

THE EXPECTED VALUE FOR THE SUM OF THE DRAWS

• In the game of Keno there are 80 balls, numbered 1 through 80. On each play, the casino chooses 20 balls at random without replacement. Suppose you bet on the number 17, i.e., that ball number 17 will be among the 20 that are chosen. The chance that you win equals _____.

• Consider the following equivalent game. Take a deck of 80 cards, numbered 1 through 80. Shuffle the deck and deal out the top 20 cards. You win if card number 17 is among those dealt out.

• Shuffling the deck brings the cards into random order, so card number 17 is equally likely to end up at any one of the 80 possible positions.

• The chance that card 17 is one of the top 20 cards is _____ in _____. So the chance of your winning is _____.

• Suppose you're betting a dollar on 17 on each play at Keno. When you win, the casino gives you your dollar back and two dollars more. When you lose, the casino keeps your dollar. Your net gain for 40 plays is like the sum of _____ draws made at random with replacement from the box

--

• In particular, the amount you should expect to win (or lose) at the casino equals the expected value for the sum of the draws from the box.

• Suppose 40 draws are to be made (at random with replacement) from the box

\$2	-\$1	-\$1	-\$1
-----	------	------	------

You'd expect the sum of the draws to be about _____.

• You'd expect to draw

\$2

 around _____ times.

• You'd expect to draw

-\$1

 around _____ times.

• You'd expect the sum of the draws to be around
 $\$2 \times \underline{\hspace{2cm}} + (-\$1) \times \underline{\hspace{2cm}} = \underline{\hspace{2cm}}$.

• Is there an easier way to reach the same conclusion?

• Yes. It's based on the fact that the average of the numbers in the box is

$$\frac{\$2 - \$1 - \$1 - \$1}{4} = \underline{\hspace{2cm}} .$$

• On the average, you lose a quarter on each draw. In 40 draws you can expect to lose around $40 \times \$0.25 = \10 .

• Why do the two methods give the same conclusion?

• Because of a mathematical identity:

$$\begin{aligned} & \$2 \times \left(\frac{1}{4} \times 40\right) + (-\$1) \times \left(\frac{3}{4} \times 40\right) \quad (\text{first method}) \\ &= 40 \times \left(\$2 \times \frac{1}{4} + (-\$1) \times \frac{3}{4}\right) \\ &= 40 \times \left(\frac{\$2 - \$1 - \$1 - \$1}{4}\right) \quad (\text{second method}) \end{aligned}$$

- In general, what's the formula for the expected value for the sum of draws made at random with replacement from a box?

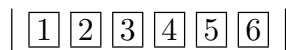
- The expected value for the sum of the draws equals

$$(\text{number of draws}) \times (\text{average of box})$$

- “average of box” is short for the average of the numbers in the box.

- What is the expected number of squares moved on a play in Monopoly?

- This is the same as the expected value for the sum of 2 draws from the box



- The number of draws equals _____ .

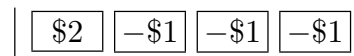
- The average of the box equals

$$\text{_____} = \text{_____} .$$

- The answer is _____ \times _____ = _____ squares.

THE STANDARD ERROR FOR THE SUM OF THE DRAWS

- Let's go back to drawing at random from the box



Suppose we plan to make 40 draws. The sum of the draws will be off from its expected value of $-\$10$ by some chance amount:

$$\text{sum of the draws} = \text{expected value} + \text{chance error}.$$

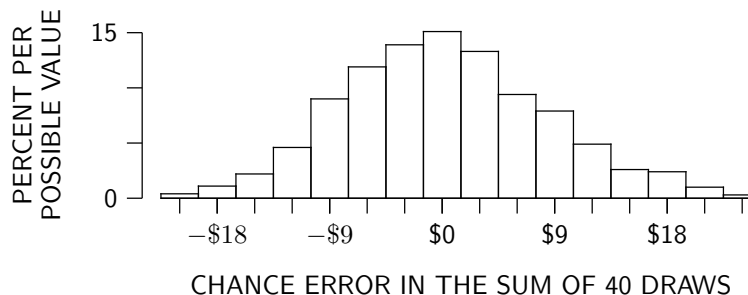
How big is the chance error likely to be?

