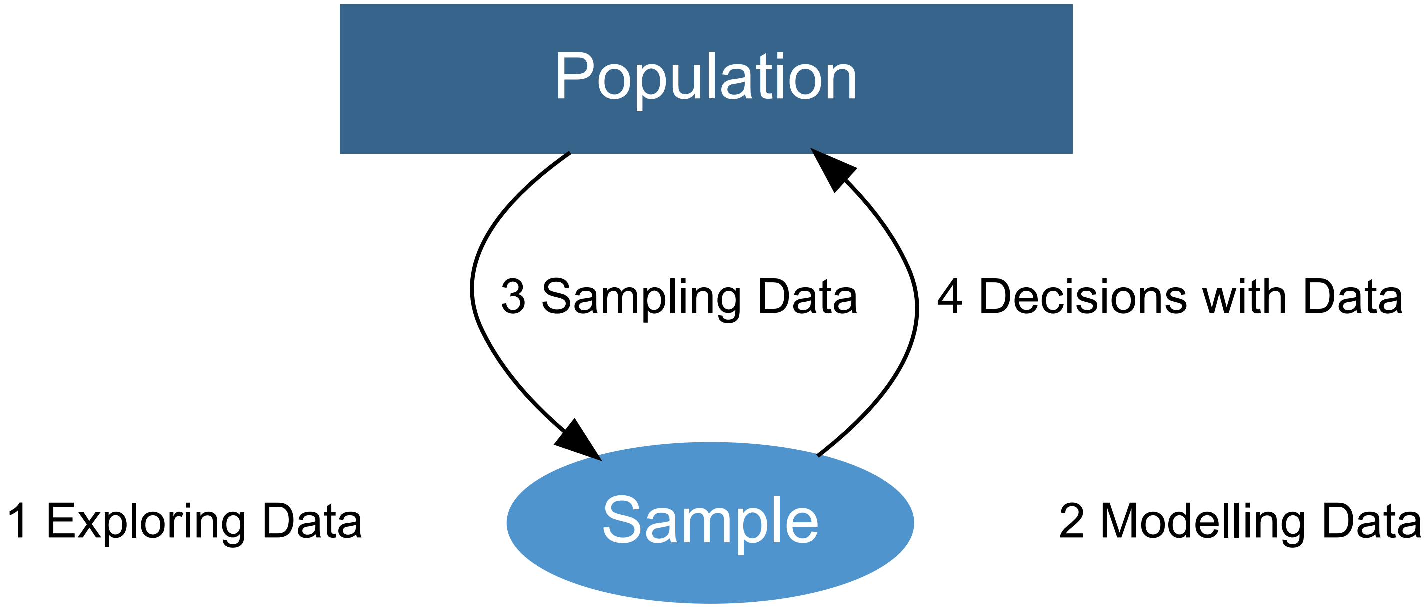
The Box Model

Sampling Data | Chance Variability

12 September 2024

Course Overview



Module3 Sampling Data

Understanding probability

What is probability?

Counting and chance simulation

How to count the number of possible outcomes?

Chance variability

How can we model chance variability by a box model?

Central limit theorem

What is the behaviour of the sample mean for a large sample size?

Today's outline

The box model

Random draws, sample sums and sample means

Expected value and Standard error

The box model

Statistical models

- A **model** is a representation of something which
  - is **simpler** but at the same time
  - captures the **key features** of the original.
- Data obtained in real life is generated by complicated processes.
- **Statistical models** are models for data-generating processes:
  - they are much simpler than the real data-generating process but
  - (hopefully) they capture the signal or key features within the data.

The box model

The **box model** is a very simple statistical model for representing a population. The box model can be thought of as:

- A collection of N objects, e.g. tickets, balls is imagined “in a box”.
- Each object bears a number.
- A **random sample** of a certain number n of the objects is taken.
- The sampling may be **with** or **without** replacement.

Random samples and random draws

- Consider all possible ways of selecting n objects from the box. A **random sample** is when each possible of these selection is equally likely.
- A **random draw** is a random sample with $n = 1$.
  - If a single draw is taken, then each object in the box has an equal chance of being picked.
- If we *completely know* the contents of the box, we can write down the chance of each possible value.
- We let X denote the **random draw**:
  - this represents the “value we might get”
  - X can take different values with different probabilities/chances.
- The **distribution** of X is a **table** with two “columns”:
  - each possible value x that X can take *and*
  - the corresponding probability/chance of that value.

