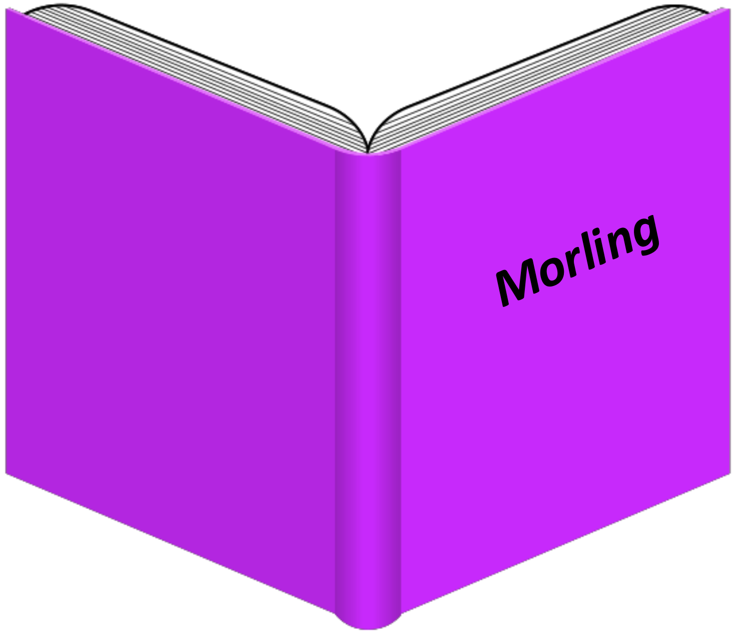
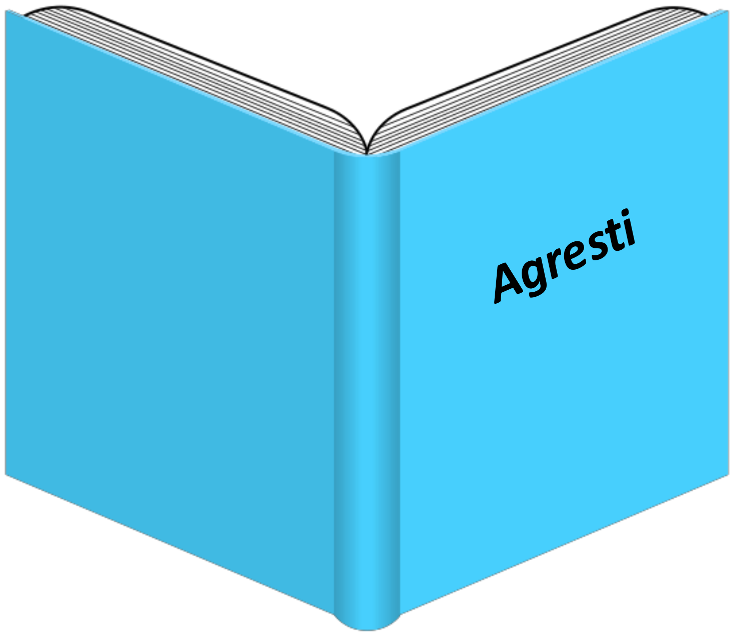


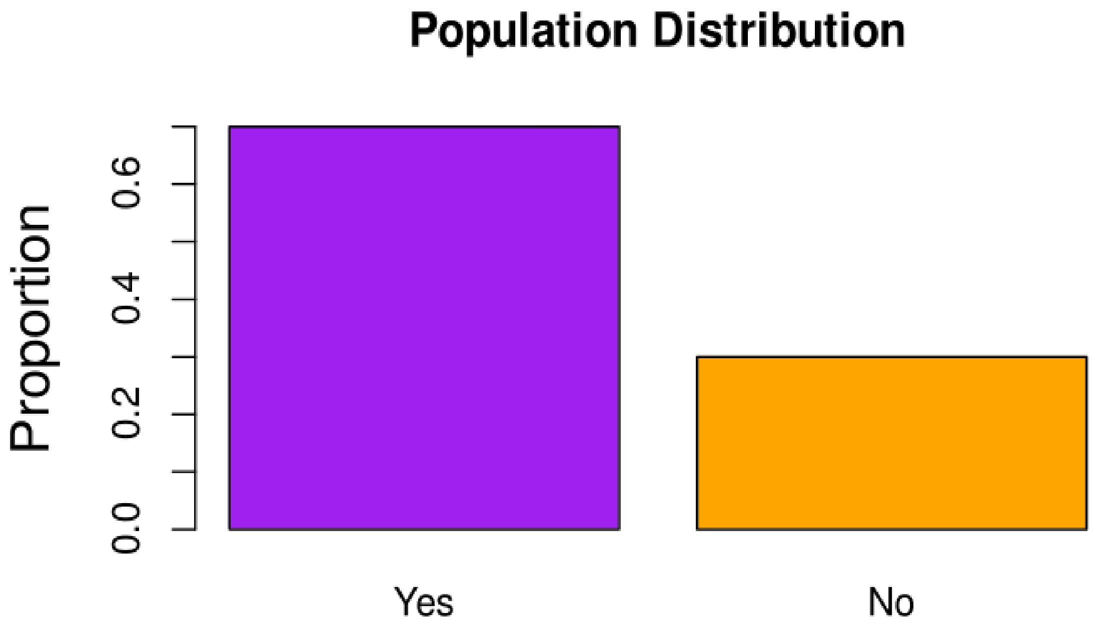


Research Methods and Statistics

Lecture 10: Sampling Distributions

Johnny van Doorn

Case 1: Voting behavior

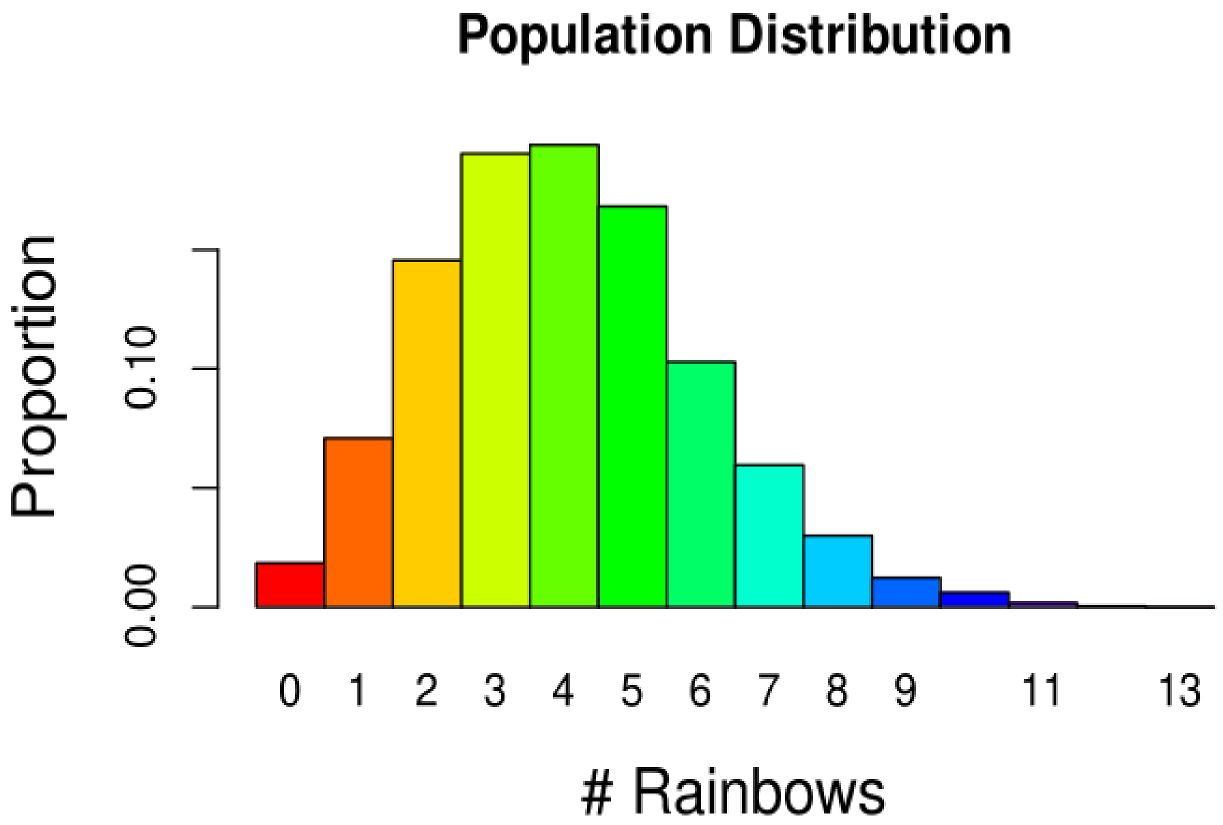


- Are you planning to vote in the upcoming elections?

Source: pixabay.com

Case 2: Rainbows

- How many rainbows have you seen this year?



Source: Pixabay.org

Overview of Today

1. **Population distribution vs data distribution**
2. Sampling distribution
    1. Distribution of a quantitative variable
    2. Distribution of a binary categorical variable
3. Central Limit Theorem
4. Excel & Demo
5. Recap
    - Next time
    - Example exam question

What is a sampling distribution?!

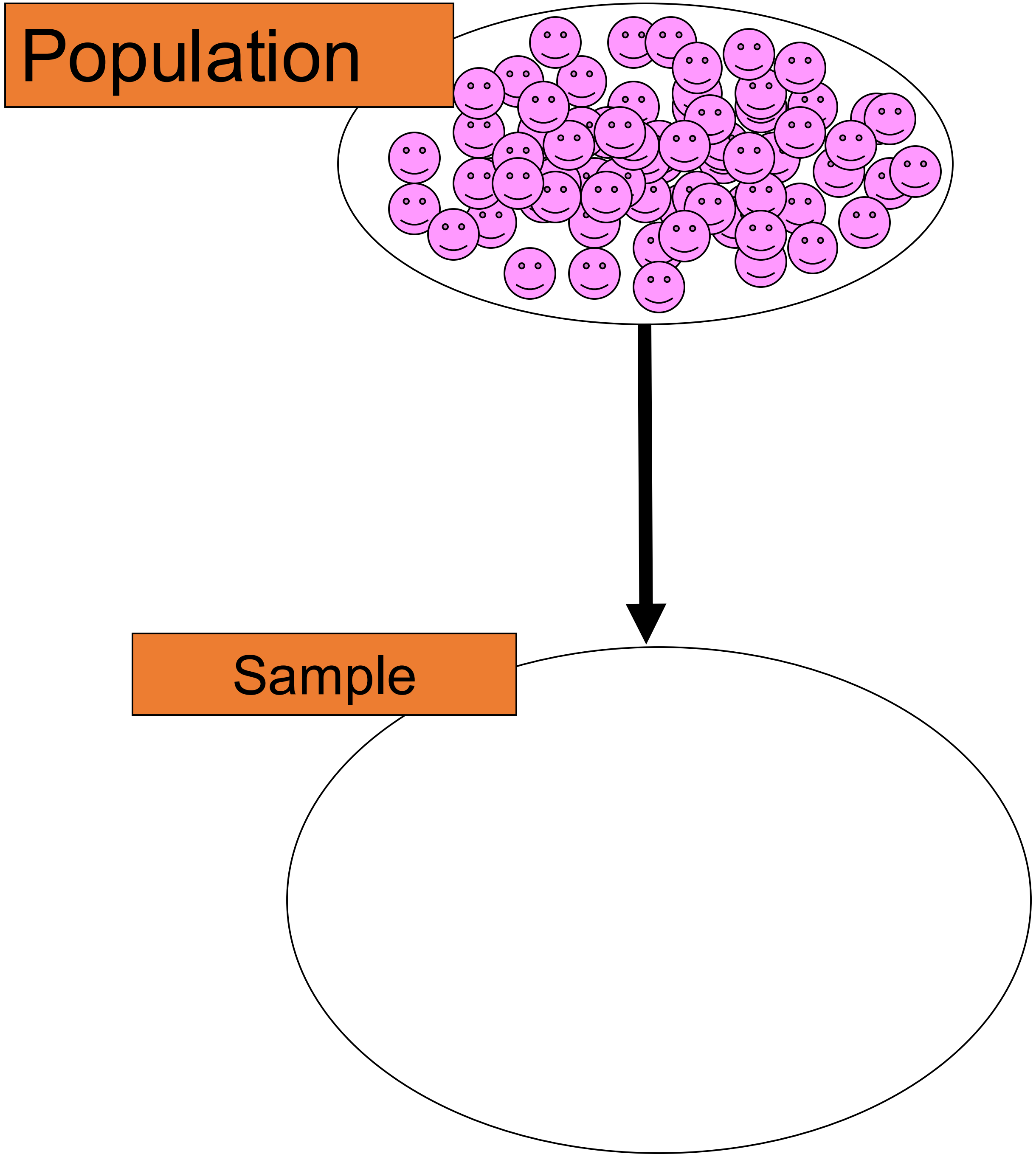
How can we use it for statistical inference?

Why are statistics needed again (Lecture 3)?

- Inferential statistics (vs. descriptive statistics)
  - Making predictions and decisions about the population

Population: **All** scores/data we are interested in
Sample: The part of the population that we have actually observed
Inference: Drawing a conclusion about the population, based on the sample

- Many things can go wrong when drawing inferences



Parameter

Statistic

Statistic vs parameter

A statistic: A numerical summary of the **data** is called a statistic

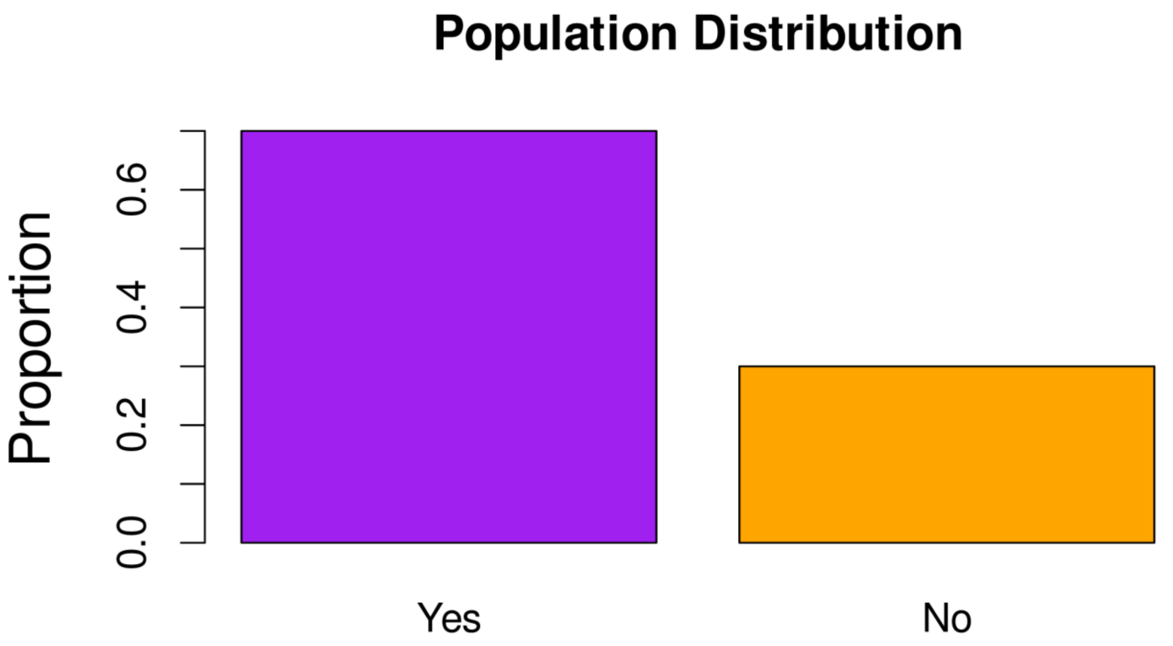
- You can compute this! (mean, median, standard deviation, etc)

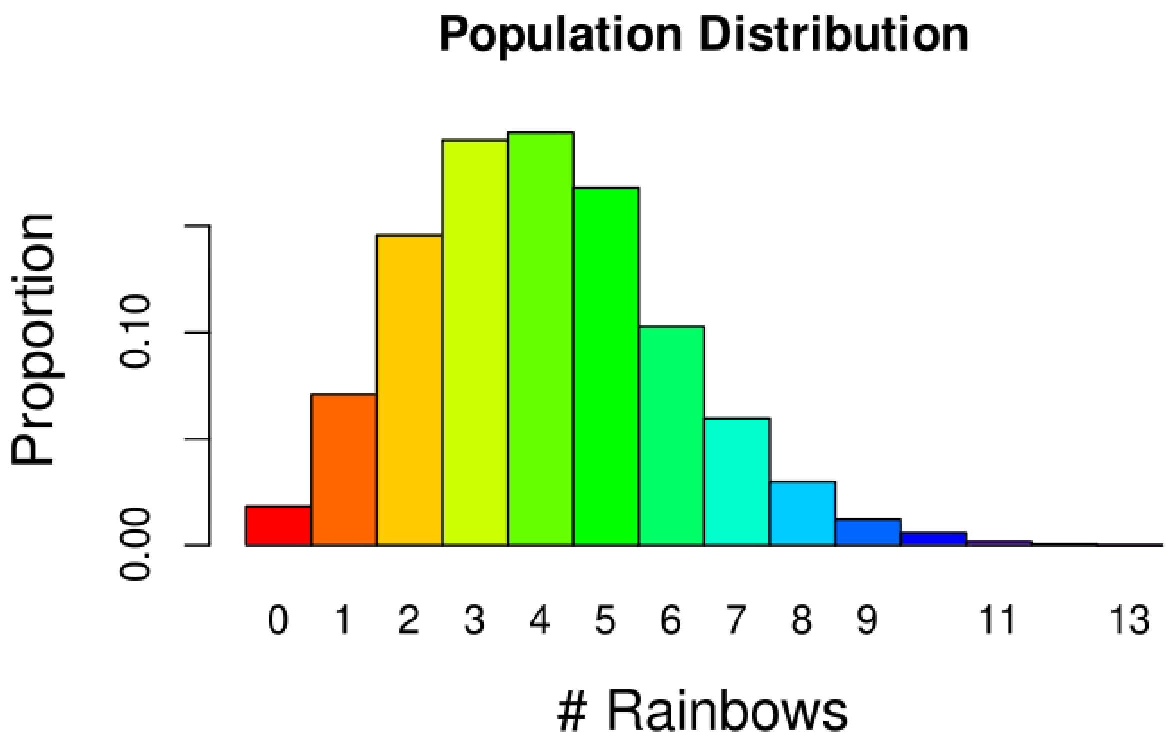
A *parameter*: A numerical summary of the **population** is called a parameter

- This cannot be computed (usually)

- ❼ **We use statistics to estimate parameters**
- *Sampling distribution* tells us something about the uncertainty of that estimate

