Think average = long-term average with continuous/infinite samples, probably

Same with variance

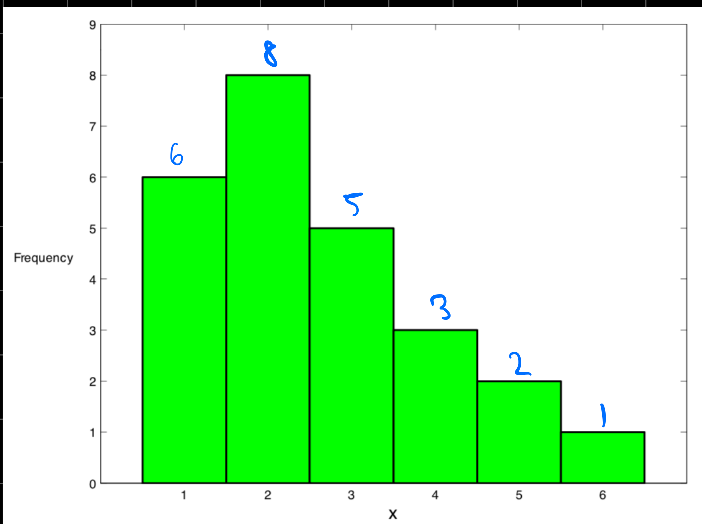
X	Frequency Count	Frequency
1	卌 |	6
2	卌 |||	8
3	卌	5
4	|||	3
5	||	2
6	|	1

Relative frequency : $6/25$

$8/25$

$\vdots$

Frequency Histogram



not symmetric → skewed (right/positively)

Skew left/right : long tail (left/right)

Arithmetic Mean

$$\bar{x} = \frac{(6)(1) + (8)(2) + \cdots + (1)(6)}{6}$$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i$$

↳ random variable : data will be different in different trials

↳ also random variable, since it's a function of random variables

$\bar{x}$ is an estimate of the population mean μ

But you'll never know $\mu - \bar{x}$ is good enough

Suppose we are counting the number of people in cars over a toll bridge. The value will probably *always* be different, so we are expecting a different sample mean every time. As such, this is a random variable, and random variables produce different results

<u>Median:</u> middle value

$x_1, \ldots, x_n$: sample

$x_{(1)}, \ldots, x_{(n)}$: ordered sample

If n is odd, median: $\left(\frac{n+1}{2}\right)$th observation

$$\text{even: } \frac{\left(\frac{n}{2}+1\right)^{th} + \left(\frac{n}{2}\right)^{th}}{2}$$

Recall that in our toll bridge example,

$$\bar{x} = \frac{(6\times1)+(8\times2)+(5\times3)+(3\times4)+(2\times5)+(1\times6)}{25}$$

Relative frequencies: $\frac{6}{25}, \frac{8}{25}, \ldots$

$$\bar{x} = 6\left(\frac{1}{25}\right) + 8\left(\frac{2}{25}\right) + \cdots$$

tallies

Now, suppose we knew that the probability function of X was actually given by:

x	1	2	3	4	5	6
$f(x) = P(X = x)$	0.30	0.25	0.20	0.15	0.09	0.01

population data : μ

Hence, in the **long-run**, if we use the p.f.

x	1	2	3	4	5	6
$f(x)$	0.30	0.25	0.20	0.15	0.09	0.01

we would expect the value of the mean to be

(0.30) x 1 + (0.25) x 2 + (0.20) x 3 + (0.15) x 4 + (0.09) x 5 + (0.01) x 6

= **2.51**

↳ expected value as we do an infinite number of trials

Thus, the expected value $\mu = E(X)$ is

$$\mu = \sum_{\forall x} x \cdot f(x)$$

Say a toll of $2 is paid per car and 50 cents per occupant. Find the average long run toll payment using the probability distribution given earlier, namely:

x	1	2	3	4	5	6
$f(x)$	0.30	0.25	0.20	0.15	0.09	0.01

That is, we are interested in the mean of the r.v. *Y* = 0.5*X* + 2.

let $Y = g(x)$

We want $E(g(x)) = \sum_{\forall x} g(x) \cdot f(x)$

$x \cdot f(x)$

all values of x transposed to $g(x)$

$$= \sum_{\forall x} (0.5x + 2) f(x)$$

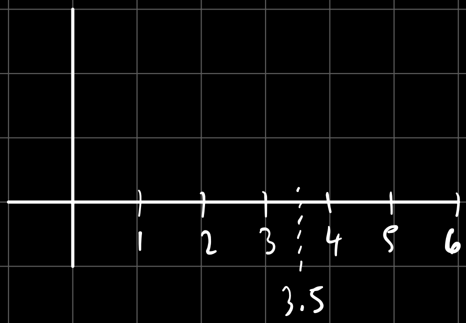
Note that μ is *not* a random variable: it is the *true*/theoretical mean of a population

In physical terms, E(X) is the *balance point* of the probability distribution of f(x)

Ex. μ for rolling a dice

$$E(X) = 1\left(\tfrac{1}{6}\right) + 2\left(\tfrac{1}{6}\right) + \cdots$$

$$= 3.5$$



Linearity property of expectation:

$$\underbrace{E[a \cdot g(x) + b]}_{\text{all values of } x \text{ transposed}} = aE(g(x)) + b$$

$$\text{Proof: } E[a \cdot g(x) + b] = \sum_{\forall x} [a \cdot g(x) + b] f(x)$$

$$= a \underbrace{\sum_{\forall x} g(x) f(x)}_{E(g(x))} + b \sum_{\forall x} f(x)$$

↳ sum of all probabilities is 1

$$= aE(g(x)) + b$$

Ex. A local television station sells 15 second, 30 second, and 60 second advertising spots. Let X denote the length of a randomly selected commercial appearing on this station, and suppose that the probability distribution of X is given by

x	15	30	60
$f(x)$	0.1	0.3	0.6

a) Find E(X)

$$E(X) = (0.1)(15) + \cdots$$
$$= 46.5$$

b) If a 15 second spot sells for \$500, a 30 second spot for \$800, and a 60 second spot for \$1000, find the average amount paid for a commercial appearing on this station.

Let g(x) = price of a commercial with duration x

$E(g(X)) = (500)(0.1) + (800)(0.3) + (1000)(0.6) = 890$