Who takes Mimy, the cat? The decisional hardships of a brother and a sister!

I want to take Mimy!

I want to keep Mimy!

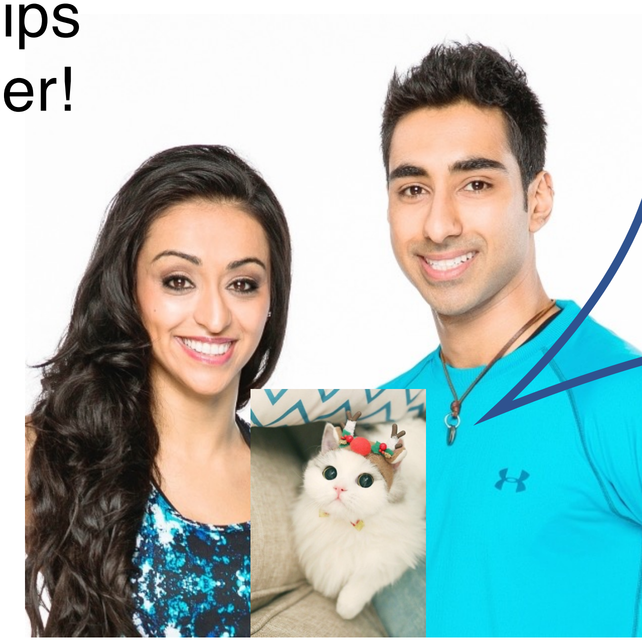
Sukhi & Jinder Atwal B.C. brother-sister team off to The Amazing Race. And they are competitive!

They share Mimy, the family cat!

But Jinder is now leaving their home town and wants to take the cat with him.

Disclaimer: the story and characters here are made up for pedagogical entertainment ☺

Who takes Mimy, the cat? The decisional hardships of a brother and a sister!



Let’ toss this coin 30 times to decide who takes Mimy. Here is a coin. The one that gets more than 15 tails takes Mimy.

Sukhi & Jinder Atwal B.C. brother-sister team off to The Amazing Race.

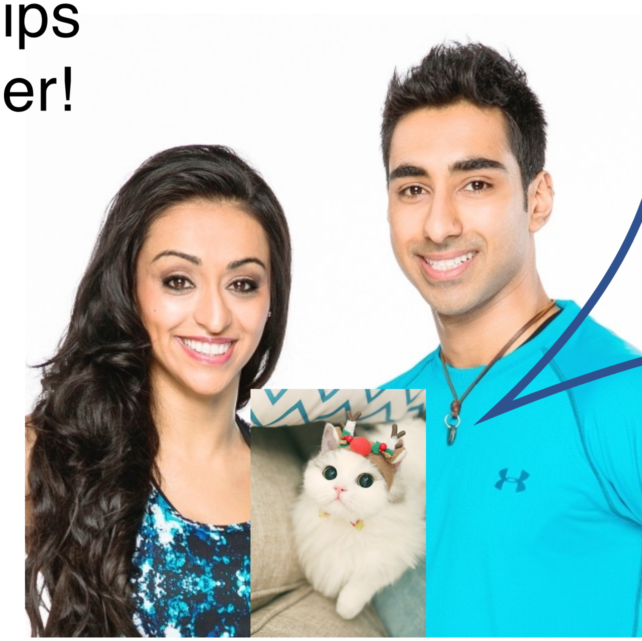
They share Mimy, the family cat!

But Jinder is now leaving their home town and wants to take the cat with him.

Disclaimer: the story and characters here are made up for pedagogical entertainment ☺

Who takes Mimy, the cat? The decisional hardships of a brother and a sister!

Let’ toss this coin 30 times to decide who takes Mimy. Here is a coin. The one that gets more than 15 tails takes Mimy.



Sukhi & Jinder Atwal B.C. brother-sister team off to The Amazing Race.

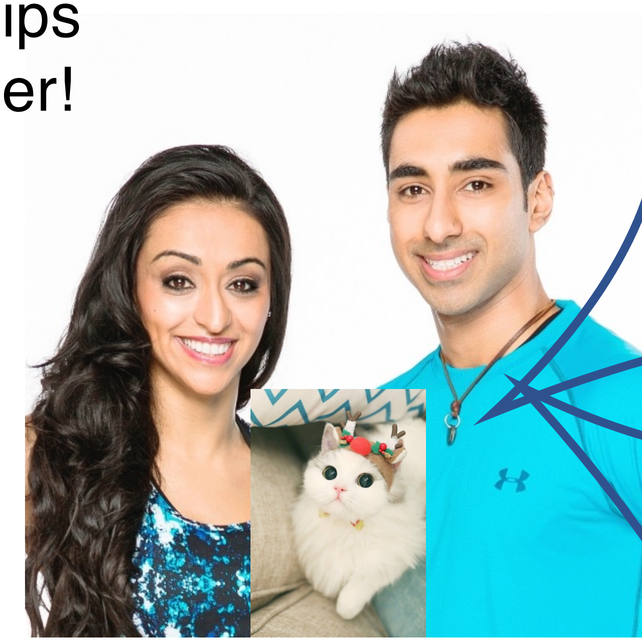
They share Mimy, the family cat!