Population distribution





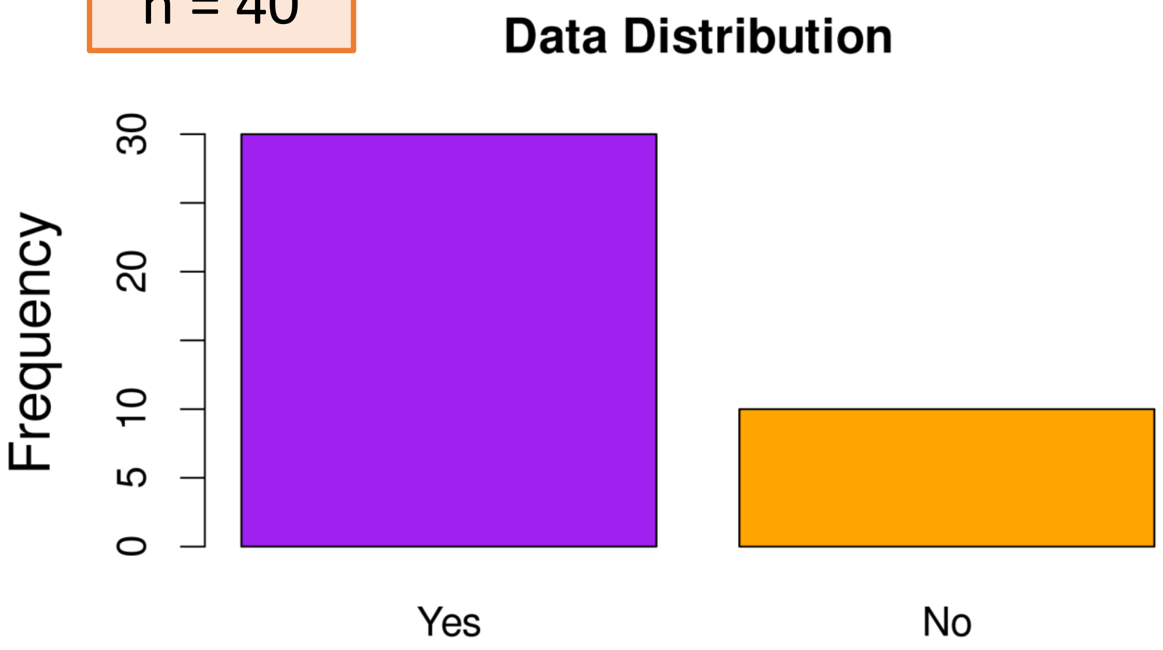
$p = 0.7$

Parameter

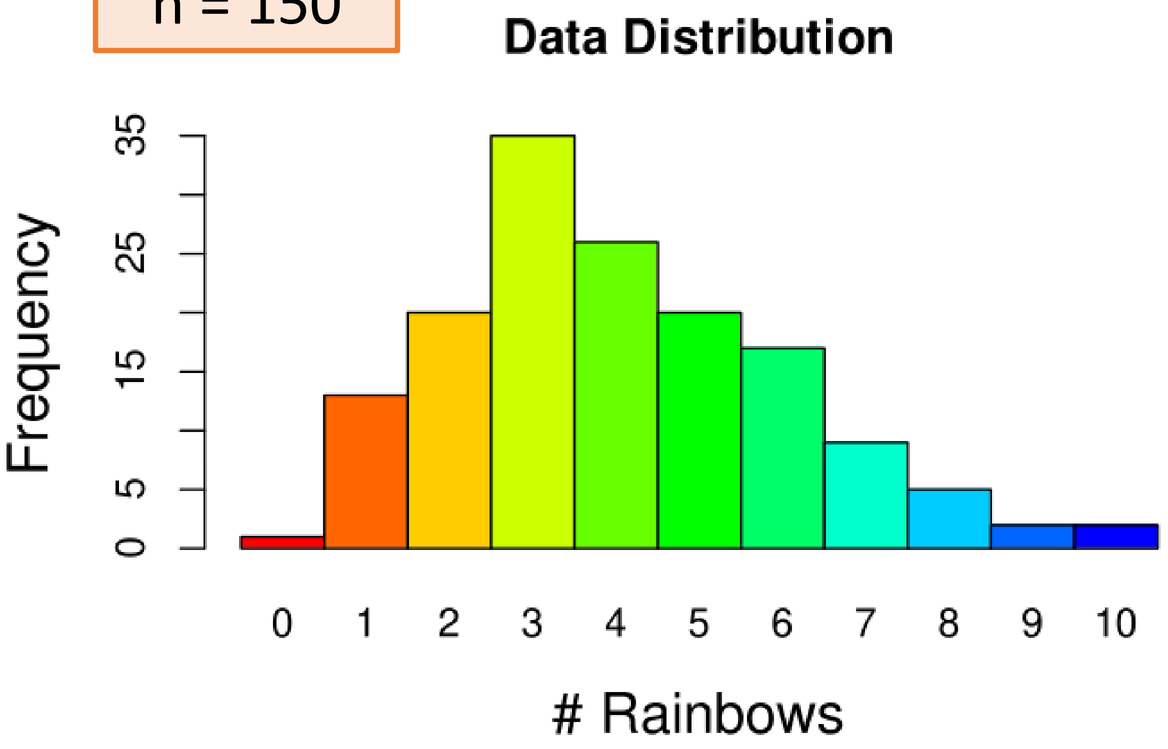
$\mu = 4$

Data distribution

n = 40



n = 150



Note the symbol ("p-hat"). The book does not introduce this until Ch 8.

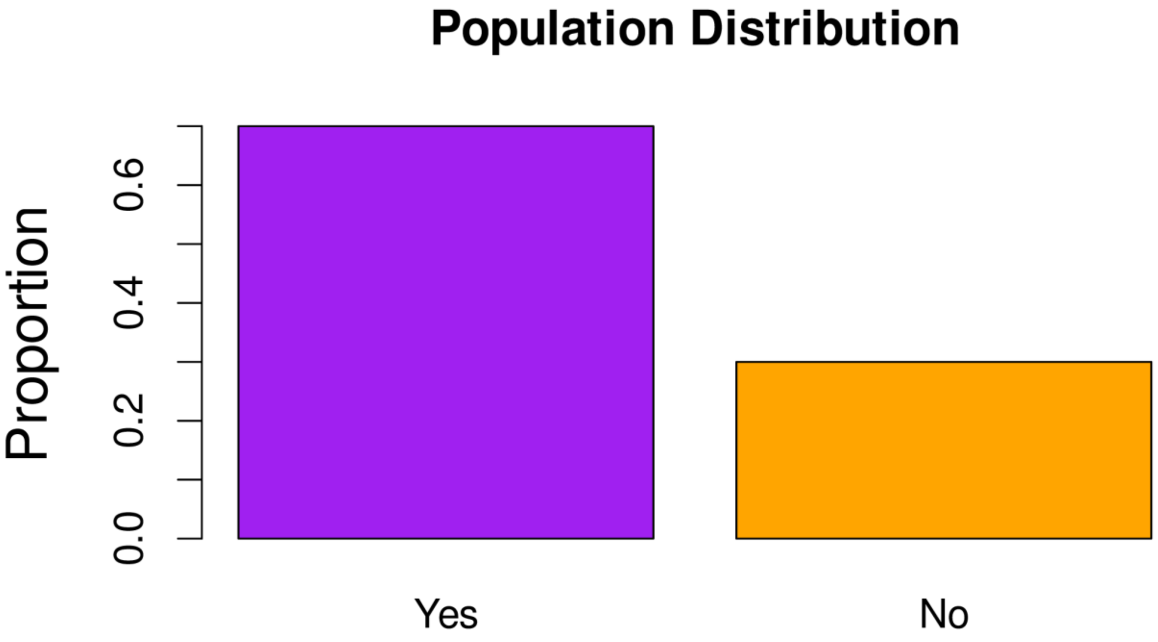
$\hat{p} = 0.75$

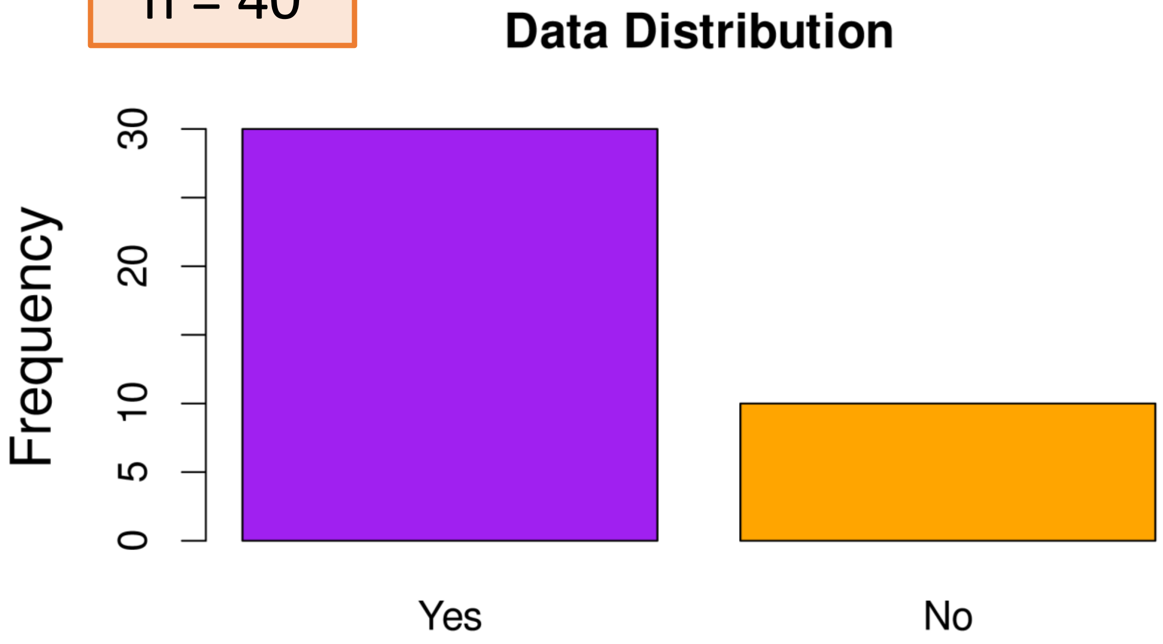
Statistic

$\bar{x} = 3.89$

Population and data distribution

n = 40





Parameter

$p = 0.7$

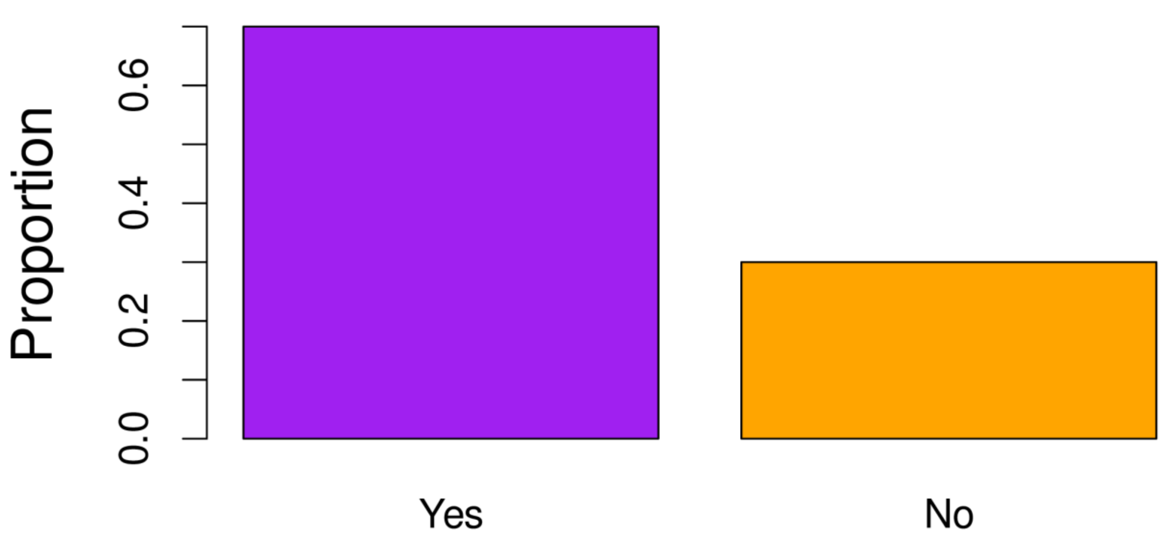
Statistic

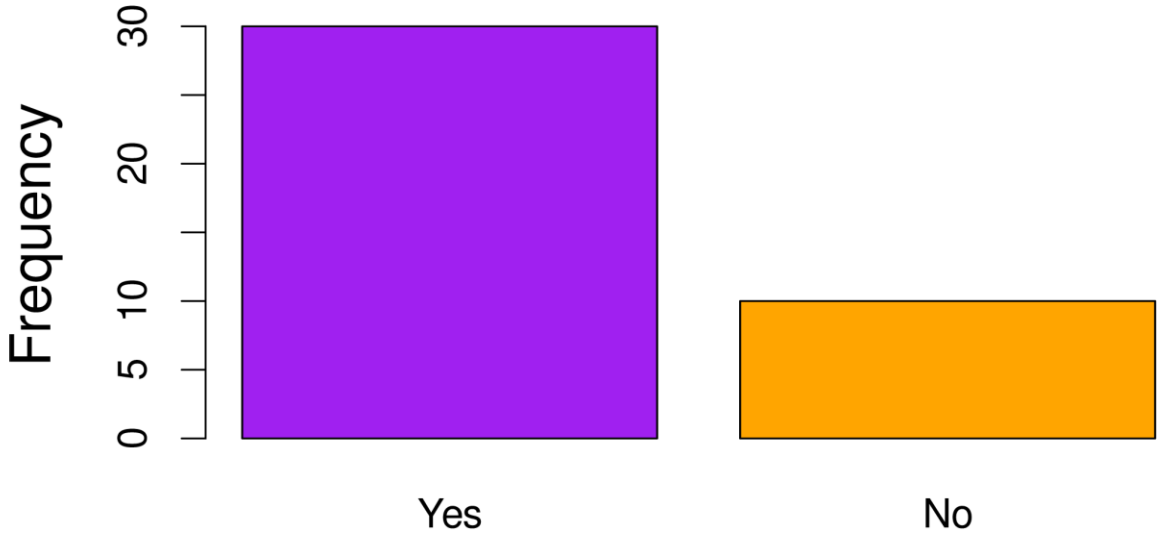
$\hat{p} = 0.75$

Population and data distribution

Spot the 3 differences!

n = 40





Parameter

$p = 0.7$

Statistic

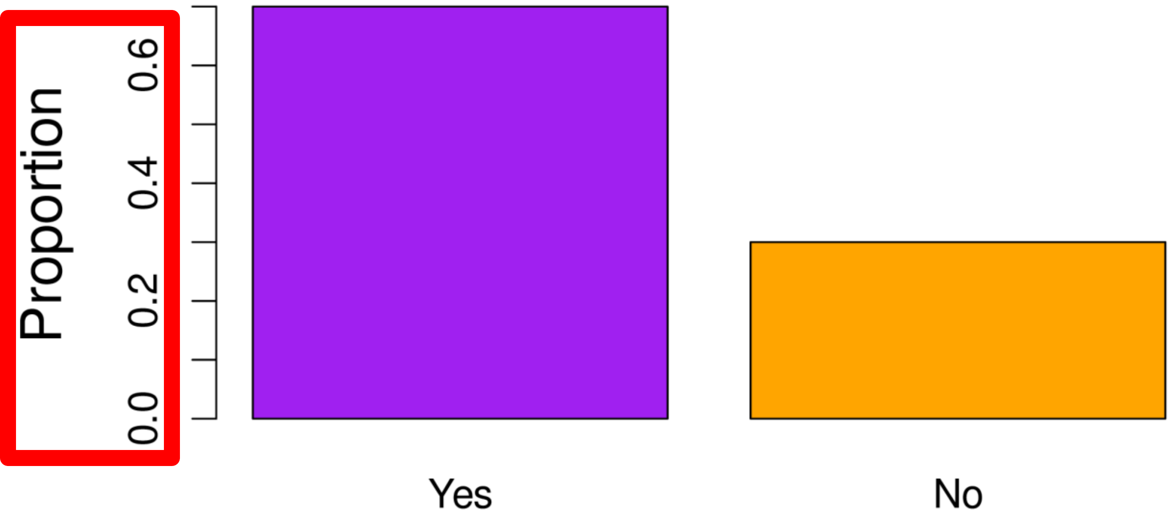
$\hat{p} = 0.75$

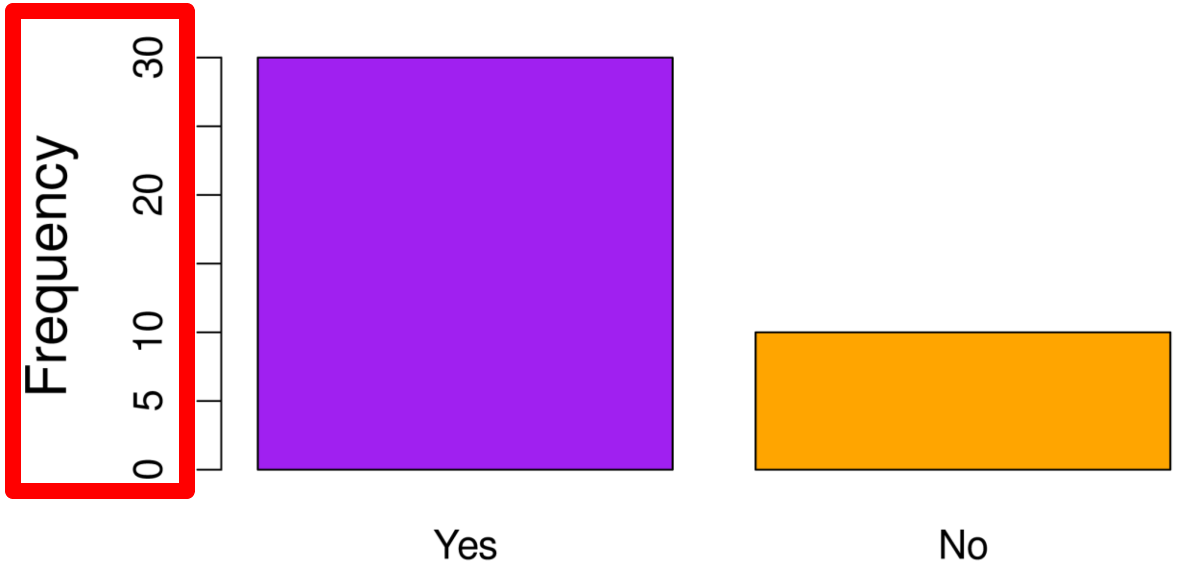
Population and data distribution

Spot the 3 differences!

Different y-axis

n = 40





Parameter

$p = 0.7$

Statistic

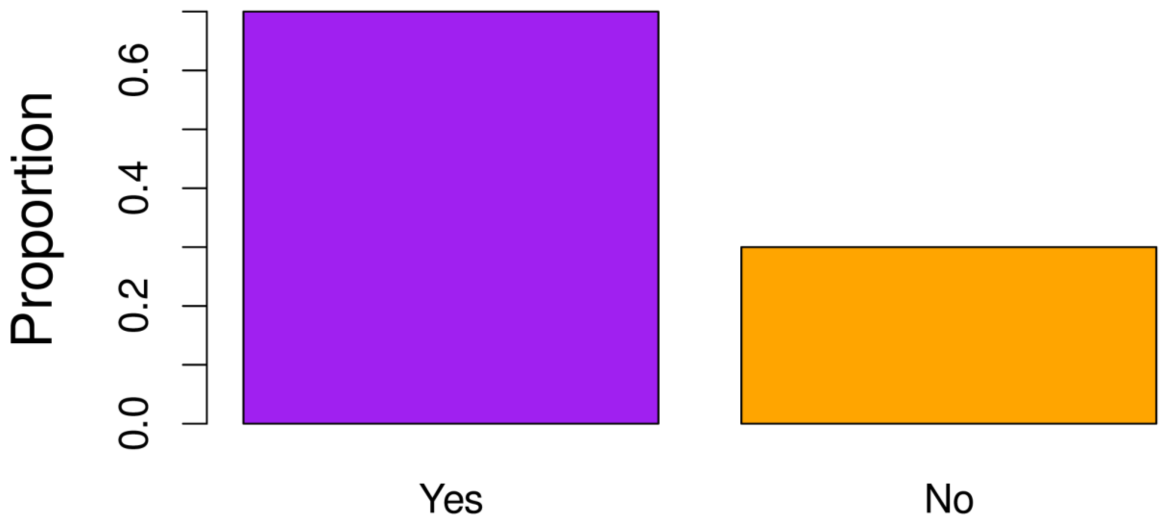
$\hat{p} = 0.75$

Population and data distribution

Spot the 3 differences!

