t-test statistics

Christoph Bartneck

Mean

$$X = \{4, 6, 8, 4, 5, 7, 4, 2\};$$

$$\overline{X} = \sum_{i=1}^{N} X_i$$

Dispersion

```
group1 = {4, 6, 7, 4, 5, 6, 4, 4};
group2 = {1, 6, 10, 4, 5, 7, 2, 5};
```

Dispersion

```
group1 = {4, 6, 7, 4, 5, 6, 4, 4};
group2 = {1, 6, 10, 4, 5, 7, 2, 5};
```

$$(4-5)+(6-5)+(7-5)+(4-5)+(5-5)+(6-5)+(4-5)+(4-5)=0$$

Variance

```
group1 = {4, 6, 7, 4, 5, 6, 4, 4};
group2 = {1, 6, 10, 4, 5, 7, 2, 5};
```

$$s^2(X) = \frac{\sum \left(X - \overline{X}\right)^2}{N}$$

Variance

```
group1 = {4, 6, 7, 4, 5, 7, 4, 4};
group2 = {1, 6, 10, 4, 5, 7, 2, 5};
```

Variance(group1)=2
Variance(group2)=7

What is the “group”?

- Population
  - Every member of the group
- Sample
  - A subset of the population

Sample Standard Deviation

$$s(X) = \sqrt{\frac{\sum\left(x - \bar{x}\right)^2}{N}}$$

Population Standard Deviation

$$\sigma(X) = \sqrt{\frac{\sum\left(x - \bar{x}\right)^2}{N - 1}}$$

Probability

- Boris and Ingrid have to get me tea
- 17 out of 30 times I asked Boris
- Is Boris at a disadvantage?

Likelihood of an event

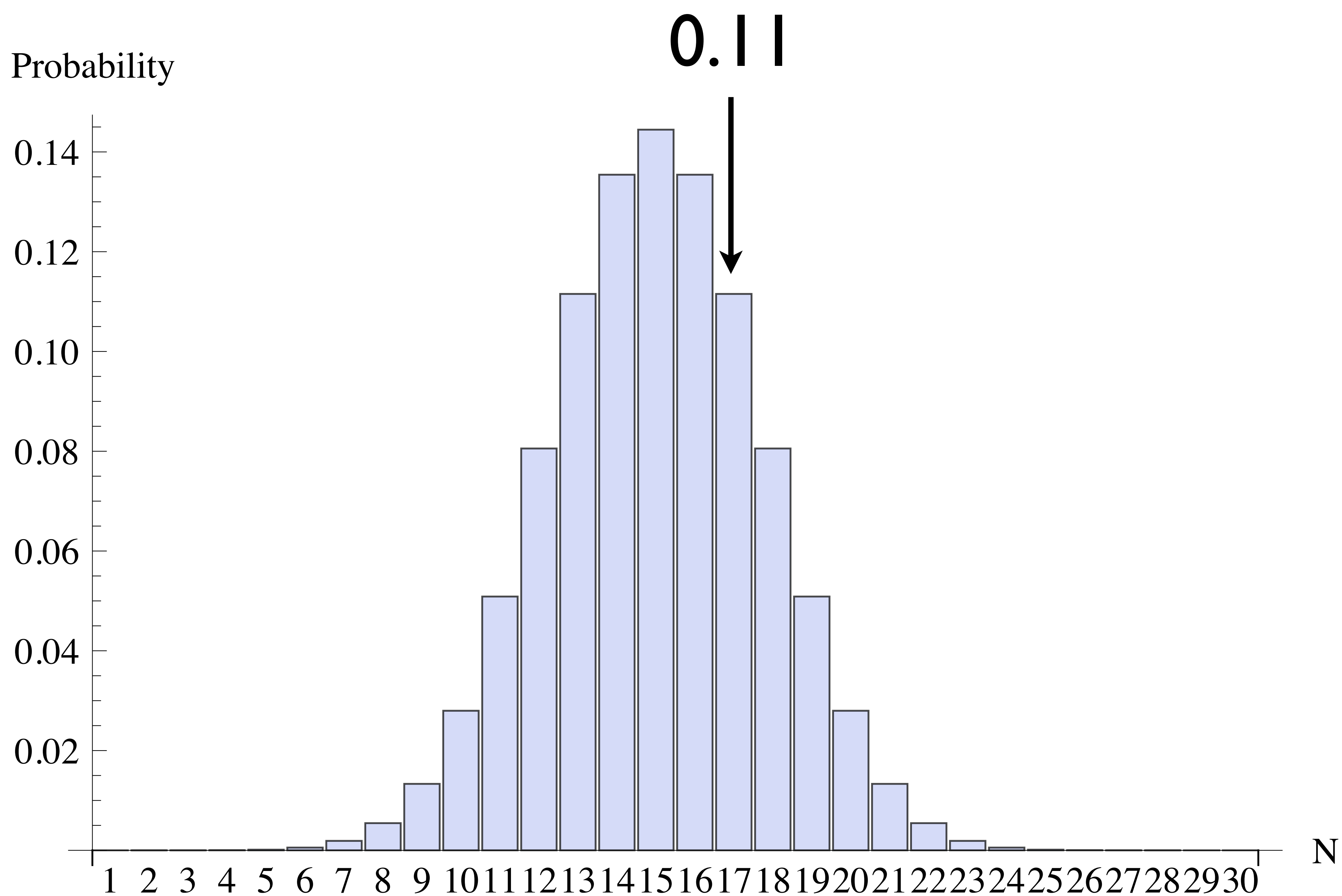
$$\texttt{likelyhood}(k, n, p) = \frac{n!}{k! \times (n - k)!} \times p^k \times (1 - p)^{n-k}$$

