1. A test for a disease correctly diagnoses a diseased person as having the disease with probability 0.85. The test incorrectly diagnoses someone without the disease as having the disease with a probability of 0.10. If 1% of the people in a population have the disease, what is the chance that a person from this population who tests positive for the disease actually has the disease?

$$\frac{.01(.85)}{.99(.1)+.01(.85)} = 0.079$$

2. A computer LCD screen contains 1 million (that is, 10^6) pixels, each of which has an independent probability p of being 'dead' due to manufacturing defects. LCD panels are considered accepted if they have at most 3 dead pixels.
   a) Suppose $p = 10^{-6}$. What is the percentage of LCD panels produced that are unacceptable?
   Hint: Use Poisson approximation.

$$\lambda = np = 10^6(10^{-6}) = 1$$

$$P(X>3) = 1 - P(X=0) - P(X=1) - P(X=2) - P(X=3) = 1 - e^{-1} - e^{-1} - \frac{e^{-1}}{2} - \frac{e^{-1}}{3!} = 0.019$$

   b) What is the largest p can be if the factory must produce on average at least 50% of panels with no dead pixels?

$$P(X=0) \geq .5 \quad \Rightarrow \quad e^{-10^6 p} \geq .5 \quad \Rightarrow -10^6 p \geq \ln(.5) \quad \Rightarrow \quad p \leq \frac{\ln(.05)}{-10^6} \leq 6.93\times 10^{-7}$$

3. In a particular year, LSAT scores were normally distributed with mean 520 and standard deviation 100.
   a) A selective law school will only admit applicants who scored in the top 9% on the LSAT. What score is required to be admitted?

$$invNorm(0.91, 520{,}100) = 654.08$$

   b) Five students were took the test are selected randomly. What is the probability that at least one student is in the top 9%?

$$1 - P(none\ in\ the\ top\ 9\%) = 1 - (.91)^5 = .376$$

4. Let X_1 and X_2 be independent random variable with the following probability density function.

$$f(x) = 2 - 2x, \qquad 0 < x < 1$$

   a) Find the probability that X_1 exceeds $1/2$.

$$P\left(X > \frac{1}{2}\right) = \int_{1/2}^{1} (2-2x)\,dx = [2x - x^2]_{1/2}^{1} = (2-1) - \left(1 - \frac{1}{4}\right) = \frac{1}{4}$$

   b) Find the probability that exactly one of the two variable exceeds $1/2$.

$$2\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{3}{8}$$

5. A manufacturing process produces widgets whose length is a random variable, with uniform distribution between 2.0 and 4.0. Let $\bar{X}$ be the average length of a batch of 400 widgets. Use central limit theorem to estimate the probability that $\bar{X}$ exceeds 3. **Note**: you may use the formula for the mean and variance of the uniform distribution derived in class.

$$\bar{X} = X_1 + \cdots + X_{400}, \qquad X_i \sim Uniform[2{,}4]$$

$$E(X_i) = \frac{2+4}{2} = 3\,, \qquad Var(X_i) = \frac{(4-2)^2}{12} = \frac{1}{3}$$

$$\bar{X} \sim N\left(3, \frac{1}{3(400)}\right) \quad \Rightarrow \quad normcdf\left(3, \infty, 3, \frac{1}{\sqrt{1200}}\right) = .5$$

Note: Some students notice a special case here: clearly the probability that a normal variable is greater than its mean is 0.5, since the normal distribution is symmetric around the mean. That observation was given full credit.

6. Suppose there are 4 cards in a box, two with number 0 and two with number 1 written on them. Two cards are randomly drawn without replacement. Let X number on the first card and Y be the number on the second card.
   a) Write the joint pdf as a 2-dimensional table of X and Y, and calculate the marginal pdf's.

$X\backslash Y$	0	1	
0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$
1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	

Blue row: marginal pdf for Y Yellow column: marginal pdf for X

b) Calculate the **covariance** of X and Y. Explain why its sign makes sense.

$$E(X)=\frac{1}{2}(0)+\frac{1}{2}(1)=\frac{1}{2}$$
$$E(Y)=\frac{1}{2}(0)+\frac{1}{2}(1)=\frac{1}{2}$$
$$E(XY)=1(1)\left(\frac{1}{6}\right)=\frac{1}{6}$$
$$Cov(X,Y)=E(XY)-E(X)E(Y)=\frac{1}{6}-\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)=-\frac{1}{12}$$

X and Y are negatively correlated. This makes sense because if the first card is a higher number, it is more likely to get the lower number in the next card.

