Stats Wednesday

Probability models for sums & means, measurement error
Midterm due next Wed (10th of May)

C.S. #8 (Pg. 88)

Possible outcome on any $1 bet

My net gain

\[ \begin{array}{c|c|c|c|c|c}
\text{Possible Outcome} & \text{Probability} \\
\hline
-8 & 0.08 \\
-1 & 0.33 \\
37 & 0.59 \\
\hline
\end{array} \]

\[ N = 1000 \]

Single #

Sample

Observed spins

\[ Y_i \]

\[ Y_n \]

I.I.D. (Independent and Identically Distributed)

Sum \( S = ? \)

(ex. = \$43)

Imaginary data set of possible values

\[ \begin{bmatrix}
-28 \\
+43 \\
\vdots \\
\end{bmatrix} \]

Note: \( S = \text{my net gain after 1000 spins} \)

Long run mean (Expected value \( E(S) \)) = \( EV \) of \( S = \) (# draws) (# mean)

\[ = (1000) (-\$0.052) \approx -\$52 \]

\( \approx \) after 1000 spins

Long run hist. of \( S \)

CLT (Central Limit Theorem)

\[ \sigma = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} \]

\( \bar{y} \approx \mu \)

Long run s.d. (standard error of \( S \)) = \( SE(S) = \frac{\sigma}{\sqrt{n}} \)

\[ = \left( \frac{\text{Pop SD}}{\# \text{draws}} \right) \]

By the mean!
Jordan Ruybal

Long Run SD

\[ \text{SE}_{\text{IID}}(\bar{X}) = \sqrt{\frac{1}{n}} = (15.76) \sqrt{1000} = 182 \]

- After 1000 draws I expect to be behind by about $52 (\text{EV}(\bar{X}))$ give or take about $182 (\text{SE} (\bar{X}))$.

Adding more "noise" makes SD ↑

will follow the normal curve pretty well, provided

\( \text{N} \) is fairly large; (2) If pop. hist. follows normal curve to begin with, then already with \( \text{N}=1 \) the hist. of \( S \) will follow normal curve; (3) the closer pop. hist. is to the normal curve to begin with, the smaller \( \text{SE} \) can be. We still get a nice normal curve for hist. of \( S \).

Examples of CLT (pg. 84)

\[ \text{SE} = 182 \]

hist. of \( S \) (single #)

\[ \frac{-52}{182} \]

\[ \frac{0}{182} \]

\[ \frac{52}{182} \]

\[ \text{SU} \]

\[ \text{CLT} \]

\[ \text{N}=1000 \]

\[ 38\% \]

= chance of coming out ahead w/1000 & 1 bet on single #

\[ \frac{0 - (-52)}{182} = \frac{52}{182} \]

\[ \frac{0 - (-52)}{182} = 0.29 \approx 0.3 \]

Split pop

\[ \text{hist.} \]

\[ \frac{3}{8} \]

\[ \frac{2}{3} \]

\[ \frac{1}{2} \]
\[
\frac{3/5}{20} \quad M = \frac{-1 + \ldots + (-1) + 2 + 17}{38} = \frac{-2}{38} = -0.052
\]

Expect to lose about the same amount of money on every bet made.

Stats Friday

Probability measurements for means (measurement error)

Relationship between SRS (Sampling at random w/ or w/o replace) & IID (Sampling w/ replace):

1. SRS is more informative than IID, b/c there's no point in sampling the same element of the pop. more than once.
2. When \( n = 1 \) then SRS = IID.
3. When \( n \ll N \) (\( n \) is a lot smaller than \( N \)) SRS \( \approx \) IID.
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 from IID.

C.S. #8 For split bet

For spin results

\( N = 38 \)

\[
M (\text{mean}) = -0.052
\]

\[
\sigma (\text{SD}) = 4.02
\]

Look at graphs

\( -1 + 11 \quad -1 + 35 \)

* Makes sense that SD smaller for split