k = number of successes,

n = number of trials,

p = probability of event

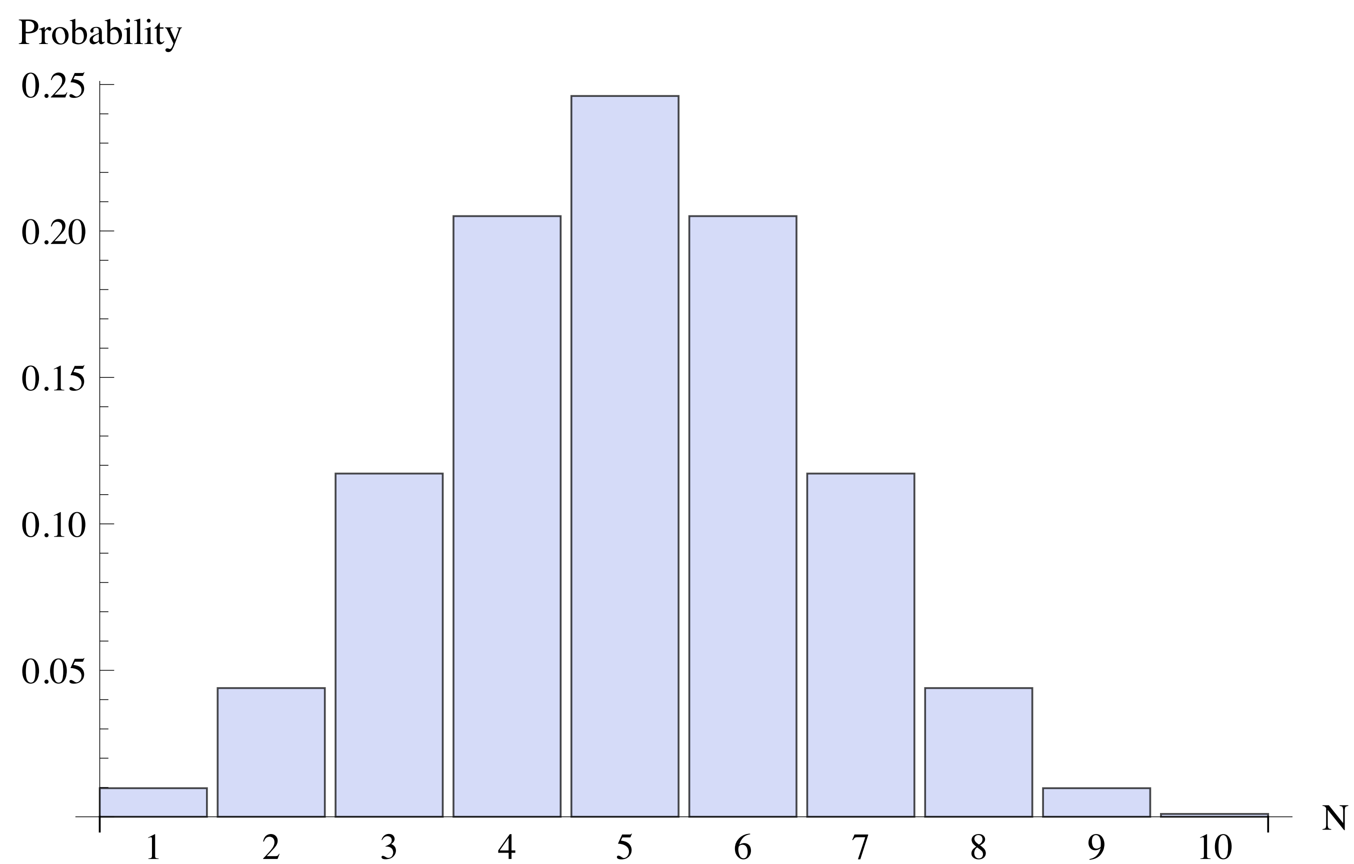
Likelihood of an event



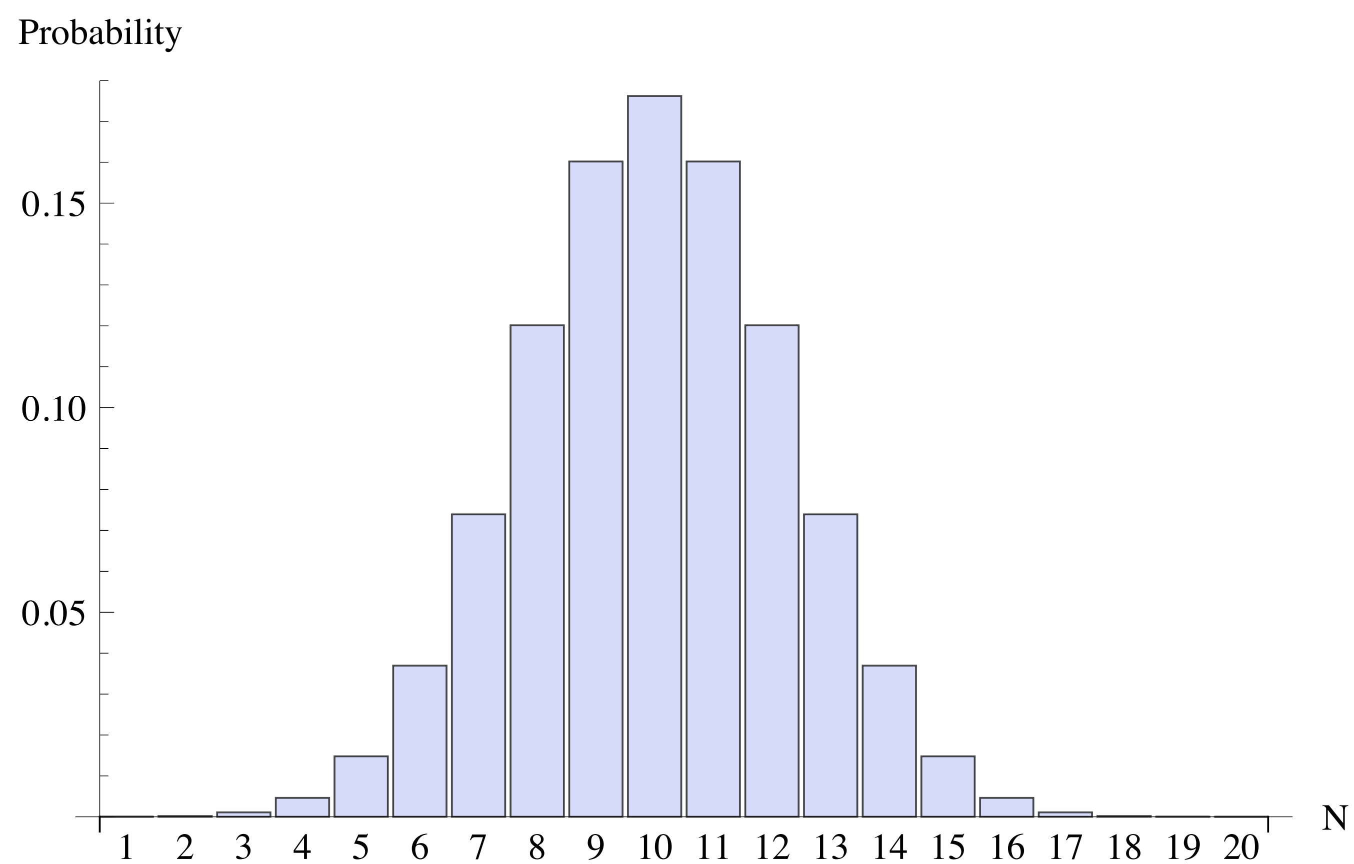
Likelihood N=3, P=0.5

Sample	Proportion of B	Probability of B
BBB	3/3	1/8
BBI	2/3	3/8
BIB	2/3	3/8
IBB	2/3	3/8
BII	1/3	3/8
IBI	1/3	3/8
IIB	1/3	3/8
III	0/3	1/8