- One way to find out is to actually make the draws and see what happens. I programmed the computer to draw 40 times at random with replacement from the box; on it's first try the computer got $\$2$ 8 times and $-\$1$ 32 times.

- The sum of these 40 draws is _____ .

- The corresponding chance error is _____ .

- Then I had the computer make another 999 repetitions. Altogether the computer got 1000 chance errors: $-\$6$, $\$9$, $\$6$, $\$9$, $-\$9$, $-\$12$, $\$15$, $-\$3$, $-\$3$, $-\$15$, ... which are conveniently exhibited by a _____:



- The likely size of a chance error is about _____. (In fact, the root-mean-square of the chance errors is _____.)

- To summarize, the sum of 40 draws from the box

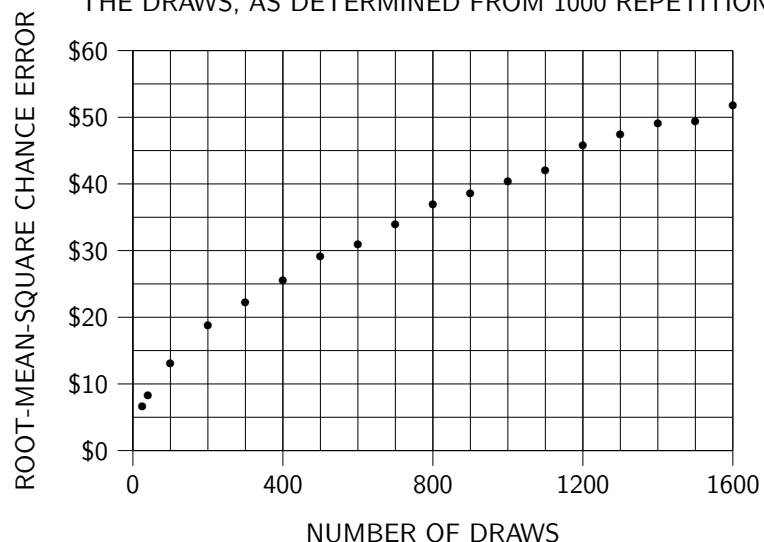
$$\boxed{\$2} \quad \boxed{-\$1} \quad \boxed{-\$1} \quad \boxed{-\$1}$$

will equal _____, give or take _____ or so.

- Still thinking about this box, how might we find out how the likely size of the chance error in the sum of the draws depends on the number of draws?

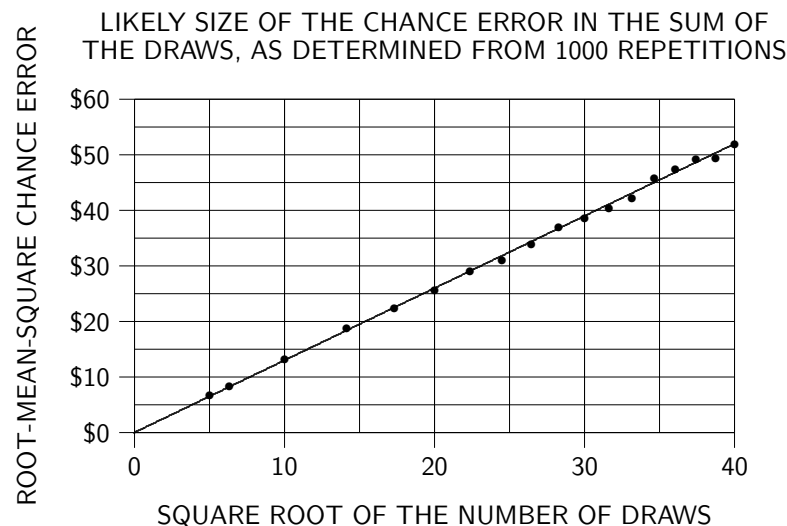
- By taking the program used to study the chance error in 40 draws, and running it for various numbers of draws. Here are the results of such an empirical investigation:

LIKELY SIZE OF THE CHANCE ERROR IN THE SUM OF THE DRAWS, AS DETERMINED FROM 1000 REPETITIONS



- As the number of draws increases, the likely size of the chance error gets _____, but ever more slowly.
 - Quadrupling the number of draws appears to _____ the chance error.