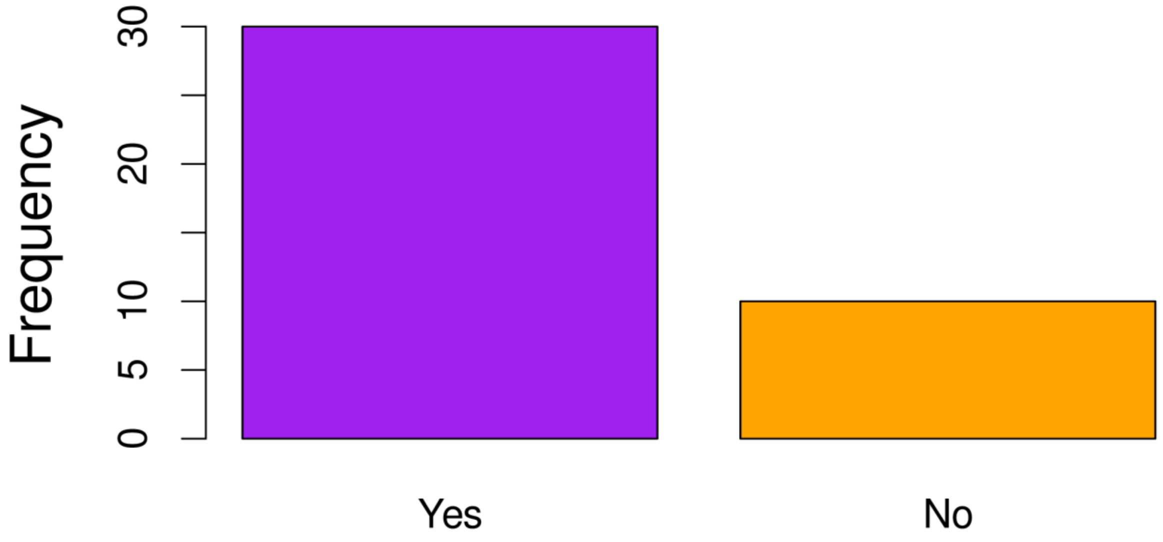
Sample size

n = 40



Parameter

$p = 0.7$



Statistic

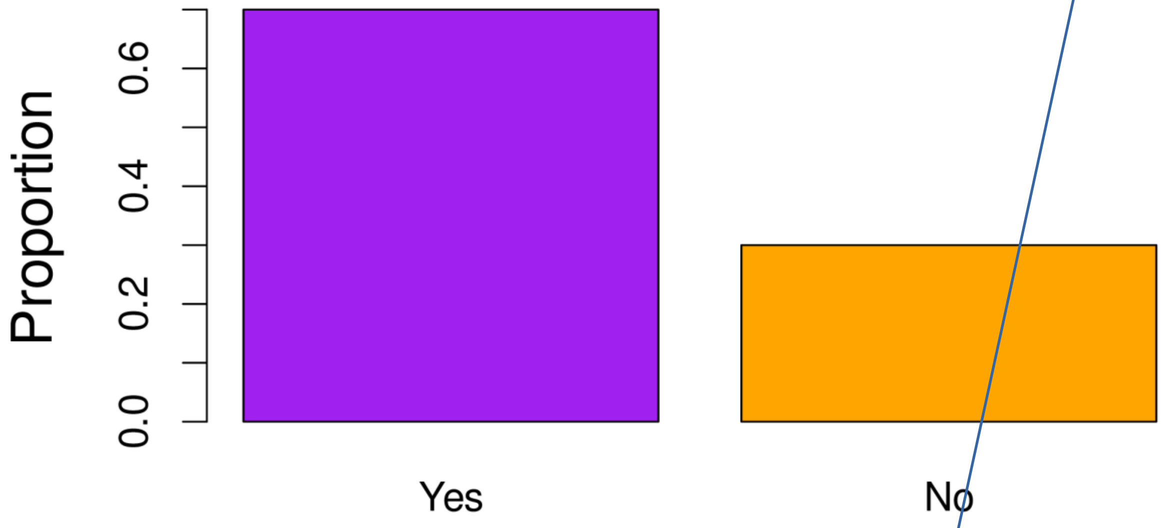
$\hat{p} = 0.75$

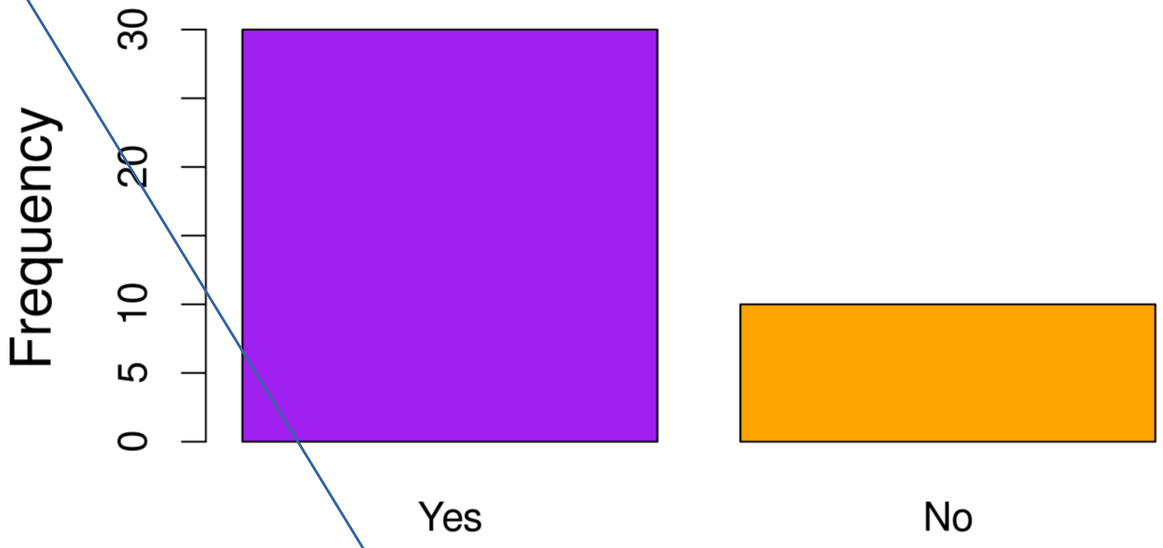
Population and data distribution

Spot the 3 differences!

Different notation

n = 40





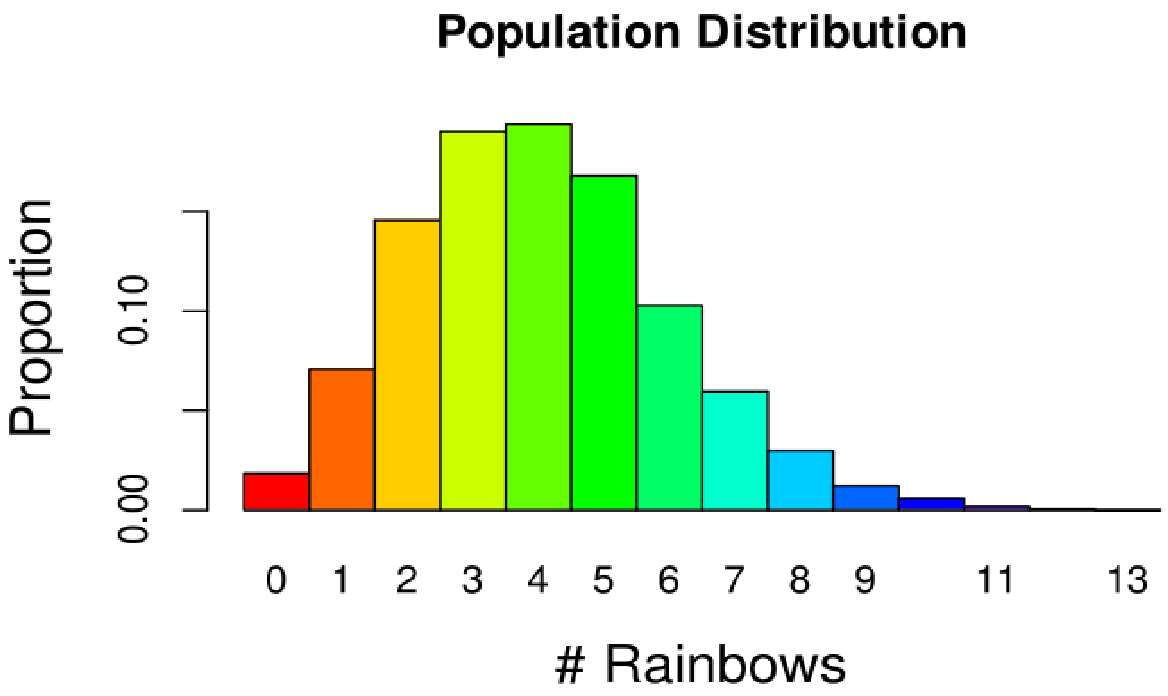
Parameter

$p = 0.7$

Statistic

$\hat{p} = 0.75$

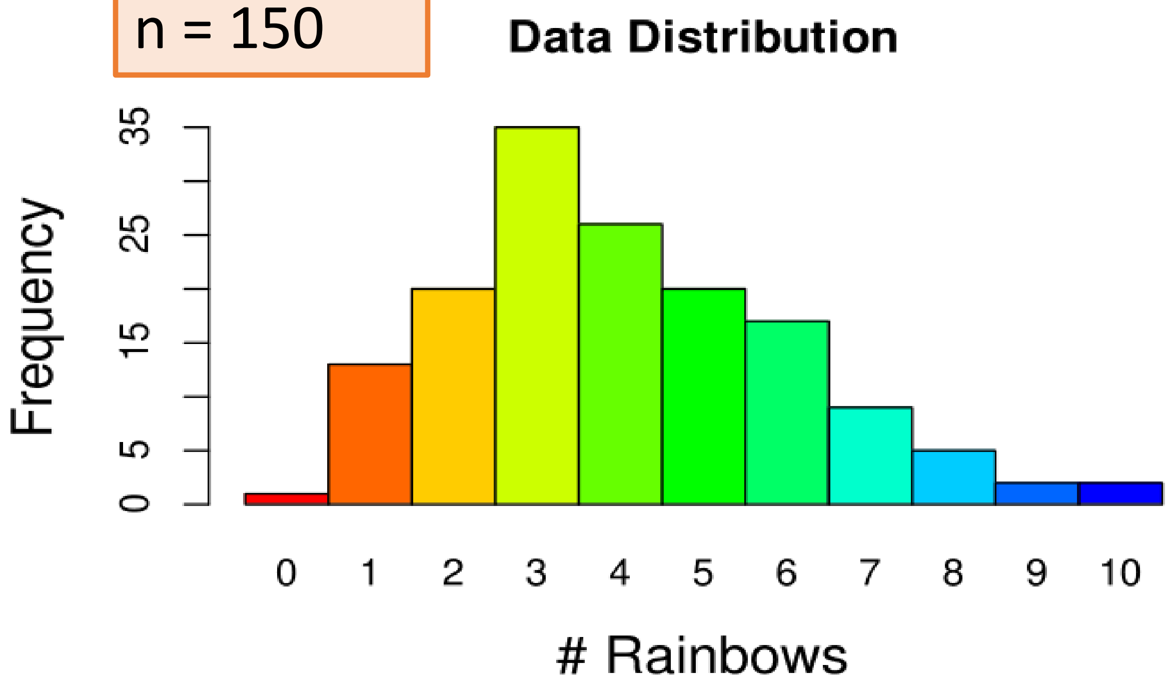
Population and data distribution



Parameter

$\mu = 4$

n = 150



Statistic

$\bar{X} = 3.89$

If we would not have known the population parameters, how certain would we be about these statistics?

Sampling distribution

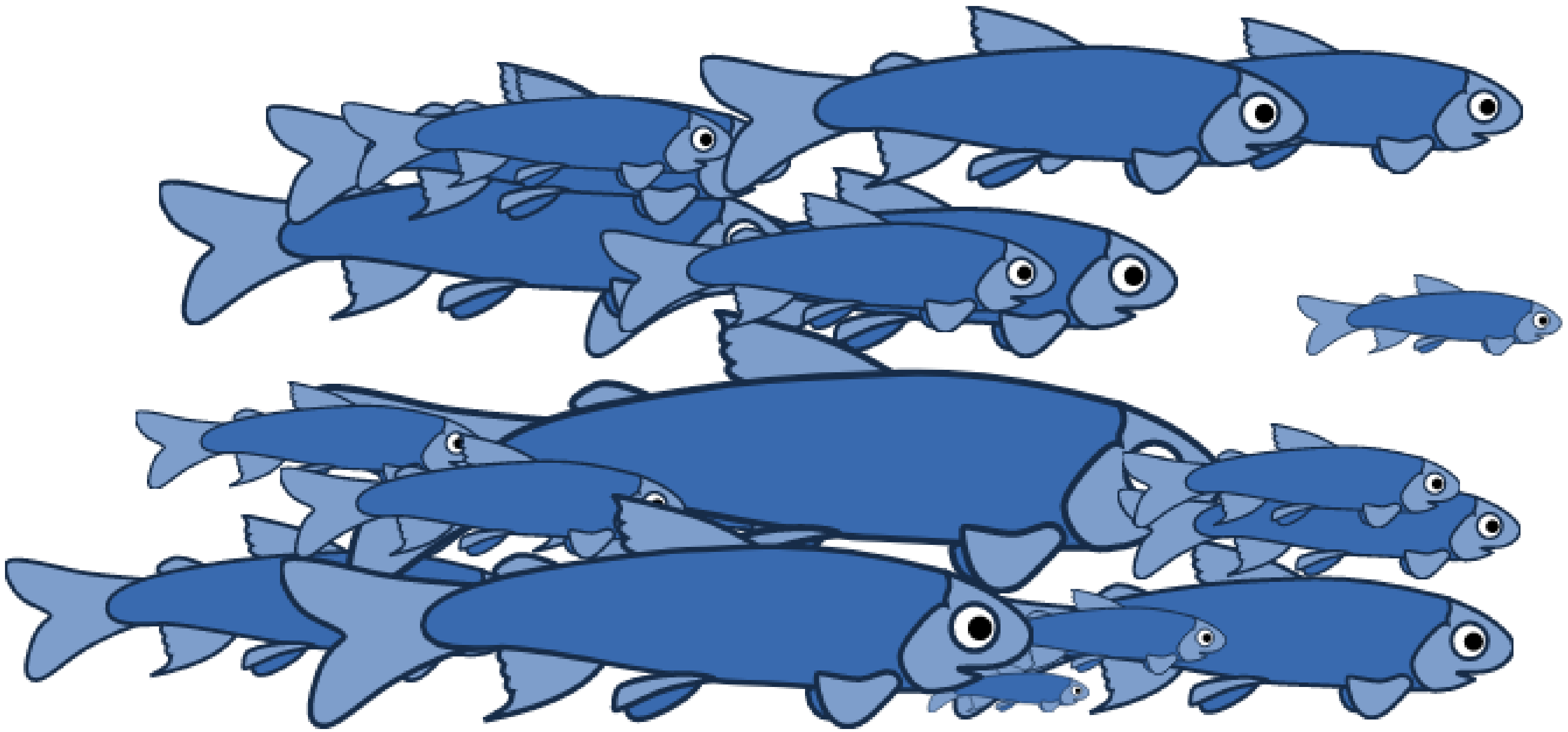
- Every statistic (e.g., proportion, mean, median, IQR, correlation, standard deviation, etc.) has a sampling distribution
- Here we focus on
  - Proportion (relevant for binary categorical variables)
  - Mean (relevant for quantitative variables)

Sampling distribution: A sampling distribution of a statistic is a probability distribution that specifies probabilities for the possible values the statistic can take (Agresti, p. 308).

Sampling distribution

- Every statistic (e.g., proportion, mean, median, IQR, correlation, standard deviation, etc.) has a sampling distribution
- Here we focus on
  - Proportion (relevant for binary categorical variables)
  - Mean (relevant for quantitative variables)

We now start reasoning at the level where we are **imagining** repeating an experiment multiple times and computing a statistic. For instance, we imagine collecting 10,000 samples of size 10. This results in 10,000 means: one for each separate experiment.



https://www.zoology.ubc.ca/~whitlock/Kingfisher/SamplingNormal.htm

Three distributions (Agresti, p. 367)

Population distribution: The distribution of the variable of interest in the population from which we sample.

Data distribution: The distribution of the variable of interest computed from the sample and the distribution we see in practice.

Sampling distribution: The distribution of the sample statistic (e.g., sample proportion or sample mean) we would get when repeatedly drawing samples of size n from the population

The sampling distribution is a bit like a unicorn: we can imagine what one looks like, we can appreciate its beauty, and we can wonder at its magical feats, but the sad truth is that you'll never see a real one. They both exist as ideas rather than physical things. You would never go out and actually collect thousands of samples and draw a frequency distribution of their means, instead very clever statisticians have worked out what these distributions look like and how they behave. Likewise, you'd be ill-advised to search for unicorns.