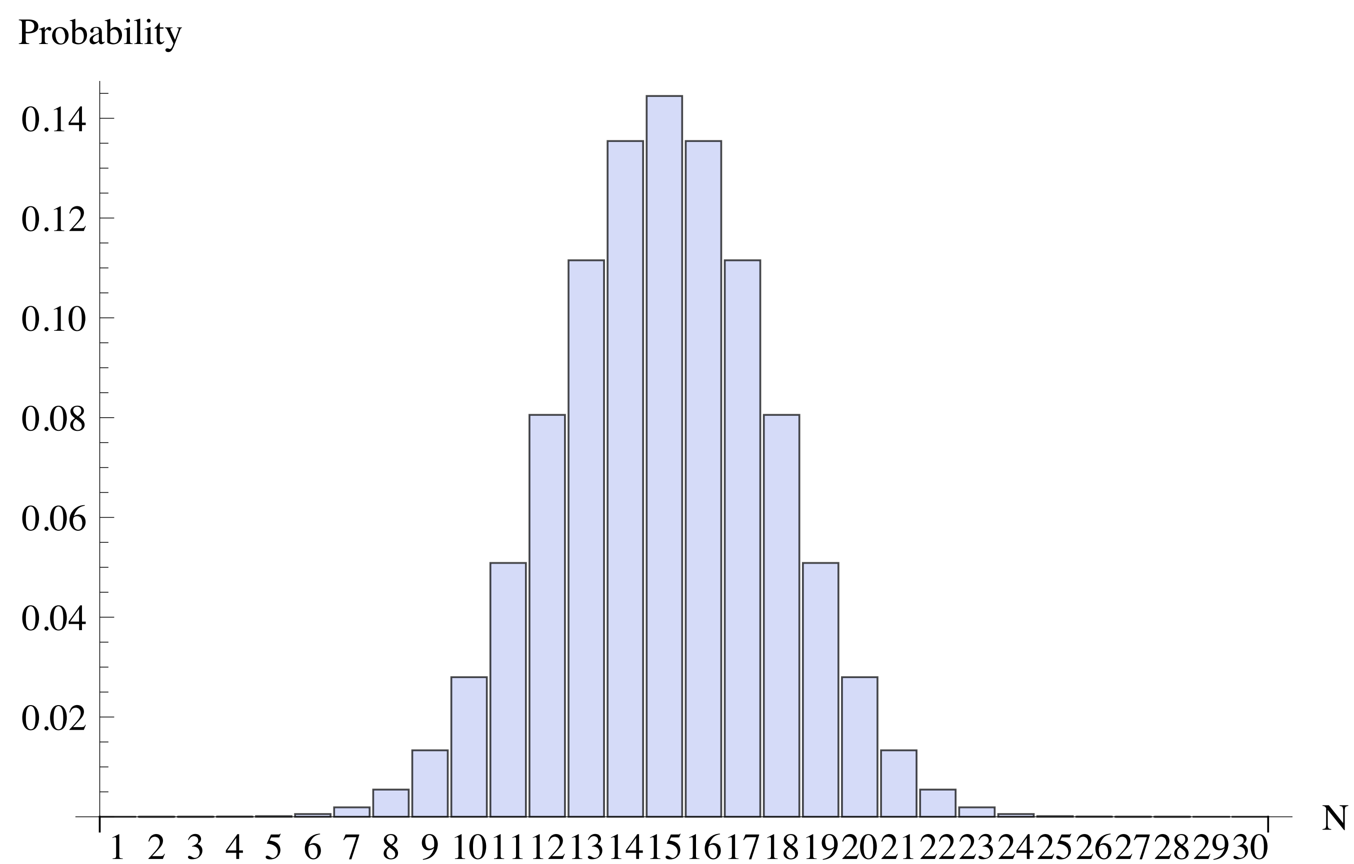
Increasing N to10



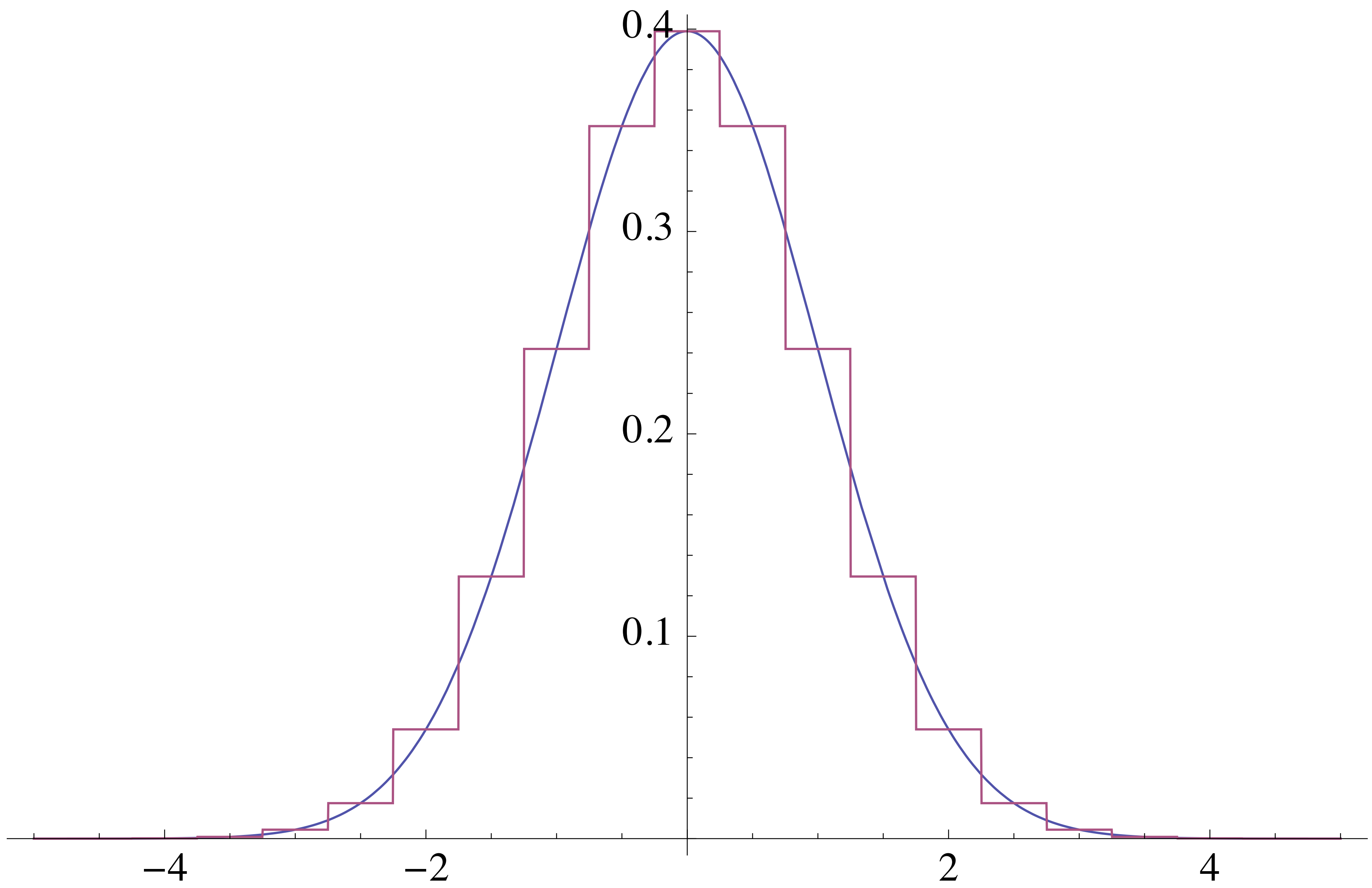
Increasing N to 20



Increasing N to 30



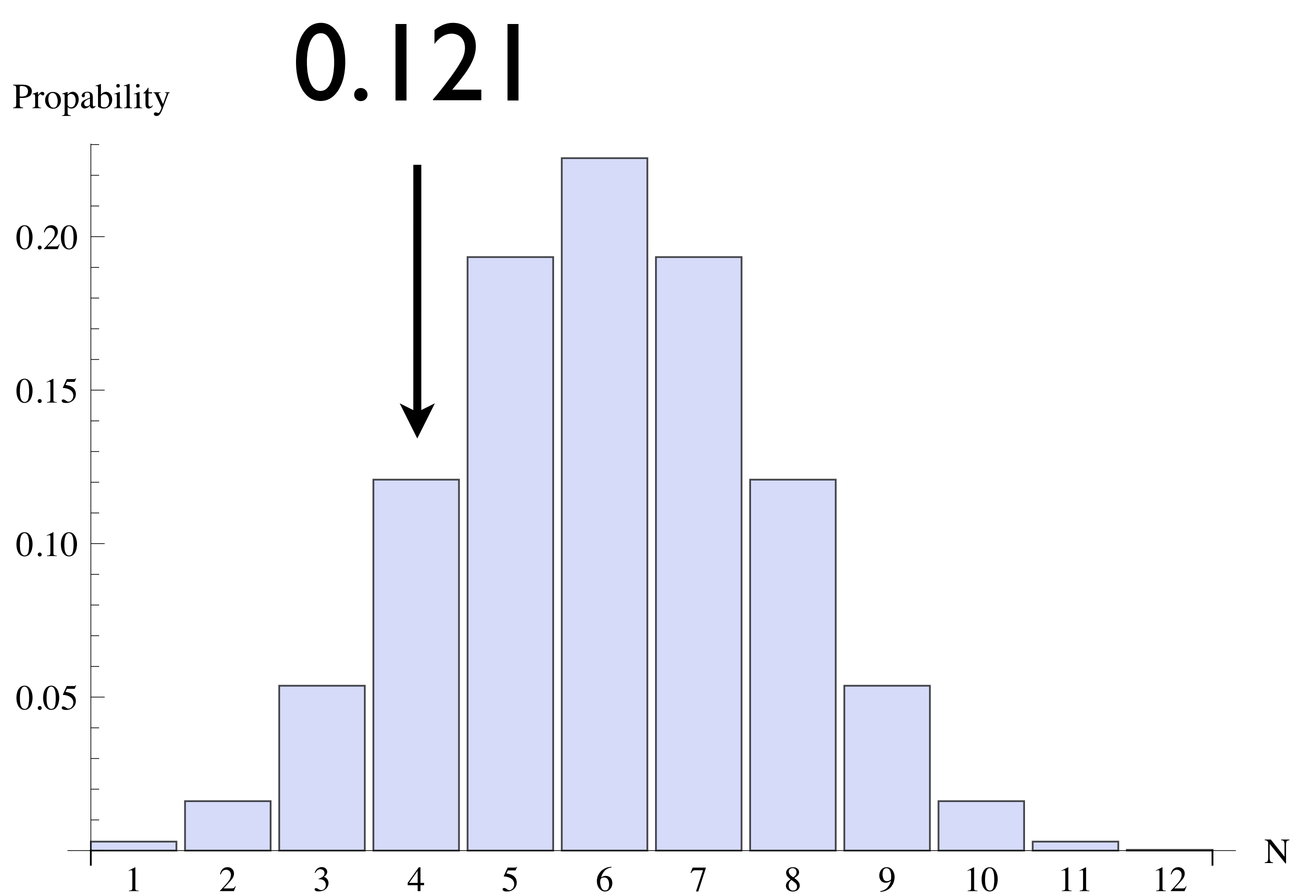
