Elementary Statistics – Principles of Hypothesis Testing

Dr. Ab Mosca (they/them)

Slides based off slides courtesy of Kaitlyn Cook (https://www.smith.edu/people/kaitlyn-cook)

Plan for Today

- Hypothesis Testing
  - Statistical Inference
  - Central Limit Theorem
  - Normal Distribution

Warm Up: Statistics and CIs

In statistics, we want to know about populations, but we only have sample data to work with.

So we estimate population parameters using ***sample statistics***.

- Sample mean: $\bar{x}$
- Sample proportion: $\hat{p}$

Sample statistics are random variables

Practice:

Let an experiment be rolling a loaded die (loaded to land on six 50% of the time) 1000 times and counting the proportion of sixes.

In this scenario, what is p, and what is $\hat{p}$?

If I repeat this experiment 100 times and construct a 95% confidence interval (CI) of $\hat{p}$ each time, how many CI's would you expect to capture p?

Models of Statistical Inference

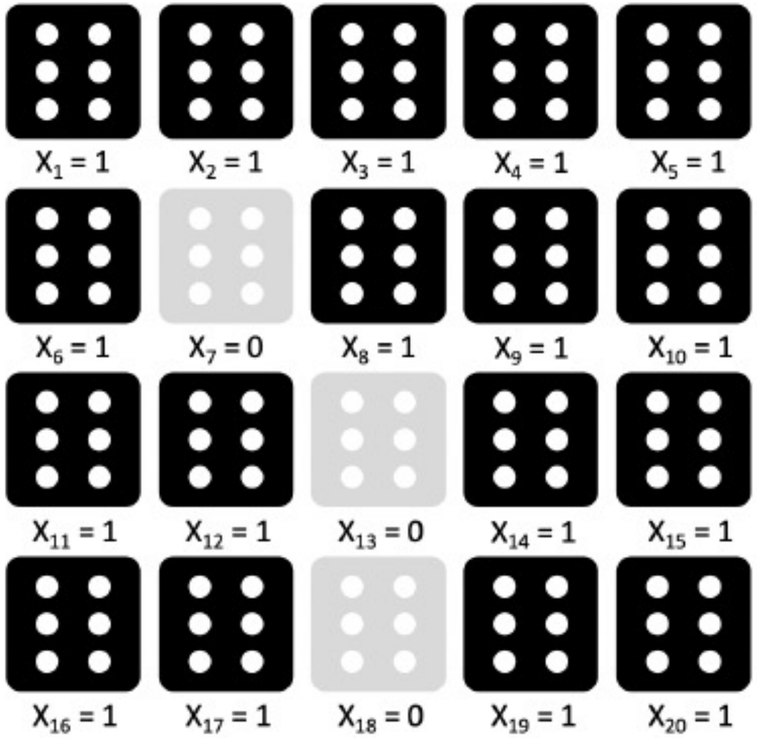
Our sample statistics represent our best guess for the true population parameter. We know this best guess is not perfect; we expect error (variability) die to the sampling process.

Because we can't know the truth directly we:

1. Construct a confidence interval
   - Expresses the uncertainty that we have in our estimate
   - Describes the range of plausible values given our observed data
2. Conduct a hypothesis test
   - Posits a specific explanation for how data were generated
   - Checks if the observed data is consistent with that explanation

Loaded Die?

Recall an experiment where we gathered data by rolling a die 20 times. We observed 17 sixes:



This feels like evidence that the die is loaded, but how compelling is it? Is there a compelling alternative explanation for this data?

The Role of Random Chance

What if the die was fair?

→ We would expect to see around 1/6 of rolls be sixes, but due to chance we wouldn't expect *exactly* 1/6 to be sixes

The Role of Random Chance

What if the die was fair?

→ We would expect to see around 1/6 of rolls be sixes, but due to chance we wouldn't expect *exactly* 1/6 to be sixes

So…

If the die was truly fair, just how unusual would be be to observe 17 sixes in 20 rolls?

We can quantify this with a hypothesis test

Pieces to a Hypothesis Test

1. ***Two competing and complementary claims about the world:***

Null Hypothesis ($\boldsymbol{H_0}$) is a statement about the population that represents the status quo (i.e. that nothing, or null, is different).

Pieces to a Hypothesis Test

1. ***Two competing and complementary claims about the world:***

Null Hypothesis ($\boldsymbol{H_0}$) is a statement about the population that represents the status quo (i.e. that nothing, or null, is different).

Ex. $H_0\text{: } the\ die\ is\ fair$

$H_0\text{: } the\ die\ rolls\ sixes\ with\ a\ probability\ of\ \frac{1}{6}$

$H_0: p = \frac{1}{6}$

Pieces to a Hypothesis Test

1. ***Two competing and complementary claims about the world:***

Null Hypothesis ($\boldsymbol{H_0}$) is a statement about the population that represents the status quo (i.e. that nothing, or null, is different).

Ex. $H_0\text{: } the\ die\ is\ fair$

$H_0\text{: } the\ die\ rolls\ sixes\ with\ a\ probability\ of\ \frac{1}{6}$

$H_0: p = \frac{1}{6}$

Alternative Hypothesis ($\boldsymbol{H_A}$) is a statement about the population that represents our research question (i.e. that something is different).

Pieces to a Hypothesis Test

1. ***Two competing and complementary claims about the world:***

Null Hypothesis ($\boldsymbol{H_0}$) is a statement about the population that represents the status quo (i.e. that nothing, or null, is different).

Ex. $H_0:\ the\ die\ is\ fair$

$H_0:\ the\ die\ rolls\ sixes\ with\ a\ probability\ of\ \frac{1}{6}$

$H_0: p = \frac{1}{6}$

Alternative Hypothesis ($\boldsymbol{H_A}$) is a statement about the population that represents our research question (i.e. that something is different).

Ex. $H_0:\ the\ die\ is\ not\ fair$

$H_0:\ the\ die\ rolls\ sixes\ with\ a\ probability\ other\ than\ 1/6$

$H_0: p \neq \frac{1}{6}$

Pieces to a Hypothesis Test

1. ***Two competing and complementary claims about the world***
   A null hypothesis (H_0) and an alternative hypothesis (H_A)
   Ex. $H_0: p = \frac{1}{6}, H_A: p \neq \frac{1}{6}$

2. ***Test Statistic***
   A metric calculated with observed data that summarized how compatible the data are with H_0

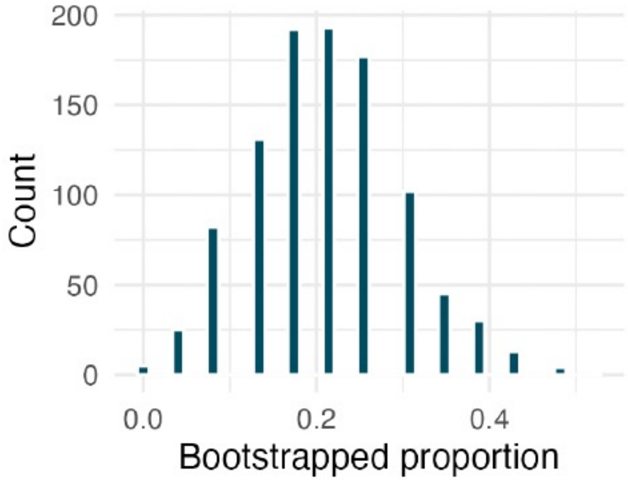
What is our test statistic for our die experiment where we observed 17 sixes in 20 rolls?

Pieces to a Hypothesis Test

1. ***Two competing and complementary claims about the world***
   A null hypothesis (H_0) and an alternative hypothesis (H_A)
   Ex. $H_0: p = \frac{1}{6}, H_A: p \neq \frac{1}{6}$

2. ***Test Statistic***
   A metric calculated with observed data that summarized how compatible the data are with H_0
   Ex. $\hat{p} = \frac{17}{20} = 0.85$

3. ***Null Distribution***
   The sampling distribution for our chosen test statistic under the assumption that our null hypothesis is true.
   Ex.

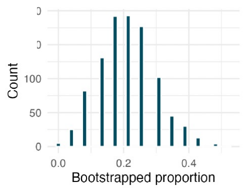


Pieces to a Hypothesis Test

1. ***Two competing and complementary claims about the world***
   A null hypothesis (H_0) and an alternative hypothesis (H_A)
   Ex. $H_0: p = \frac{1}{6}, H_A: p \neq \frac{1}{6}$

2. ***Test Statistic***
   A metric calculated with observed data that summarized how compatible the data are with H_0
   Ex. $\hat{p} = \frac{17}{20} = 0.85$

3. ***Null Distribution***
   The sampling distribution for our chosen test statistic under the assumption that our null hypothesis is true.
   Ex.