Simple example

- For example, suppose X is a random draw from the following simple box:

$$\boxed{\boxed{1}\,\boxed{2}\,\boxed{3}}$$

- There are then three possible tickets: $\boxed{1}$, $\boxed{2}$ and $\boxed{3}$ and each has equal chance of $\frac{1}{3}$ of being picked, so:

$$P(X = 1) = P(X = 2) = P(X = 3) = \frac{1}{3}\,.$$

Here we write $P(\cdot)$ to denote the probability of each event.

- The distribution of X:
  -

x	1	2	3
$P(X = x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Non-equal chance example

- We can have box models where the different possible *values* are not necessarily equally likely.
- For the box

$$\boxed{\boxed{1}\ \boxed{2}\ \boxed{2}\ \boxed{3}\ \boxed{3}\ \boxed{3}}$$

if each “ticket” is equally likely, we have

$$P(X = 1) = \frac{1}{6}, \quad P(X = 2) = \frac{2}{6} = \frac{1}{3}, \quad P(X = 3) = \frac{3}{6} = \frac{1}{2}.$$

- The distribution of X:
  -

x	1	2	3
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

Larger box example

Consider the box defined by the file `y.dat` in the R code below:

```
box = scan("y.dat")
box
```

```
##   [1]  3  4  5  6  7  8  4  5  6  7  8  9  5  6  7  8  9 10  6  7  8  9 10 11  7  8  9 10 11 12
##  [31]  8  9 10 11 12 13  4  5  6  7  8  9  5  6  7  8  9 10  6  7  8  9 10 11  7  8  9 10 11 12
##  [61]  8  9 10 11 12 13  9 10 11 12 13 14  5  6  7  8  9 10  6  7  8  9 10 11  7  8  9 10 11 12
##  [91]  8  9 10 11 12 13  9 10 11 12 13 14 10 11 12 13 14 15  6  7  8  9 10 11  7  8  9 10 11 12
## [121]  8  9 10 11 12 13  9 10 11 12 13 14 10 11 12 13 14 15 11 12 13 14 15 16  7  8  9 10 11 12
## [151]  8  9 10 11 12 13  9 10 11 12 13 14 10 11 12 13 14 15 11 12 13 14 15 16 12 13 14 15 16 17
## [181]  8  9 10 11 12 13  9 10 11 12 13 14 10 11 12 13 14 15 11 12 13 14 15 16 12 13 14 15 16 17
## [211] 13 14 15 16 17 18
```

```
table(box)  # note: first two rows below are only labels: the 'real' output is the third line
```

```
## box
##  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18
##  1  3  6 10 15 21 25 27 27 25 21 15 10  6  3  1
```

Find the probability $P(X < 8)$

```
sum(table(box))  # gives total freq, i.e. size of the box
```

```
## [1] 216
```

```
length(box) == sum(table(box))
```

```
## [1] TRUE
```

```
head(box < 8)  # reports the first 6 values of 'box<8'
```

```
## [1]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
```

```
sum(box < 8)  # reports the total numer of TRUE values in 'box<8'
```

```
## [1] 35
```

Find the *proportion* less than 8

```
sum(box < 8)/length(box)
```

```
## [1] 0.162037
```

```
mean(box < 8)  # mean of TRUEs in 'box<8'
```

```
## [1] 0.162037
```

- The chance of drawing a value less than 8 is $\frac{35}{216} \approx 16\%$.
- Note: $35 = 1 + 3 + 6 + 10 + 15$ (the frequencies of 3, 4, 5, 6 and 7 respectively).