binomial distribution approaches normal distribution



normal distribution

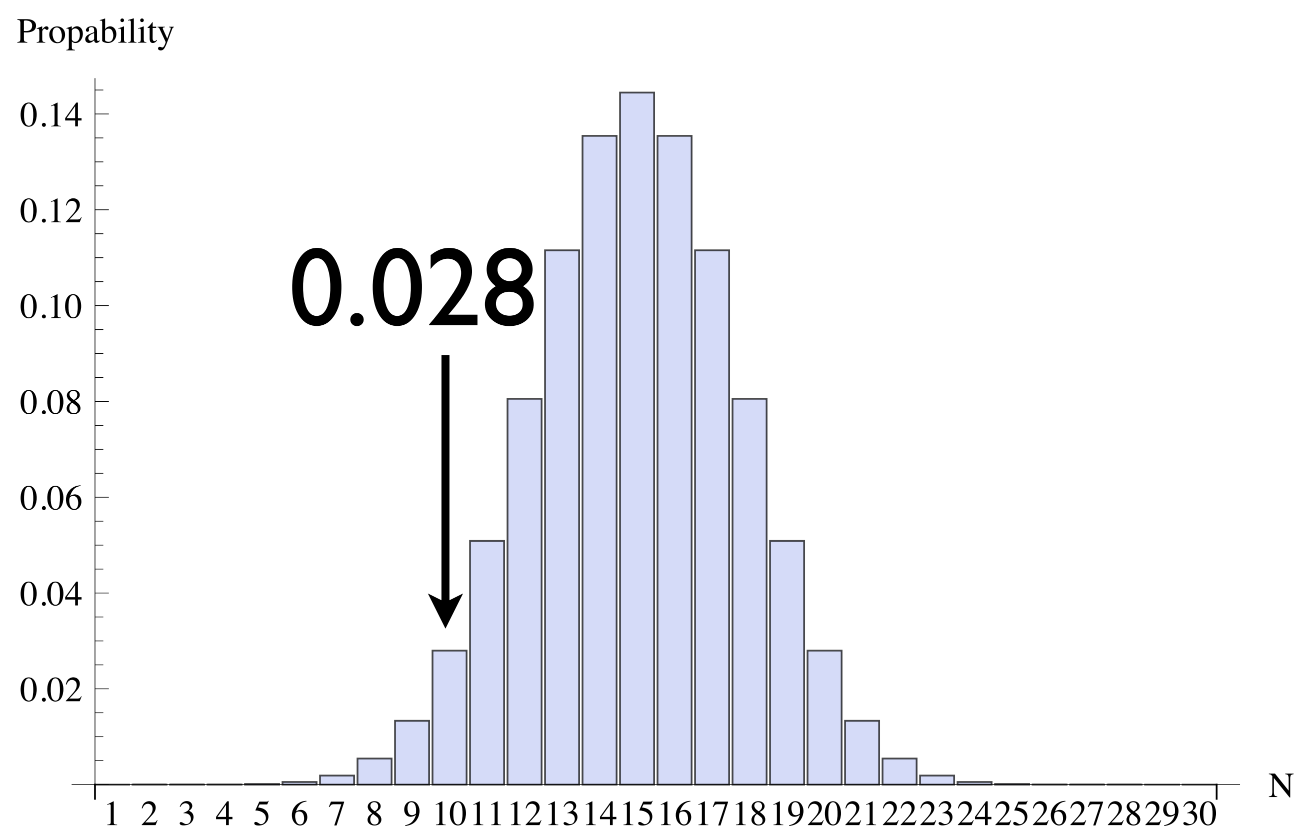
$$\texttt{normDist}(x) = \frac{1}{\sqrt{2\pi}} \times e^{\frac{-x^2}{2}}$$

