Chapter 7: Central Limit Theorem Discovery Lab

Rolling a Single Die

4. Set up PLOT1 to do a histogram with xList = L1 and Freq = L2
5. Enter the faces (1,2,3,4,5,6) in L1 and the Tally in L2
6. Go into GRAPH. Check y(x)= and clear out any functions. Check to make sure that only PLOT1 is on.
7. Set up your WINDOW:

   xMin = 0.75
   xMax = 6.75
   xScl = 1
   yMin = 0
   yMax = your instructor will give you this number. = 10
   yScl = 50

8. Graph your histogram. Use a ruler to sketch the histogram below. Label and scale your axes.

   *In gen, with enough rolls, we get a uniform (even) distr.

9. Calculate the mean $\overline{x}$ and standard deviation s_x using One Variable Statistics on your calculator. Round your values to 2 decimal places.

   $\overline{x}$ =_______ s_x =_______

Average of 2 Dice

1. Clear out your data in L1 and L2.
2. Enter the Average of the Faces (1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6) into L1 and the Tally for 2 dice into L2.
3. Press GRAPH to get the histogram. Sketch what you see using a ruler.

4. Calculate the mean $\overline{x}$ and standard deviation s_x using One Variable Statistics on your calculator. Round your values to 2 decimal places.

   $\overline{x}$ =_______ s_x =_______

Average of 5 dice

1. Clear out your data in L2 only.
2. Enter the Tally for 5 dice into L2.
3. Graph a histogram by pressing GRAPH. Sketch what you see.

4. Calculate One Variable Statistics on your calculator. Round values to 2 decimal places.
   $\bar{x}$ =________ s_x =________

Summary Questions

1. Describe the shapes of the three histograms. List 3 things that you notice about the changes in the histograms for 2 dice and 5 dice as compared to the histogram for a single die.
   With 1 die we had distributions all over the place, as we averaged more dice, we started getting a normal distribution with smaller variance
2. Compare the sample means $\bar{x}$ for each part. What do you notice about the approximate values for the sample means?
   $\bar{x} \approx 3.5$
3. Compare the standard deviations s_x for each part. What do you notice about the values of the standard deviations as we average more dice?
   S_x gets smaller the more dice we avg.
4. Suppose I average n dice. Complete the following:

- The shape of the histogram of averages is like a bell curve.
- As I average more dice, the histogram becomes more sym., more bell-like, more narrow, more centered
- The means are approximately 3.5 no matter how many dice I average.
- As I average more dice, the standard deviation gets closer to 0

Chapter 7 **The Central Limit Theorem**

The Central Limit theorem for means or averages.

(a) Start with any data set with either a known or unknown distribution.

- μ = mean
- σ = std. dev.

(b) Begin to choose samples of size n, **with replacement** , with n "big enough". The size of n depends on how normal the original data set is to begin with. But once n is chosen, it remains fixed.

(c) Find the mean of each sample.

In our Discovery we made histograms of the means of the samples w/ $n=2$ and $n=5$

(d) The set of sample means will always form

a bell curve

Important: The orig. distribution of the population does not matter. We don't even need to know what it is. We just need to know the pop. mean μ & std. dev. σ

The Central Limit Theorem for Means (formally written)

(a) X = a random variable w/ <u>any</u> distribution (known or unknown)

- μ_x = mean of X
- σ_x = population std. dev. for X

(b) Now, suppose we begin to draw samples of size n.

(c) We can define a new random variable:

$\overline{X}$ = sample means

(d) Then as ~~n~~ n increases, the random variable $\overline{X}$ tends toward a <u>normal distribution</u> and

$$\overline{X} \sim N(\mu_{\overline{X}}, \sigma_{\overline{X}})$$

$$\boxed{\overline{X} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right)}$$

(e) The values of $\overline{X}$ have associated z-scores. The formula would be

$$\boxed{z = \frac{\overline{x} - \mu_x}{\left(\sigma_x / \sqrt{n}\right)}}$$

Ch. 7 CLT

Example: The length of human pregnancies is approximately normally distributed with a mean of 266 days and a standard deviation of 16 days.

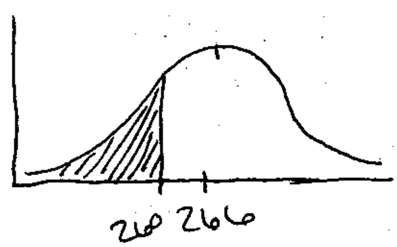
a. Define the random variable in words and write the distribution to use.

X = length of a pregnancy in days

b. What is the probability that one randomly selected pregnancy lasts less than 260 days?

*Ch. 6

$P(X < 260)$ = normalcdf (-1E99, 260, 266, 16)
= 0.3538



$n=20$

Suppose that a random sample of 20 pregnancies is obtained and the average calculated.

c. Define the random variable in words and write the distribution to use.

*Ch7

$\bar{X}$ = avg. length of the 20 pregnancies

$\bar{X} \sim N\left(266, \frac{16}{\sqrt{20}}\right) = N(266, 3.5777)$

d. What is the probability that the average of 20 pregnancies is less than 260 days?

$P(\bar{X} < 260)$

= normalcdf(-1E99, 260, 266, 3.5777)

= 0.0468

Example: According to the IRS, the average length of time for an individual to complete IRS form 1040 is 10.53 hours with standard deviation 2 hours. The distribution, however, is unknown. Suppose we randomly sample 36 taxpayers.

(a) In words, define X = length of time for 1 individual to complete Form 1040

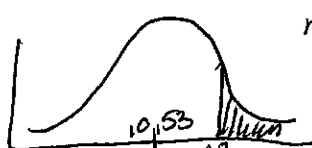
(b) In words, define $\overline{X}$ = avg. time for the 36 people to complete the form

(c) $\overline{X} \sim N\left(10.53, \frac{2}{\sqrt{36}}\right) = N\left(10.53, \frac{1}{3}\right)$

(d) Find the probability that the average time to complete form 1040 for the sample of 36 was more than 12 hours.

$P(\bar{x} > 12) =$ normalcdf(12, 1E99, 10.53, $\frac{1}{3}$)

$= 0.000005$



(e) Would you be surprised if a <u>single taxpayer</u> finished his or her 1040 in more than 12 hours? Explain why or why not using z-scores.

X

$$z_x = \frac{x-\mu}{\sigma_x} = \frac{12-10.53}{2} = 0.735$$

So 12 is less than 1 std. dev. from the mean. So we shouldn't be surprised

By comparison $z_{\bar{x}} = \frac{12-10.53}{(\sigma_x/\sqrt{n})} = \frac{12-10.53}{(\frac{1}{3})} = 4.41$

So this an outlier

Chapter 7 Practice 2

The attention span of a typical two year old is exponentially distributed with a mean of 8 minutes. Suppose we randomly survey 60 two year olds.

(a) In words, define X = ______________________________

(b) ~~Mn~~ $\mu_x = 8$ + $\sigma_x = 8$

(c) In words, define $\overline{X}$ = ______________________________

(d) $\overline{X} \sim$ ____________________

(e) Before doing any calculations, which do you think is higher?

a. The probability that the attention span of one individually chosen two year old is less than 10 minutes.
b. The probability that average attention span for the group of 60 two year olds is less than 10 minutes.

WHY?

(f) Calculate P(X < 10) = 0.7135

(g) Calculate P($\overline{X}$ < 10)

(h) Was your guess to part (e) correct or not?

(i) Explain why the distribution for $\overline{X}$ is not exponential even though the distribution for X is.