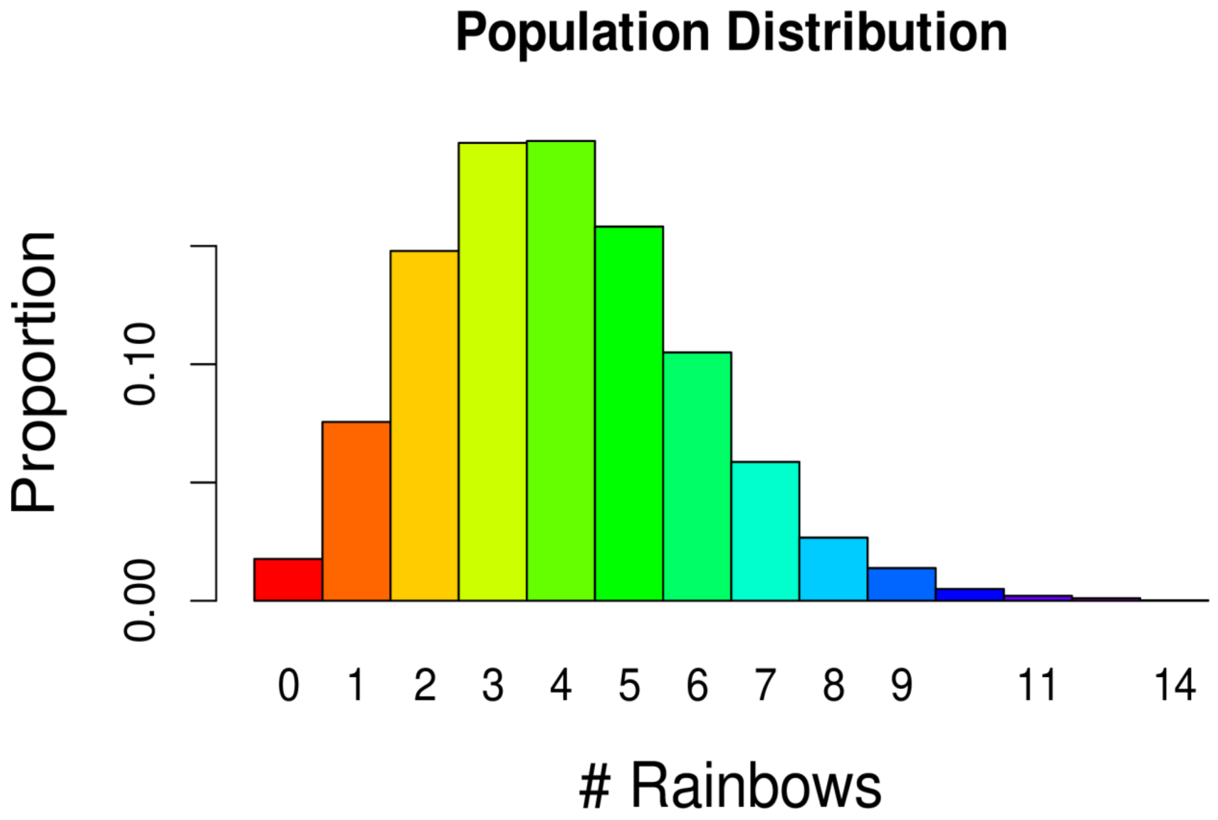
Andy Field (Discovering Statistics Using JASP, p110)

Overview of Today

1. Population distribution vs data distribution
2. Sampling distribution
    1. **Distribution of a quantitative variable**
    2. Distribution of a binary categorical variable
3. Central Limit Theorem
4. Excel & Demo
5. Recap
    - Next time
    - Example exam question

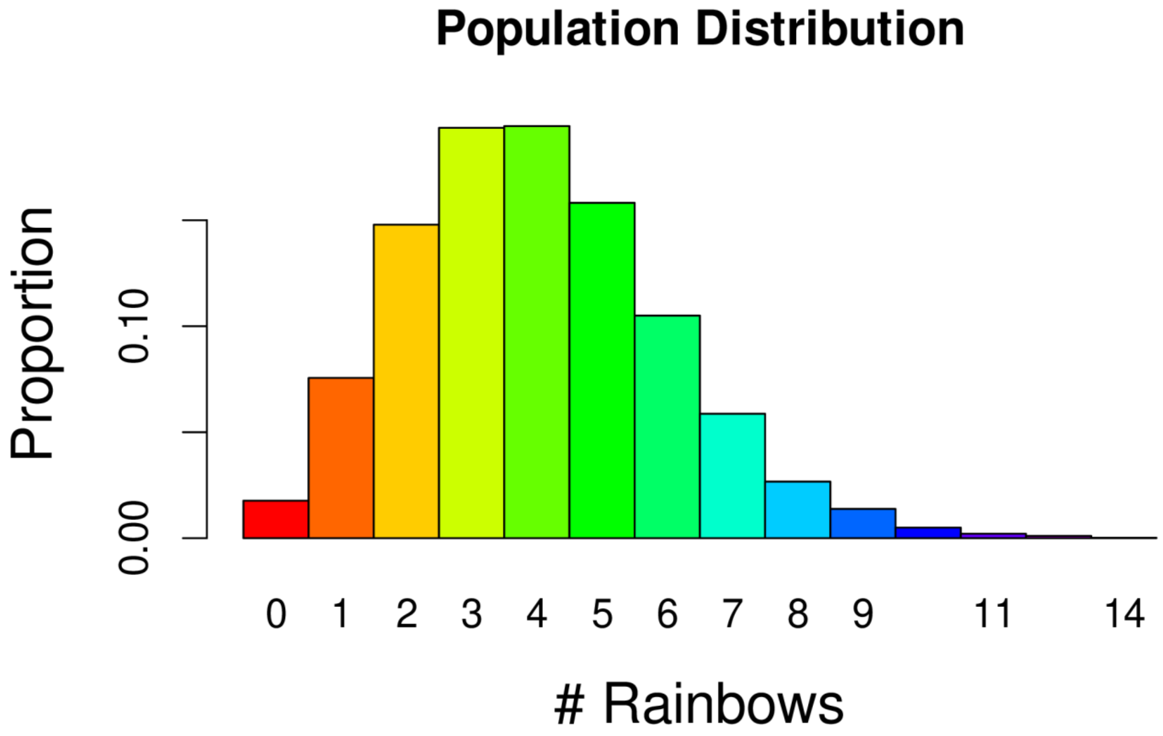
Case 1: Rainbows

- How many rainbows have you seen this year?



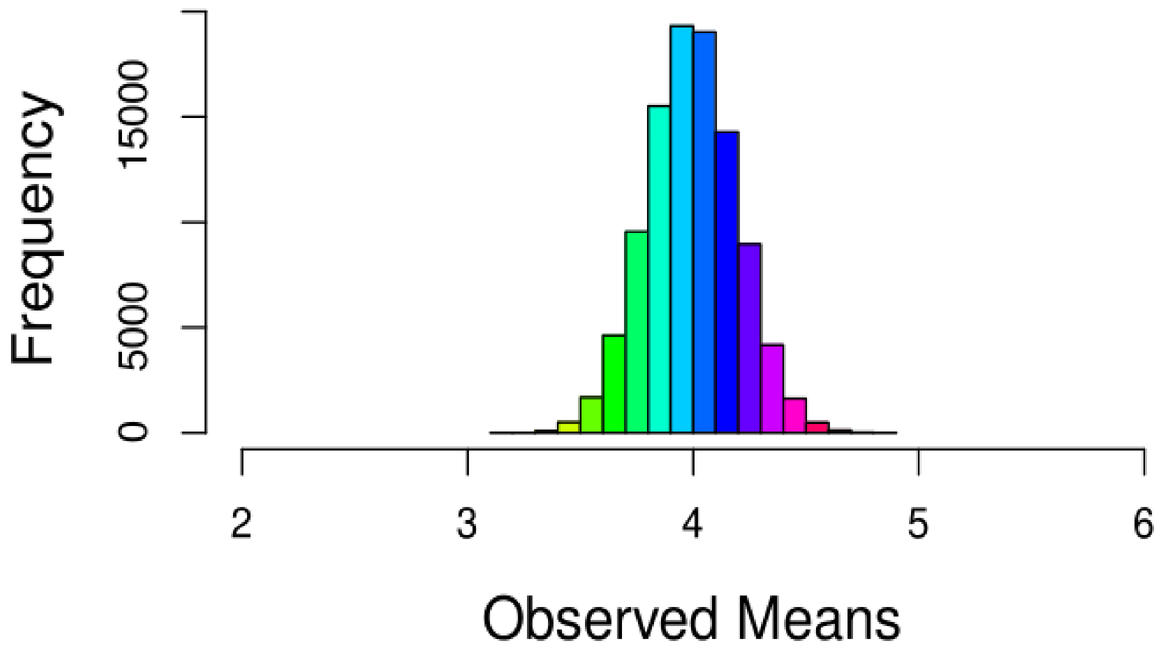
Source: Pixabay.org

What is the shape of the sampling distribution of means?



$\mu = 4, \sigma = 2$

Sampling distribution mean
After 100,000 samples (n=100)



What is the shape of the sampling distribution of means?

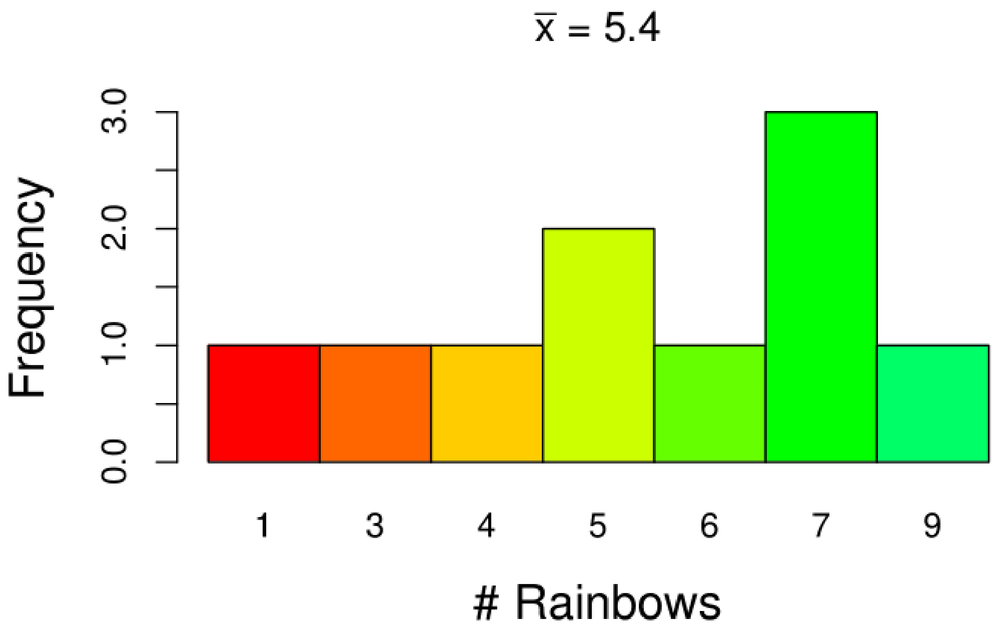
- It turns out, that this is always a normal distribution with
  - Mean: μ
  - Standard deviation: $\sigma / \sqrt{n}$

- For now, let's assume we know the parameters in our example:
  - $\mu = 4$
  - $\sigma = 2$

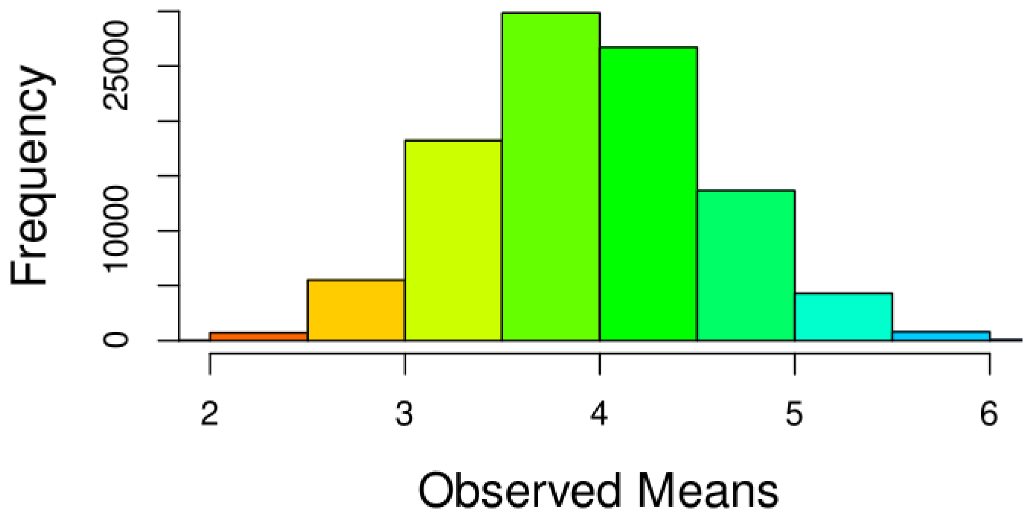
Next week, we see how parameter estimation in practice, when we do **not** know the parameter values!

What is the shape of the sampling distribution of means?

One observed sample and its observed mean (n=10)



Sampling distribution mean
After 100,000 samples (n=10)

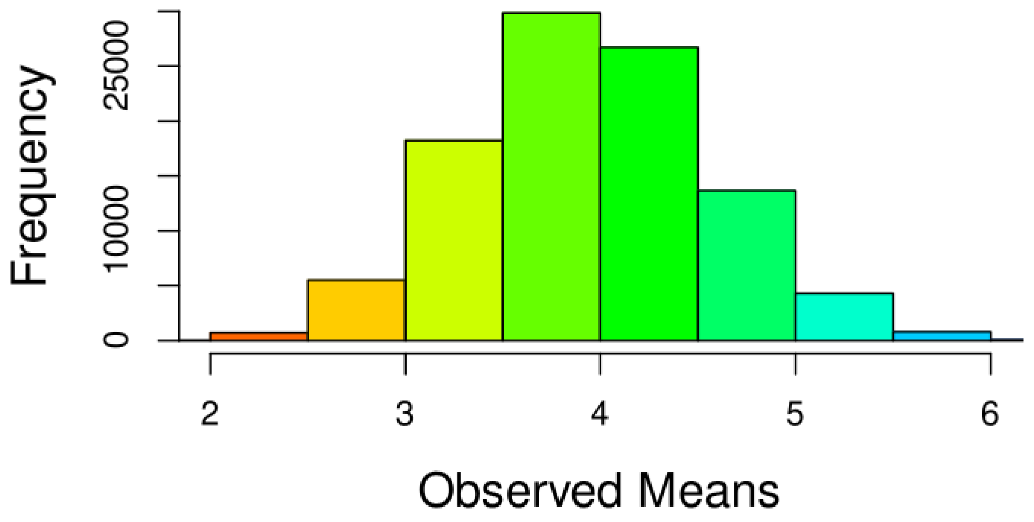


What is the shape of the sampling distribution of means?

$$\text{mean} = \mu = 4$$

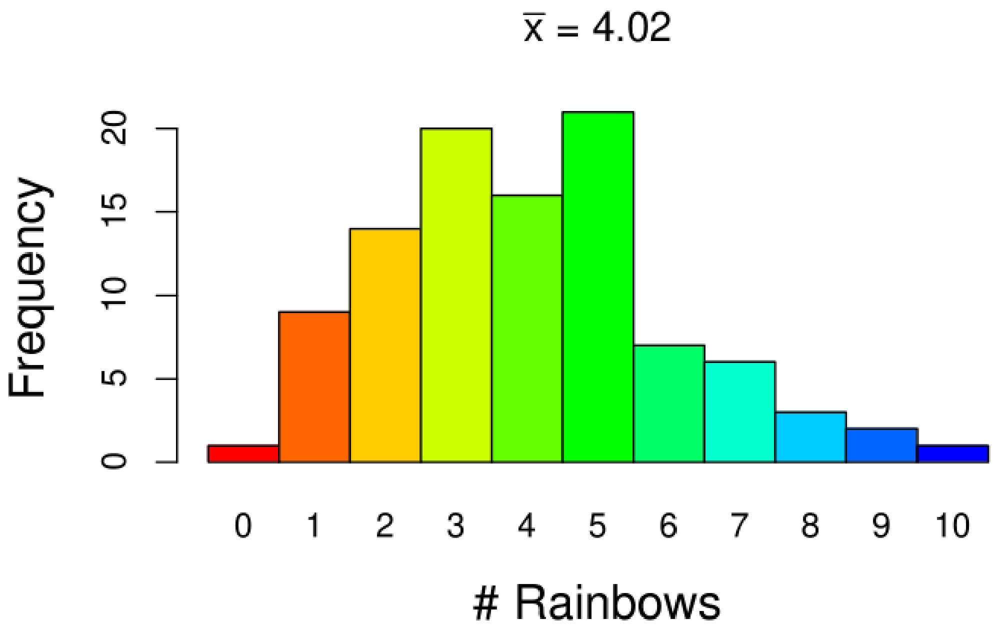
$$sd = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{10}} = 0.63$$

Sampling distribution mean
After 100,000 samples (n=10)

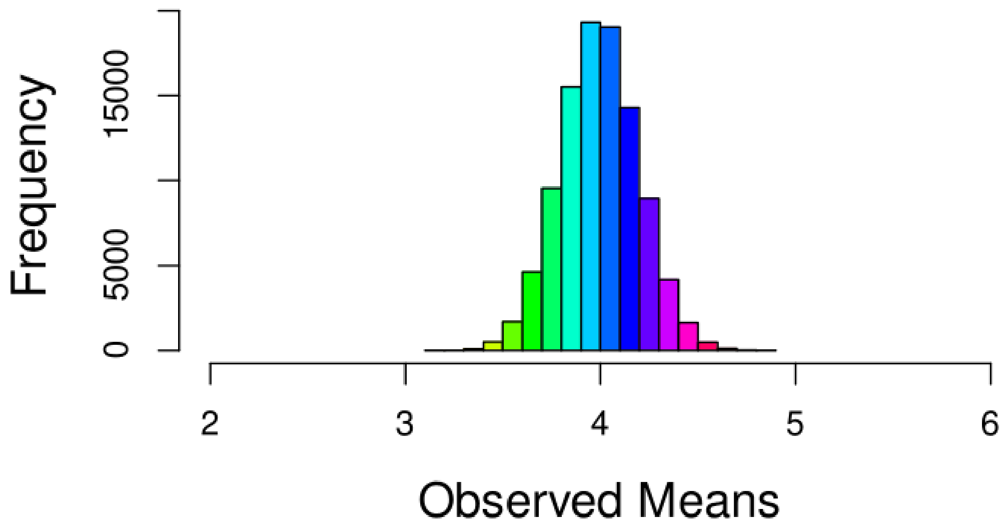


What is the shape of the sampling distribution of means?