Sample error N=12, P=0.5



0.33%

Sample error N=30, P=0.5



0.33%

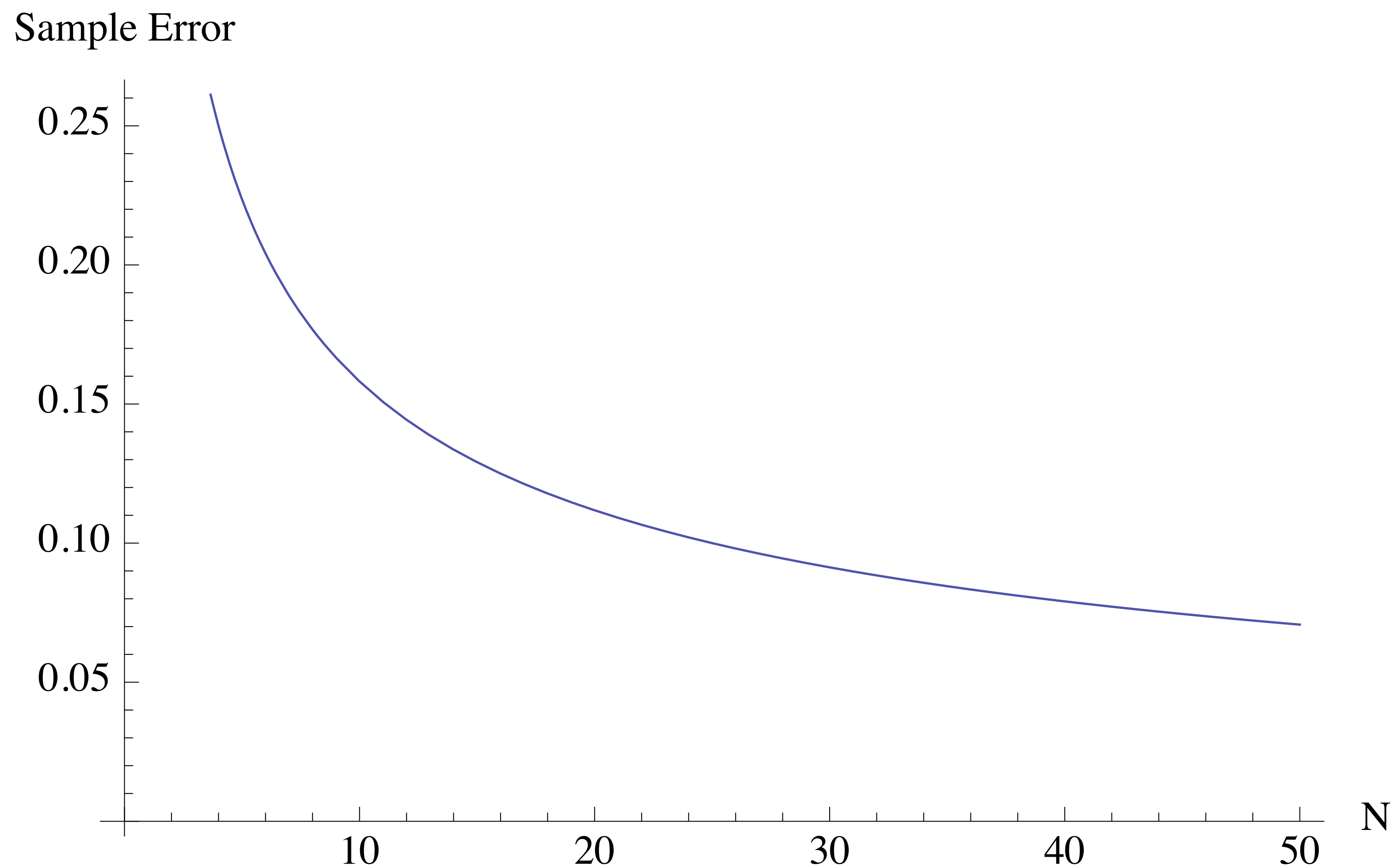
standard error of proportion

$$s_P = \sqrt{\frac{P\,(1 - P)}{N}}$$

standard error of mean

$$s_{\bar{x}} = \frac{s}{\sqrt{N}}$$

Sample error for proportions at P=0.5



Definitions

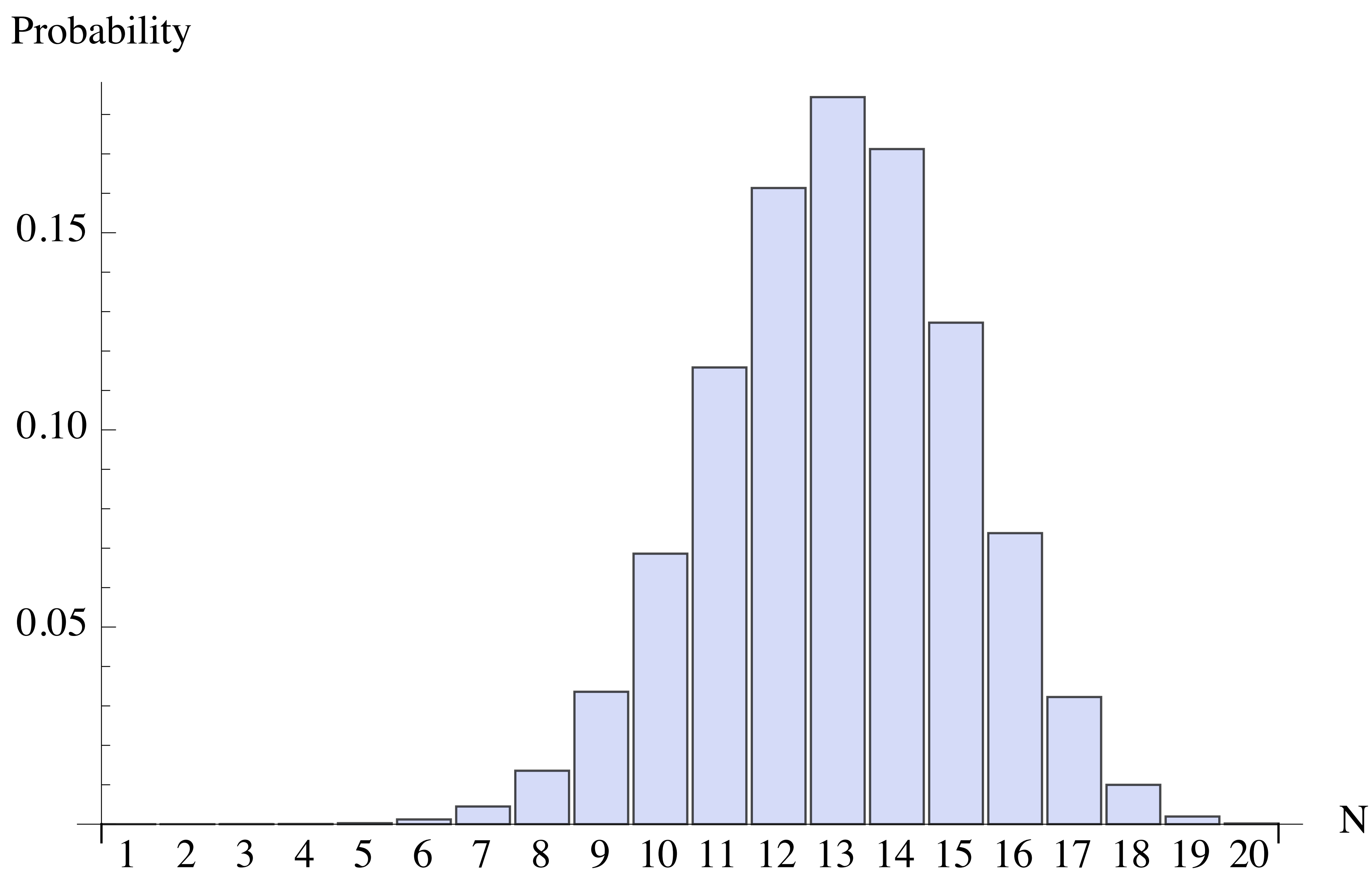
	sample	population
mean	$\bar{x}$	μ
proportion	p	Π
standard deviation	s	σ

