Today: Probability models for means (measurement error) Read: DD ch. 11

Monday: Statistical models for means FPP ch 13, 24

At the last minute Ken was unable to give the Thu 4pm section yesterday. We apologize for any inconvenience; he will cover the section material in his office hours this week.

Extra office hours today: 2-4pm

Relationship between SRS (sampling at random without replacement) & IID (at random with replacement):

1) SRS is more informative than IID, because there's no point in sampling the same element of the pop more than once.
2) When \( n = 1 \), SRS = IID
3) When \( n \ll N \) (\( n \) is a lot smaller than \( N \)), SRS = IID (is about the same).
4) The math is easier with IID, but SRS is what usually occurs in the real world.
5) When \( n \ll N \), the sample may be SRS but we will use formulas than IID.

<table>
<thead>
<tr>
<th>Possible results on a single spin</th>
<th>Sample observed 3 spins</th>
<th>Imaginary Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 36</td>
<td></td>
<td>N = 1000</td>
</tr>
<tr>
<td>$$1$</td>
<td>$$6$</td>
<td>$$46$</td>
</tr>
<tr>
<td>$$10$</td>
<td>$$12$</td>
<td>$$8$</td>
</tr>
</tbody>
</table>

Mean (\( \mu \)) = $0.852$  
SD(\( \sigma \)) = $9.02$

Long Run Mean: EV of \( S \)  
\[ \mu = \text{E}_{\text{IID}}(S) = -$5.82$ \]

Long Run SD = \( SE \) of \( S \)  
\[ SE = $1.27 \]

\( \sigma = $5.76 \)  
Pop Hist for split

\( \sigma = $5.76 \)  
Pop Hist

Long Run Hist:  
\( SE \approx $1.27 \)  
CLT
\[
\mu = \frac{36}{36} + \frac{36}{36} + \frac{2(-0.05)}{36} = \frac{3.2}{36} = -\frac{0.082}{36} \text{ (same as single #)}
\]

\[
\sigma = \sqrt{\frac{36}{36} \times (-0.05)^2 + \frac{36}{36} \times (-0.05)^2 + 2 \times 17 \times (-0.05)^2 = \frac{2.462}{36}}
\]

\[
E_{\text{IID}}(S) = \mu_n = \frac{4(0.000)}{36} = -\frac{0.082}{36} \text{ (same as single #)}
\]

\[
SE_{\text{IID}}(S) = \frac{4(0.000)}{\sqrt{36}} = -\frac{0.082}{36} \text{ (same as single #)}
\]

So: Single # has higher chance of coming out ahead than split
(also than any other gamble on roulette: "bold play is optimal")
but it also has higher chance of losing $100 or more.

Look at Case Study 9

Weight of "1 lb" of butter:

<table>
<thead>
<tr>
<th>To nearest oz.</th>
<th>1/16 oz</th>
<th>1/100 oz</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>16.0</td>
<td>16.0</td>
</tr>
<tr>
<td>16</td>
<td>15.8</td>
<td>15.8</td>
</tr>
<tr>
<td>16</td>
<td>15.6</td>
<td>15.6</td>
</tr>
<tr>
<td>16</td>
<td>15.4</td>
<td>15.4</td>
</tr>
</tbody>
</table>

Any measurement, no matter how carefully made, could come out differently
the next time. Why? If thing being measured doesn't change, then
it's because of "measurement error"
Basic Measurement Error Model: \[ \text{Mean} = 0 \]
\[ \text{Random SD } \sigma = 0.2 \]
\[ \text{(Each Individual Measurement)} = (\text{exact value}) + (\text{bias}) + ("error") \]
\[ 3.7 = \text{Potassium on Measurement 1} = 3.8 + 0 + 0.1 \]
\[ 4.0 = \text{Potassium on Measurement 2} = 3.8 + 0 + 0.2 \]

### Population
- **Conceptual:** All Possible Potassium Readings on my blood
- **Real:** The Observed Measurements

\[ N = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad \text{Like } \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \]

- Mean \( \mu \approx 3.8 \)
- SD \( \sigma = 0.2 \)

### Imaginary Data

- Pop. Histogram

\[ \Pr(\text{Misdiagnosis w/ a single obs.}) \approx \]
\[ \Phi(0.7) \approx 0.76 \]

- 6.5%