Expected value and standard error

- In some situations, we may not know the exact contents of the box. Indeed, boxes are used to model populations and we might not know everything about the population.
- Instead we might have access to summary information about the box.
- For a random draw X from a box, we define the following two quantities:
  - We denote the **expected value** $E(X)$ as the mean of the box
  - We denote the **standard error** $SE(X)$ as the standard deviation of the box

Interpreting the expected value $E(X)$

- The random draw may be "decomposed" into two pieces:

$$X = E(X) + [X - E(X)] = E(X) + \varepsilon\,.$$

- The first part $E(X)$ is *not random*.
- All randomness is included in the chance error ε, which is itself can be represented by a random draw from an **error box** (a box with mean zero).
- **Example**: a random draw X from the box

$$\boxed{\boxed{1}\,\boxed{2}\,\boxed{3}}$$

(which has mean 2) may instead be thought of as $X = 2 + \varepsilon$ where the chance error ε is a random draw from the error box

$$\boxed{\boxed{-1}\,\boxed{\;0\;}\,\boxed{+1}}\,.$$

Interpreting the standard error $SE(X)$

- The standard error measures the typical size of the error ε. It is a measure of random variation in the outcome of X.
- For two different random draws, one with the larger SE is likely to differ from its expected value by a larger amount.
- The standard error is the root-mean-square of the error box.

$$SE(X) = SD(\text{box}) = \sqrt{\frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3}} \approx 0.816$$

$$SE(X) = RMS(\text{error box}) = \sqrt{\frac{(-1)^2 + 0^2 + 1^2}{3}} \approx 0.816$$

Sums of random draws

New interpretation of mean and SD

- We have introduced the concepts of
  - a random draw X from a box
  - its expected value $E(X)$
  - its standard error $SE(X)$
- The expected value and standard error are not “new” things, rather, they are new interpretations of old things.
  - It is really “worth the effort” to introduce these new names for these things are already know about?
  - The expected value and standard error become very useful when we have **more than one draw**.

Sum of two random draws

- Consider the two boxes

$$\boxed{\boxed{1}\,\boxed{2}\,\boxed{3}} \text{ and } \boxed{\boxed{2}\,\boxed{4}\,\boxed{6}\,\boxed{8}}.$$

  - The first box has mean 2 and SD $\sqrt{\frac{1}{3}\left[(-1)^2 + 0^2 + 1^2\right]} = \sqrt{\frac{2}{3}} \approx 0.816.$
  - The second box has mean 5 and SD

$$\sqrt{\frac{1}{4}\left[(-3)^2 + (-1)^2 + 1^2 + 3^2\right]} = \sqrt{5} \approx 2.236\,.$$

- Suppose we are going to take a random draw from each, X from the first box, Y from the second box, in such a way that **each possible pair of values is equally likely**. What is the behaviour of the (random) **sum** $S = X + Y$?

All possible pairs/sums

- There are 12 possible pairs:

$$\left(\boxed{1},\boxed{2}\right),\ \left(\boxed{1},\boxed{4}\right),\ \left(\boxed{1},\boxed{6}\right),\ \left(\boxed{1},\boxed{8}\right),$$

$$\left(\boxed{2},\boxed{2}\right),\ \left(\boxed{2},\boxed{4}\right),\ \left(\boxed{2},\boxed{6}\right),\ \left(\boxed{2},\boxed{8}\right),$$

$$\left(\boxed{3},\boxed{2}\right),\ \left(\boxed{3},\boxed{4}\right),\ \left(\boxed{3},\boxed{6}\right),\ \left(\boxed{3},\boxed{8}\right).$$

Table of all possible pairs and their sums

Sample	Sum
(1,2)	3
(1,4)	5
(1,6)	7
(1,8)	9
(2,2)	4
(2,4)	6
(2,6)	8
(2,8)	10
(3,2)	5
(3,4)	7
(3,6)	9
(3,8)	11

Single random draw from a "bigger" box

Thus getting a random pair (X, Y) and forming the sum $S = X + Y$ is **equivalent** to a *single random draw* from the bigger box

$$\boxed{\boxed{3}\,\boxed{4}\,\boxed{5}\,\boxed{5}\,\boxed{6}\,\boxed{7}\,\boxed{7}\,\boxed{8}\,\boxed{9}\,\boxed{9}\,\boxed{10}\,\boxed{11}}$$

What are the mean and SD of this "bigger" box?