Statistical model

$$\mathbf{P} = \Pi + \mathbf{e}$$

$$\overline{\mathbf{X}} = \mu + \mathbf{e}$$

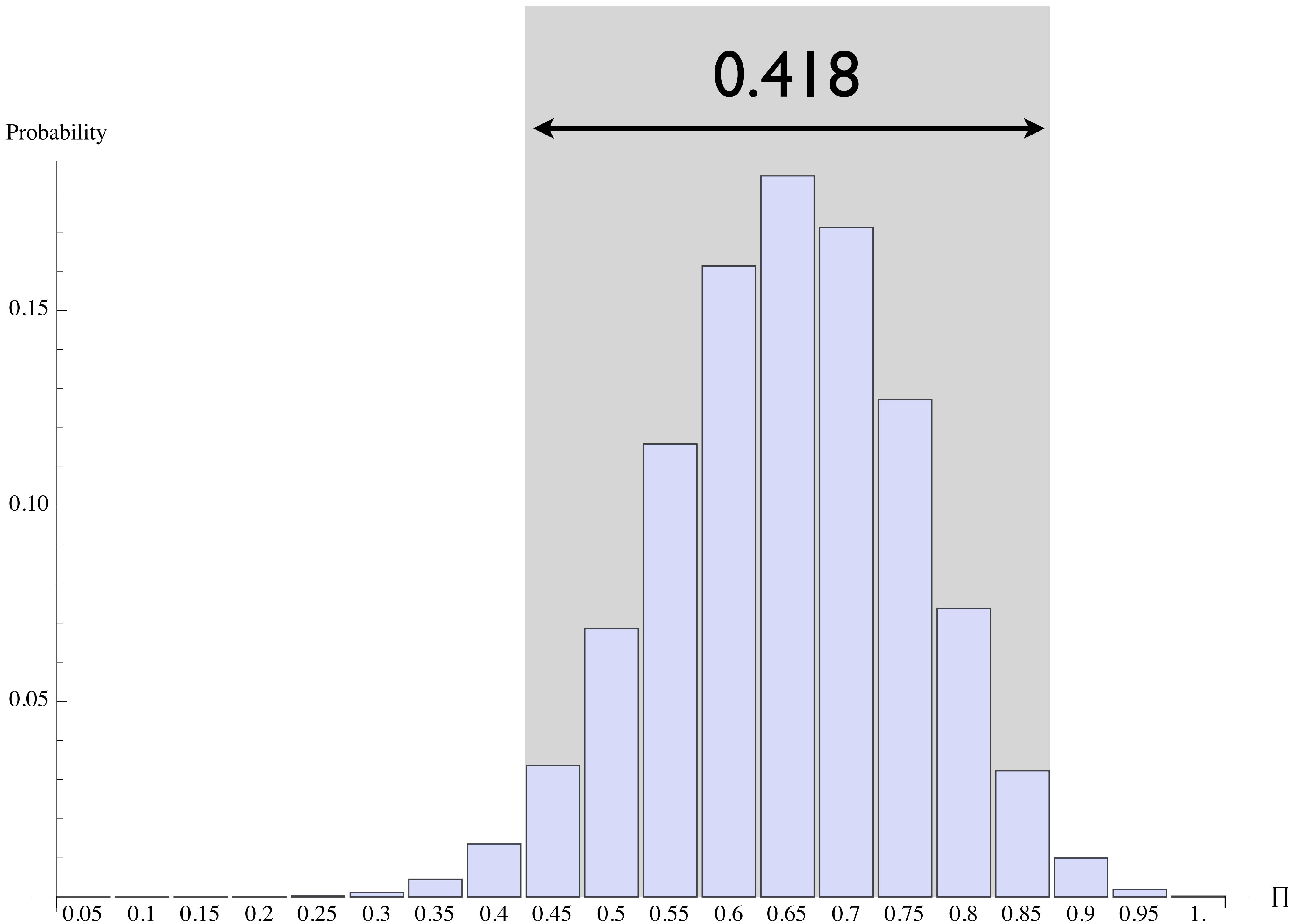
Boris gets 13/20 times tea



N=20

Successes=13

confidence interval of 95%

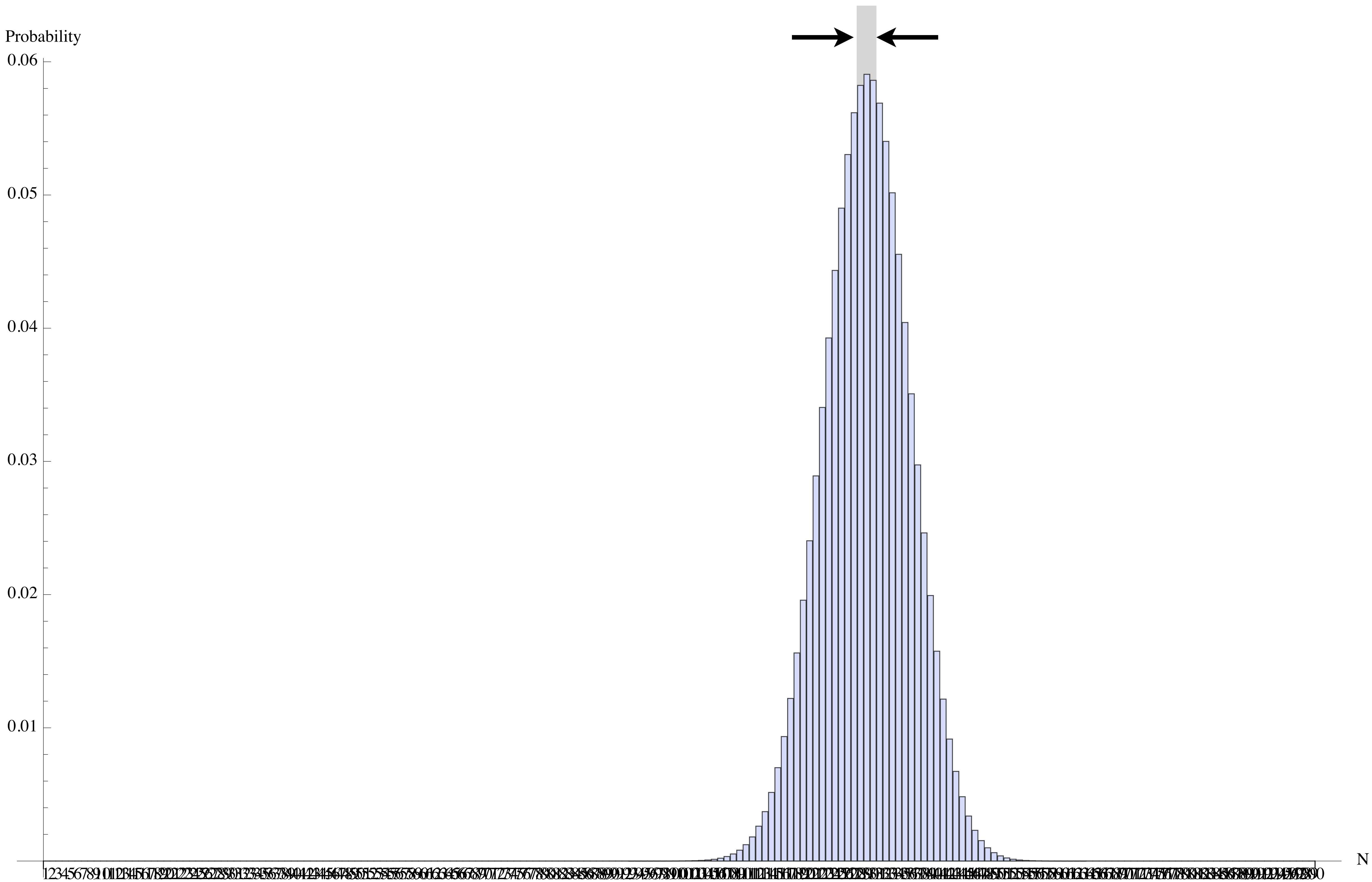


N=200

Successes=130

confidence interval of 95%

0.13



Comparison

	sample *a* mean	sample *a* proportion
fixed value	One sample t-test	Binomial test
sample *b* value	Independent samples t-test	Independent samples t-test

Comparison

	sample *a* mean	sample *a* proportion
fixed value	One sample t-test	Binomial test
sample *b* value	Independent samples t-test	Independent samples t-test

Binomial test, N>20

1. Collect a sample of N observations

2. Calculate the sample proportion and standard error

3. set α to 0.05

4. Using table A2, find the value $\alpha/2$ in the column labeled "area beyond z", and read off the value of z in the same row. Denote this z value by:

   $$z_{\frac{\alpha}{2}}$$

5. Compute the half-width of the confidence interval as:

   $$\omega = z_{\frac{\alpha}{2}} \times s_P$$

6. Put everything into this formula:

$$\Pr(P - \omega < \Pi < P + \omega) = 1 - \alpha$$