4. ***P-value***
   The probability of obtaining a test statistic as rare or more rare than our observed test statistic if the null hypothesis were true.

P-Values

P-value is a conditional probability that tells us how unusual our test statistic would be *given the null hypothesis is true*.

Ex. p-value = $P(\hat{p} \geq 0.85 \mid p = \frac{1}{6})$

A *high p-value* implies it is *highly likely to observe our sample* statistic if the null hypothesis it true.

A *low p-value* implies it is *highly unlikely to observe our sample* statistic if the null hypothesis is true.

P-Values

P-value is a conditional probability that tells us how unusual our test statistic would be *given the null hypothesis is true*.

Ex. p-value = $P(\hat{p} \geq 0.85 \mid p = \frac{1}{6})$

A *high p-value* implies it is *highly likely to observe our sample* statistic if the null hypothesis it true.

A *low p-value* implies it is *highly unlikely to observe our sample* statistic if the null hypothesis is true.

Because we assume H_0 is true to calculate our p-value, we cannot use it in support of H_0. ***A high p-value suggests $\boldsymbol{H_0}$ is true but does not prove $\boldsymbol{H_0}$ is true.***

P-Values

P-value is a conditional probability that tells us how unusual our test statistic would be *given the null hypothesis is true*.

Ex. p-value = $P(\hat{p} \geq 0.85 \mid p = \frac{1}{6})$

A ***high p-value*** implies it is *highly likely to observe our sample* statistic if the null hypothesis it true.

When we see a high p-value we "**Fail to reject $\boldsymbol{H_0}$**"

A ***low p-value*** implies it is *highly unlikely to observe our sample* statistic if the null hypothesis is true.

When we see a low p-value we "**Reject $\boldsymbol{H_0}$**"

Because we assume H_0 is true to calculate our p-value, we cannot use it in support of H_0. ***A high p-value suggests $\boldsymbol{H_0}$ is true but does not prove $\boldsymbol{H_0}$ is true.***

P-Values

What is the delineation between a high and low p-value?

Before we perform hypothesis testing we choose a threshold for high vs low p-values. We call this threshold α

If **p-value $\leq \alpha$ we reject H_0**
If **p-value $> \alpha$ we fail to reject H_0**

When a p-value $\leq \alpha$ we say ***a statistically significant difference*** exists.

α can be any probability, but the most common is 5%.

P-Values

What is the delineation between a high and low p-value?

Before we perform hypothesis testing we choose a threshold for high vs low p-values. We call this threshold α

If **p-value $\leq \alpha$ we reject H_0**
If **p-value $> \alpha$ we fail to reject H_0**

When a p-value $\leq \alpha$ we say ***a statistically significant difference*** exists.

α can be any probability, but the most common is 5%.

Let $\alpha = 0.05$. What would we do with respect to H_0 if we calculated a p-value of 0.01? 0.15?

P-Values

What is the delineation between a high and low p-value?

Before we perform hypothesis testing we choose a threshold for high vs low p-values. We call this threshold α

If **p-value $\leq \alpha$ we reject H_0**
If **p-value $> \alpha$ we fail to reject H_0**

When a p-value $\leq \alpha$ we say ***a statistically significant difference*** exists.

α can be any probability, but the most common is 5%.

Let $\alpha = 0.01$. What would we do with respect to H_0 if we calculated a p-value of 0.01? 0.15?

P-Values

What is the delineation between a high and low p-value?

Before we perform hypothesis testing we choose a threshold for high vs low p-values. We call this threshold α

If **p-value $\leq \alpha$ we reject H_0**
If **p-value $> \alpha$ we fail to reject H_0**

When a p-value $\leq \alpha$ we say ***a statistically significant difference*** exists.

α can be any probability, but the most common is 5%.

Let $\alpha = 0.1$. What would we do with respect to H_0 if we calculated a p-value of 0.01? 0.15?

P-Values

What is the delineation between a high and low p-value?

Before we perform hypothesis testing we choose a threshold for high vs low p-values. We call this threshold α

If **p-value $\leq \alpha$ we reject H_0**
If **p-value $> \alpha$ we fail to reject H_0**

When a p-value $\leq \alpha$ we say ***a statistically significant difference*** exists.

α can be any probability, but the most common is 5%.

Let $\alpha = 0.1$. What would we do with respect to H_0 if we calculated a p-value of 0.01? 0.15?

P-Values

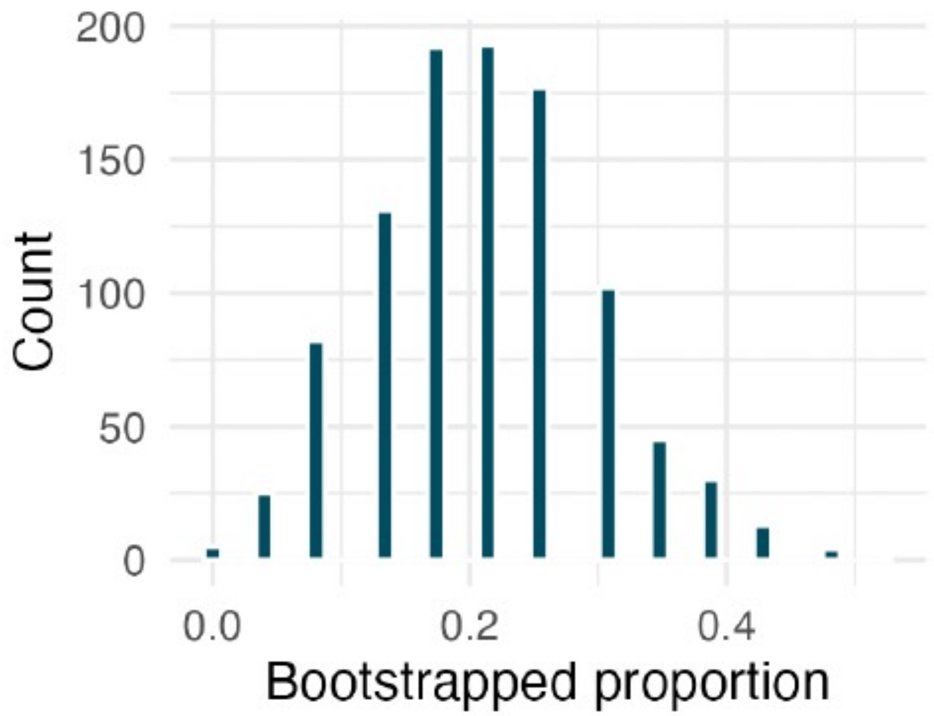
A ***p-value*** is the probability of obtaining a test statistic as rare or more rare than our observed test statistic if the null hypothesis were true.

How do we get the p-value?

P-Values

A ***p-value*** is the probability of obtaining a test statistic as rare or more rare than our observed test statistic if the null hypothesis were true.