Using `outer()`

- The R function `outer()` forms a two-way array by applying an operation to each pair of elements from two vectors:

```
bx = c(1, 2, 3)
by = c(2, 4, 6, 8)
bs = outer(bx, by, "+")
bs
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    3    5    7    9
## [2,]    4    6    8   10
## [3,]    5    7    9   11
```

```
mean(bs)
```

```
## [1] 7
```

```
mean((bs - mean(bs))^2)
```

```
## [1] 5.666667
```

Expected value and standard error of the sum

- So we have that $E(S) = 7$ and $SE(S) = \sqrt{5\frac{2}{3}} \approx 2.38$.
- As it turns out

$$7 = E(S) = E(X+Y) = E(X) + E(Y) = 2 + 5\,.$$

$$5\frac{2}{3} = SE(S)^2 = SE(X+Y)^2 = SE(X)^2 + SE(Y)^2 = \frac{2}{3} + 5\,.$$

- So in this case we have
  - expected value of sum is sum of expected values;
  - *squared* SE of the sum is the sum of the *squared* SEs
- These results hold in general.

Sum of two random draws.

- Consider two boxes

$$\boxed{\boxed{x_1}\ \boxed{x_2}\ \cdots\ \boxed{x_M}} \text{ and } \boxed{\boxed{y_1}\ \boxed{y_2}\ \cdots\ \boxed{y_N}}$$

- Suppose we are going to take a random draw from each: X from the first box, Y from the second box, in such a way that **each possible pair of values is equally likely**.

All possible sums

- There are MN possible sums, we may arrange them in a two-way array with M (horizontal) rows and N (vertical) columns.

- Noting that $\sum_{i=1}^{M} x_i = M\bar{x}$, we may write the column sums below the line:

$$
\begin{array}{cccc}
x_1 + y_1 & x_1 + y_2 & \cdots & x_1 + y_N \\
x_2 + y_1 & x_2 + y_2 & \cdots & x_2 + y_N \\
\vdots & \vdots & \ddots & \vdots \\
x_M + y_1 & x_M + y_2 & \cdots & x_M + y_N \\
\hline
M\bar{x} + My_1 & M\bar{x} + My_2 & \cdots & M\bar{x} + My_N
\end{array}
$$

The sum of column sums is

$$\underbrace{M\bar{x} + \cdots + M\bar{x}}_{N \text{ terms}} + M(y_1 + \cdots + y_N) = NM\bar{x} + MN\bar{y}$$

Thus the average of all possible sums is

$$\frac{\text{sum of all possible sums}}{\text{no. of all possible sums}} = \frac{NM\bar{x} + MN\bar{y}}{MN} = \bar{x} + \bar{y} = E(X) + E(Y)\,.$$

That is,

$$E(X + Y) = E(X) + E(Y)\,.$$

Computing formula for SD

- For a list of numbers $x_1, x_2, \dots, x_M$, the square of the SD may be written as

$$SD^2 = \frac{1}{M}\sum_{i=1}^{M}(x_i - \bar{x})^2 = \left(\frac{1}{M}\sum_{i=1}^{M} x_i^2\right) - \bar{x}^2$$

  the “mean square minus the square of the mean”.

- To see why, recall that $\sum_{i=1}^{M} x_i = M\bar{x}$ and so:

$$\begin{aligned}
\sum_{i=1}^{M}(x_i - \bar{x})^2 &= (x_1^2 - 2\bar{x}x_1 + \bar{x}^2) + \cdots + (x_M^2 - 2\bar{x}x_M + \bar{x}^2) \\
&= (x_1^2 + \cdots + x_M^2) - 2\bar{x}(x_1 + \cdots + x_M) + \underbrace{\bar{x}^2 + \cdots + \bar{x}^2}_{M \text{ terms}} \\
&= \sum_{i=1}^{M} x_i^2 - 2\bar{x}M\bar{x} + M\bar{x}^2 = \sum_{i=1}^{M} x_i^2 - M\bar{x}^2
\end{aligned}$$

Easy way to compute SD in R

- The computing formula above can be used to write a quick-and-easy R function to compute the (population) SD of a list of numbers.

```
popsd = function(x) sqrt(mean(x^2) - (mean(x)^2))
```

- Let's try it out:

```
x = 1:10
x  # this list has mean 5.5
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10
```

```
sqrt(mean((x - 5.5)^2))
```

```
## [1] 2.872281
```

```
popsd(x)
```

```
## [1] 2.872281
```

SE of a sum (not examinable)