But Jinder is now leaving their home town and wants to take the cat with him.

Your coin? You always cheated in games when we were young! Let’s use my coin

Disclaimer: the story and characters here are made up for pedagogical entertainment ☺

Who takes Mimy, the cat? The decisional hardships of a brother and a sister!



Let' toss this coin 30 times to decide who takes Mimy. Here is a coin. The one that gets more than 15 tails takes Mimy.

Your coin? You always cheated in games when we were young! Let's use my coin

But should I trust you?

Sukhi & Jinder Atwal B.C. brother-sister team off to The Amazing Race.

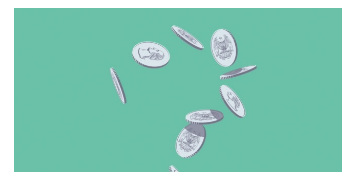
They share Mimy, the family cat!

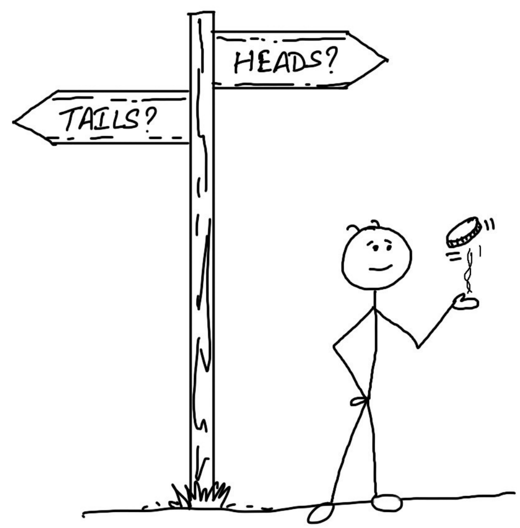
But Jinder is now leaving their home town and wants to take the cat with him.

Disclaimer: the story and characters here are made up for pedagogical entertainment ☺

Sukhi, who took BIOL322, proposes that they use statistics to judge each other coins.

All right, I propose a statistical experiment to test our coins.





Disclaimer: the story and characters here are made up for pedagogical entertainment ☺

Sukhi proposes that Jinder takes her coin & she takes his.

Each of us toss each other coins many sets of 30 times and graph the results!

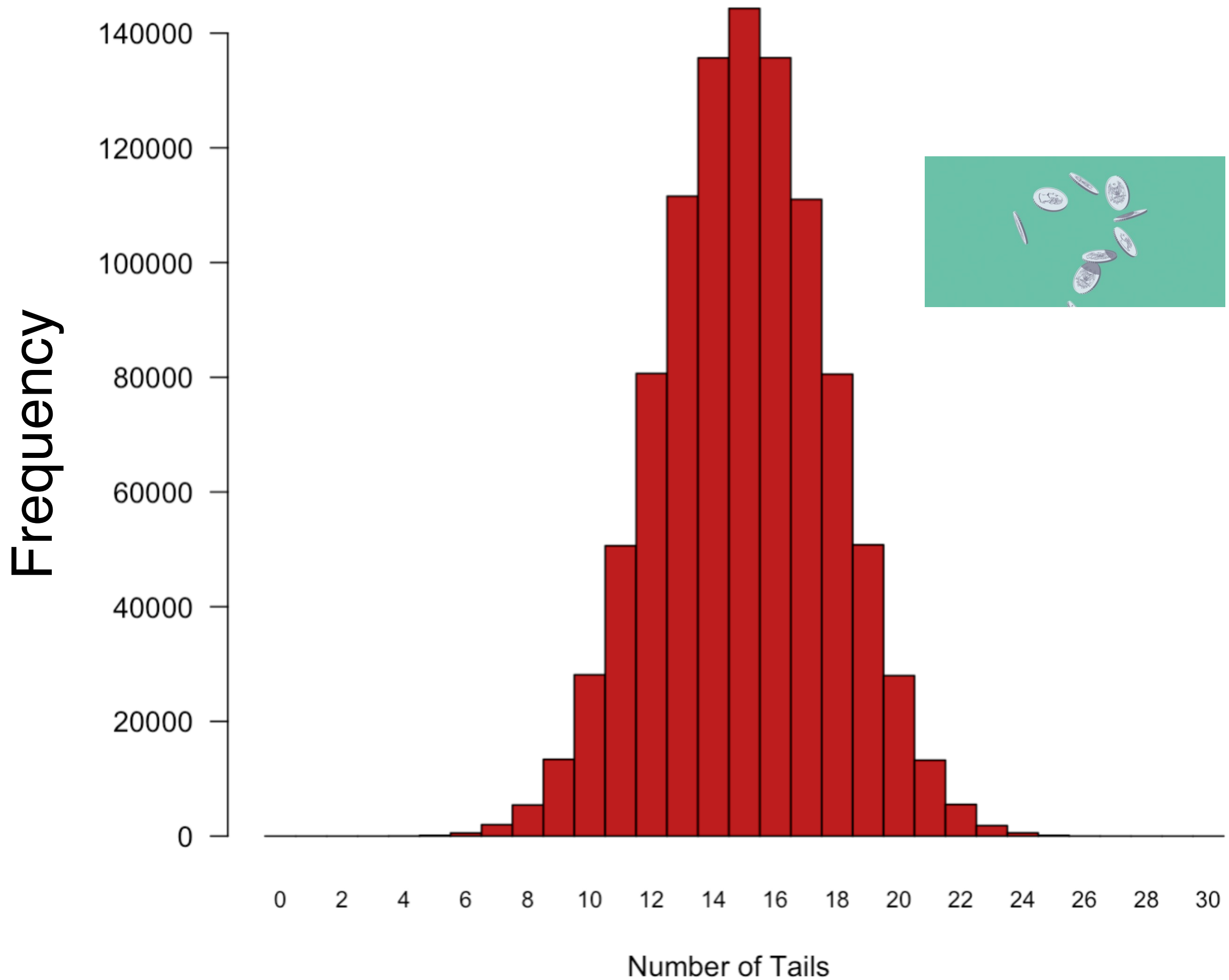
Sukhi proposes that Jinder takes her coin & she takes his.