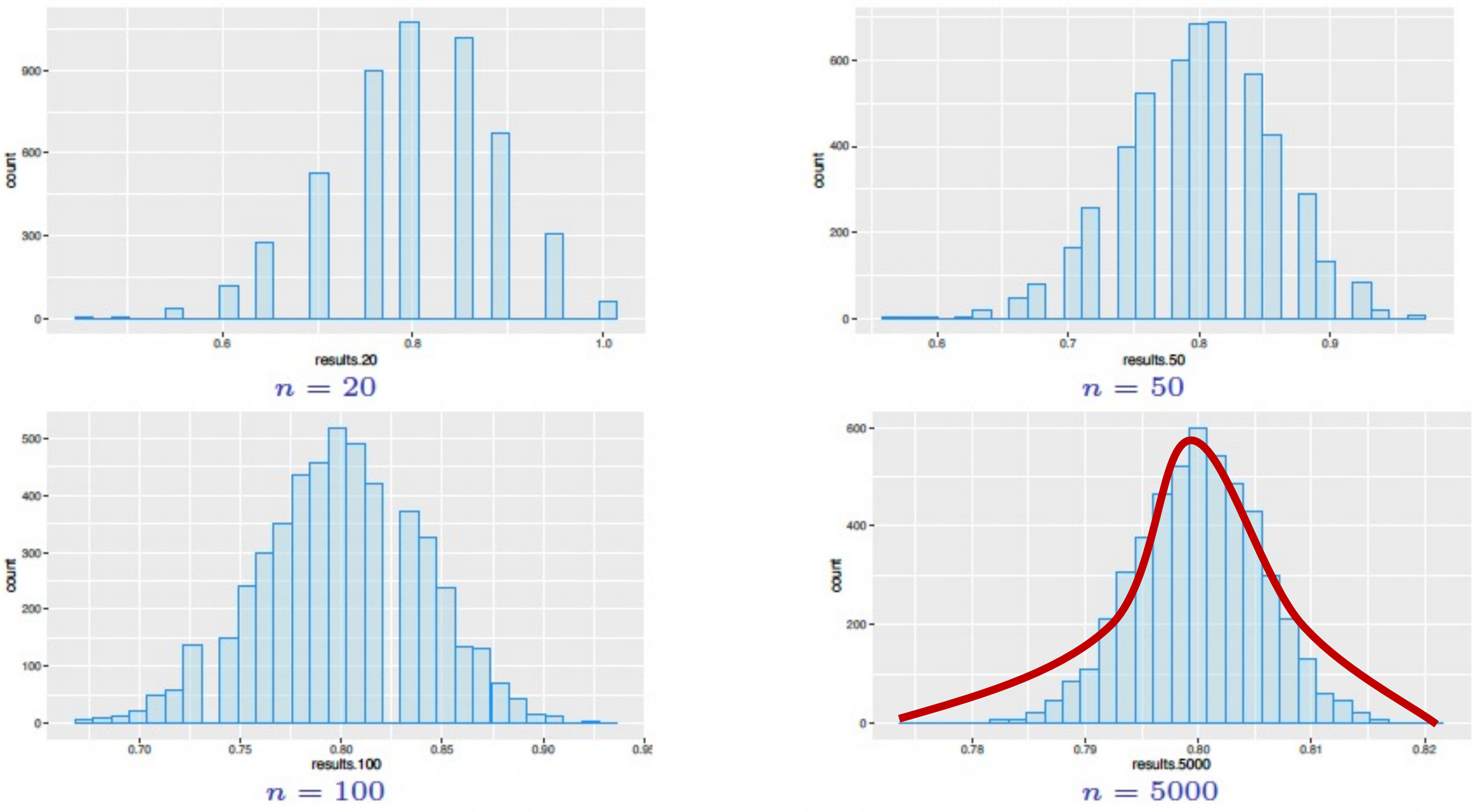
How do we get the p-value? From our null distribution, generated with bootstrap.



Null Distribution

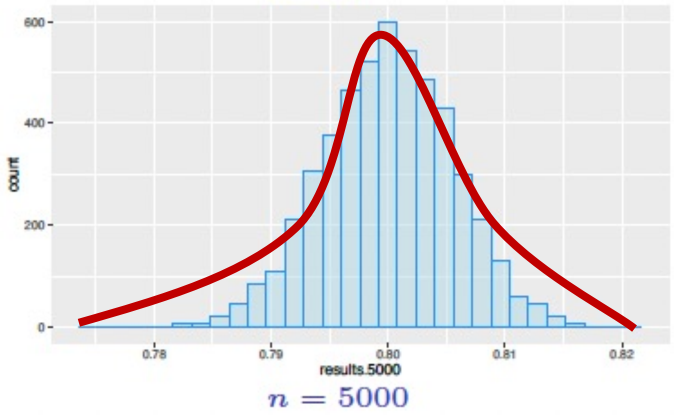
Under increased sample sizes, sampling distributions (of continuous random variables) form a ***density curve***.

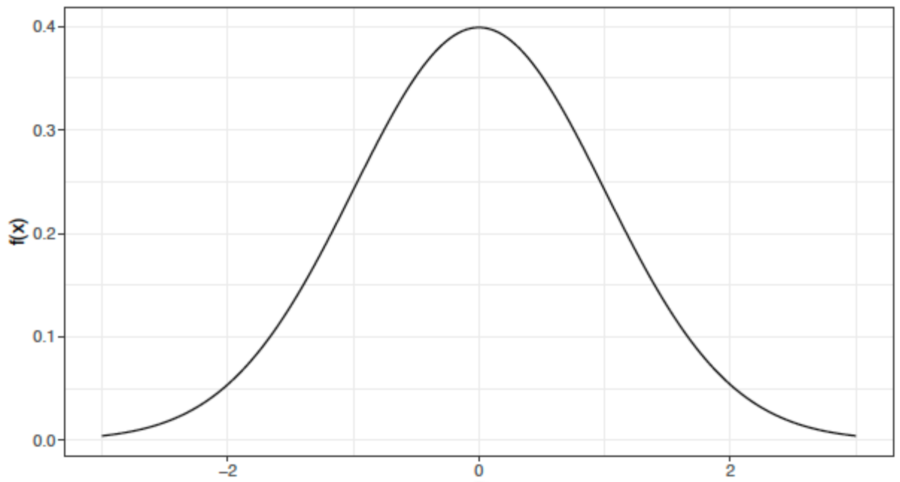
Ex. n = rolls, $\hat{p}$ = proportion of sixes



Central Limit Theorem

The ***central limit theorem*** states if observations in a sample are independent and the size of the sample is large, then the sampling distribution of a parameter is well-approximated the Normal distribution.

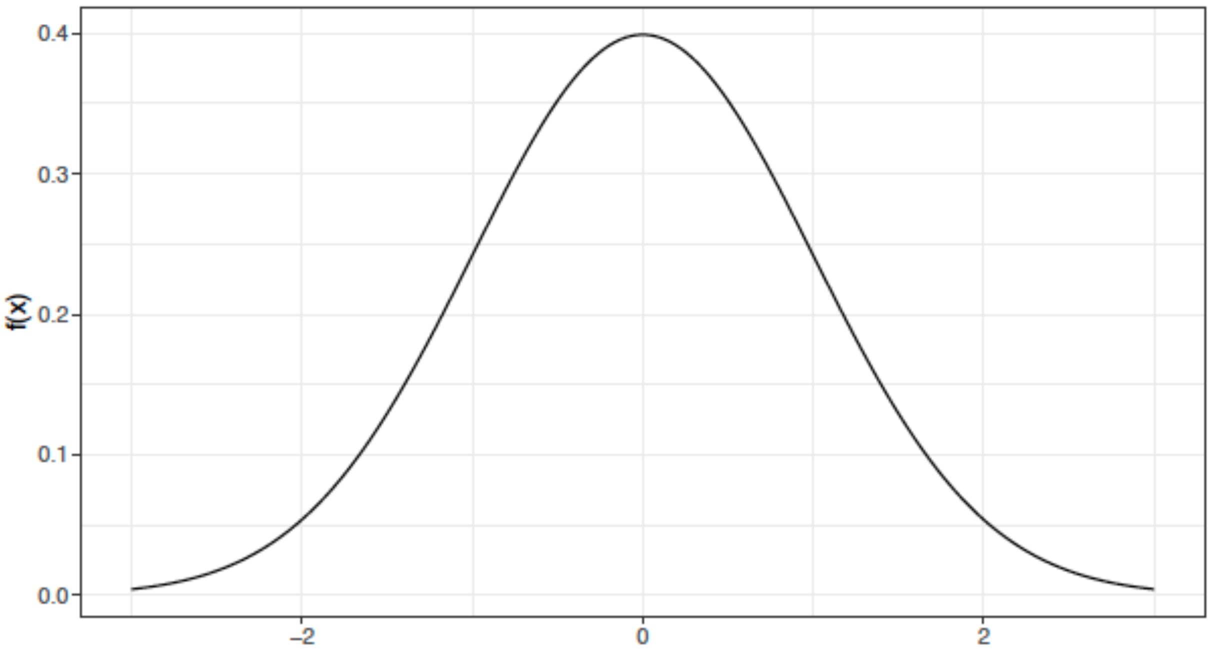




Symmetric, unimodal, bell-shaped

Normal Distribution

The ***Normal distribution*** describes a continuous random variable whose density curve is symmetric, unimodal, and bell-shaped.