- It is possible to deduce the SE of our general sum $S = X + Y$.
- We do so by first working out the mean-square of the bigger box of all possible sums.
- Write each squared sum $(x_i + y_j)^2 = x_i^2 + 2x_i y_j + y_j^2$ in an array and add over columns:

$$
\begin{array}{ccc}
x_1^2 + 2x_1 y_1 + y_1^2 & \cdots & x_1^2 + 2x_1 y_N + y_N^2 \\
x_2^2 + 2x_2 y_1 + y_1^2 & \cdots & x_2^2 + 2x_2 y_N + y_N^2 \\
\vdots & \ddots & \vdots \\
x_M^2 + 2x_M y_1 + y_1^2 & \cdots & x_M^2 + 2x_M y_N + y_N^2 \\
\hline
\sum_i x_i^2 + 2M\bar{x}y_1 + My_1^2 & \cdots & \sum_i x_i^2 + 2M\bar{x}y_N + My_N^2
\end{array}
$$

SE of a sum (not examinable)

- The sum of squares (of all possible sums) is then

$$
\begin{aligned}
&\left(\sum_i x_i^2 + 2M\bar{x}y_1 + My_1^2\right) + \cdots + \\
&\qquad \left(\sum_i x_i^2 + 2M\bar{x}y_N + My_N^2\right) = N\sum_i x_i^2 + 2M\bar{x}(y_1 + \cdots + y_N) + \\
&\qquad\qquad M(y_1^2 + \cdots + y_N^2) \\
&\qquad\qquad = N\sum_i x_i^2 + 2MN\bar{x}\bar{y} + M\sum_j y_j^2 .
\end{aligned}
$$

- Since there are MN possible sums, the mean square is

$$
\frac{1}{M}\sum_i x_i^2 + 2\bar{x}\bar{y} + \frac{1}{N}\sum_j y_j^2 .
$$

SE of a sum (not examinable)

- Since mean of all possible sums is $\bar{x} + \bar{y}$, the squared SD of all possible sums is

$$SE(S)^2 = \underbrace{\frac{1}{M}\sum_i x_i^2 + 2\bar{x}\bar{y} + \frac{1}{N}\sum_j y_j^2}_{mean\ sq.} - \underbrace{\left(\bar{x}^2 + 2\bar{x}\bar{y} + \bar{y}^2\right)}_{sq.\ of\ mean}$$

$$= \frac{1}{M}\sum_i x_i^2 - \bar{x}^2 + \frac{1}{N}\sum_j y_j^2 - \bar{y}^2$$

$$= \frac{1}{M}\sum_i (x_i - \bar{x})^2 + \frac{1}{N}\sum_j (y_j - \bar{y})^2$$

$$= SE(X)^2 + SE(Y)^2 .$$

Random samples with replacement of size $n = 2$

- A special case of our general sum is where we have a **single** box

$$\boxed{\boxed{x_1}\ \boxed{x_2}\ \cdots\ \boxed{x_N}}$$

  but take two random draws with replacement.
  - This means each of the N^2 possible pairs $(x_1, x_1), \ldots, (x_1, x_n), \ldots, (x_n, x_1), \ldots, (x_n, x_n)$ is **equally likely**.
- This is where both boxes are (effectively) the same, so $E(X) = E(Y)$ and $SE(X) = SE(Y)$.
- If we write the mean of the box as μ and the SD of the box as σ, then the sum S of the two random draws has
  - $E(S) = E(X) + E(Y) = \mu + \mu = 2\mu$
  - $SE(S)^2 = SE(X)^2 + SE(Y)^2 = \sigma^2 + \sigma^2 = 2\sigma^2 \implies SE(S) = \sqrt{2}\sigma.$

Sums and averages of random samples of size n

Random samples of size n

- We may easily extend the results to any $n \geq 2$. Suppose:
  - we have a box with mean μ and SD σ;
  - we are going to take a random sample of size n from the box **with replacement**;
  - so each possible sample of size n is equally likely.
- Let us write
  - the random draws as $X_1, X_2, \ldots, X_n$;
  - the sum as $S = X_1 + \cdots + X_n$;
  - the *sample average* as $\bar{X} = \frac{S}{n} = \frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$.
- What are the expected value and standard error of both S and $\bar{X}$?