You really don't trust me!

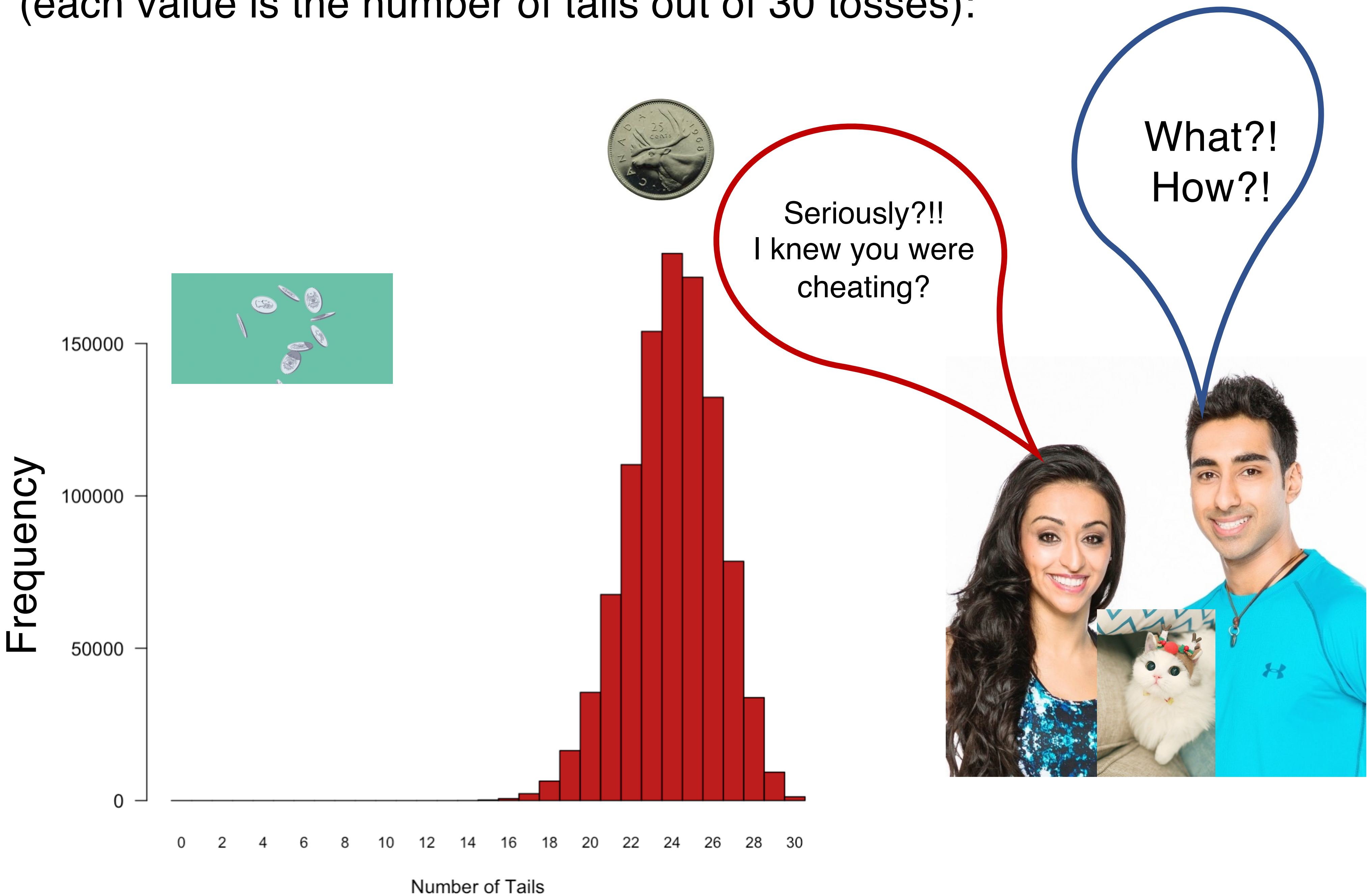
Each of us toss each other coins many sets of 30 times and graph the results!