One observed sample and its observed mean (n=100)



Sampling distribution mean After 100,000 samples (n=100)

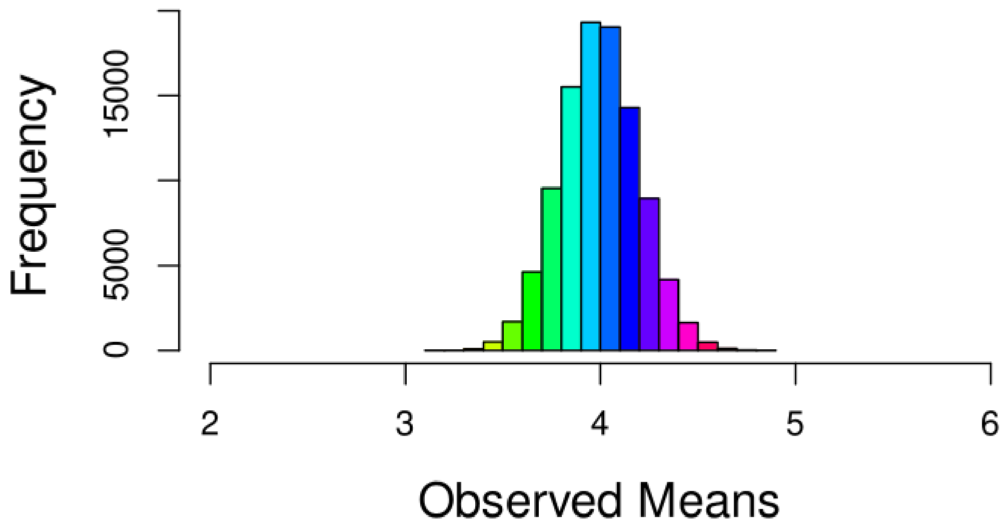


What is the shape of the sampling distribution of means?

$$\text{mean} = \mu = 4$$

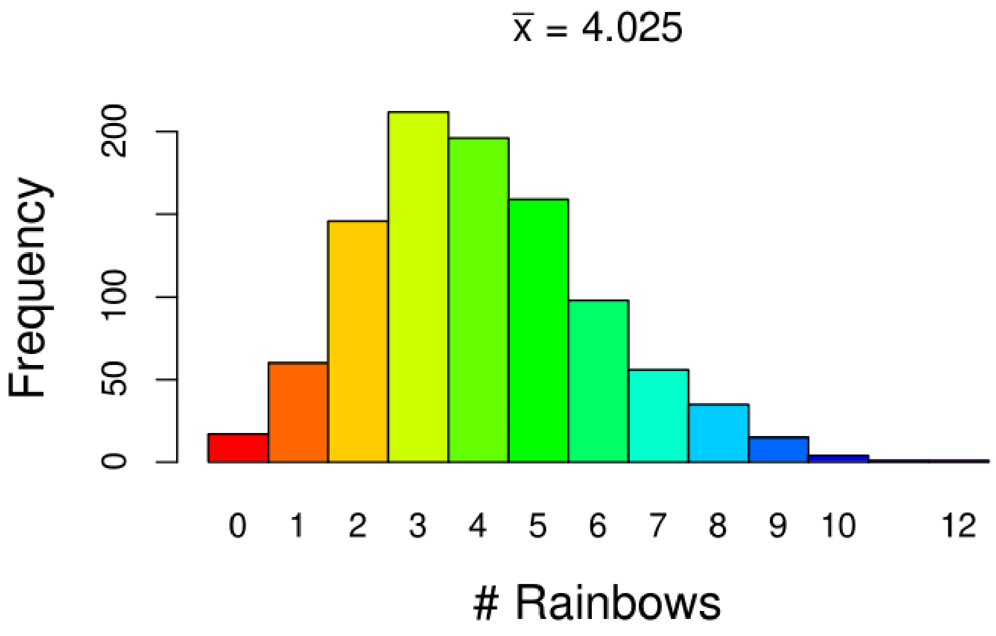
$$sd = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{100}} = 0.2$$

Sampling distribution mean
After 100,000 samples (n=100)

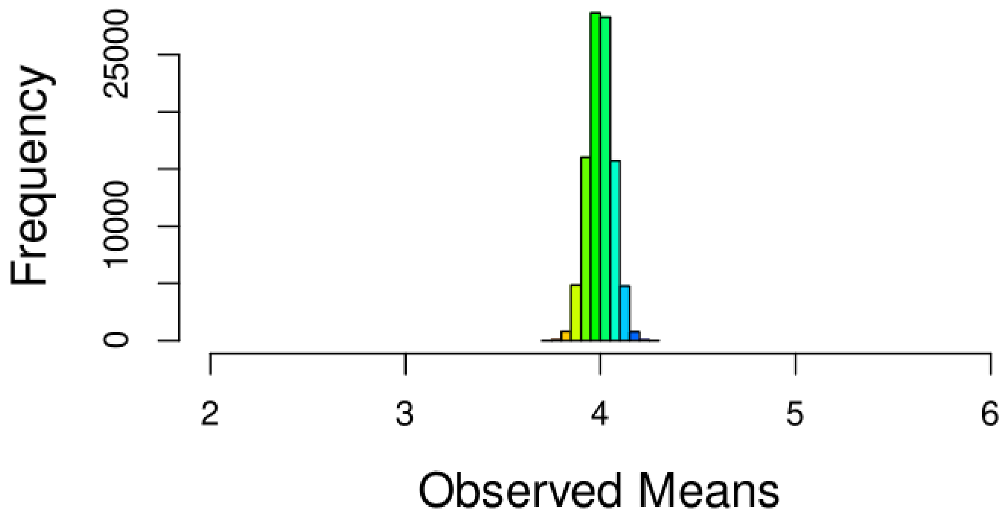


What is the shape of the sampling distribution of means?

One observed sample and its observed mean (n=1,000)



Sampling distribution mean
After 100,000 samples (n=1,000)

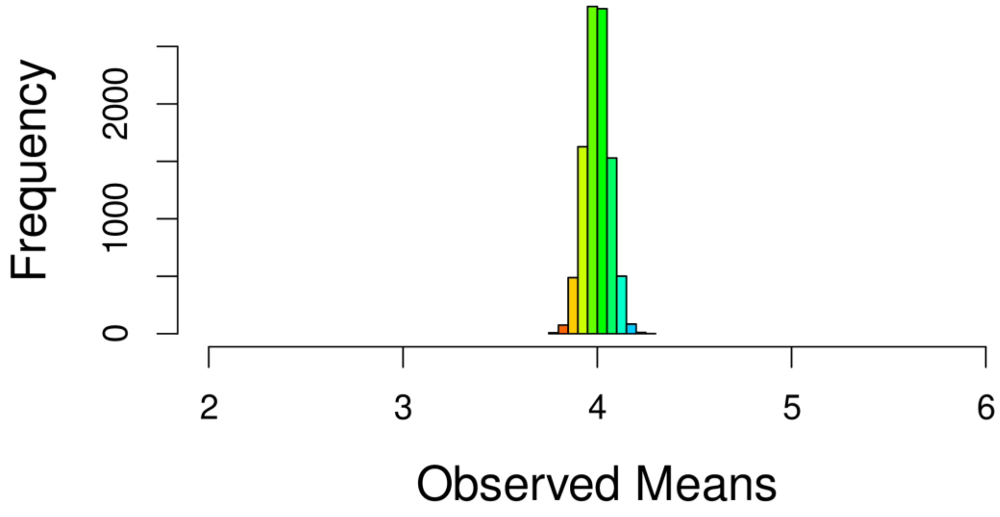


What is the shape of the sampling distribution of means?

$$\text{mean} = \mu = 4$$

$$sd = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{1000}} = 0.063$$

Sampling distribution mean
After 100,000 samples (n=1,000)



Overview of Today

1. Population distribution vs data distribution
2. Sampling distribution
    1. Distribution of a quantitative variable
    2. **Distribution of a binary categorical variable**
3. Central Limit Theorem
4. Excel & Demo
5. Recap
    - Next time
    - Example exam question

What is the shape of the sampling distribution of proportions?

- We are drawing repeatedly and independently from a population that has two categories (e.g., yes/no on the voting question)
- The probability of one category (yes) is P(yes) = p = 0.7
  - The population proportion p
  - P(no) = 1 – p = 1 – 0.7 = 0.3
- For every draw, p is the same
- What is the distribution of the proportion of people that will vote, after n draws?

❼ **Binomial distribution**

What is the shape of the sampling distribution of proportions?

- We are drawing repeatedly and independently from a population that has two categories (e.g., yes/no on the voting question)
- The probability of one category (yes) is P(yes) = $p = 0.7$
  - The population proportion p
  - P(no) = $1 - p = 1 - 0.7 = 0.3$
- For every draw, p is the same
- What is the distribution of the proportion of people that will vote after n draws?

 Binomial distribution

Note: this is only true if we draw *with replacement.*
In practice, for large populations this does not make a difference.

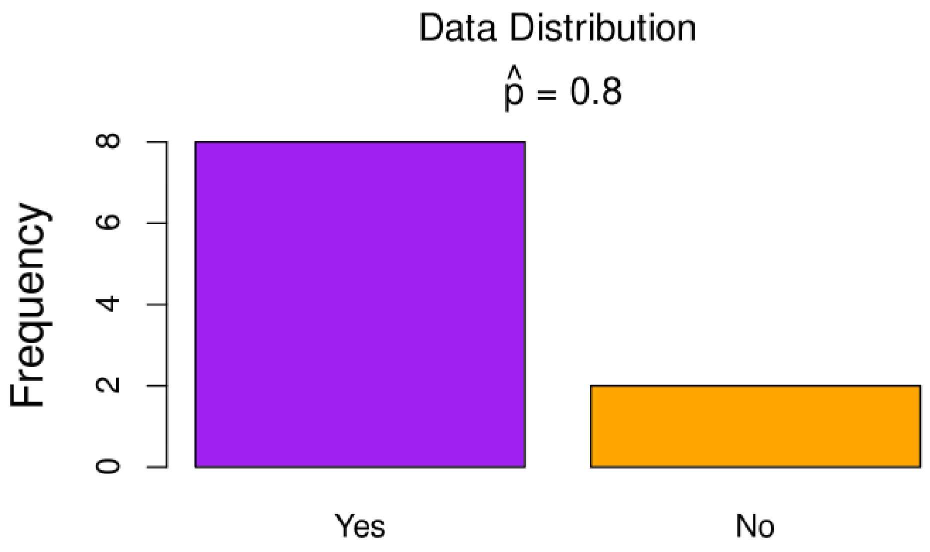
What is the shape of the sampling distribution of proportions?

- We are drawing repeatedly and independently from a population that has two categories (e.g., yes/no on the voting question)
- The probability of one category (yes) is P(yes) = p = 0.7
  - The population proportion
  - P(no) = 1 – p = 1 – 0.7 = 0.3
- For every draw, p is the same
- What is the distribution of the proportion of people that will vote after n draws?

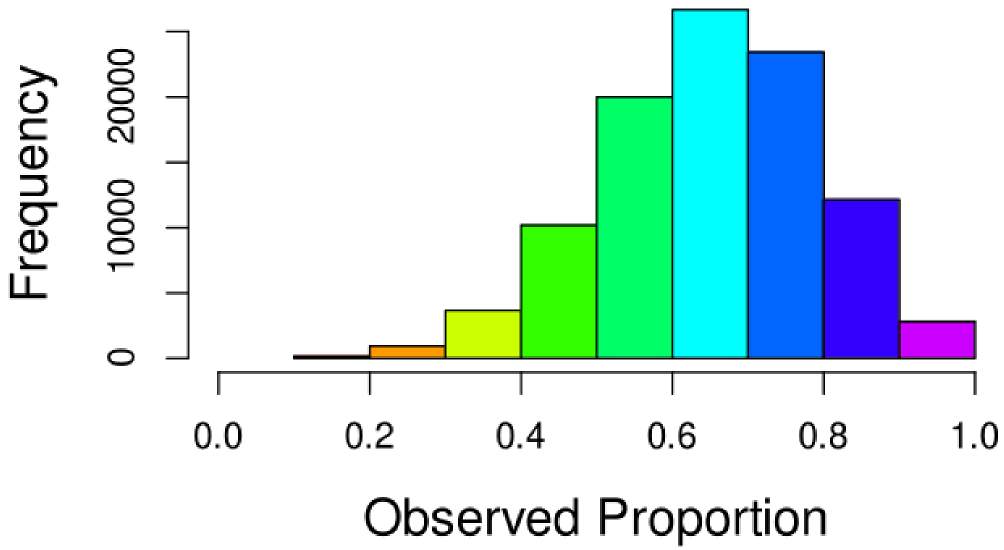
 Binomial distribution

What is the shape of the sampling distribution of proportions?

One observed sample and its observed proportion (n=10)

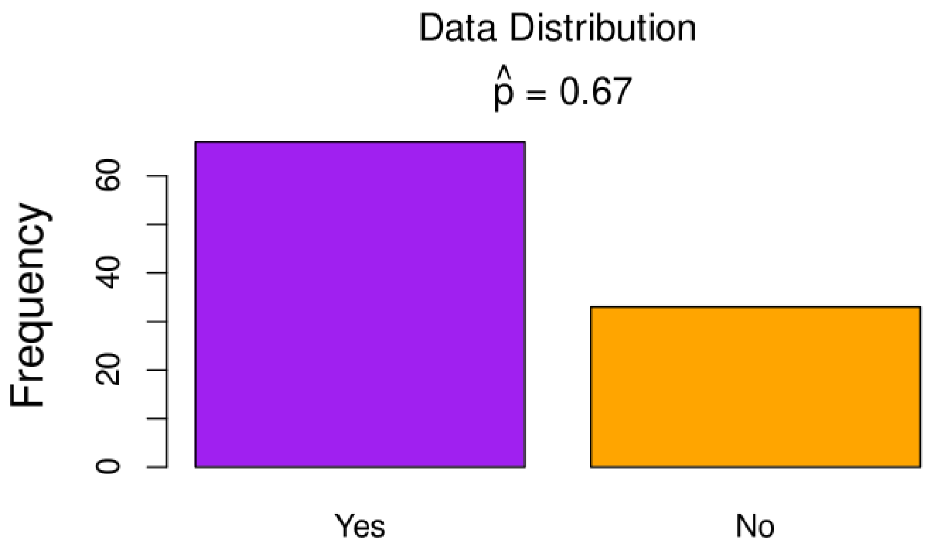


Sampling distribution proportion After 100,000 samples (n=10)

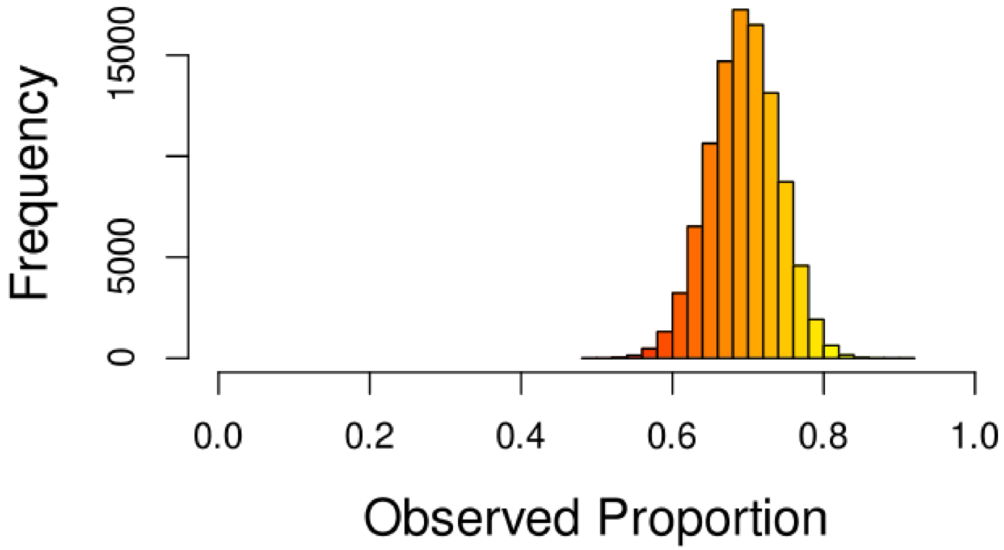


What is the shape of the sampling distribution of proportions?

One observed sample and its observed proportion (n=100)

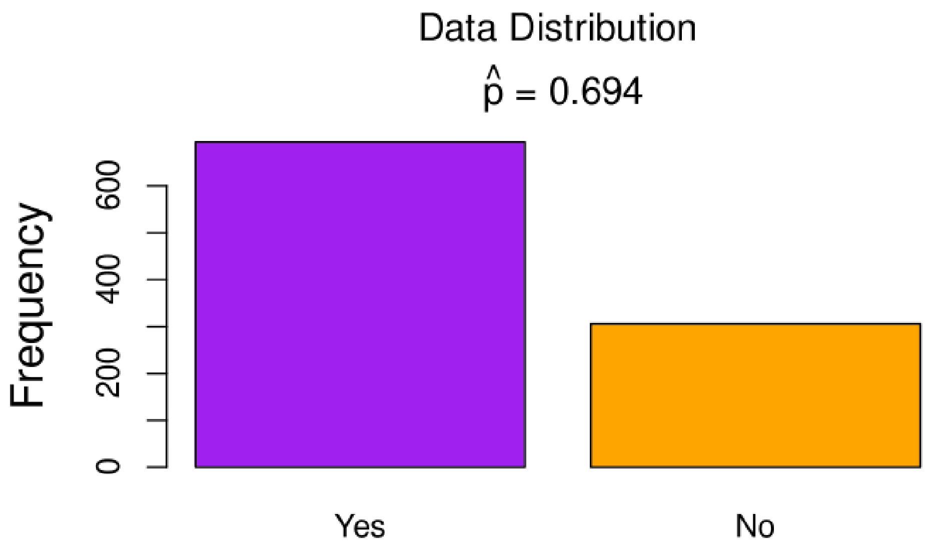


Sampling distribution proportion
After 100,000 samples (n=100)

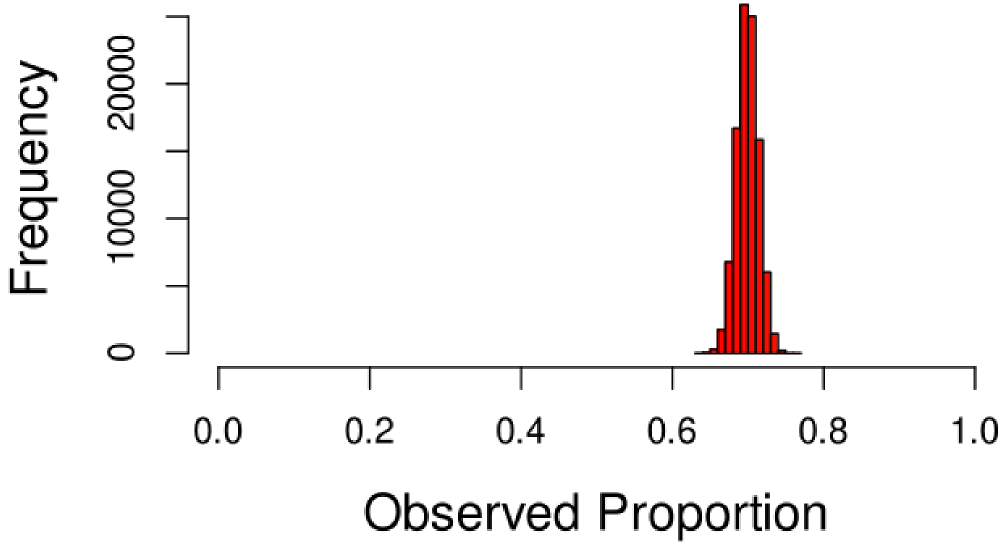


What is the shape of the sampling distribution of proportions?

One observed sample and its observed proportion (n=1,000)



Sampling distribution proportion
After 100,000 samples (n=1,000)



What is the shape of the sampling distribution of proportions?

The shape of the sampling distribution of proportions $\hat{p}$: A binomial distribution with

mean: p

standard deviation: $\sqrt{\frac{p(1-p)}{n}}$

- The mean of the sampling distribution is the population parameter
- The standard deviation indicates the variability of the sampling distribution: How far can a statistic deviate from the parameter?
- A larger sample (*n*) means that the standard deviation of the sampling distribution decreases

What is the shape of the sampling distribution of proportions?