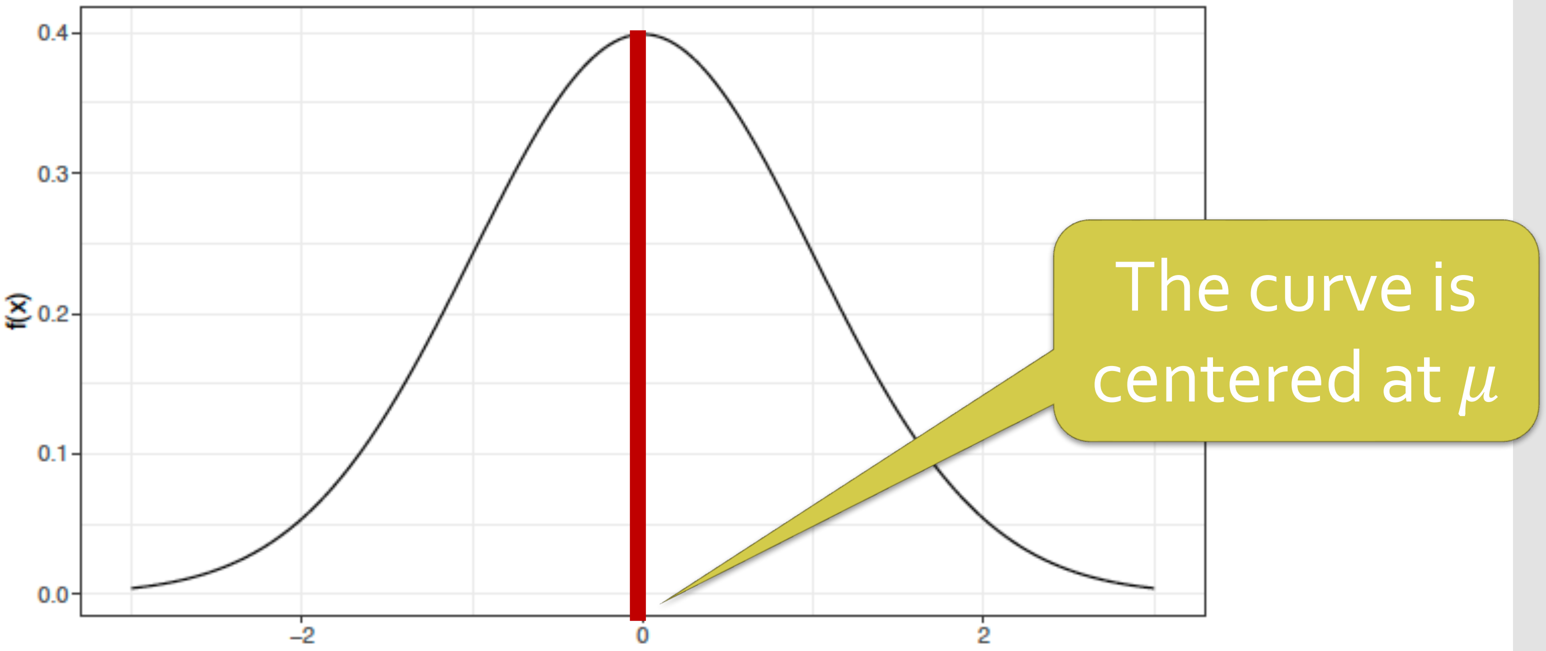
We write $X \sim Norm(\mu, \sigma)$ where μ is the mean of the random variable, X, and σ is its standard deviation.

When $\mu = 0, \sigma = 1$ we call the distribution the ***standard Normal distribution***.

Normal Distribution

The ***Normal distribution*** describes a continuous random variable whose density curve is symmetric, unimodal, and bell-shaped.

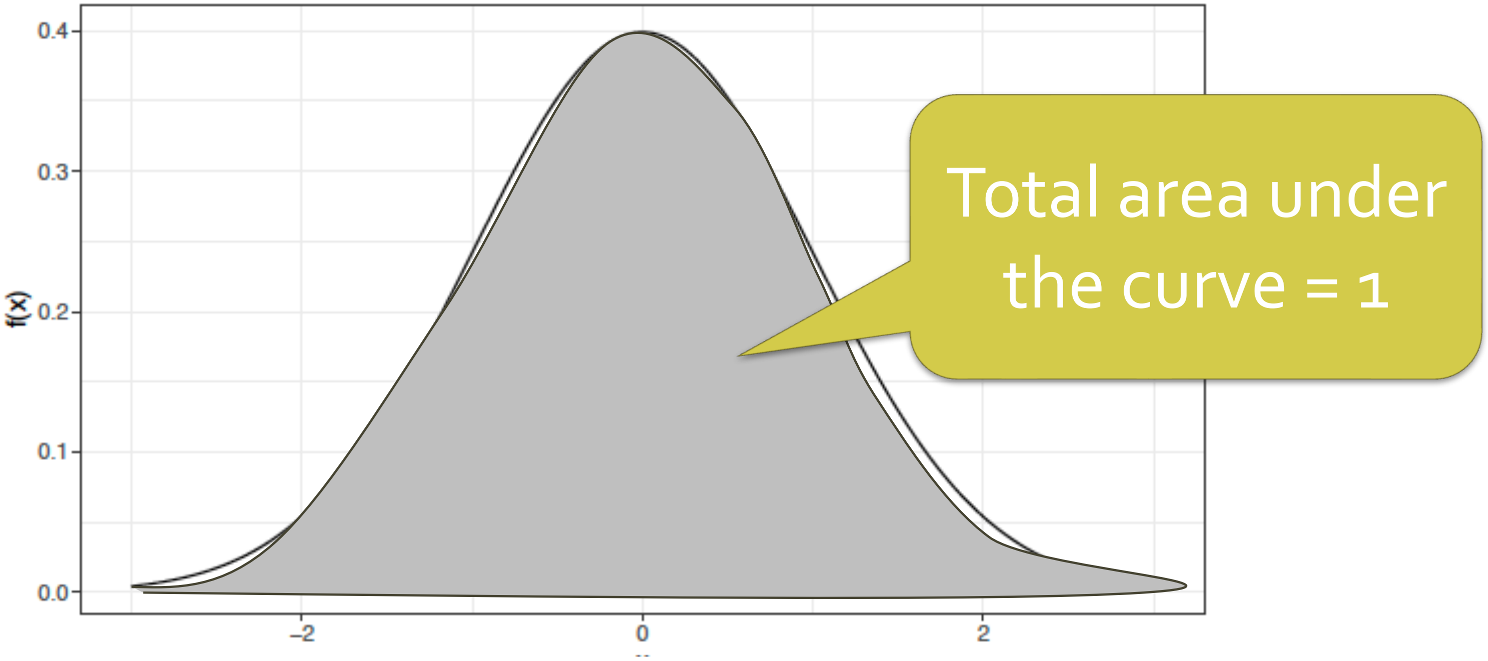
We write $X \sim Norm(\mu, \sigma)$ where μ is the mean of the random variable, X, and σ is its standard deviation.



Normal Distribution

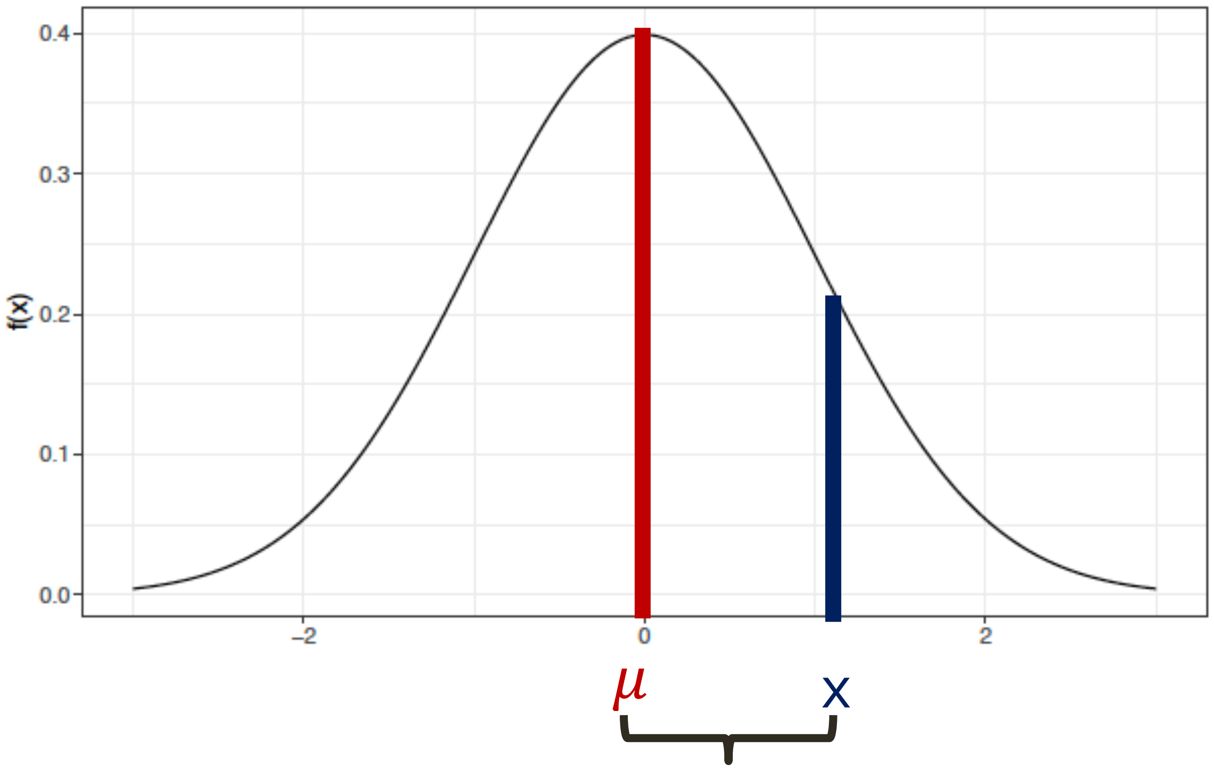
The ***Normal distribution*** describes a continuous random variable whose density curve is symmetric, unimodal, and bell-shaped.