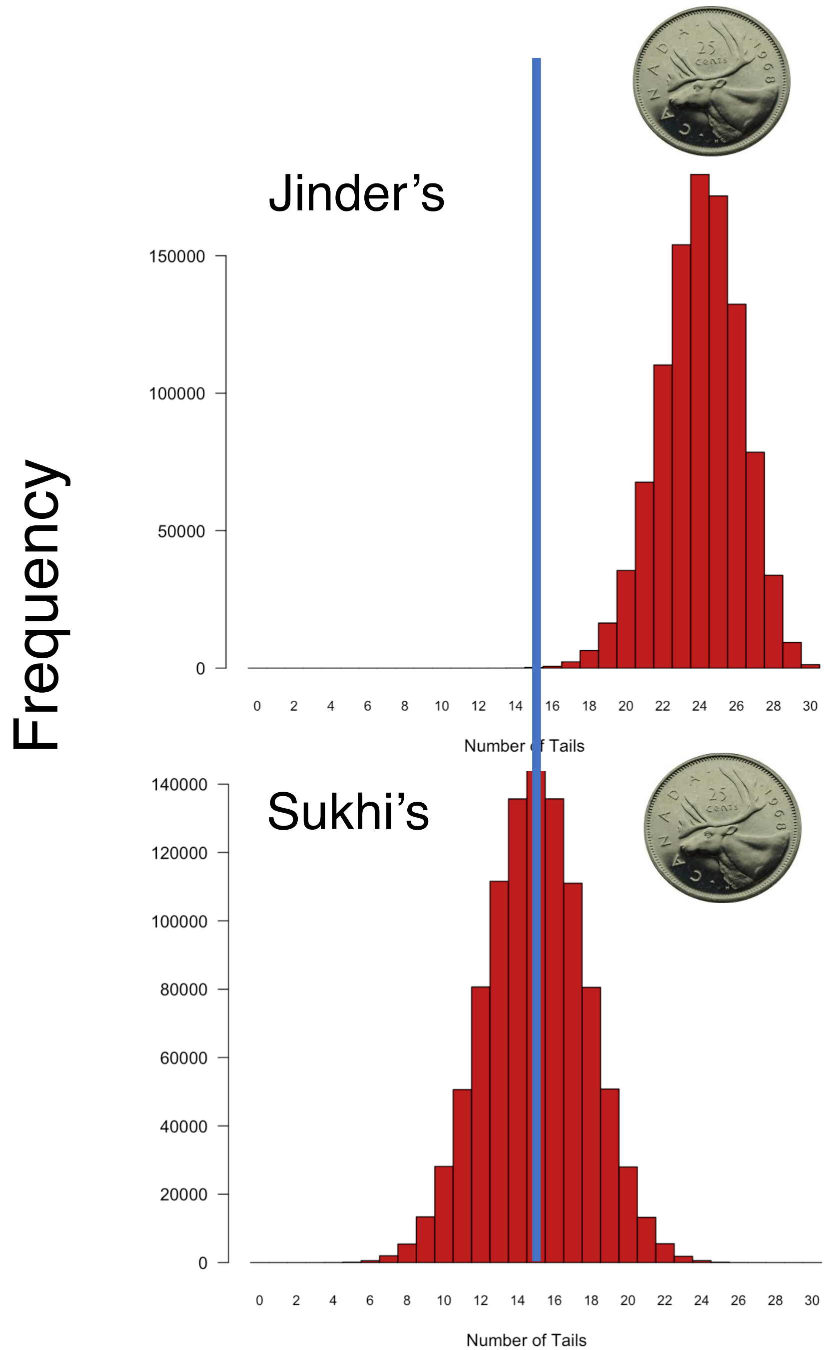
Sukhi proposes that Jinder takes her coin & she takes his.

This is the distribution of Sukhi's coin generated by Jinder (each value is the number of tails out of 30 tosses):

