

9. Suppose that the average number of calls that come in to a certain switchboard each minute is 0.5.

   (a) What is the probability that there will be one call in the next minute?
   This is Poisson with $\lambda = 0.5$;
   $p(1; 0.5) \approx 0.3033$,

   (b) We make 100 random observations of the number of calls in a minute at this switchboard.What is the probability that the mean number of calls per minute for this sample will be between 0.5 and 1.5?
   (Use CLT: the means are nearly normally distributed with: $\mu_{\bar{X}} = 0.5$ and $\sigma_{\bar{X}} = \frac{\sqrt{0.5}}{\sqrt{100}}$)
   $P(0.5 < \bar{X} < 1.5) = \text{normalcdf}(0.5, 1.5, 0.5, \frac{\sqrt{0.5}}{10}) \approx 0.5$

10. Let X be an exponentially distributed random variable with mean $\mu = 2$.

   (a) What is the probability that X is between 1 and 3?
   $f(x) = \frac{1}{2}e^{-x/2}$
   $P(1 < X < 3) = \displaystyle\int_1^3 \frac{1}{2}e^{-x/2}\mathrm{d}x = e^{-1/2} - e^{-3/2} \approx 0.3834$

   (b) We take a random sample of size 36 from X. What is the probability that the mean of that sample will be between 1 and 3?
   (Use CLT: the means are nearly normally distributed with: $\mu_{\bar{X}} = 2$ and $\sigma_{\bar{X}} = \frac{2}{\sqrt{36}} = \frac{1}{3}$)
   $P(1 < \bar{X} < 3) = \text{normalcdf}(1, 3, \mu = 2, \sigma = \frac{1}{3}) \approx 0.9973$

11. The city is considering the purchase of a different brand of lightbulbs for public buildings. The salesman claims that the new lightbulbs will last longer than the ones the city is currently using. Let μ represent the mean lifetime of the new lightbulbs, and suppose the mean lifetime of the bulbs currently being

used is 10,000 hours, so that we can take the null hypothesis $\mu = 10,000$. What alternative hypothesis would you take if the city is willing to purchase the new brand of lightbulbs unless they are actually worse than the old ones?
$\mathrm{H_A} : \mu < 10000$

12. Tar is a substance in tobacco smoke that can cause health problems. "Low-tar" cigarettes are supposed to contain no more than 6 mg of tar.
We wish to test a random sample of a certain brand of cigarettes which claim they are low-tar, to see if the claim is true.
   (a) What null and alternative hypotheses would we use in this case? Why?
   $\mathrm{H_0} : \mu = 6$
   $\mathrm{H_A} : \mu > 6$ because we are3 concerned that the level of tar may be too high.
   (b) State the decision rule we will use to test the claim.
   Reject the null hypothesis if $p < 0.05$ (see next question).
   (c) We take a random sample of 36 cigarettes from this brand, and find that the sample mean amount of tar per cigarette is 6.2 mg, and the sample standard deviation is 0.5 mg. Using the 0.05 level of significance, what conclusion do we reach regarding the claim that the cigarettes are low-tar? Explain.
   $p \approx 0.0082$: we reject the null hypothesis. The data support the conclusion that there is more than 6 mg of tar per cigarette on average

13. In the past a machine has produced washers having a mean thickness of 0.050 inches. To determine whether the machine is in proper working order, a sample of 10 washers is taken at random.
   (a) What null and alternative hypotheses would we use in this case? Why?
   $\mathrm{H_0} : \mu = 0.050$
   $\mathrm{H_A} : \mu \neq 0.050$ because we don't want the washers to be either too thin or too thick.
   (b) State the decision rule we will use to test the claim.
   Reject the null hypothesis if $p < 0.05$
   (c) We take a random sample of 10 of these washers, and find that the sample mean thickness is 0.053 inches, and the sample standard deviation is 0.003 mg. Test the hypothesis that the machine is in good working order at the 0.05 level of significance.
   $p \approx 0.0016$: reject the null hypothesis. The data support the conclusion that the machine is no longer in good working order.

14. We flip a coin 100 times to test the claim that the coin is balanced, against the alternate hypothesis that there is a 60% probability that heads will show when it is flipped. We will accept the claim if the number of heads that show in our 100 flips is less than 55, and reject it otherwise.
   (a) What is the probability that 55 or more heads will show, if the coin is really balanced?
   $P(X \geq 55) = (b(55; 100, 0.5) + ... + b(100; 100, 0.5) = 1-$ binomialcdf$(n = 100, p = 0.5, X = 54) \approx 0.1841$
   (b) What is the probability that fewer than 55 heads will show, if the probability that heads show is really 60%?
   $P(X < 55) = b(0; 100, 0.6) + ... + b(54; 100, 0.6) =$binomialcdf$(n = 100, p = 0.6, X = 54) \approx 0.1311$

15. We wish to test the claim that 3% of the lightbulbs produced at a certain factory are defective. We will reject the claim if the proportion of defective lightbulbs in a random sample of 100 lightbulbs is greater than 4%, and accept it otherwise. What is the probability that the sample will contain more than 4% defective lightbulbs, if the claim is true?
$P(X > 4) = b(5; 100, 0.03) + ... + b(100; 100, 0.03) = 1-$ binomialcdf$(n = 100, p = 0.03, X = 4) \approx 0.1821$

16. We count the number of calls which come in to a customer service center during a single one-hour period to test the claim that the average number of calls per hour is 5. We will reject the claim if the observed number of calls is greater than 7, and accept it otherwise. What is the probability that there will be more than 7 calls during that hour, if the claim is true?
$P(X > 7|\lambda = 5) = p(8; 5) + p(9; 5) + ... = 1-$ poissoncdf$(\lambda = 5, X = 7) \approx 0.1334$

17. A single observation of the time between calls which come in to a customer service center is made to test the claim that the average time between calls is 12 minutes. We will reject the claim if the observed time between calls is less than 9 minutes, and accept it otherwise. What is the probability that there will be less than 9 minutes between calls, if the claim is true?
This is an exponential density probability (waiting time for Poisson).
$P(X < 9) = \int_0^9 \frac{1}{12} e^{-x/12} \mathrm{d}x = 1 - e^{-3/4} \approx 0.5276$