We write $X \sim Norm(\mu, \sigma)$ where μ is the mean of the random variable, X, and σ is its standard deviation.



Normal Distribution

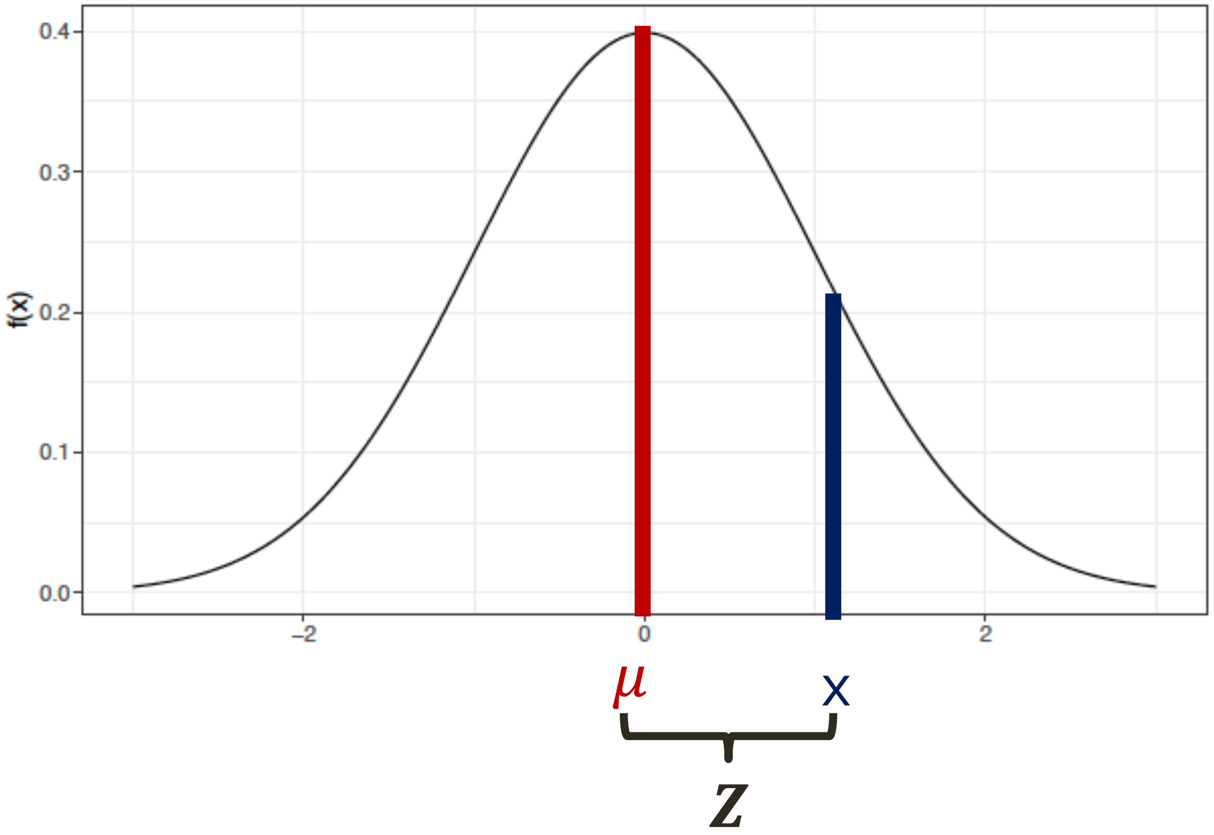
Normal distribution: $X \sim Norm(\mu, \sigma)$, μ is the mean of the random variable, X, and σ is its standard deviation.



For any observation, x, we can quantify how unusual it is by looking at how many σ's away from μ it falls.

Normal Distribution

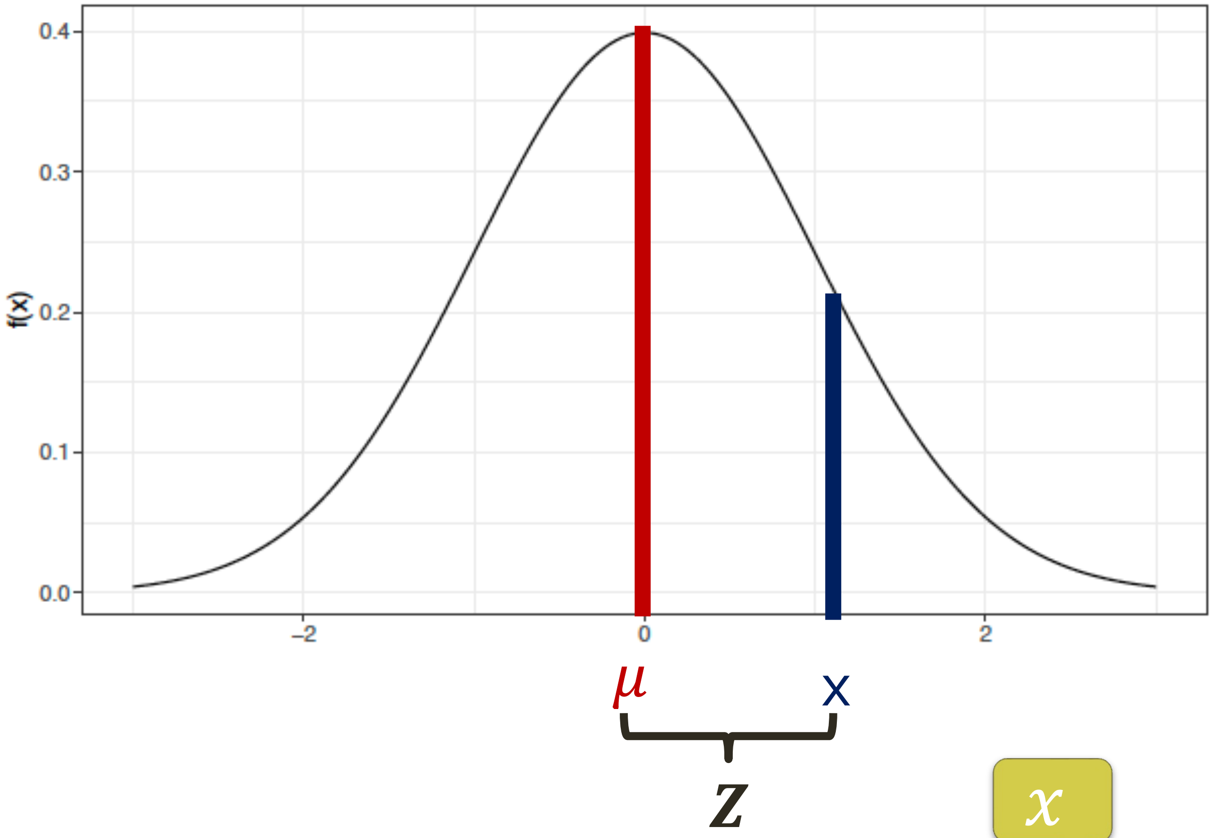
Normal distribution: $X \sim Norm(\mu, \sigma)$, μ is the mean of the random variable, X, and σ is its standard deviation.



For any observation, x, we can quantify how unusual it is by looking at how many σ's away from μ it falls, this is called a ***z-score***.