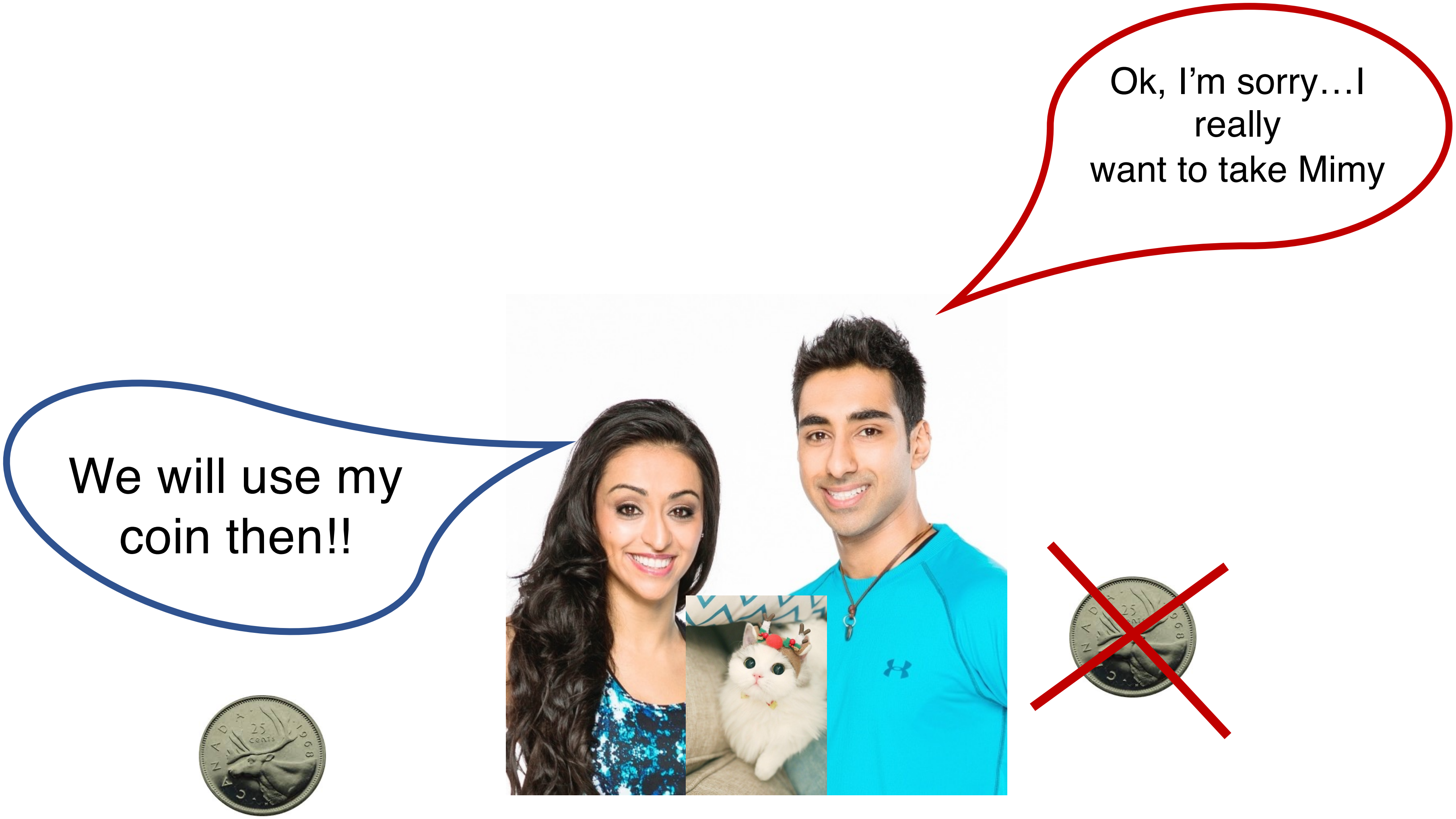


This is the distribution of Jinder’s coin generated by Sukhi (each value is the number of tails out of 30 tosses):

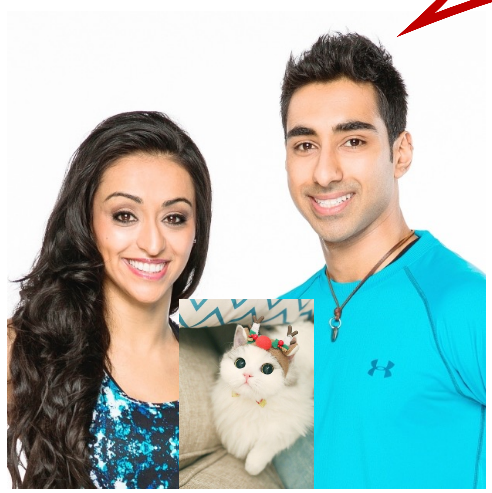


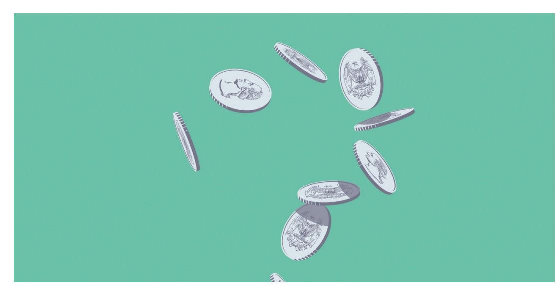

Frequency
Jinder's
150000
100000
50000
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Number of Tails
Sukhi's
140000
120000
100000
80000
60000
40000
20000
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Number of Tails


Seriously? I knew you were cheating?
What?! How?!


Ok, I'm sorry…I really want to take Mimy
We will use my coin then!!

Nice, I got 27 tails!!! Coin was yours…and Mimy is mine!