The sum S

- Each single draw has the same behaviour. That is each $X_1, \ldots, X_n$ is a single random draw from the same box with $E(X_1) = \mu$ and $SE(X_1) = \sigma$.
- Expected value of sum is sum of expected values:

$$E(S) = E(X_1 + \cdots + X_n) = E(X_1 + \cdots + X_{n-1}) + E(X_n) =$$
$$\cdots = E(X_1) + \cdots + E(X_n) = \underbrace{\mu + \cdots + \mu}_{n \text{ terms}} = n\mu \,.$$

- Also,

$$SE(S)^2 = SE(X_1 + \cdots + X_{n-1})^2 + SE(X_n)^2 =$$
$$\cdots = SE(X_1)^2 + \cdots + SE(X_n)^2 = \underbrace{\sigma^2 + \cdots + \sigma^2}_{n \text{ terms}} = n\sigma^2$$
$$\implies SE(S) = \sqrt{n}\sigma$$

What if we divide by N?

- Consider the box

$$\boxed{\boxed{x_1}\ \boxed{x_2}\ \cdots\ \boxed{x_N}}$$

  What is the expected value and standard error of a random draw if we divide each x_i by N?

- This gives us a new box

$$\boxed{\boxed{y_1}\ \boxed{y_2}\ \cdots\ \boxed{y_N}}$$

  where $y_i = \frac{x_i}{N}$.

What if we divide by N?

- If Y is a random draw from this new box then we can work out $E(Y)$ as:

$$E(Y) = \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{N}\sum_{i=1}^{N} \frac{x_i}{N} = \frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) = \frac{\bar{x}}{N} = \frac{E(X)}{N}$$

- We can also work out the standard error:

$$SE(Y)^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 = \frac{1}{N}\sum_{i=1}^{N}(\frac{x_i}{N} - \frac{\bar{x}}{N})^2$$

$$= \frac{1}{N^2}\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 = \frac{SE(X)^2}{N^2}$$

$$\implies SE(Y) = \frac{SE(X)}{N}$$

The sample average $\bar{X}$

- The sample average $\bar{X}$ is just $\frac{S}{n}$, so we can immediately work out the expected value and standard error.
- We thus obtain immediately that for the average,

$$E(\bar{X}) = \frac{E(S)}{n} = \frac{n\mu}{n} = \mu\,;$$

- As for the standard error we have

$$SE(\bar{X}) = \frac{SE(S)}{n} = \frac{\sigma\sqrt{n}}{n} = \frac{\sigma}{\sqrt{n}}\,.$$

Example: 6-sided die

- Consider rolling a fair 6-sided die.
- In this case each of the numbers 1,2,3,4,5,6 are equally likely.
- This is equivalent to a random draw from the box

$$\boxed{\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}}$$

- The mean is $\mu = 3.5 = \frac{7}{2}$, mean-square $\frac{1+4+9+16+25+36}{6} = \frac{91}{6}$ and thus SD

$$\sigma = \sqrt{\frac{91}{6} - \left(\frac{7}{2}\right)^2} = \sqrt{\frac{91}{6} - \frac{49}{4}} = \sqrt{\frac{182 - 147}{12}} = \sqrt{\frac{35}{12}} \approx 1.71$$

Rolling the die 3 times: sum of rolls

- Suppose we roll the die (independently) 3 times. What is the random behaviour of the **sum** of the values of the three rolls?
- Let X_1, X_2, X_3 denote 3 random draws with replacement from the box

$$\boxed{\boxed{1}\,\boxed{2}\,\boxed{3}\,\boxed{4}\,\boxed{5}\,\boxed{6}}$$

- Then the sum of the 3 rolls $S = X_1 + X_2 + X_3$ has $E(S) = 3\mu = \frac{21}{2} = 10.5$ and

$$SE(S) = \sigma\sqrt{3} = \sqrt{\frac{35}{12} \times 3} = \sqrt{\frac{35}{4}} = \frac{\sqrt{35}}{2} \approx 2.958\,.$$

- The box of all possible sums here is exactly the dataset `y.dat` from earlier in the lecture!