Binomial Example

1. Bert and Ingrid got me tea. Bert got it 17 out of 30 times. N=30
2. Sample proportion P=17/30=0.566, sample error = 0.9047 ⟶ $s_P = \sqrt{\frac{P(1-P)}{N}}$
3. α=0.05
4. Lookup $z_{\frac{\alpha}{2}}$ =1.96
5. ω=1.96*0.09047=0.1773 ⟶ $\omega = z_{\frac{\alpha}{2}} \times s_P$
6. Pr (0.388<π<0.743)=0.95 ⟶ $Pr(P - \omega < \Pi < P + \omega) = 1 - \alpha$

Comparison

	sample *a* mean	sample *a* proportion
fixed value	One sample t-test	Binomial test
sample *b* value	Independent samples t-test	Independent samples t-test

Confidence Interval for Mean, N<20

1. Collect a sample of N observations
2. Calculate the sample mean, standard deviation and standard error
3. set α to 0.05
4. Using table A3, find the cell entry in the row corresponding do df = N-1 and the column corresponding to a two-tailed area of α. Denote this cell entry by:

   $$t_{\frac{\alpha}{2}}$$

5. Compute the half-width of the confidence interval as:

   $$\omega = t_{\frac{\alpha}{2}} \times s_{\overline{X}}$$

6. Put everything into this formula:

   $$\Pr\left(\overline{X} - \omega < \mu < \overline{X} + \omega\right) = 1 - \alpha$$