7. Let Y be a continuos random variable with the pdf
$$f_Y(y;\theta)=\frac{1}{\theta+1}e^{-y/(\theta+1)},\qquad y>0$$
where $\theta>-1$ is an unknown parameter. Find the maximum likelihood estimate (**MLE**) for θ based on a random sample of size n: $y_1,\dots,y_n$.

$$L(\theta)=\prod\frac{1}{\theta+1}e^{-y_i/(\theta+1)}\quad\Rightarrow\quad \ln L(\theta)=-\sum\ln(\theta+1)-\frac{y_i}{\theta+1}=-n\ln(\theta+1)-\sum\frac{y_i}{\theta+1}$$
$$\frac{d\ln L(\theta)}{d\theta}=-\frac{n}{\theta+1}+\frac{1}{(\theta+1)^2}\sum_{i=1}^{n}y_i=0\quad\Rightarrow\quad n=\frac{1}{\theta+1}\sum_{i=1}^{n}y_i$$
$$\theta+1=\frac{1}{n}\sum_{i=1}^{n}y_i\quad\Rightarrow\quad\theta=\bar{y}-1$$

8. A manufacturer of automatic washers offers a model in two colors: white or black. Of the first 1000 washers sold, 567 are white.

a) Would you conclude that customers in general have a preference for a color? Test with a 99% confidence interval for the unknown proportion (of the population who prefer white).

From calculator: 1-PropZInt $\Rightarrow$ (0.52664, 0.60736)

Since 0.5 is NOT contained in the confidence interval, the customers seem to have a preference for white.

b) To reduce by half the margin of error of the estimate for the unknown proportion, what should be the sample size?

$$n=\frac{{z_{\alpha/2}}^2}{4d^2}$$

To reduce d to $d/2$, n should be 4 times as large, i.e. $n=4000$.

9. Two sets of elementary schoolchildren were taught to read by using different methods, 15 by each method. At the conclusion of the instructional period, a reading test yielded the results $\bar{y}_1=76$, $\bar{y}_2=71$, $s_1=9$, and $s_2=10$.

a) Find the **test statistic** and test at $\alpha=0.05$ to see if the two methods yield different results.

Since not both samples are large than 30, we use a 2-sample t-test with pooled variance. From the calculator, the test statistic is 1.4394, which is less than the t-value $t_{28,0.025}=2.0484$. So we fail to reject the null hypothesis, i.e. the two methods may yield the same result.

b) Find a 95% confidence interval for the difference of the means.

From calculator: 2-SampTint $\Rightarrow$ $(-2.116, 12.116)$

10. The output voltage for an electric circuit is specified to be 130. A sample of 35 independent readings on the voltage for this circuit gave a sample mean 129.3 and standard deviation of 2.1.

a) (2 pts) Explain why we may assume that the distribution of sample mean is normal.

Because the sample size is large (>35), and we may assume $\sigma=s$, and we may use CLT.

b) (3 pts) Find the lowest significance level at which the null hypothesis will be rejected.

Note: This problem should have specified a left-sided test.

This is the P-value of the data, i.e, 0.0243, from calculator Z-test.

If you assume a two-sided test, then the answer is $0.0243\times 2=.0486$

c) (5 pts) Suppose $\alpha=0.01$ and the true mean is 129. Find the Type II error β.

<u>Left-sided</u>: The critical value to reject H_0 is $InvNorm\left(.99, 130, \frac{2.1}{\sqrt{35}}\right)=129.17$

If the true mean is 129, then $\beta=P(\bar{X}>129.17)=normcdf\left(129.17, \infty, 129, \frac{2.1}{\sqrt{35}}\right)=.316$

<u>Two-sided</u>: The critical values (on two sides) to reject H_0 is $InvNorm\left(.005, 130, \frac{2.1}{\sqrt{35}}\right)=129.086$ and $InvNorm\left(.995, 130, \frac{2.1}{\sqrt{35}}\right)=130.194$. So

$$\beta=P(129.086\le\bar{X}\le 130.914)=normcdf\left(129.086, 130.914, 129, \frac{2.1}{\sqrt{35}}\right)=.4$$