- Changing the scale on the horizontal axis makes the pattern clearer:



- Conclusion: the likely size of the chance error in the sum of the draws appears to be _____ to the _____ of the number of draws.
- The straight line in the plot rises from \$0 to _____ as the square root of the number of draws runs from 0 to 40, and so has a slope of _____. How is the slope related to the box?
 - It's the SD of the numbers 2, -1, -1, -1 in the box:
 - Those numbers average to _____.
 - The deviations from average are $\frac{9}{4}$, $\frac{-3}{4}$, $\frac{-3}{4}$, and $\frac{-3}{4}$.
 - The r.m.s. deviation is $\sqrt{\left(\left(\frac{9}{4}\right)^2 + \left(\frac{-3}{4}\right)^2 + \left(\frac{-3}{4}\right)^2 + \left(\frac{-3}{4}\right)^2\right)/4} \approx 1.30$

- For the “Keno” box, our empirical investigation has shown that the likely size of the chance error in the sum of the draws is given by

$$\sqrt{\hspace{10em}} \times (\hspace{10em}).$$

- In particular, the formula says that the likely size of the chance error for a single draw is just the SD of the box. Here’s the intuition for that.

- On a single draw, you’d expect to get the average of the box.
- So the chance error in a single draw is the difference between the number drawn and the average of the box.
- The likely size of such a chance error is the amount by which the tickets in the box typically deviate from their average.
- That’s just the SD of the box.

- I just argued that the variability in a single draw is the same as the variability of the tickets in the box. It’s intuitively clear that the sum of n draws should be more variable than a single draw. Why is there only \sqrt{n} times as much variability, and not n times as much variability?

- Intuition can only provide a qualitative answer. On each draw there is an individual chance error, equal to the number drawn minus the average of the box. The chance error in the sum of the draws equals the sum of these individual chance errors. The point is that some of the individual errors are positive, and some negative. By and large, the positive individual errors cancel out the negative individual errors, so that the chance error in the sum grows slowly with respect to n .

- The discussion so far has been for the box

$$\boxed{\$2} \quad \boxed{-\$1} \quad \boxed{-\$1} \quad \boxed{-\$1} \quad |.$$

But the results apply to any box.

- Think about drawing *at random with replacement* from a box of numbered tickets.

- The sum of the draws will be off from its expected value by some chance error. The chance error is:

- positive (+), when the sum is _____ than expected;
- negative (−), when the sum is _____ than expected; and
- zero, when the sum is _____ what’s expected.

- The likely size of the chance error — the so-called *standard error* for the sum of the draws — is given by the formula

$$\text{SE of the sum} = \sqrt{\text{number of draws}} \times (\text{SD of the box}).$$

- “SD of the box” is short for the “standard deviation of the list of numbers in the box.”
- The formula has this interpretation:
 - The SD of the box measures the _____ of the tickets in the box, and hence the _____ in a single draw.
 - The sum of n draws from the box is more variable than a single draw by a factor of _____ .
- The formula is called the *square root law*.
- According to this law, the SE for the sum of 1600 draws from the box at the top of the page equals

$$\sqrt{\hspace{10em}} \times (\hspace{10em}) \approx 40 \times \$1.30 = \$52.00 .$$

A SHORT CUT FOR FINDING THE SD OF A BOX

- When the tickets in a box show *only two* different numbers, the SD of the box works out to

$$\left(\frac{\text{bigger number} - \text{smaller number}}{\text{number}} \right) \times \sqrt{\frac{\text{fraction with bigger number}}{\text{number}} \times \frac{\text{fraction with smaller number}}{\text{number}}}$$

- For example, consider the box

$$\boxed{2} \boxed{-1} \boxed{-1} \boxed{-1}$$

- The short cut can be used since the tickets show only two different numbers, 2 and -1 .
- The bigger number equals _____, and the fraction of tickets in the box with this number equals _____.
- The smaller number equals _____, and the fraction of tickets in the box with this number equals _____.
- The SD of the box equals

$$\left(2 - (-1) \right) \times \sqrt{\frac{1}{4} \times \frac{3}{4}} \approx 1.30,$$

in agreement with our previous calculation.