The shape of the sampling distribution of proportions $\hat{p}$: A binomial distribution with

mean: p

standard deviation: $\sqrt{\frac{p(1-p)}{n}}$

Note the differences with the "standard" binomial distribution:

mean: np

standard deviation: $\sqrt{np(1-p)}$

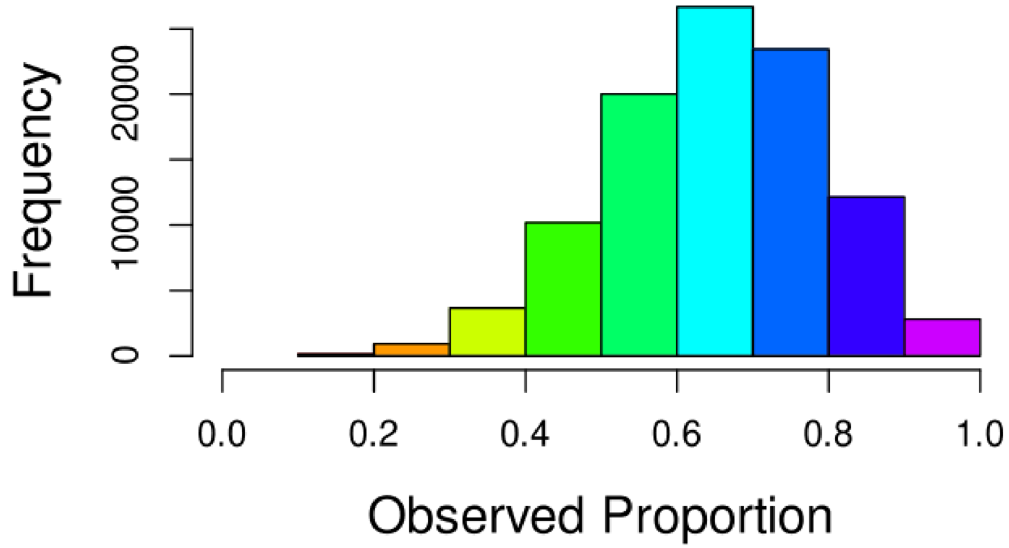
- The mean of the sampling distribution is the population parameter
- The standard deviation indicates the variability of the sampling distribution: How far can a statistic deviate from the parameter?
- A larger sample (*n*) means that the standard deviation of the sampling distribution decreases

mean = p = 0.7

Standard deviation = $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{p(1-p)}{10}} = 0.145$

Sampling distribution proportion
After 100,000 samples (n=10)

Mean = 0.7
Sd = 0.145



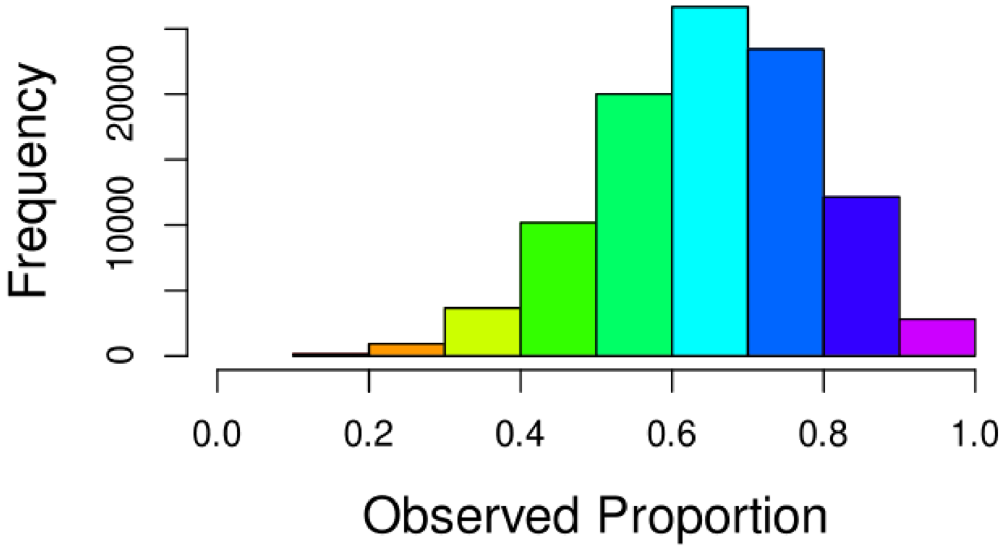
42

mean = p = 0.7

Standard deviation = $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{p(1-p)}{10}} = 0.145$

Sampling distribution proportion
After 100,000 samples (n=10)

Mean = 0.7
Sd = 0.145

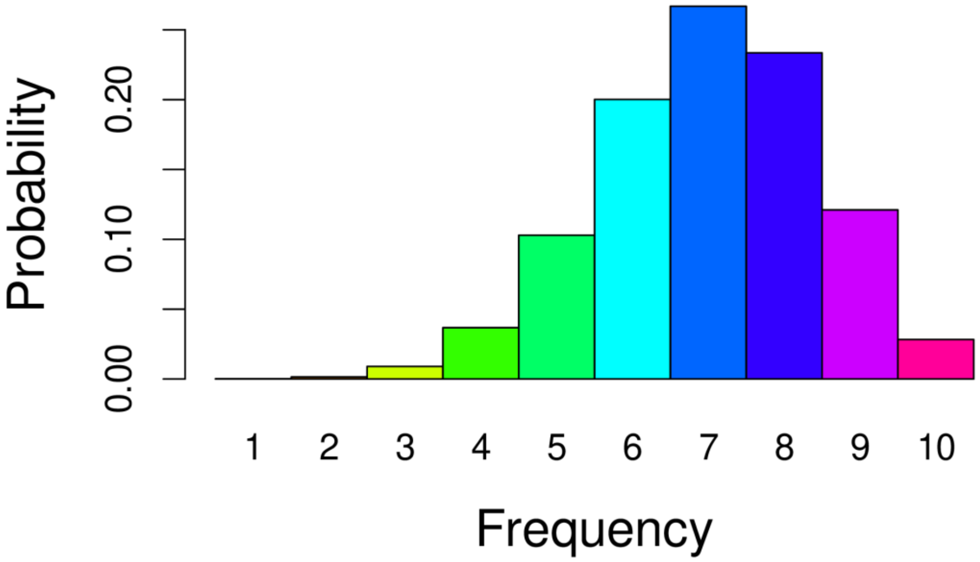


mean = np = 7

Standard deviation = $\sqrt{np(1-p)} = 1.45$

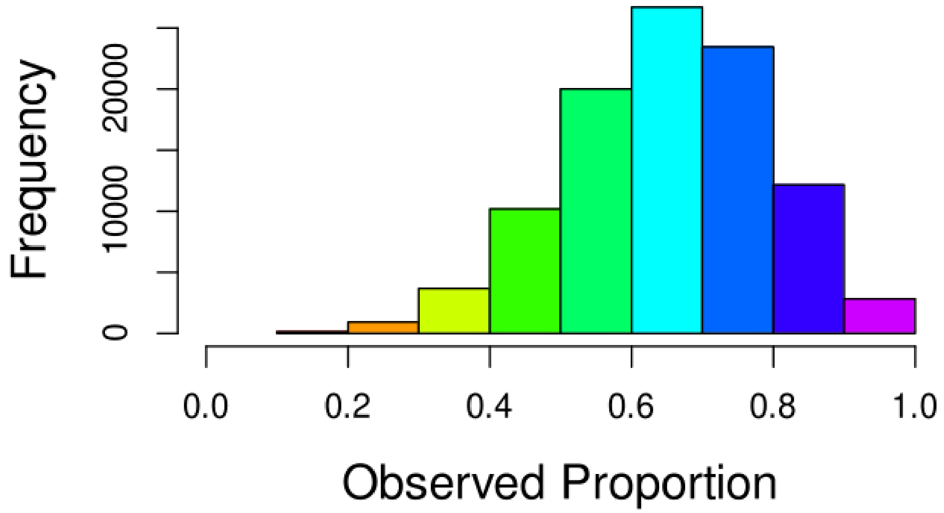
Binomial distribution
n=10, p=0.7

Mean = 7
Sd = 1.45

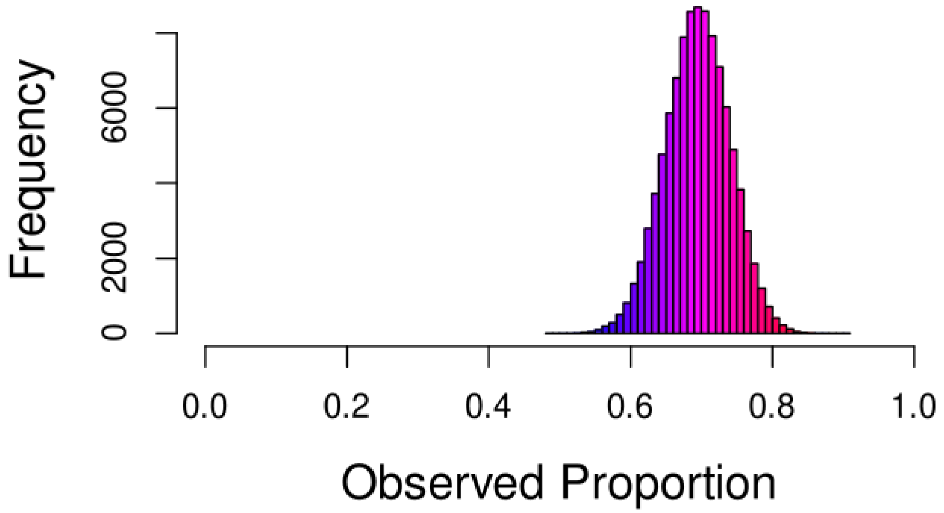


What is the shape of the sampling distribution of proportions?

Sampling distribution proportion
After 100,000 samples (n=10)



Sampling distribution proportion
After 100,000 samples (n=100)



mean = p = 0.7

Standard deviation = $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{p(1-p)}{10}} = 0.145$

mean = p = 0.7

Standard deviation = $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{p(1-p)}{100}} = 0.046$

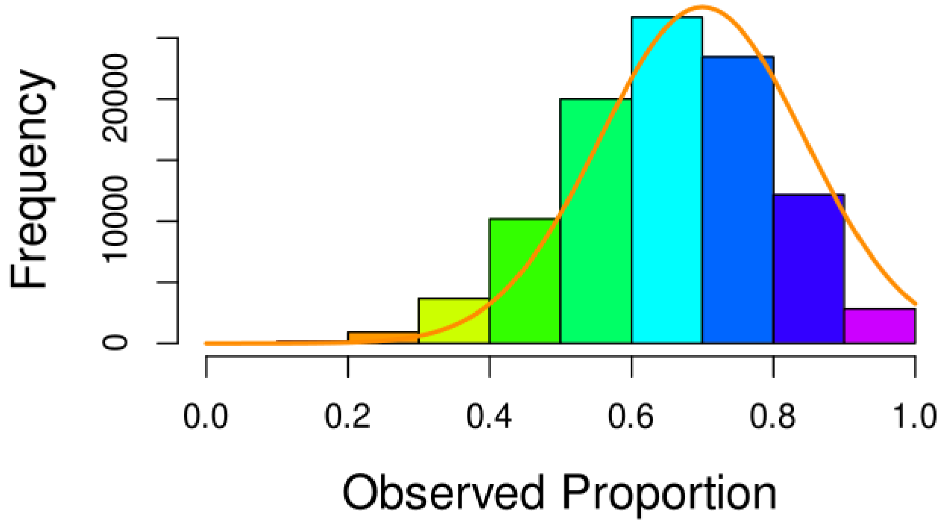
What is
distribu

Approximately normally distributed when:

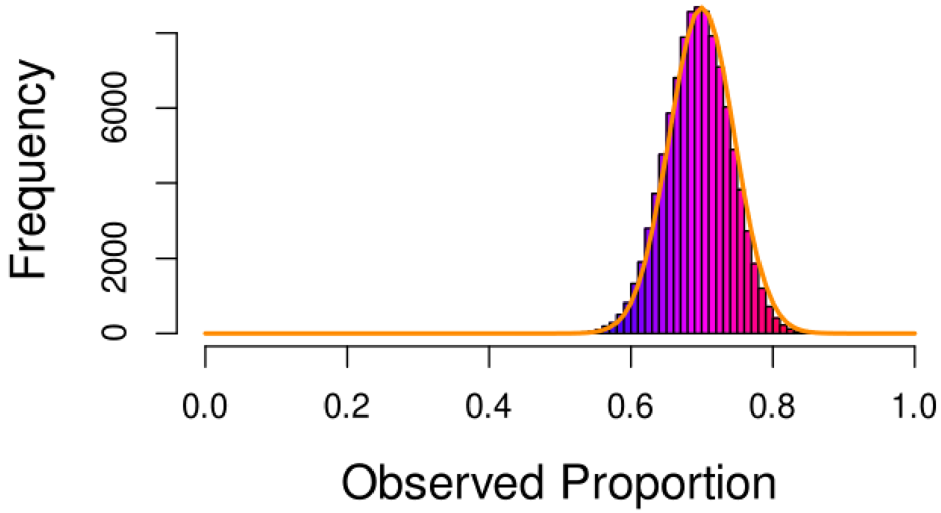
$np \geq 15$

$n(1-p) \geq 15$

Sampling distribution proportion
After 100,000 samples (n=10)



Sampling distribution proportion
After 100,000 samples (n=100)



mean = p = 0.7

Standard deviation = $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{p(1-p)}{10}} = 0.145$

mean = p = 0.7

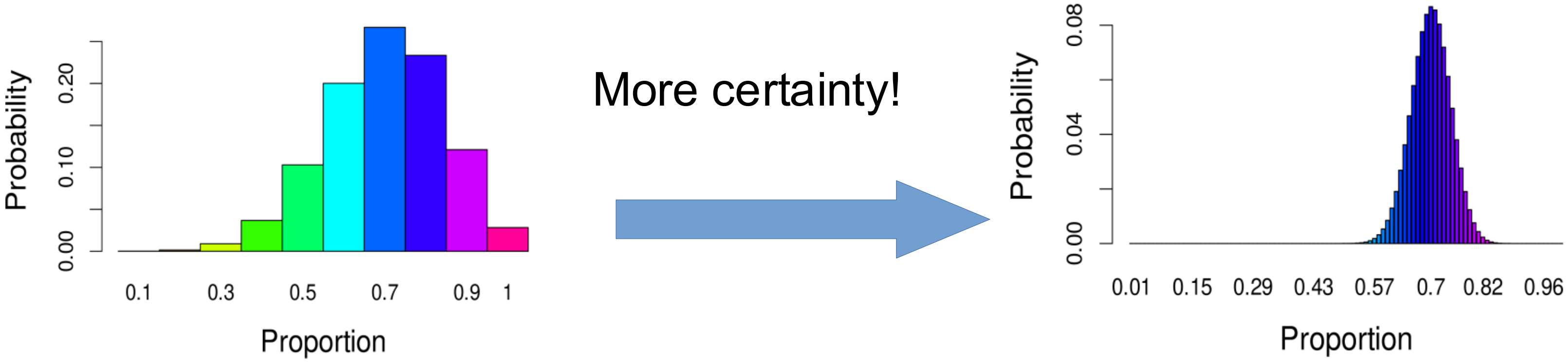
Standard deviation = $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{p(1-p)}{100}} = 0.046$

So, what have we learned now?

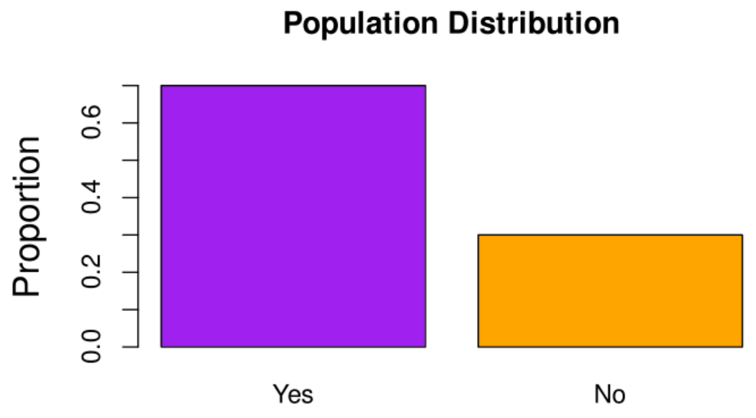
- We can now estimate how likely it is that the statistics we get from our experiment or observational study are good representations of the population

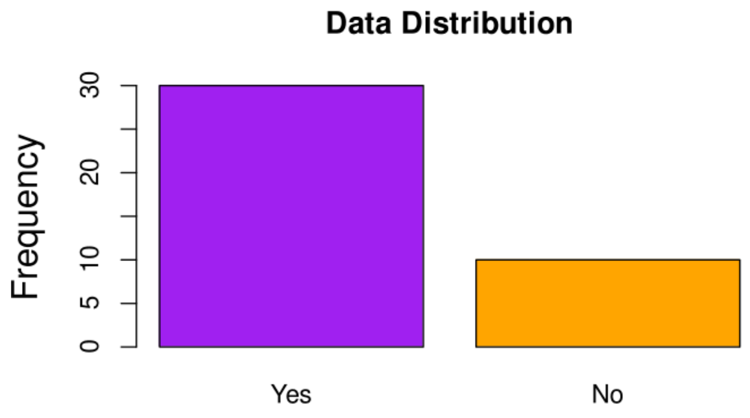
- We know how likely it is that we will get the same statistics if we would have collected another random sample

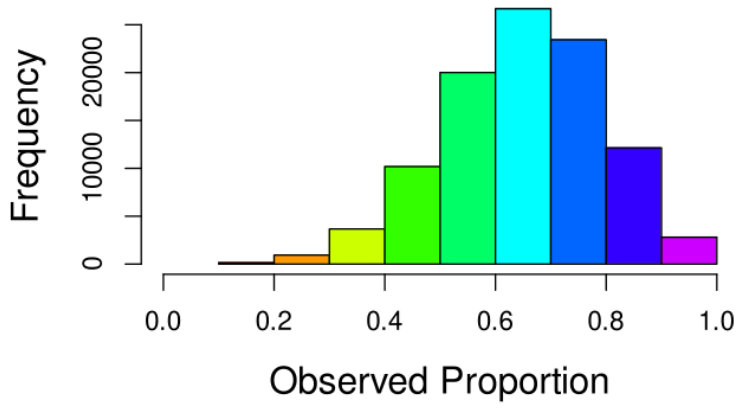
- We know to what extent we can trust our inference to the population → Validities & proper methodology



Sampling distribution proportion
After 100,000 samples (n=10)







Overview of Today

1. Population distribution vs data distribution
2. Sampling distribution
    1. Distribution of a binary categorical variable
    2. Distribution of a quantitative variable
3. **Central Limit Theorem**
4. Recap
    - Next time
    - Example exam question

48

Central Limit Theorem

Central Limit Theorem: For any large enough sample, the sampling distribution of the mean is approximately normally distributed with

mean: μ

Standard deviation: $\sigma/\sqrt{n}$

- This is *irrespective* of the shape of the distribution
- http://www.zoology.ubc.ca/~whitlock/Kingfisher/CLT.htm

Central Limit Theorem

This is not an all-or-nothing threshold, but we generally need around 30 observations

Central Limit Theorem: For any large enough sample, the sampling distribution of the mean is approximately normally distributed with

mean: μ
Standard error (SE): $\sigma/\sqrt{n}$

From now, we denote the standard deviation of the sampling distribution with SE: Standard error (from chapter 8, p405)

Central Limit Theorem

- This means that we can study all kinds of variables (with different distributions) because the sampling distributions of their means converge to a normal distribution

- This means that we only need a limited set of statistical tools to draw inferences about our data
  - Parametric statistics

	Quantitative	**Categorical**
Quantitative	Correlation Regression	t-test ANOVA
Categorical	Logistic regression	Contingency table

What can we do now?