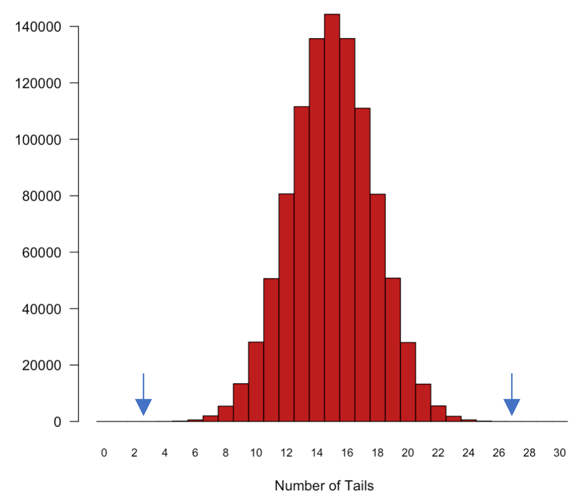


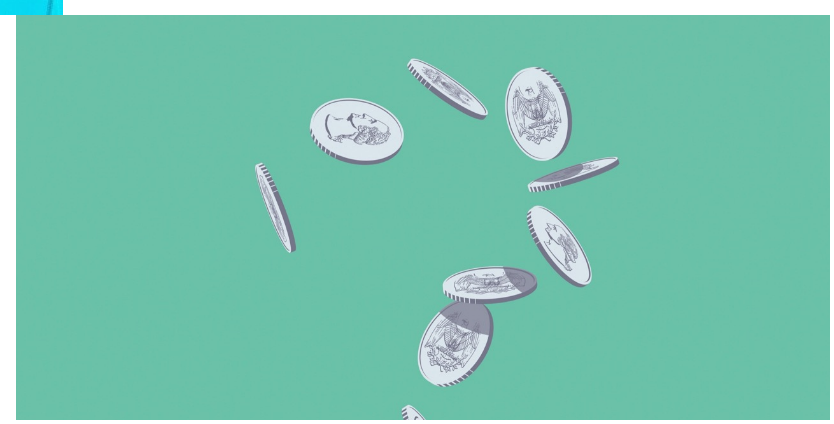


Hummm……
The probability of getting 27 values or more to be tails or heads is P=0.00000003 according to my fair coin!

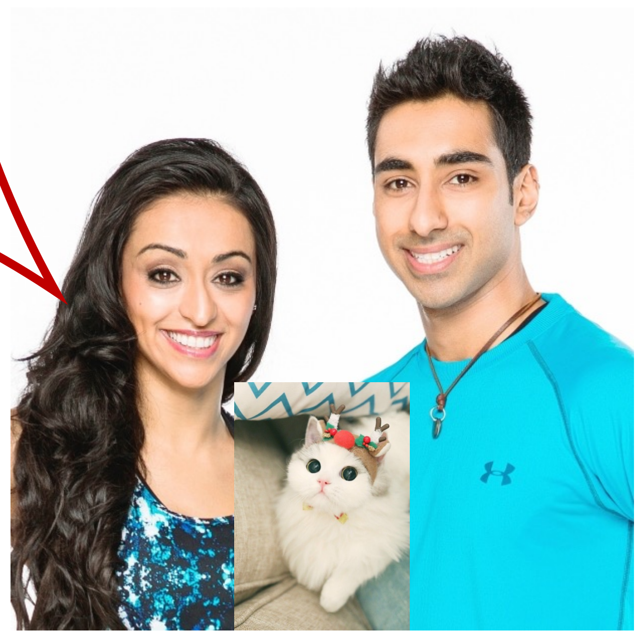
Nice, I got 27 tails!!! Coin was yours…and Mimy is mine!

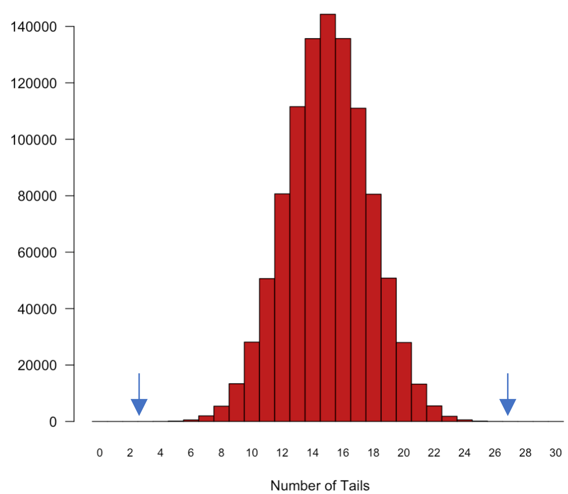
I'm just that lucky! I guess!