Rolling the die 3 times: average of rolls

- What is the random behaviour of the **average** of the values of the three rolls?
- Writing $\bar{X} = \frac{X_1+X_2+X_3}{3} = \frac{S}{3}$, we have

$$E(\bar{X}) = \frac{E(S)}{3} = \frac{3\mu}{3} = \mu = 3.5$$

and

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{3}} = \sqrt{\frac{35}{12} \times \frac{1}{3}} = \sqrt{\frac{35}{36}} = \frac{\sqrt{35}}{6} \approx 0.956\,.$$

Demonstration

- Let us simulate 3 rolls of a 6-sided die 1000 times, and look at the corresponding 1000 sums and averages of each triplet.

```
d = 1:6
S = 0  # empty vector to catch the sums
for (i in 1:1000) {
    rolls = sample(d, size = 3, replace = T)
    S[i] = sum(rolls)
}
mean(S)
```

```
## [1] 10.476
```

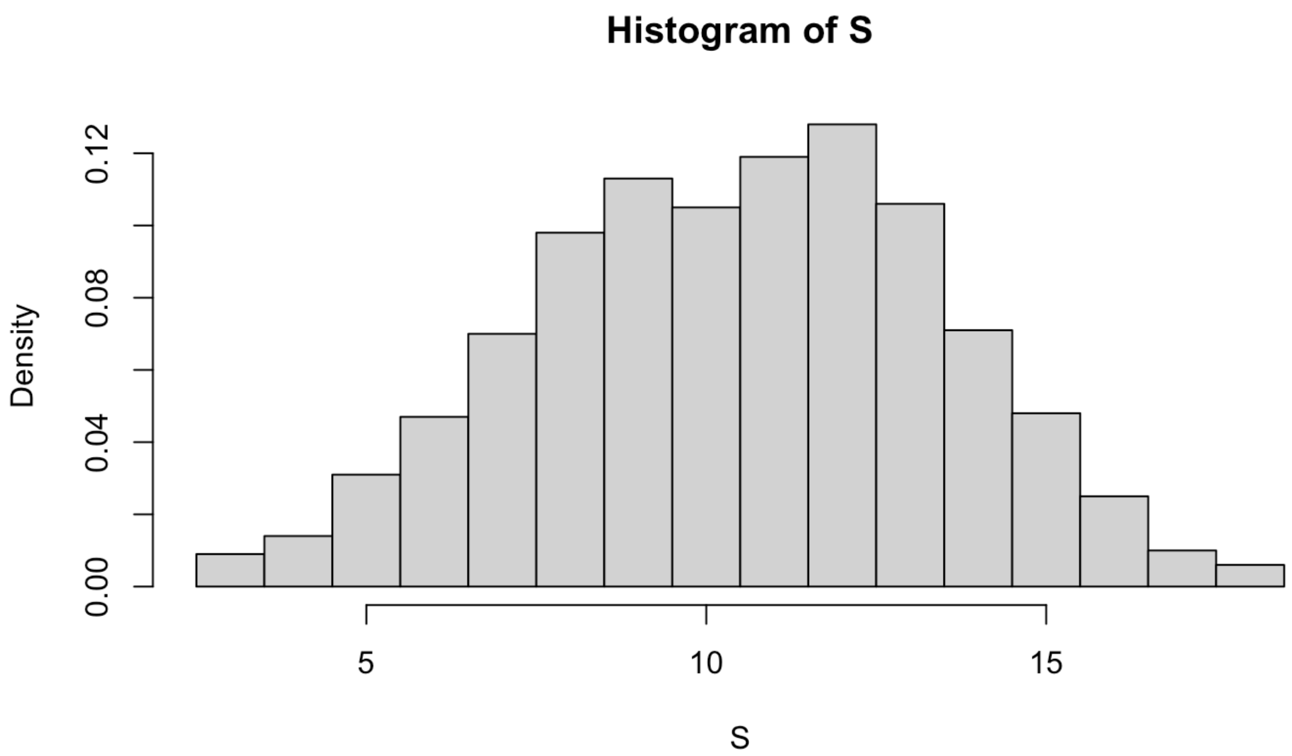
```
sd(S)
```

```
## [1] 3.014052
```

```
popsd(S)
```

```
## [1] 3.012544
```

```
hist(S, pr = T, breaks = br)
```



Note these proportions are *close* to (but not *exactly* equal to) the corresponding proportions in `y.dat`.

