Lecture #10

Prof. John W. Sutherland

Sept. 19, 2005
Second Example

\[ P \left( 75 \leq X \leq 115 \right) = ? \]

\[ \sigma_X = 10 \]

How many std. devs. from the mean?

75??

115??
Example #2 Continued

<table>
<thead>
<tr>
<th># of Std. Devs., $z$</th>
<th>Cum. Prob. - area under curve from $-\infty$ to $z$, $F(z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>0.00135</td>
</tr>
<tr>
<td>-2</td>
<td>0.0228</td>
</tr>
<tr>
<td>-1</td>
<td>0.1587</td>
</tr>
<tr>
<td>0</td>
<td>0.50</td>
</tr>
<tr>
<td>1</td>
<td>0.8413</td>
</tr>
<tr>
<td>2</td>
<td>0.9772</td>
</tr>
<tr>
<td>3</td>
<td>0.99865</td>
</tr>
</tbody>
</table>

Table A.1 lists $F(z)$ for various std. dev., $z$

$F(-2.5)=0.0062$  \hspace{2cm}  $F(1.5)=0.9332$
Example # 2 Summary

\[ P (-2.5 \leq Z \leq 1.5) = P (Z \leq 1.5) - P (Z \leq -2.5) \]

\[ = 0.9332 - 0.0062 = 0.927 \]
How are the $\bar{X}$’s distributed?

- Central tendency
- Spread
- Shape - distribution of sample means
Distribution of Sample Means

\[ \mu_x \]

\[ \sigma_x \]

\[ \bar{x} \]
Distribution of Sample Means

\[ \mu_{\bar{X}} = E[\bar{X}] \]

\[ \sigma_{\bar{X}}^2 = Var[\bar{X}] \]
Central Limit Theorem

Averages (in fact, any linear combination of data) tend to be normally distributed regardless of the distribution of $X$. Tendency towards normality improves as $n$ increases. If $X$’s are normal, averages are also normal.
Example
(2000 throws of a die)
Example
(Throw 2 dice -- find avg )
Example
(Throw 5 dice 2000 times -- find avg )
Shape of the Distribution of X’s

Process $\rightarrow$ X’s $\rightarrow$ \( \overline{X} \) estimates, \( \mu_x \)

\( \frac{R}{d_2} \) estimates, \( \sigma_x \)

\( s_x \) estimates, \( \sigma_x \)

when \( n=5, \ d_2 = 2.326 \)

n=10

Can’t judge distribution shape - not enough data
Here’s Some Data

44, 62, 39, 53, 80, 33, 57, 22, 49, 68

What distribution describes this data?

Are the data normally distributed??
Cumulative Histogram (Plot)

Rank X’s in Ascending Order

<table>
<thead>
<tr>
<th>i</th>
<th>Ranked X’s</th>
<th>Cum. Prob., P_i</th>
</tr>
</thead>
</table>

Smallest X must represent lowest 10% of underlying distribution

P_i = (i-.5)/n
Cumulative Histogram

Which "S" shaped cdf best matches the cumulative probability plot of the data?
Normal Probability Paper

Hard to make the comparison since comparing S shapes

Cumulative Prob. Plot of Data

Let's create a special vertical scale for the graph so that the Normal cdf appears as a straight line.
Normal Probability Paper