Z-scores

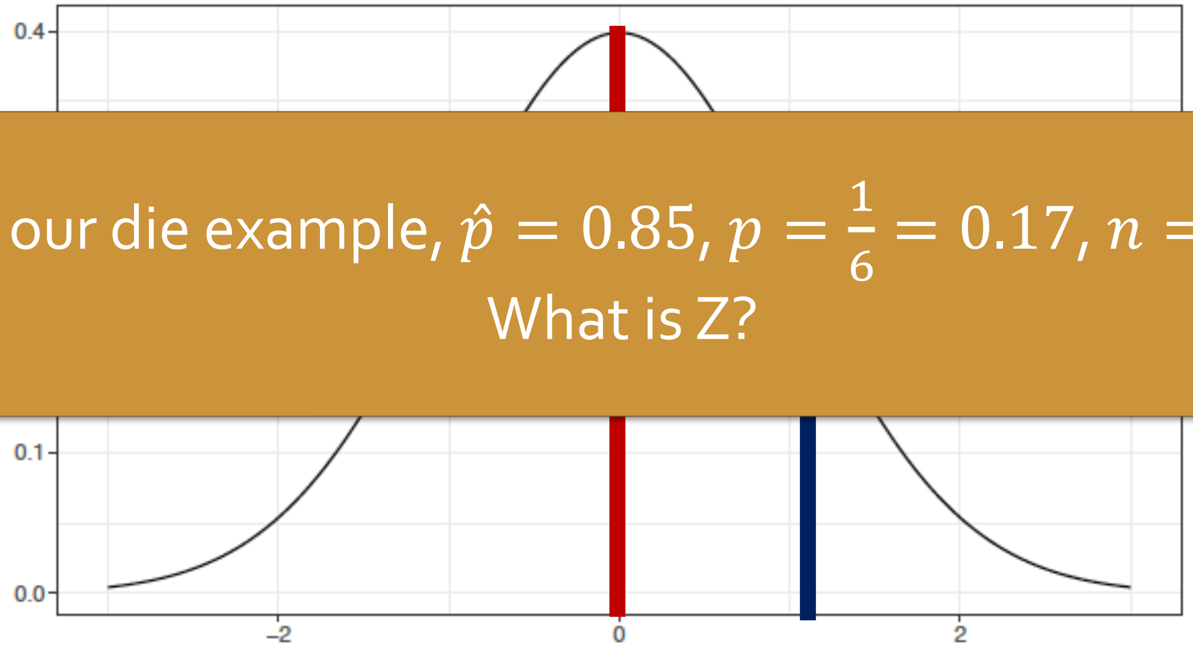
The ***z-score*** of an observation characterizes the number of standard deviations it falls above or below the mean.



$Z = \frac{x-\mu}{\sigma}$, for a sample proportion, $\hat{p}$, $Z = \dfrac{\underbrace{\hat{p}}_{x}-\underbrace{p}_{\mu}}{\underbrace{\sqrt{\frac{p(1-p)}{n}}}_{\sigma}}$

Z-scores

The ***z-score*** of an observation characterizes the number of standard deviations it falls above or below the mean.



In our die example, $\hat{p} = 0.85$, $p = \frac{1}{6} = 0.17$, $n = 20$

What is Z?

μ x

$\underbrace{\qquad\qquad}_{Z}$

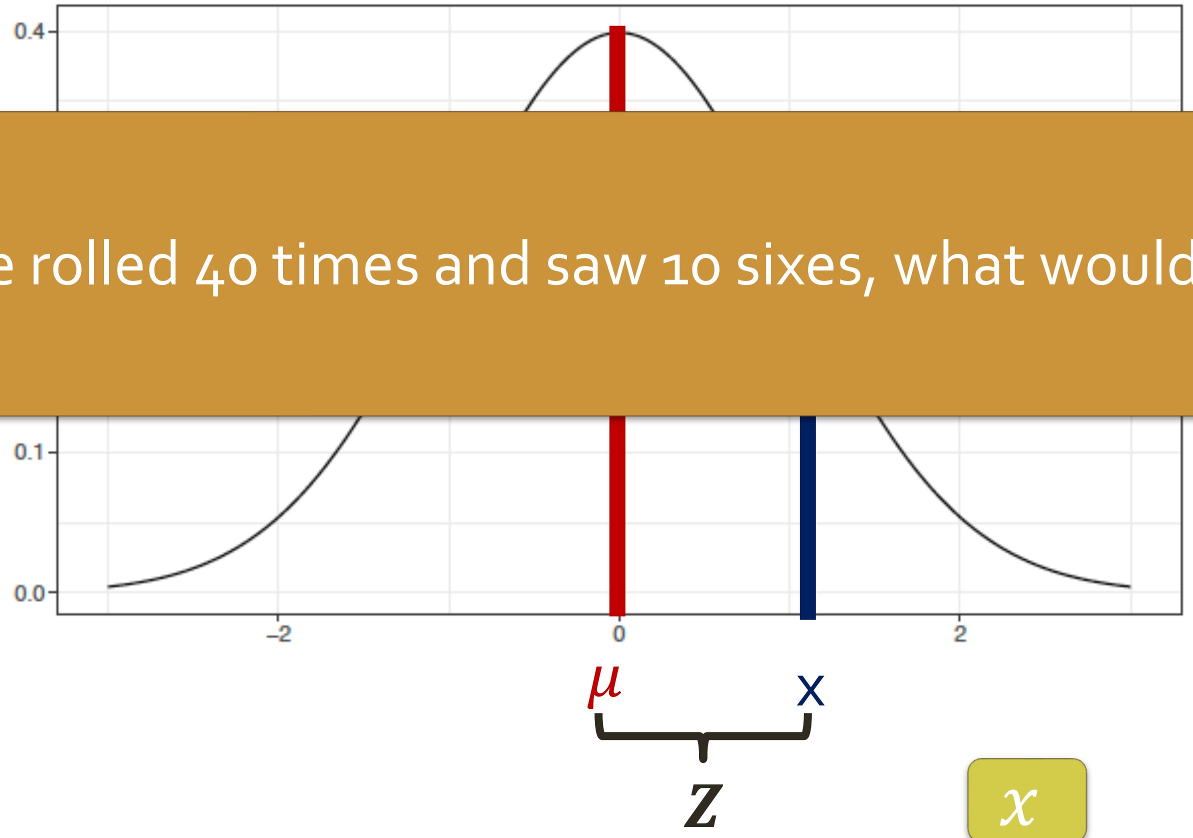
x μ

$Z = \frac{x-\mu}{\sigma}$, for a sample proportion, $\hat{p}$, $Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$

σ

Z-scores

The ***z-score*** of an observation characterizes the number of standard deviations it falls above or below the mean.



If we rolled 40 times and saw 10 sixes, what would Z be?

$Z = \frac{x-\mu}{\sigma}$, for a sample proportion, $\hat{p}$, $Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$

x

μ

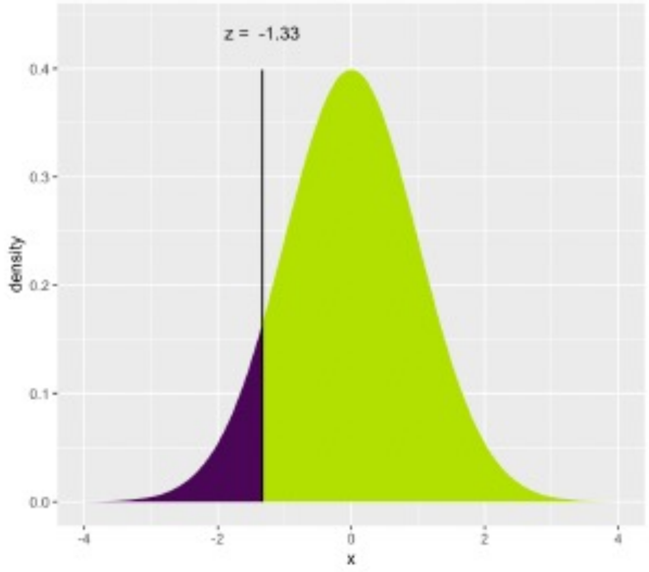
σ

Z-scores

Given a ***z-score***, we can use the standard normal distribution to find several probabilities.