Averages

```
Xbar = S/3
mean(Xbar)
```

```
## [1] 3.492
```

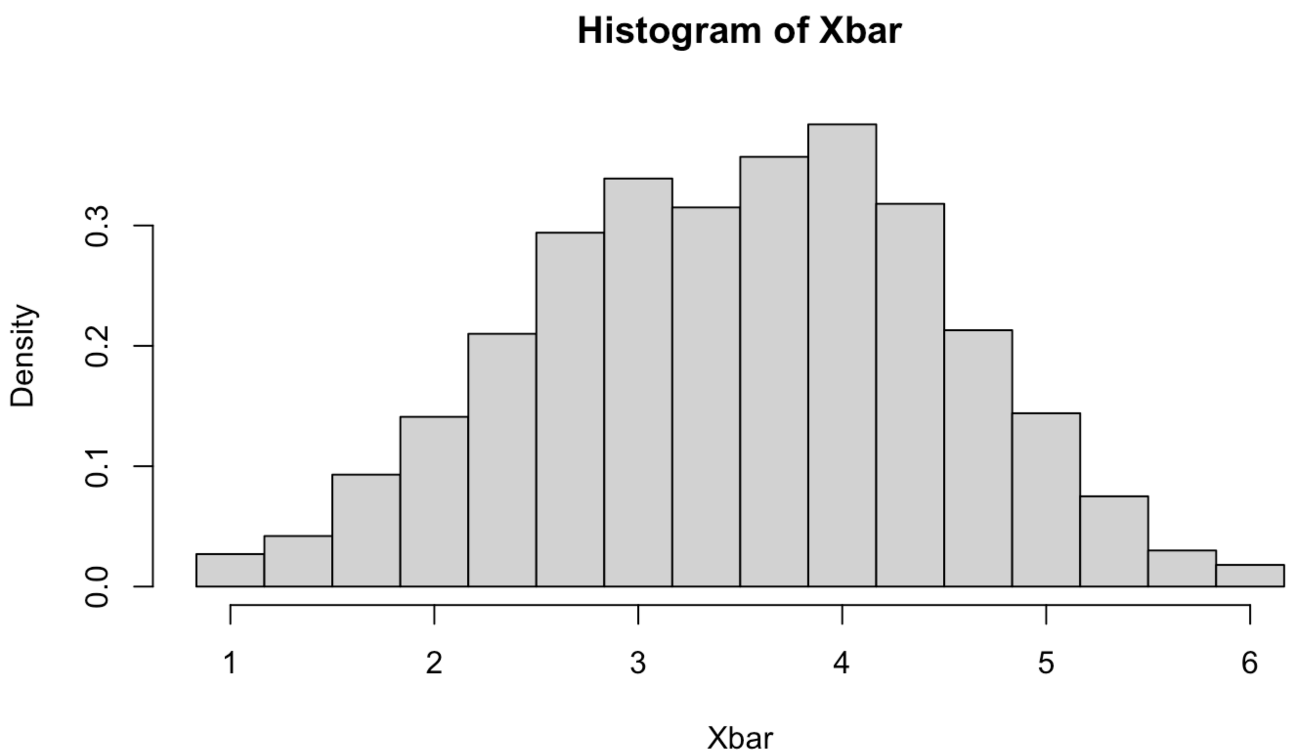
```
sd(Xbar)
```

```
## [1] 1.004684
```

```
popsd(Xbar)
```

```
## [1] 1.004181
```

```
hist(Xbar, pr = T, breaks = br/3)
```



Same shape as for the sums, but centred on 3.5 and less spread-out.

Closing remarks: n getting larger

- We have seen that for n random draws (with replacement) from a box with mean μ and SE σ
  - the *sum* of draws S has $E(S) = n\mu$ and $SE(S) = \sigma\sqrt{n}$;
  - the *average* of the draws $\bar{X}$ has $E(\bar{X}) = \mu$ and $SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$.
- What happens to the SE of each as n gets bigger?
  - for the sum, $\sigma\sqrt{n}$ gets larger **but**
  - for the average, $\frac{\sigma}{\sqrt{n}}$ gets **smaller**.
- In particular, for the average $\bar{X}$, the random variability about $E(\bar{X}) = \mu$ gets less as the sample size n increases.