- True or false, and explain:

- The SD of the box $\boxed{0} \boxed{0} \boxed{0} \boxed{1} \boxed{1}$ equals $\sqrt{2 \times 3}$.
- The SD of the box $\boxed{-1} \boxed{-1} \boxed{1} \boxed{1} \boxed{1}$ equals $\sqrt{3/5 \times 2/5}$.
- The SD of the box $\boxed{0} \boxed{1} \boxed{1}$ equals $\sqrt{2/3 \times 1/3}$.

SUMMARY

- So far we have seen that when drawing at random with replacement from a box of numbered tickets:

- The expected value for the sum of the draws equals
(number of draws) \times (average of box).

- This was “expected.”

- The standard error for the sum of the draws equals

$$\sqrt{\text{number of draws}} \times (\text{SD of the box}).$$

- This is remarkable.

- There is one more important fact — provided the number of draws is sufficiently large, the normal curve can be used to figure chances for the sum of the draws, by using the expected value and SE to convert to standard units.

- This is truly miraculous! The next chapter discusses what is meant by “sufficiently large.”

USING THE NORMAL CURVE

- At Nevada roulette tables, the “house special” is a bet on the numbers 0, 00, 1, 2, and 3. The bet pays 6 to 1, and there are 5 chances in 38 to win.

- For all other bets at Nevada roulette tables, the house expects to make about 5 cents out of every dollar put on the table. How much does it expect to make per dollar on the house special?

- From the point of view of the house, a dollar bet on the house special is like _____ draw from the box



- The average of the box is

$$\frac{\times () + \times}{38} = \$0.08 .$$

So the house expects to make about _____ cents per dollar bet.

- As far as the house is concerned, this is a great bet.

- Someone plays roulette 100 times, betting a dollar on the house special each time. Estimate the chance that this person comes out ahead.

- The player’s net gain is like the sum of _____ draws at random with replacement from the box



- The average of the box

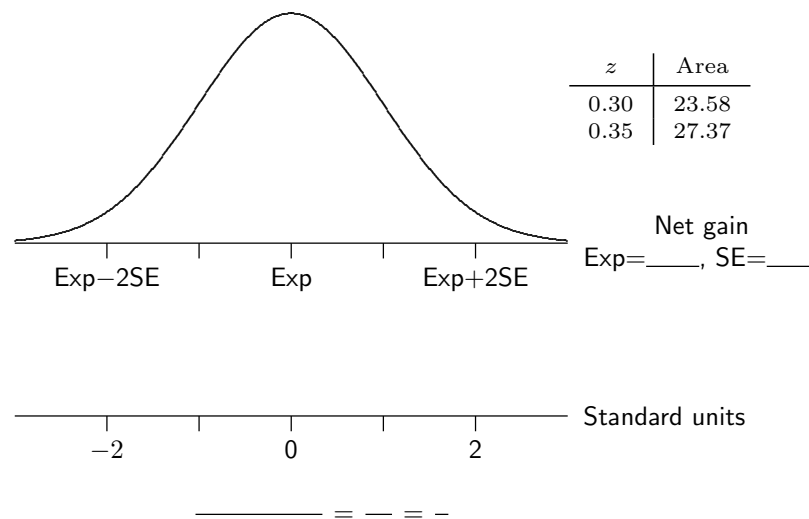


equals _____ , and it’s SD equals

$$(\text{SD of box}) = (\quad - \quad) \times \sqrt{ / \quad \times / \quad } = \$2.37 .$$

- The player’s net gain in 100 plays equals _____ , give or take _____ or so.

- The chance that the player comes out ahead is approximately equal to the area under the normal curve to the _____ of \$0 on the net gain axis:



- Chance \approx shaded area \approx _____ .