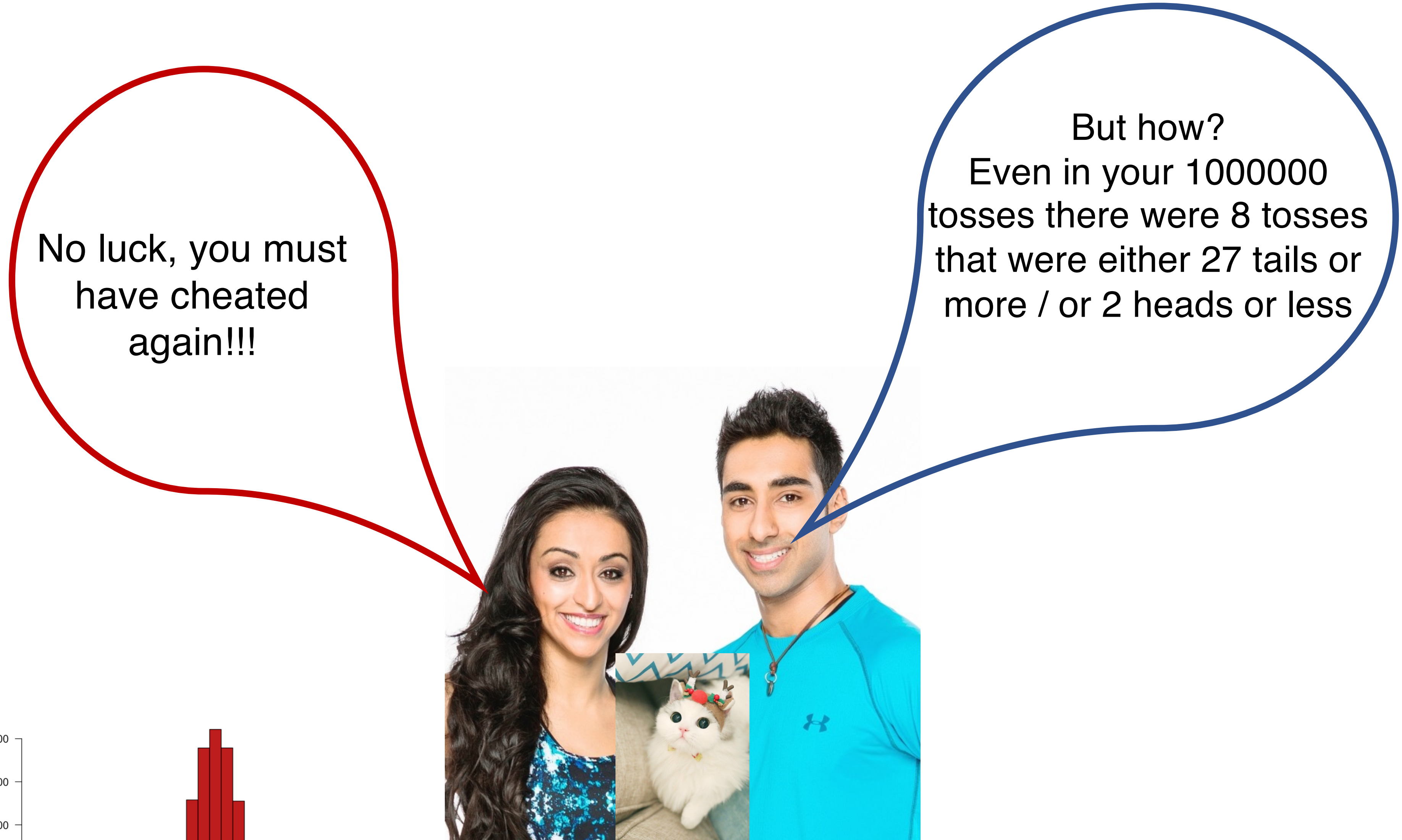


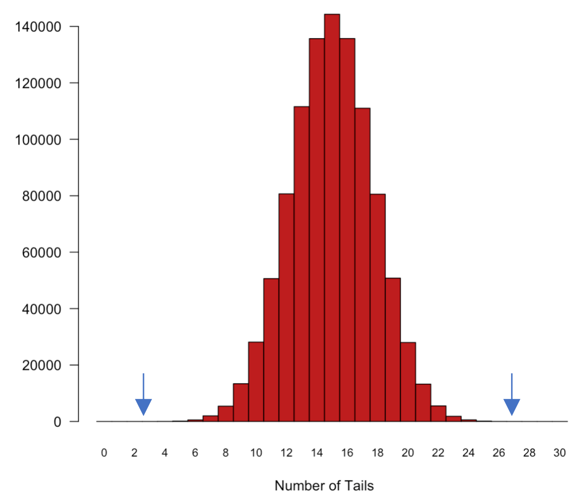


No luck, you must have cheated again!!!

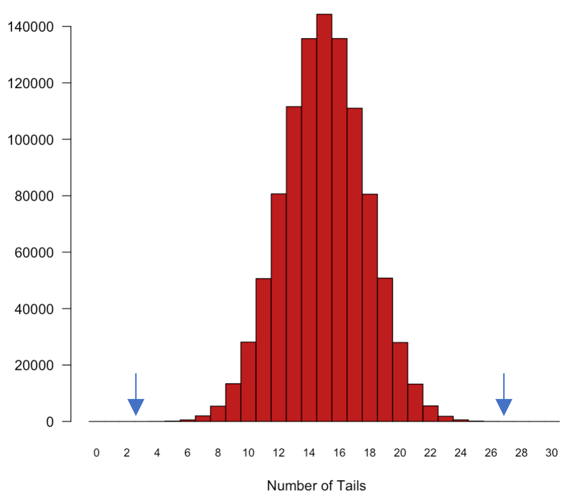





No luck, you must have cheated again!!!
But how? Even in your 1000000 tosses there were 8 tosses that were either 27 tails or more / or 2 heads or less

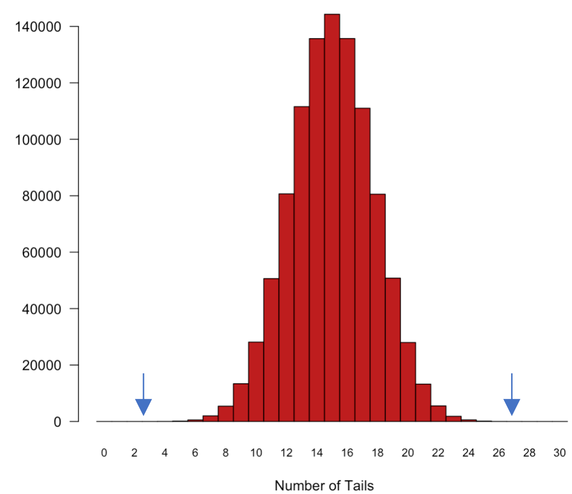

140000
120000
100000
80000
60000
40000
20000
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Number of Tails

Yes, but it’s almost impossible to get the result you got if you didn’t cheat!