Probability that of seeing an observation less than or equal to x:

$$P(Z \leq -1.33)$$



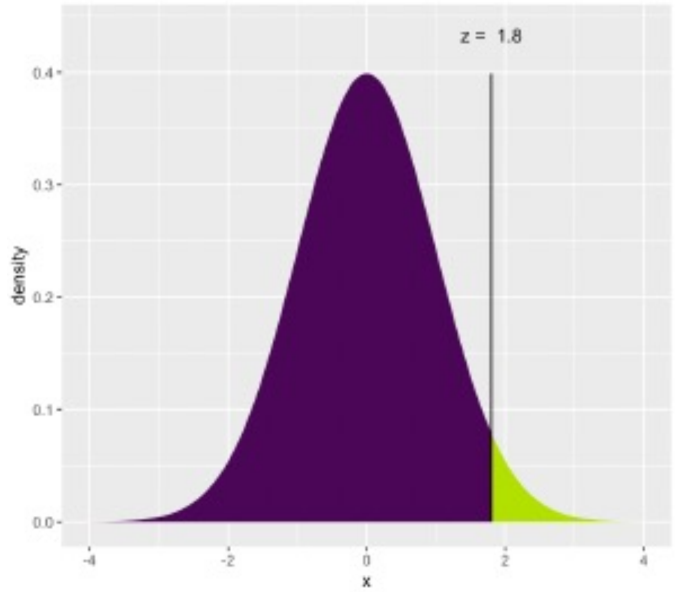
Area under the curve to the left of the z-score = probability of seeing a smaller z-score. Often called the ***percentile.***

Z-scores

Given a ***z-score***, we can use the standard normal distribution to find several probabilities.

Probability that of seeing an observation greater than x:

$$P(Z > 1.8)$$



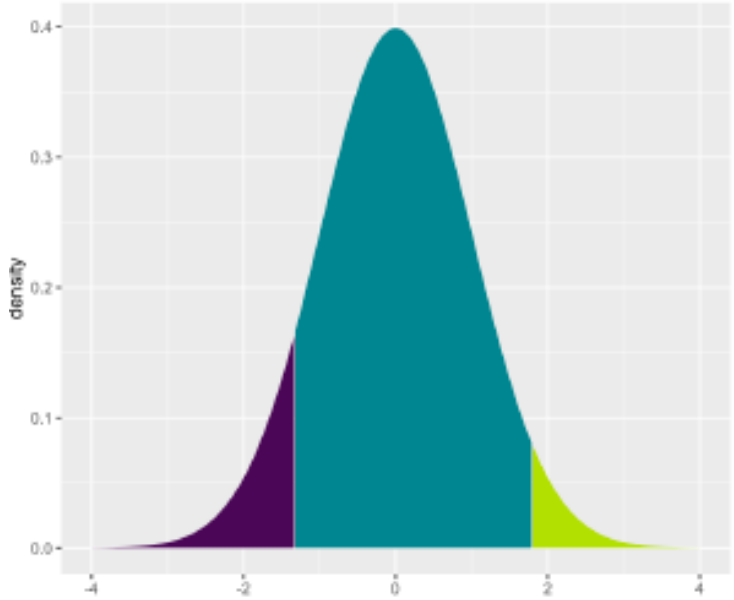
Area under the curve to the right of the z-score = probability of seeing a larger z-score.

Z-scores

Given a ***z-score***, we can use the standard normal distribution to find several probabilities.

Probability that of seeing an observation between two x's:

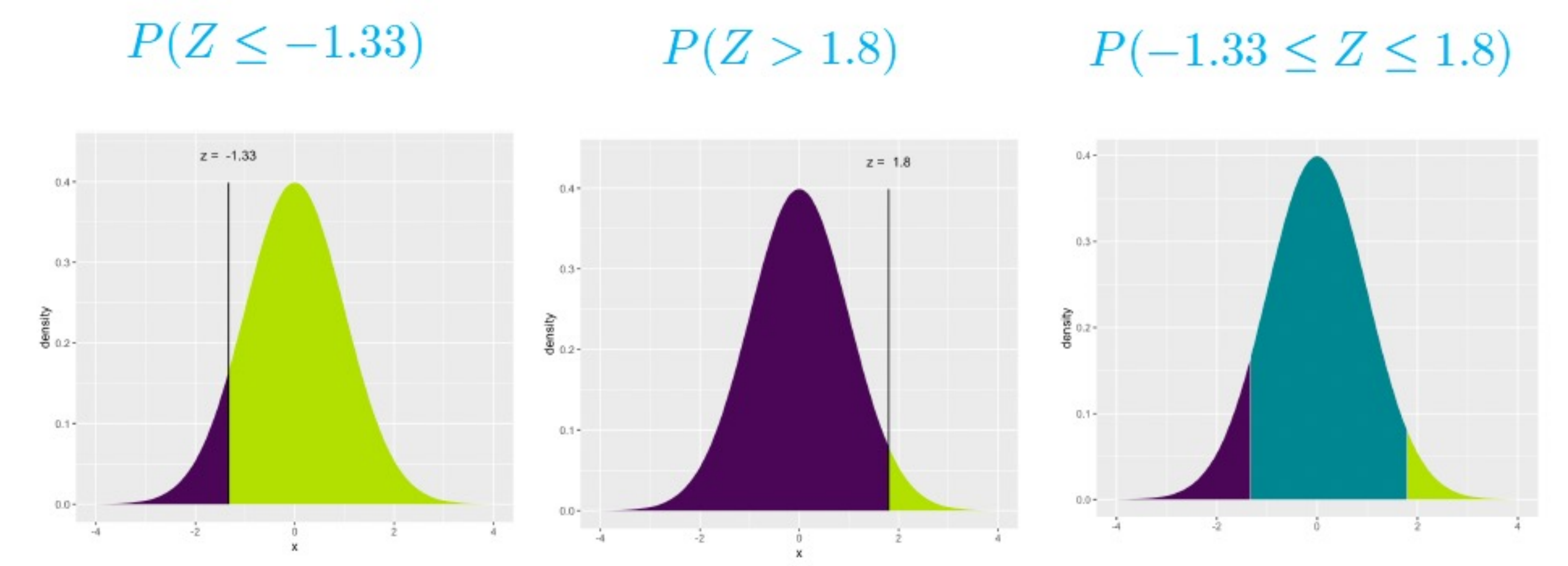
$$P(-1.33 \leq Z \leq 1.8)$$



Area under the curve between the two values. Equivalent to $P(Z \leq 1.8) - P(Z \leq$. -1.33).

Z-scores

Given a ***z-score***, we can use the standard normal distribution to find several probabilities.



Z-scores

Given a ***z-score***, we can use the standard normal distribution to find several probabilities.

Calculating area under the standard normal distribution ($Norm(0, 1)$) for a z-score, z.

Option 1:

- z-score table: https://www.z-table.com/

Option 2:

- online calculator: https://www.calculator.net/z-score-calculator.html

Option 3:

- regular scientific calculator

Option 4:

- excel =NORMSDIST(Z)

Z-scores

What str the percentiles for z = 8.1, and for z = 1.34?

Calculating area under the standard normal distribution ($Norm(0,1)$) for a z-score, z.

Option 1:

- z-score table: https://www.z-table.com/

Option 2:

- online calculator: https://www.calculator.net/z-score-calculator.html

Option 3:

- regular scientific calculator

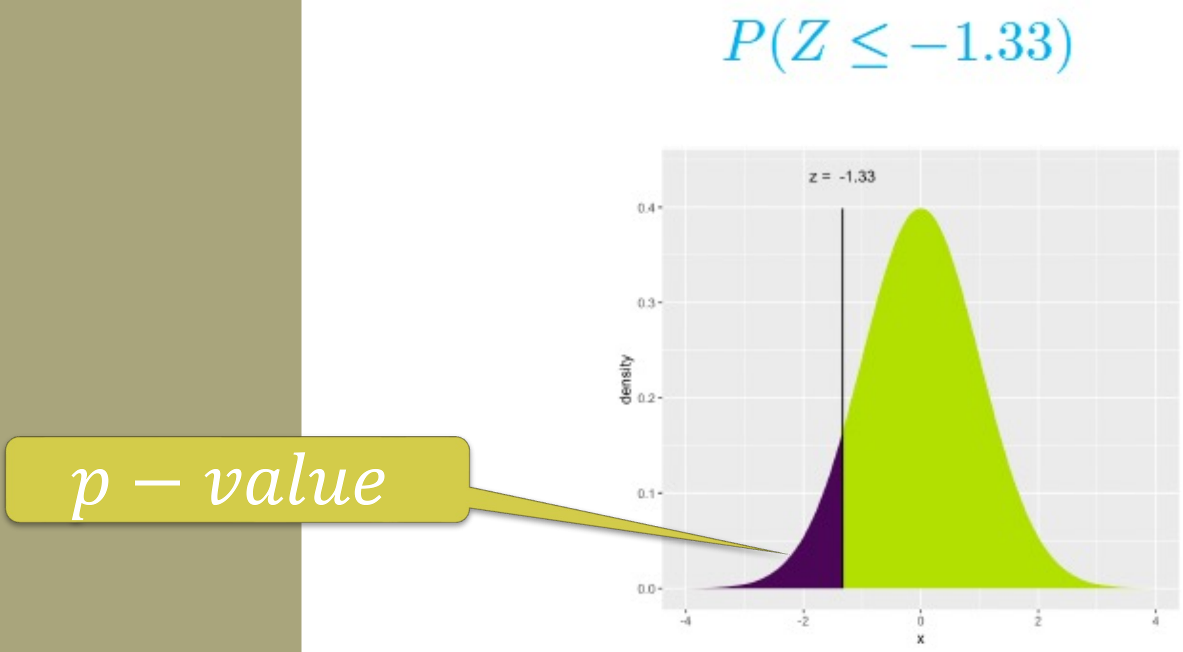
Option 4:

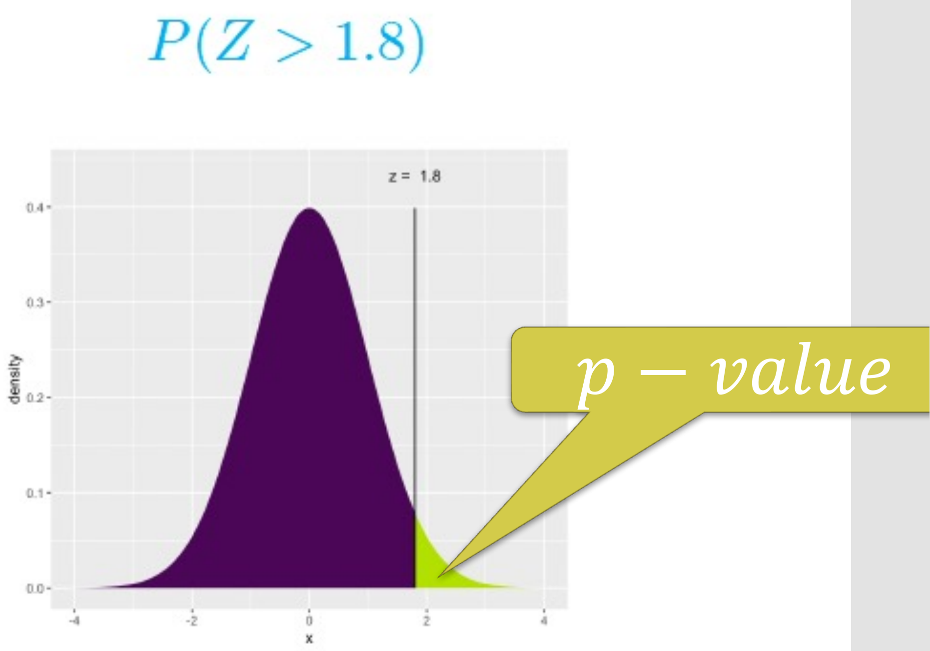
- excel =NORMSDIST(Z)

Z-scores

Given a ***z-score***, we can use the standard normal distribution to find several probabilities.

The probability of seeing an observation as or more extreme than the one we saw given our null hypothesis is the probability of seeing a z-score as or more extreme than the z-score we got for our observation.



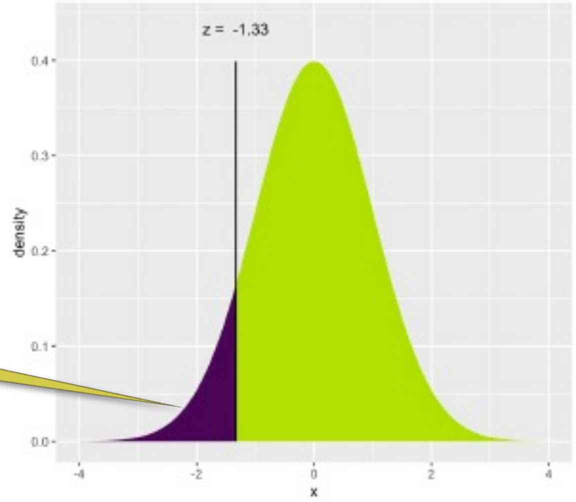


Given the percentile 1 for z = 8.1 (from $\hat{p} = 0.85$), what is the p-value?
Given the percentile 0.91 for z = 1.34. (from $\hat{p} = 0.25$), what is the p-value?

Z-scores

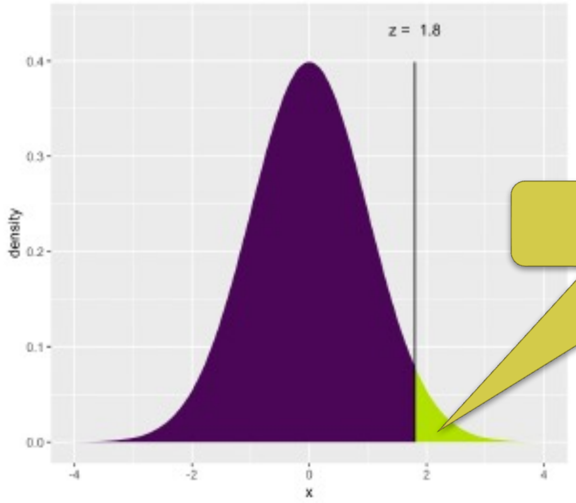
The probability of seeing an observation as or more extreme than the one we saw given our null hypothesis is the probability of seeing a z-score as or more extreme than the z-score we got for our observation.

$P(Z \leq -1.33)$



$p - value$

$P(Z > 1.8)$



$p - value$

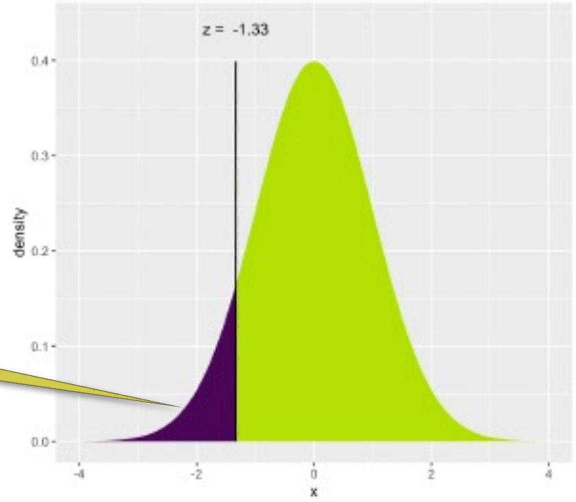
Given the percentile 1 for z = 8.1 (from $\hat{p} = 0.85$), what is the p-value?
Given the percentile 0.91 for z = 1.34. (from $\hat{p} = 0.25$), what is the p-value?

If we use $\alpha = 0.05$, do we reject or fail to reject our H_0 that the die is fair in each of these experiments?

Z-scores

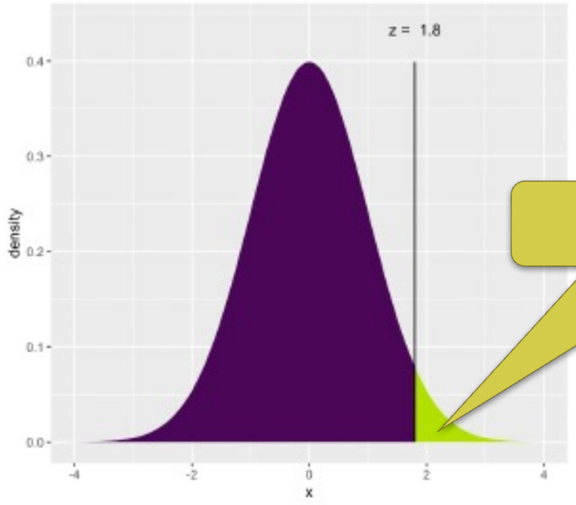
for our observation:

$P(Z \leq -1.33)$



$p - value$

$P(Z > 1.8)$



$p - value$

Z-scores

For any normally-distributed random variable:

- about 68% of the distribution is within 1 σ of μ
- about 95% of the distribution is within 2 σ's of μ
- about 99.7% of the distribution is within 3 σ's of μ