One sample t-test

- Traditional t-test

$$\mathbf{t\ (df)} = \frac{\overline{\mathbf{X}} - \mu}{\mathbf{s}_{\overline{\mathbf{X}}}}$$

$$\mathbf{Pr}\left(-\mathbf{t}_{\frac{\alpha}{2}} < \mathbf{t\ (df)} < +\mathbf{t}_{\frac{\alpha}{2}}\right) = \mathbf{1} - \alpha$$

- Report: $t(15)=1.33$, $p>0.05$

Example

1. We assembled 16 IQ tests
2. The sample mean was 103, standard deviation of 9, → $s(X) = \sqrt{\frac{\sum (X - \overline{X})^2}{N - 1}}$
   standard error of 2.25 → $s_{\overline{X}} = \frac{s}{\sqrt{N}}$
3. α set to 0.05
4. lookup df=N-1, $t_{\frac{\alpha}{2}}$=2.1314
5. t(df)=(103-100)/2.25=1.333 → $t(df) = \frac{\overline{X} - \mu}{s_{\overline{X}}}$
6. Pr(-2.1314<t(df)<+2.1314)=0.95 → $Pr\left(-t_{\frac{\alpha}{2}} < t(df) < +t_{\frac{\alpha}{2}}\right) = 1 - \alpha$
7. t(15)=1.333, p>0.05

Within subjects t-test

- same as one sample t-test, but use the difference in the score for the mean
- Participant score after - participant score before

Confidence Interval for Mean, N>20

1. Collect a sample of N observations
2. Calculate the sample mean, standard deviation and standard error
3. set α to 0.05
4. Using table A2, find the value $\alpha/2$ in the column labeled "area beyond z", and read off the value of z in the same row. Denote this z value by:

   $$z_{\frac{\alpha}{2}}$$

5. Compute the half-width of the confidence interval as:

   $$\omega = z_{\frac{\alpha}{2}} \times s_{\overline{X}}$$

6. Put everything into this formula:

   $$\Pr\left(\overline{X} - \omega < \mu < \overline{X} + \omega\right) = 1 - \alpha$$

Comparison

	sample *a* mean	sample *a* proportion
fixed value	One sample t-test	Binomial test
sample *b* value	Independent samples t-test	Independent samples t-test

Independent samples t-test for means

$$\mathbf{Pr}\left(\overline{X} - \omega < \mu < \overline{X} + \omega\right) = 1 - \alpha$$

$$\mathbf{Pr}\left(\overline{X}_1 - \overline{X}_2 - \omega < \mu_1 - \mu_2 < \overline{X}_1 - \overline{X}_2 + \omega\right) = 1 - \alpha$$

$$\omega = t_{\frac{\alpha}{2}} \times s_{\overline{X}}$$

$$s^2_{pooled} = \frac{(N_1 - 1) \times s^2_1 + (N_2 - 1) \times s^2_2}{(N_1 - 1) + (N_2 - 1)}$$

$$s_{err} = \sqrt{s^2_{pooled} \times \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}$$

Example

1. Collect random samples

X_1	X_2
8	5
6	6
7	8
6	3
8	4
	6
	3

2. Calculate means, variance

$N_1 = 5,\ \bar{X}_1 = 7,\ s_1^2 = 1$

$N_2 = 7,\ \bar{X}_2 = 5,\ s_2^2 = \frac{10}{3}$

$\bar{X}_1 - \bar{X}_2 = 2$

3. set alpha to 0.05

Example

4. Find half-width confidence interval

- Using table A3 find the cell entry in the row corresponding to df=N_1+N_2-2 and the column corresponding to a two-tailed area of alpha. Denote this cell entry by:

  $$t_{\frac{\alpha}{2}}$$

- The half-width of the confidence interval is:

  $$\omega = t_{\frac{\alpha}{2}} \times s_{err}$$

Example

5. Enter the values in the formula

$$\mathbf{Pr}\left(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 - \omega < \mu_1 - \mu_2 < \overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2 + \omega\right) = 1 - \alpha$$

Example: Confidence Interval

$$s^2_{pooled} = \frac{(N_1 - 1) \times s_1^2 + (N_2 - 1) \times s_2^2}{(N_1 - 1) + (N_2 - 1)} =$$

$$\frac{(5 - 1) \times (1) + (7 - 1) \times (3.3)}{(5 - 1) + (7 - 1)} = \frac{4 + 20}{4 + 6} = 2.4$$

2.4

$$s_{err} = \sqrt{s^2_{pooled} \times \left(\frac{1}{N_1} + \frac{1}{N_2}\right)} = \sqrt{2.4 \times \left(\frac{1}{5} + \frac{1}{7}\right)} = 0.907$$

0.907

look up

$$\omega = t_{\frac{\alpha}{2}} \times s_{err} = (2.228)\ (0.907) = 2.021$$

$$Pr\left(\overline{X}_1 - \overline{X}_2 - \omega < \mu_1 - \mu_2 < \overline{X}_1 - \overline{X}_2 + \omega\right) = 1 - \alpha$$

$$Pr\ (2 - 2.021 < \mu_1 - \mu_2 < 2 + 2.021) = 1 - 0.05$$

$$Pr\ (-0.021 < \mu_1 - \mu_2 < 4.021) = 0.95$$

Example: t-test mean

$$t\,(df) = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{s_{err}} = \frac{2 - 0}{0.907} = 2.205$$

$$Pr\left(-t_{\frac{\alpha}{2}} < t\,(df) < t_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$Pr\,(-2.228 < t\,(10) < 2.228) = 0.95$$

$$t\,(10) = 2.205,\ p > 0.05$$

previous slide

Comparison

	sample *a* mean	sample *a* proportion
fixed value	One sample t-test	Binomial test
sample *b* value	Independent samples t-test	Independent samples t-test

t-test for proportions

$$\mathbf{Pr}\ (\mathbf{P}_1 - \mathbf{P}_2 - \omega < \Pi_1 - \Pi_1 < \mathbf{P}_1 - \mathbf{P}_2 + \omega) = 1 - \alpha$$

$$\omega = \mathbf{t}_{\frac{\alpha}{2}} \times \mathbf{s}_{\mathbf{err}}$$

$$\mathbf{s}_{\mathbf{err}} = \sqrt{\frac{\mathbf{P}_1\ (1 - \mathbf{P}_1)}{\mathbf{N}_1} + \frac{\mathbf{P}_2\ (1 - \mathbf{P}_2)}{\mathbf{N}_2}}$$

Study setup recommendations

- No more than two conditions
- Measurements can only be nominal (proportion) or interval (mean)
- Within or between subject design

Next Week Presentation

- 5 minutes per group
- Presentation should include:
  - Research question
  - Method
  - Experiences