- When we know the sampling distribution (for instance because of the C.L.T., or because we know or assume it), we can speak about the *probability* of observing a certain value for a statistic.
- For instance, the probability of observing a sample of 9 people, who have seen an average of **5 rainbows *or more***, when the population average is 4, with a standard deviation of 2 (i.e., $\mu = 4$ & $\sigma = 2$).

$$P(\bar{x} \geq 5 | \mu = 4)$$

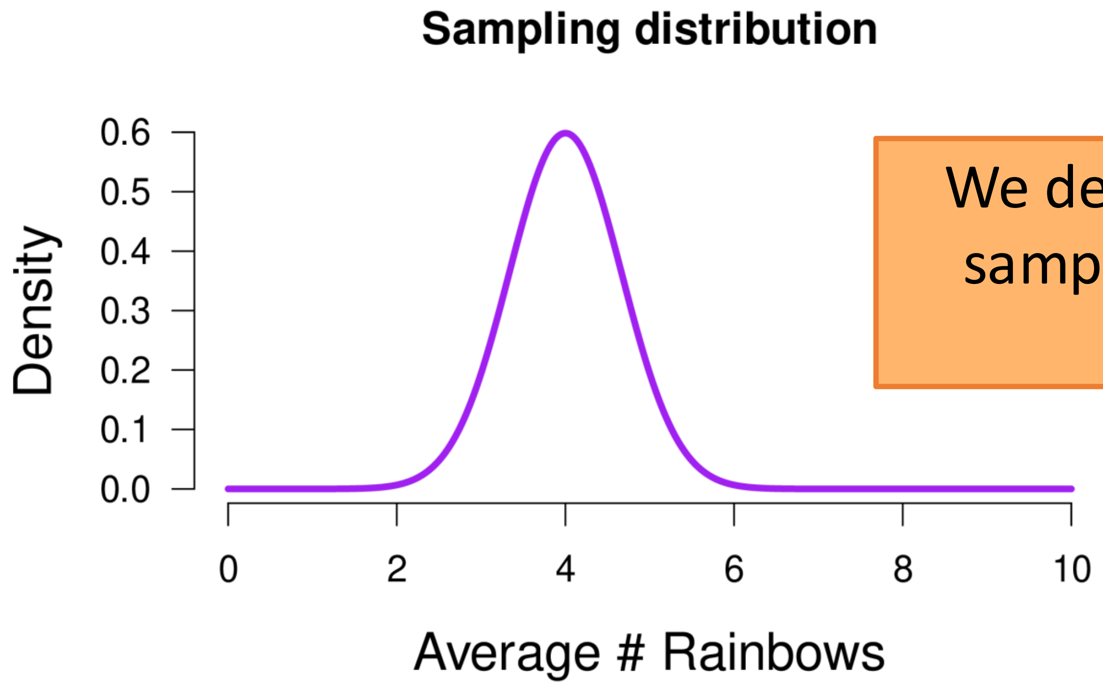
Same steps as in Riet's lecture about probability distributions!

What can we do now?

$$P(\bar{x} \geq 5 | \mu = 4)$$

- We know that the mean is normally distributed:

$$SE = \frac{2}{\sqrt{(9)}} = 0.667$$



We denote the standard deviation of the sampling distribution with SE: Standard error (from chapter 8, p356)

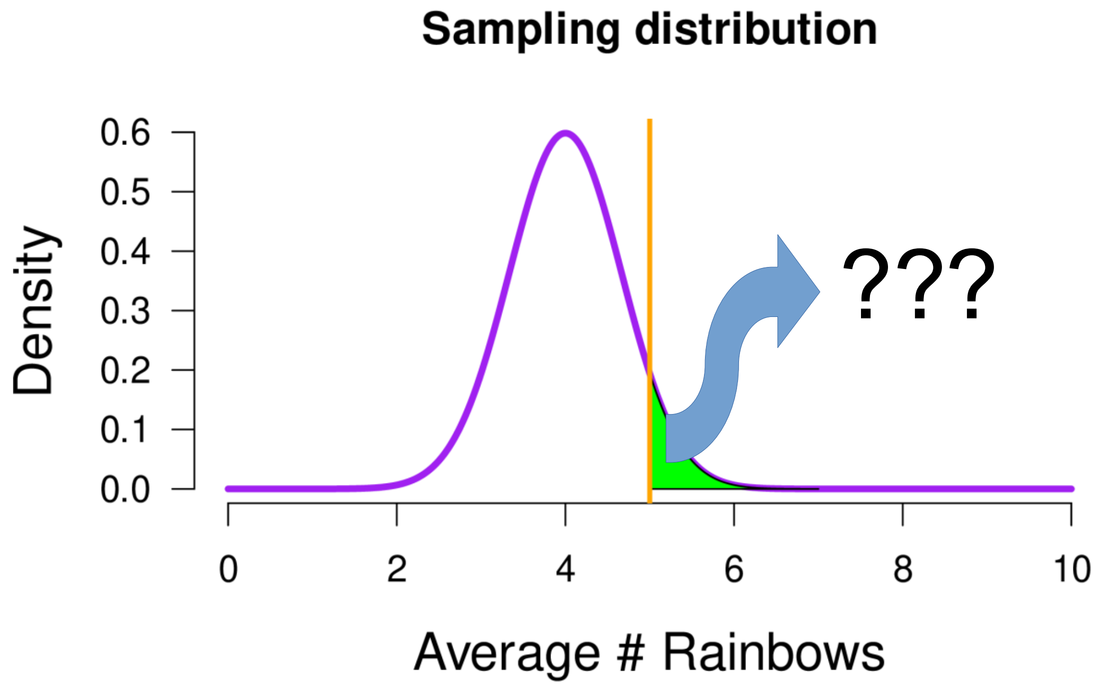
What can we do now?

$$P(\bar{x} \geq 5 | \mu = 4)$$

- We know that the mean is normally distributed:

$$SE = \frac{2}{\sqrt{(9)}} = 0.667$$

$$\mu = 4$$



What can we do now?

We are converting the observed mean to a z-score, so we use the **mean of the mean**, and the **standard error**

- To know this area, we can transform to z-scores and use Excel

$$SE = \frac{2}{\sqrt{(9)}} = 0.667$$

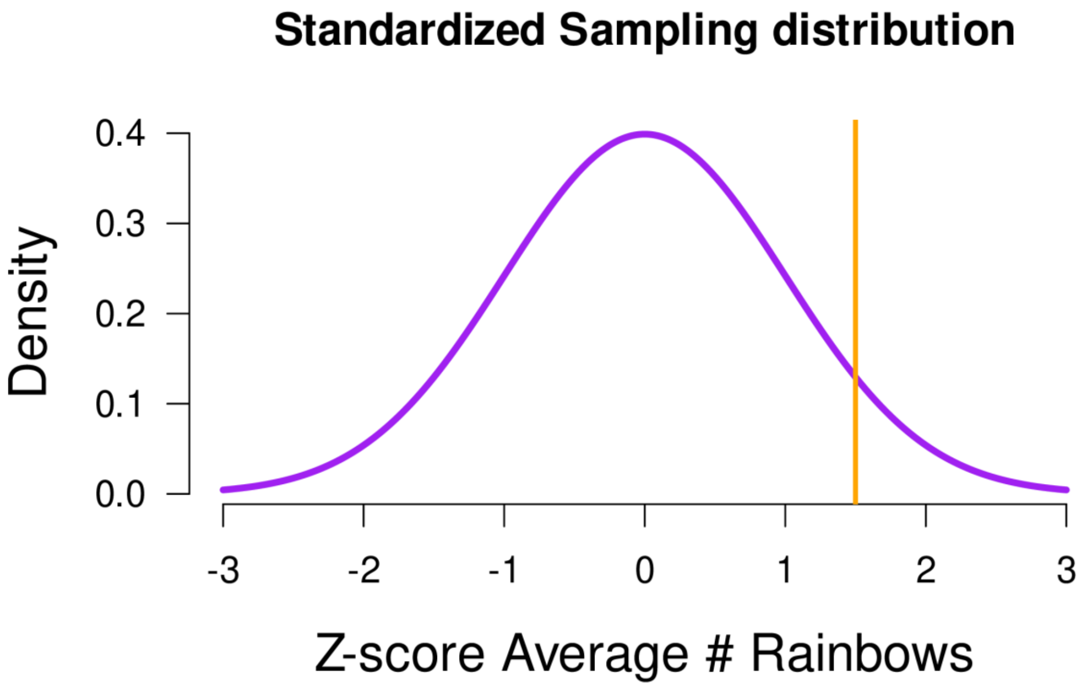
$$\mu = 4$$

$$z = \frac{\bar{x} - \mu}{SE} = \frac{\bar{x} - 4}{0.667} = 1.5$$

$$P(\bar{x} \geq 5 | \mu = 4) = P(z \geq 1.5)$$

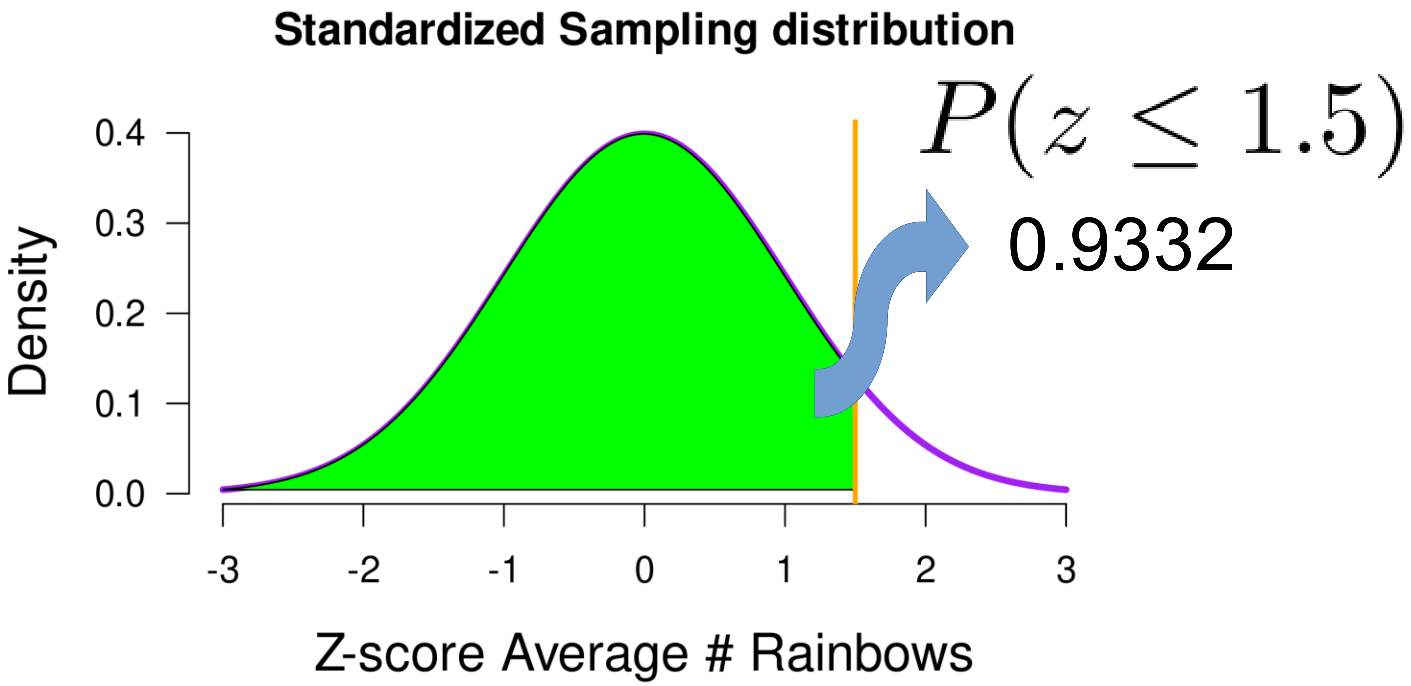
$$z = \frac{\bar{x} - \mu}{SE} = \frac{\bar{x} - 4}{0.667} = 1.5$$

- To know this area, we can transform to z-scores and then use Excel



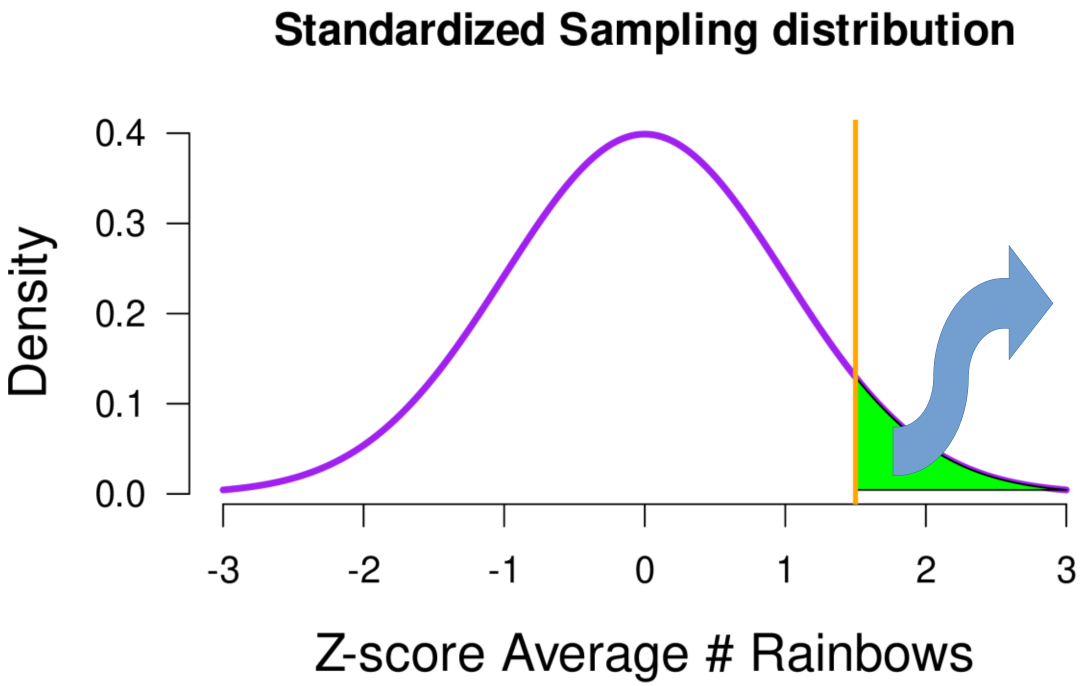
- To know this area, we can use Excel!
  - NORM.DIST(1.5, 0, 1, TRUE) = 0.9332
- Or use the mean and sd directly:
  - NORM.DIST(5, 4, 0.667, TRUE) = 0.9332

$$z = \frac{\bar{x} - \mu}{SE} = \frac{\bar{x} - 4}{0.667} = 1.5$$



$$z = \frac{\bar{x} - \mu}{SE} = \frac{\bar{x} - 4}{0.667} = 1.5$$

- NORM.DIST(1.5, 0, 1, TRUE) = 0.9332



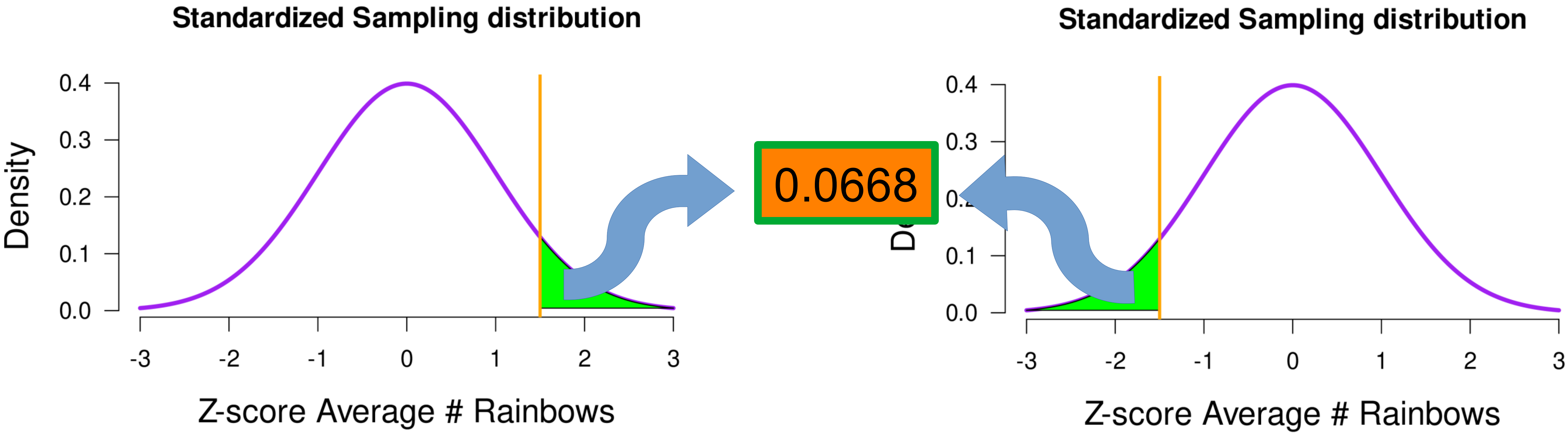
Total probability is always 1:

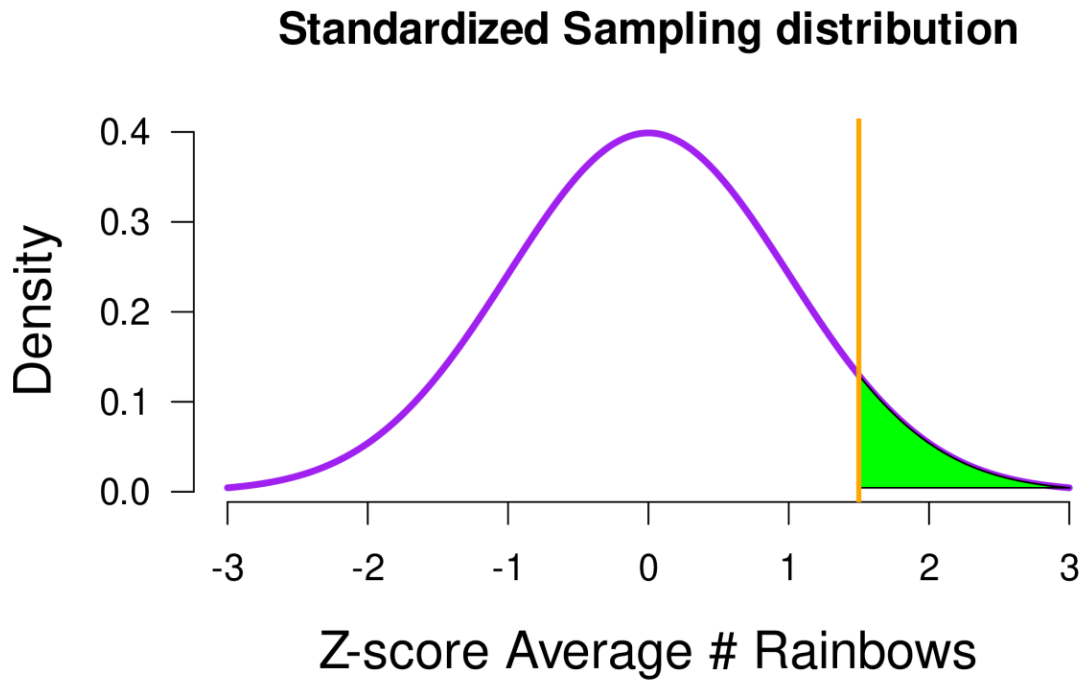
1-0.9332 = 0.0668

$$P(z \geq 1.5)$$

Symmetry

$$P(z \leq -1.5) = P(z \geq 1.5)$$





$P(\bar{x} \geq 5|\mu = 4) = P(z \geq 1.5)$ = 0.0668 ⟶ +- 6.6%

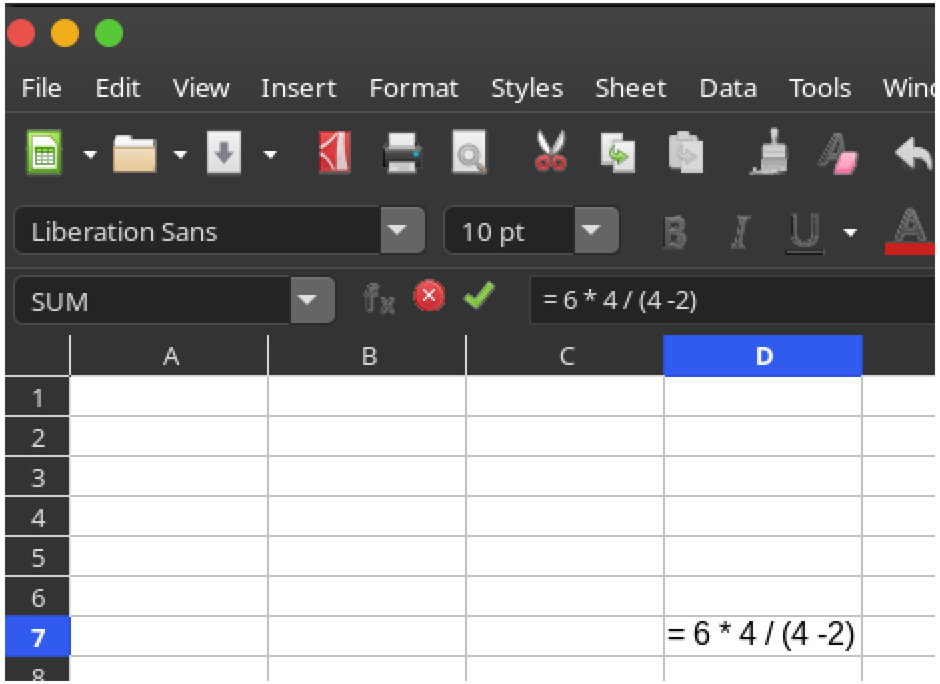
Overview of Today

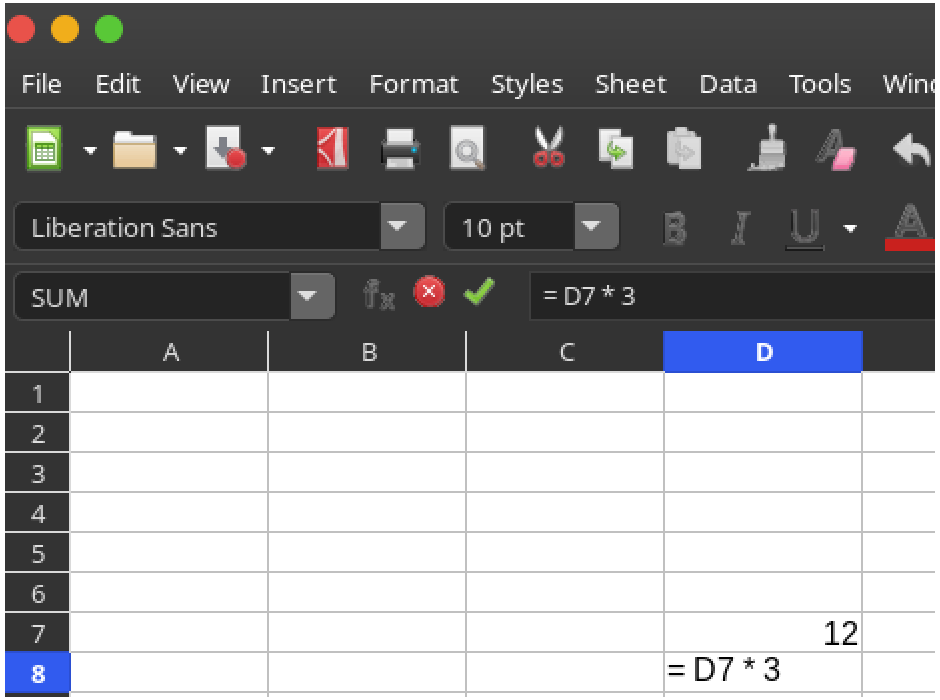
1. Population distribution vs data distribution
2. Sampling distribution
    1. Distribution of a binary categorical variable
    2. Distribution of a quantitative variable
3. Central Limit Theorem
4. **Excel & Demo**
5. Recap
    - Next time
    - Example exam questions

Excel commands for calculations:

Regular numeric operations:
= 2.3 + 3 - 7 * SQRT(10) / 4^2

→ Paste these commands into any cell and press enter!





→ 36

Source: Pixabay.org

Excel commands for calculations:

Regular numeric operations:
= 2.3 + 3 - 7 * SQRT(10) / 4^2

Microsoft office is available for free here, but you can also take a look at the open-source (/free) LibreOffice!

https://www.libreoffice.org/

Overview of Today

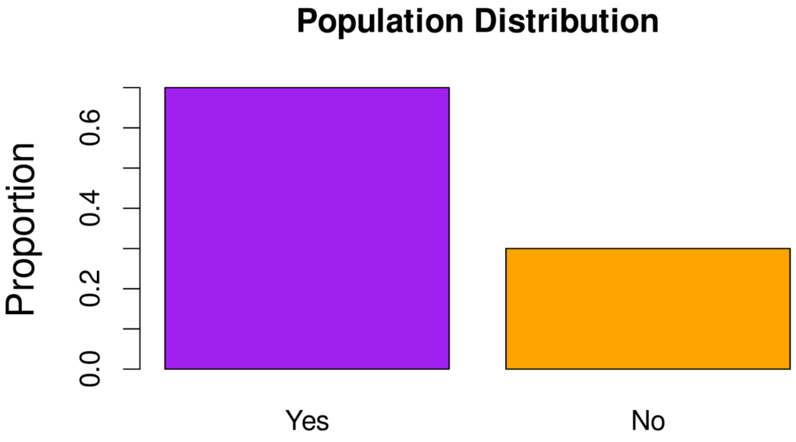
1. Population distribution vs data distribution
2. Sampling distribution
    1. Distribution of a binary categorical variable
    2. Distribution of a quantitative variable
3. Central Limit Theorem
4. Excel & Demo
5. **Recap**
    - Next time
    - Example exam questions

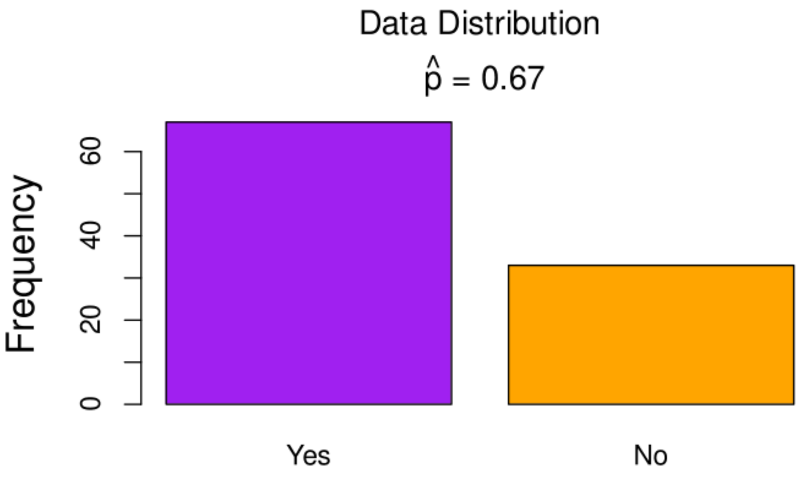
Recap

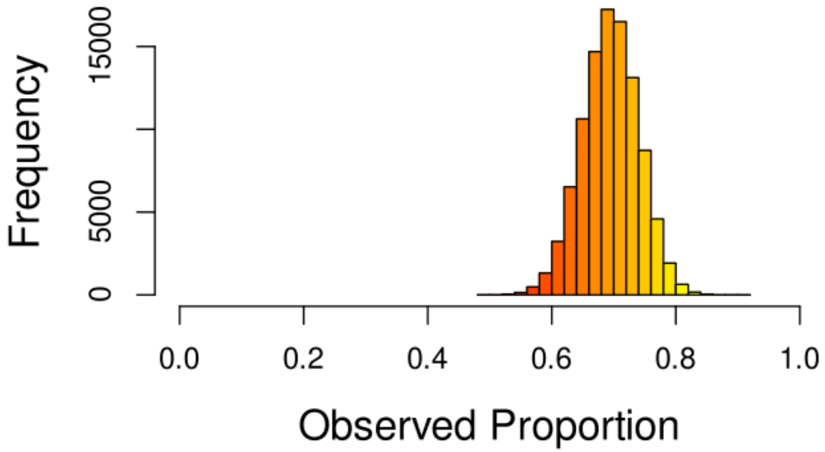
Sample (n=100)

Sampling distribution

Binomial distribution
n=100, p=0.7







$p = 0.7$

$\hat{p} = 0.67$

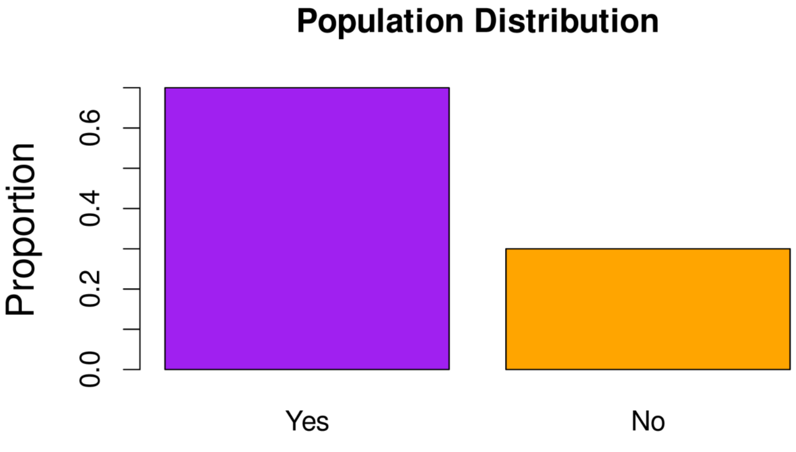
Sampling distribution: What values can you expect, if you would repeat an experiment?

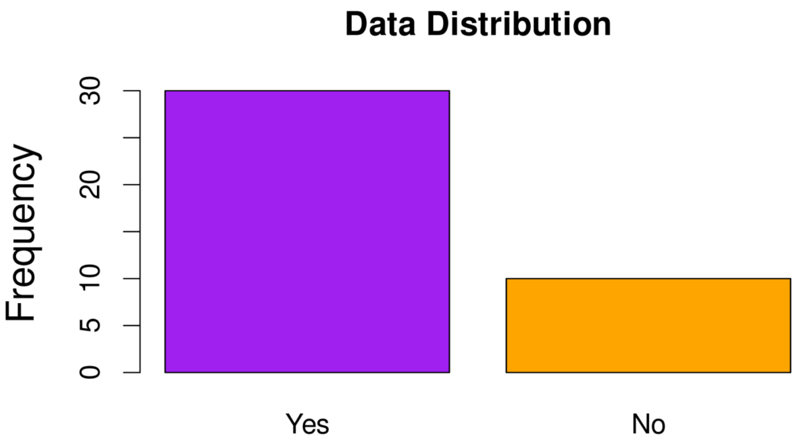
Recap

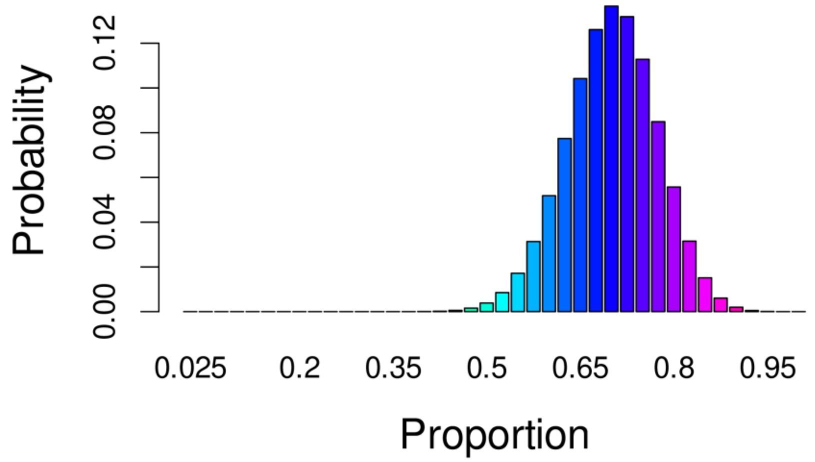
Sample (n=40)

Sampling distribution

Binomial distribution
n=40, p=0.7







$p = 0.7$

$\hat{p} = 0.75$

Behavior of the **variable**

Behavior of the **statistic** (on x-axis)

Recap

- Each *statistic* has a sampling distribution
- Based the sampling distribution, we can say something about the population
  - For the mean, the sampling distribution is the normal distribution
  - For the proportion, the sampling distribution is the binomial distribution
- OR the other way around: we assume/know/hypothesize the sampling distribution, so we can say something about the sample
  - Later in the course we use the sampling distribution to test hypotheses

Next Time

- This lecture: Confidence in parameter *estimates* (i.e., statistics)
- Next week: Using this confidence to draw inferences
- Problem is that we have *not observed a sampling distribution*, but *only one sample*
- → *Confidence interval*

Overview of Today

1. Population distribution vs data distribution
2. Sampling distribution
    1. Distribution of a binary categorical variable
    2. Distribution of a quantitative variable
3. Central Limit Theorem
4. Recap
    - Next time
    - **Example exam questions**

Example Exam Question 1

- IQ is normally distributed with a mean (μ) of 100 and a standard deviation (σ) of 15.
- If you would measure IQ in 100 random people, what is the probability of a sample mean of 103 or more?

a) 0.023
b) 0.12
c) 0.42

Solution

- Given the sample size and the Central Limit Theorem, you can assume that the sampling distribution is normally distributed with μ=100 and SE = $\frac{15}{\sqrt{100}} = 1.5$
- From this, you can compute the z-score, which indicates the probability:

  P($\bar{x}$>103) = P($\frac{\bar{x}-\mu}{\sigma} > \frac{103-100}{1.5}$)

  (Because total probability is always 1)

NORM.DIST(2; 0; 1; TRUE) =0.977, therefore P(z>2) = 1 – 0.977 = 0.023

- Alternative: P(z > 2) = P (z < -2) ← Because of symmetry

NORM.DIST(103; 100; 1.5; TRUE) also possible -> without conversion to z

Example Exam Question 2

- An old study showed that 4% of Dutch children under the age of 16 have ADHD. A researcher wants to understand whether that percentage is still accurate. She takes a random sample of 500 children from the appropriate population. 5.5% of the children in the sample have ADHD.
- What is the probability of finding a percentage of 5.5% or more in a sample, if the population percentage is 4%?

a) 0.043
b) 0.143
c) 0.469

Solution

We know that the sampling distribution is binomially distributed with *p* = 0.04 and

SE = $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.04(1-0.04)}{500}}$ = 0.0087

- Because of the sample size and the Central Limit Theorem, we can assume that the sampling distribution of the proportion is approximately normal

  - $np > 15$ 500 * 0.04 = 20

  - $n(1-p) > 15$ 500 * (1– 0.04) = 480

Solution

- We can calculate the z-score and look it up in Excel:
- Z-score = (0.055 – 0.04) / 0.0087 = 1.71

Solution

- From this, you can compute the z-score, which indicates the probability:

  $P(\hat{p}>0.055) = P(\frac{\hat{p}-\mu}{SE} > \frac{0.055-0.04}{0.00876}) = P(z > 1.71) = 1 - P(z<1.71)$

  (Because total probability is always 1)

- P(z<1.71)= NORM.DIST(1.71, 0, 1, TRUE) = 0.957
  - therefore P(z>1.71) = 1 – 0.957 = 0.043

- Alternative: P(z > 1.71) = P (z < -1.71) → Because of symmetry
  - NORM.DIST(-1.71, 0, 1, TRUE) = 0.043

Solution

- You can also skip the converting to z-scores and plug in mu and se directly into your Excel command:
  - NORM.DIST(0.055, 0.04, 0.00876, TRUE) = 0.957
  - 1-NORM.DIST(0.055, 0.04, 0.00876, TRUE) = 0.043

Questions?

Thank you for your attention

THE PROBLEM WITH AVERAGING STAR RATINGS

https://www.xkcd.com/937/

Highlighted exercises from the book

- 7.40
- 7.46
- 7.50

→ try yourself first, then check
next slides for answers

7.40

a) n = 50, p = 0.7; mean = p, sd = sqrt((0.7 * 0.3) / 50) = 0.0648 $\sqrt{\frac{p(1-p)}{n}}$

b) It will approximate the normal distribution, since

- np = 50 * 0.7 = 35
- n(1-p) = 50 * 0.3 = 15
- Both are greater than or equal to 15

c) See next slide

7.40

a)
b)
c)Convert the observation (0.6) to z-score, using the sampling distribution mean and standard deviation:

- (0.6 – 0.7) / 0.0648 = -1.5432
- Then look up this z-value in the table, to get the probability of this observation or lower (i.e., the area under the curve to the left of z)
- This is approximately equal to 0.0618.
- So if the population mean is 0.7, observing 0.6 or less (with 50 questions) has a probability of +- 6%, which is pretty low/surprising (meaning something might have gone wrong in your exam preparations!)

7.46

a) The mean is 900, and standard deviation 300, for the population (see the symbols – these indicate the parameters)

b) The data statistics are also reported in the story: mean is 980 and standard deviation is 276 – again, these are about the **variable**, so this is about the spread of the daily sales.

c) See next slide

7.46

a)

b)

c) Mean and standard deviation of sampling distribution are about the **statistic**. The mean is equal to the population mean, and the standard deviation (i.e., standard error) is equal to the population sd divided by sqrt(n). The average is computed over 7 days, so the SE is equal to 300 / sqrt(7) = 113.39. The SE describes how much the **mean** varies if we were to keep taking 7-day samples and compute the mean for those 7 days.

7.50

a) Since this is about the z-score for **a single value** of PDI (and not a statistic), we use the data standard deviation instead of the SE:
Z-score: (90 – 100) / 15 = 0.6667

b) Now we are talking about the **mean** PDI, so we use the SE for the z-score calculation:
Z-score: (90 – 100) / (15 / sqrt(225)) = 10

c) See next slide

7.50

This wall of text is just meant for intuition – your answer does not need to completely overlap

a)

b)

c) The z-score for the mean of 225 observations is 10, whereas the z-score for a single observation is 0.6667. In the latter case, you can see the score of 90 as the mean PDI for a sample of 1 (so the standard error is equal to sigma). For a sample size of 1, the sampling distribution is much more wide than for a sample size of 225, since these values will fluctuate a lot more. It is therefore much more unlikely to observe a mean of 90 based on 225 participants. (extra intuition: to observe a mean of 90, a lot of those 225 need to score even lower than that 90, which is more unlikely than scoring 90)

Source: Pixabay.org