Homework 2.

3.37 This is a binomial distribution with parameter $p = 0.02$ and $n = 50$.

$$\Pr\{\hat{p} \leq 0.04\} = \Pr\{x \leq 2\} = \sum_{x=0}^{2} \binom{50}{x} (0.02)^x (1-0.02)^{50-x}$$

$$= \binom{50}{0} (0.02)^0 (1-0.02)^{50} + \binom{50}{1} (0.02)^1 (1-0.02)^{49} + \binom{50}{2} (0.02)^2 (1-0.02)^{48} = 0.922$$

3.38 This is a binomial distribution with parameter $p = 0.01$ and $n = 100$.

$$\sigma = \sqrt{0.01(1-0.01)/100} = 0.0100$$

$$\Pr\{\hat{p} > k\sigma + p\} = 1 - \Pr\{\hat{p} \leq k\sigma + p\} = 1 - \Pr\{x \leq n(k\sigma + p)\}$$

Relevant binomial terms are:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$\binom{100}{x} 0.01^x 0.99^{100-x}$</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.36603</td>
<td>0.36603</td>
</tr>
<tr>
<td>1</td>
<td>0.36973</td>
<td>0.73576</td>
</tr>
<tr>
<td>2</td>
<td>0.18486</td>
<td>0.92063</td>
</tr>
<tr>
<td>3</td>
<td>0.06100</td>
<td>0.98163</td>
</tr>
<tr>
<td>4</td>
<td>0.01494</td>
<td>0.99657</td>
</tr>
</tbody>
</table>

$k = 1$

$$1 - \Pr\{x \leq n(k\sigma + p)\} = 1 - \Pr\{x \leq 100(1(0.0100) + 0.01)\} = 1 - \Pr\{x \leq 2\}$$

$$= 1 - [0.921] = 0.079$$

$k = 2$

$$1 - \Pr\{x \leq n(k\sigma + p)\} = 1 - \Pr\{x \leq 100(2(0.0100) + 0.01)\} = 1 - \Pr\{x \leq 3\}$$

$$= 1 - [0.982] = 0.018$$

$k = 3$

$$1 - \Pr\{x \leq n(k\sigma + p)\} = 1 - \Pr\{x \leq 100(3(0.0100) + 0.01)\} = 1 - \Pr\{x \leq 4\}$$

$$= 1 - [0.992] = 0.003$$

Continuation. The Poisson approximation uses $\lambda = np = 100*0.01 = 1$. So the probabilities are as follows (we already calculated those for the binomial):

<table>
<thead>
<tr>
<th>$x$</th>
<th>Binomial</th>
<th>Poisson</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.36603</td>
<td>0.36788</td>
</tr>
<tr>
<td>1</td>
<td>0.36973</td>
<td>0.36788</td>
</tr>
<tr>
<td>2</td>
<td>0.18486</td>
<td>0.18394</td>
</tr>
<tr>
<td>3</td>
<td>0.06100</td>
<td>0.061313</td>
</tr>
</tbody>
</table>

2.36 This is a hypergeometric distribution with $N = 30$, $n = 5$, and $D = 3$. 
\[
\Pr\{x = 1\} = p(1) = \binom{3}{1} \binom{30 - 3}{5 - 1} \binom{30}{5} = \frac{(3)(17,550)}{(142,506)} = 0.369
\]

\[
\Pr\{x \geq 1\} = 1 - \Pr\{x = 0\} = 1 - p(0) = 1 - \binom{3}{0} \binom{27}{5} \binom{30}{5} = 1 - 0.567 = 0.433
\]

If we sampled with replacement, we’d have a binomial distribution with \(n=5, \ p=0.1\), \(\Pr\{x=0\} = 0.95 = 0.59049\), \(\Pr\{x=1\} = 0.32805\), \(\Pr\{x \geq 1\} = 1-0.59049 = 0.40951\).

If the sampling fraction \(n/N\) were small (which it is not), then we could use the binomial as an approximation to the hypergeometric. The not-great agreement between the hypergeometric and the binomial probabilities illustrates the point that we don’t want to be using the approximation here, with a sampling fraction of 1/6.

2.35 This is a hypergeometric distribution with \(N = 25\) and \(n = 5\), without replacement.

(a) Given \(D = 2\) and \(x = 0\):

\[
\Pr\{\text{Acceptance}\} = p(0) = \frac{\binom{2}{0} \binom{25 - 2}{5 - 0}}{\binom{25}{5}} = \frac{(1)(33,649)}{(53,130)} = 0.633
\]

(b) For the binomial approximation to the hypergeometric, \(p = D/N = 2/25 = 0.08\) and \(n = 5\).

\[
\Pr\{\text{acceptance}\} = p(0) = \binom{5}{0} (0.08)^0 (1 - 0.08)^5 = 0.659
\]

This approximation, though close to the exact solution for \(x = 0\), violates the rule-of-thumb that \(n/N = 5/25 = 0.20\) be less than the suggested 0.1. The binomial approximation is not satisfactory in this case.

