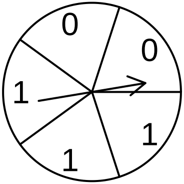
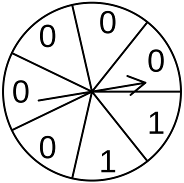
Probability

Bernoulli Random Variable

- A Bernoulli random variable is a binary variable with two possible outcomes: 0 and 1.
- e.g. A Bernoulli random variable with $p = \frac{3}{5}$

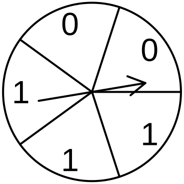


- e.g. A Bernoulli random variable with $p = \frac{2}{7}$



Population mean and standard deviation

- e.g. A Bernoulli variable with $p = \frac{3}{5}$



- Mean

$$\mu = \frac{0+0+1+1+1}{5} = \frac{3}{5} = 0.6$$

- Standard Deviation

$$\sigma = \sqrt{\frac{\left(0-\frac{3}{5}\right)^2 + \left(0-\frac{3}{5}\right)^2 + \left(1-\frac{3}{5}\right)^2 + \left(1-\frac{3}{5}\right)^2 + \left(1-\frac{3}{5}\right)^2}{5}}$$

Simplification...

$$\sigma = \sqrt{\frac{\left(0-\frac{3}{5}\right)^2+\left(0-\frac{3}{5}\right)^2+\left(1-\frac{3}{5}\right)^2+\left(1-\frac{3}{5}\right)^2+\left(1-\frac{3}{5}\right)^2}{5}}$$

$$= \sqrt{\frac{\left(\frac{3}{5}\right)^2+\left(\frac{3}{5}\right)^2+\left(\frac{2}{5}\right)^2+\left(\frac{2}{5}\right)^2+\left(\frac{2}{5}\right)^2}{5}}$$

$$= \sqrt{\frac{2\left(\frac{3}{5}\right)^2+3\left(\frac{2}{5}\right)^2}{5}}$$

$$= \sqrt{\frac{2}{5}\left(\frac{3}{5}\right)^2+\frac{3}{5}\left(\frac{2}{5}\right)^2}$$

Continuation of simplification...

$$\sigma = \sqrt{\frac{2}{5}\left(\frac{3}{5}\right)^2 + \frac{3}{5}\left(\frac{2}{5}\right)^2}$$

$$= \sqrt{\frac{2}{5} \cdot \frac{3}{5} \cdot \left(\frac{3}{5} + \frac{2}{5}\right)}$$

$$= \sqrt{\frac{2}{5} \cdot \frac{3}{5} \cdot 1}$$

$$= \sqrt{\frac{2}{5} \cdot \frac{3}{5}} = 0.4898979$$

Generalization

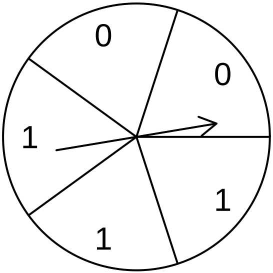
- You can show that any Bernoulli variable has simple formulas for mean and standard deviation.

$$\mu = p$$

$$\sigma = \sqrt{p(1-p)}$$

Sample mean and standard deviation

- If we spin this Bernoulli variable 25 times...



sample:

1	0	1	1	0
1	1	0	0	1
1	1	1	1	1
1	1	0	1	0
1	1	1	1	0

- We calculate the sample mean...

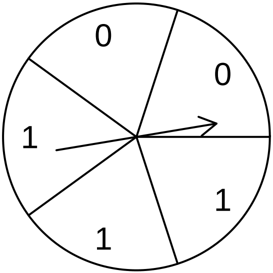
$$\bar{x} = \frac{18}{25} = 0.72$$

- We calculate the sample standard deviation...

$$s = \sqrt{\frac{7 \cdot (0 - 0.72)^2 + 18 \cdot (1 - 0.72)^2}{25 - 1}} = 0.4582576$$

Sample mean and standard deviation again

- If we spin this Bernoulli variable 25 times...



sample:

1	1	0	1	0
1	0	1	0	1
0	0	0	0	1
0	1	0	1	0
0	1	0	1	1

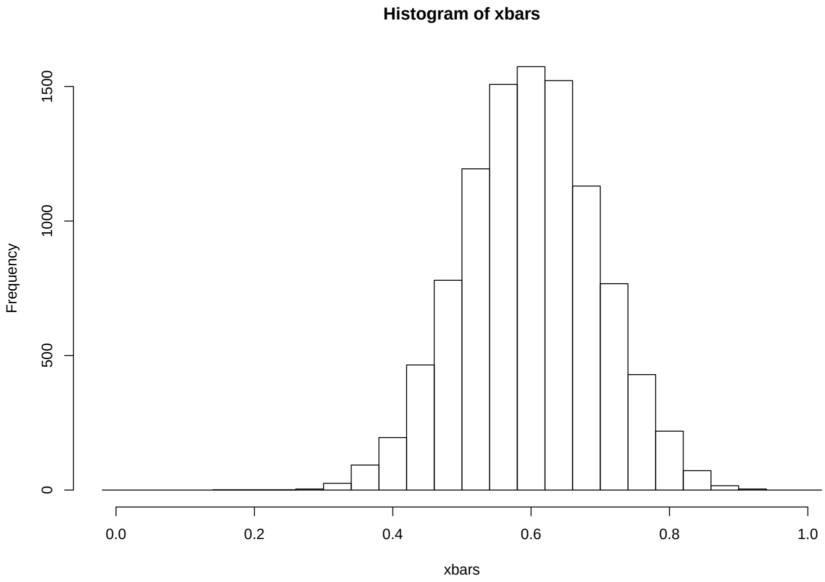
- We calculate the sample mean...

$$\bar{x} = \frac{12}{25} = 0.48$$

- We calculate the sample standard deviation...

$$s = \sqrt{\frac{13 \cdot (0 - 0.48)^2 + 12 \cdot (1 - 0.48)^2}{25 - 1}} = 0.509902$$

Repeat many times (10000 iterations of samples of size 25)



Interval of typical sample means

We will define the **interval of typical sample means** as:

$$\text{interval of typical means} = \left(\mu - \frac{2\sigma}{\sqrt{n}},\, \mu + \frac{2\sigma}{\sqrt{n}}\right)$$

- We expect about 95% of sample means to be in the interval of typical sample means.
- This is more true when n is larger or when the random variable is nearly normal.

Back to example.

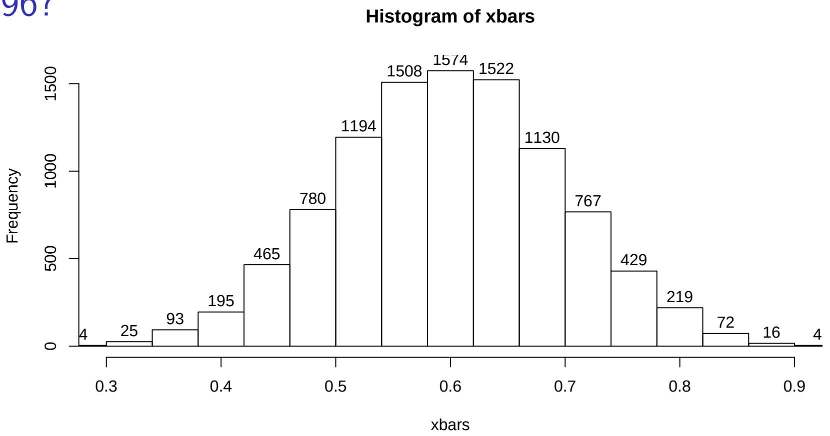
$$\mu = p = \frac{3}{5} = 0.6$$

$$\sigma = \sqrt{\frac{3}{5} \cdot \frac{2}{5}} = 0.4898979$$

$$n = 25$$

$$\begin{aligned}
\text{interval of typical means} &= \left(\mu - \frac{2\sigma}{\sqrt{n}},\, \mu + \frac{2\sigma}{\sqrt{n}}\right) \\
&= \left(0.6 - 2 \cdot \frac{0.4898979}{\sqrt{25}},\, 0.6 + 2 \cdot \frac{0.4898979}{\sqrt{25}}\right) \\
&= (0.404,\, 0.796)
\end{aligned}$$

What percent of sample means were between 0.404 and 0.796?



$$\frac{465 + 780 + 1194 + 1508 + 1574 + 1522 + 1130 + 767 + 429}{10000}$$

$$0.9369 = 93.69\%$$