(c) For \(N = 150\), \(n/N = 5/150 = 0.033 \leq 0.1\), so the binomial approximation would be a satisfactory approximation the hypergeometric in this case.

(d) Find \(n\) to satisfy \(\Pr\{x \geq 1 \mid D \geq 5\} \geq 0.95\), or equivalently \(\Pr\{x = 0 \mid D = 5\} < 0.05\).

\[
p(0) = \frac{\binom{5}{0} \binom{25 - 5}{n - 0}}{\binom{25}{n}} = \frac{\binom{5}{0} \binom{20}{n}}{\binom{25}{n}}
\]
try $n = 10$

$$p(0) = \binom{5}{0} \frac{20}{10} = \frac{(1)(184,756)}{(3,268,760)} = 0.057$$

Let sample size $n = 11$.

$$p(0) = \binom{5}{0} \frac{20}{11} = \frac{(1)(167,960)}{(4,457,400)} = 0.038$$

2.40 This is a Poisson distribution with a true mean of $\lambda = 0.01$ errors/bill.

$$\Pr\{x = 1\} = p(1) = \frac{e^{-0.01}(0.01)^1}{1} = 0.0099$$

2.42 This is a Pascal distribution with $\Pr\{\text{defective weld}\} = p = 0.01$, $r = 3$ welds, and $x = 1 + (5000/100) = 51$.

$$\Pr\{x = 51\} = p(51) = \binom{51-1}{3-1}(0.01)^3(1-0.01)^{51-3} = \binom{1225}{0.000001}(0.617290) = 0.0008$$

$$\Pr\{x > 51\} = \Pr\{r = 0\} + \Pr\{r = 1\} + \Pr\{r = 2\}$$

$$= \binom{50}{0}0.01^00.99^{50} + \binom{50}{1}0.01^10.99^{49} + \binom{50}{2}0.01^20.99^{48} = 0.9862$$

**Probability plotting exercise:**

The data series and auxiliary columns are:

<table>
<thead>
<tr>
<th>X</th>
<th>LNX</th>
<th>4ROOT</th>
<th>FX</th>
<th>YPLOT</th>
<th>PHINV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
<td>1.000</td>
<td>0.050</td>
<td>-2.970</td>
<td>-1.645</td>
</tr>
<tr>
<td>4</td>
<td>1.386</td>
<td>1.414</td>
<td>0.150</td>
<td>-1.817</td>
<td>-1.036</td>
</tr>
<tr>
<td>9</td>
<td>2.197</td>
<td>1.732</td>
<td>0.250</td>
<td>-1.246</td>
<td>-0.674</td>
</tr>
<tr>
<td>12</td>
<td>2.485</td>
<td>1.861</td>
<td>0.350</td>
<td>-0.842</td>
<td>-0.385</td>
</tr>
<tr>
<td>19</td>
<td>2.944</td>
<td>2.088</td>
<td>0.450</td>
<td>-0.514</td>
<td>-0.126</td>
</tr>
<tr>
<td>30</td>
<td>3.401</td>
<td>2.340</td>
<td>0.550</td>
<td>-0.225</td>
<td>0.126</td>
</tr>
<tr>
<td>34</td>
<td>3.526</td>
<td>2.415</td>
<td>0.650</td>
<td>0.049</td>
<td>0.385</td>
</tr>
<tr>
<td>45</td>
<td>3.807</td>
<td>2.590</td>
<td>0.750</td>
<td>0.327</td>
<td>0.674</td>
</tr>
<tr>
<td>70</td>
<td>4.248</td>
<td>2.893</td>
<td>0.850</td>
<td>0.640</td>
<td>1.036</td>
</tr>
<tr>
<td>99</td>
<td>4.595</td>
<td>3.154</td>
<td>0.950</td>
<td>1.097</td>
<td>1.645</td>
</tr>
</tbody>
</table>

Here are the plots. I did them using software; students should use hand calculators to make sure they know the details of the plotting.
Highly curved. X is clearly not normal

Weibull plot. This looks pretty linear, indicating that the Weibull is a good fit. The line shown is of 
log(-log(1-F(x))) = -3.04 + 0.872 log(x), so the estimates of the Weibull parameters are
\( \hat{\beta} = 0.872 \) and \( -\hat{\beta} \log(\hat{\theta}) = -3.045 \), so \( \hat{\theta} = 32.9 \). Recall that \( \theta \) is the value that has a fraction \( e^{-1} \approx 0.37 \) above it; the actual frequency above this estimate is 4 out of the sample size of 10, a frequency in good agreement.
This plot looks pretty linear. This suggests that if you have Weibull data, transforming to some power (in this case the 0.25 power) might give you a measure that is ‘normal enough’ for many practical purposes.