Yes, but it’s almost impossible to get the result you got if you didn’t cheat!

Sorry…you are right… I did it again…you didn’t see it, but I exchanged our coins without you seeing it!



Take Mimy…if you cheated like this, I take that you really love and want Mimy!

Thanks sister!!!!!

These guys are nuts! I'm not going with either of them

Criminal trial analogy to help you understand further statistical decisions regarding statistical hypothesis testing: pedagogical support for long-lasting learning

In the US, a criminal trial makes two contradictory (mutually exclusive) claims about the defendant, either “not guilty” or “guilty”. Despite these claims, the following happens:

1. The defendant is truly either “innocent” or “guilty.” [but is assumed we never know the truth in most cases].

2. The defendant is then presumed “innocent until proven guilty.”

3. The defendant is found guilty only if there is *strong evidence* that the defendant is guilty. The phrase “beyond a reasonable doubt” is often used as a guideline for determining a cut off for when enough evidence exists to find the defendant guilty.

4. The defendant is found to be either “not guilty” or “guilty” in the ultimate verdict.

Adapted from: https://moderndive.com/9-hypothesis-testing.html#ht-interpretation

The lower the p-value, the more surprising the evidence is, the more ridiculous the null hypothesis looks possible.

[that Jinder was not guilty].

And what did Sukhy do when she felt ridiculous about her null hypothesis?

She rejected that and choose the alternative hypothesis instead [Jinder was guilty].

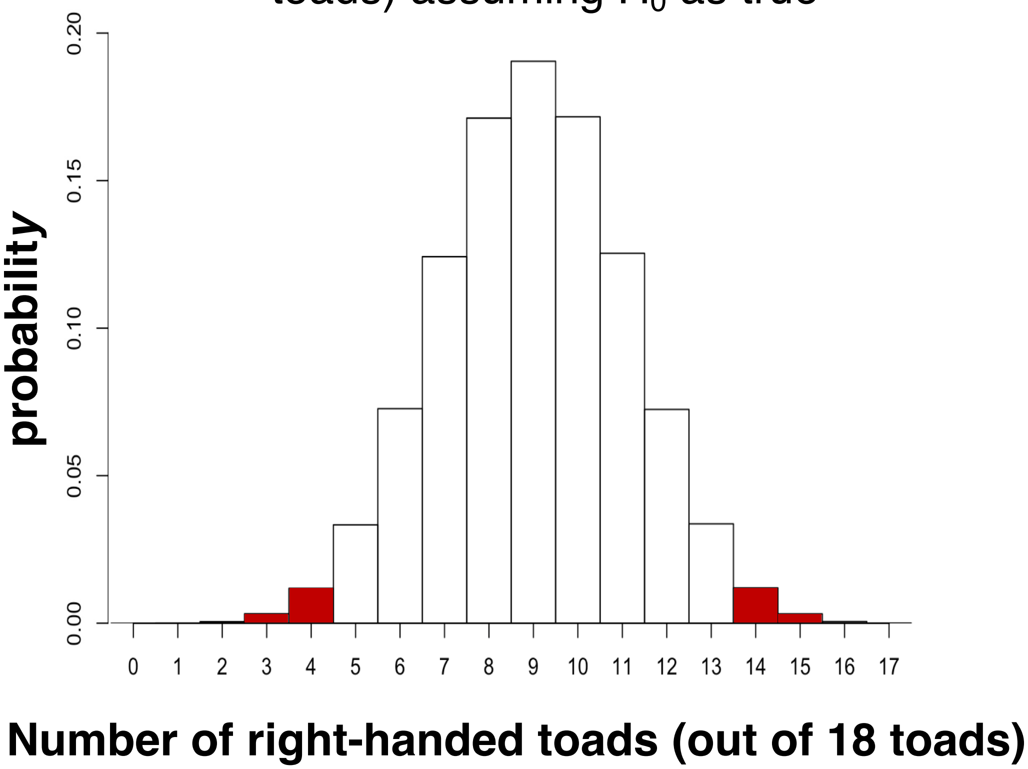
Adapted from:https://towardsdatascience.com/p-values-explained-by-data-scientist-f40a746cfc8

Let’s take a break - 2 minutes

Back to toads

Remember that we assume the null hypothesis to be true to build the sampling distribution under the null hypothesis (H_0). As such, all values in the sampling distribution of the theoretical populations (including the one of your sample) are also possible. So when using an alpha to reject H_0, we can make a mistake because those values below alpha are possible even when H_0 is true.

Sampling distribution of the test statistic (here number of right-handed toads) assuming H_0 as true

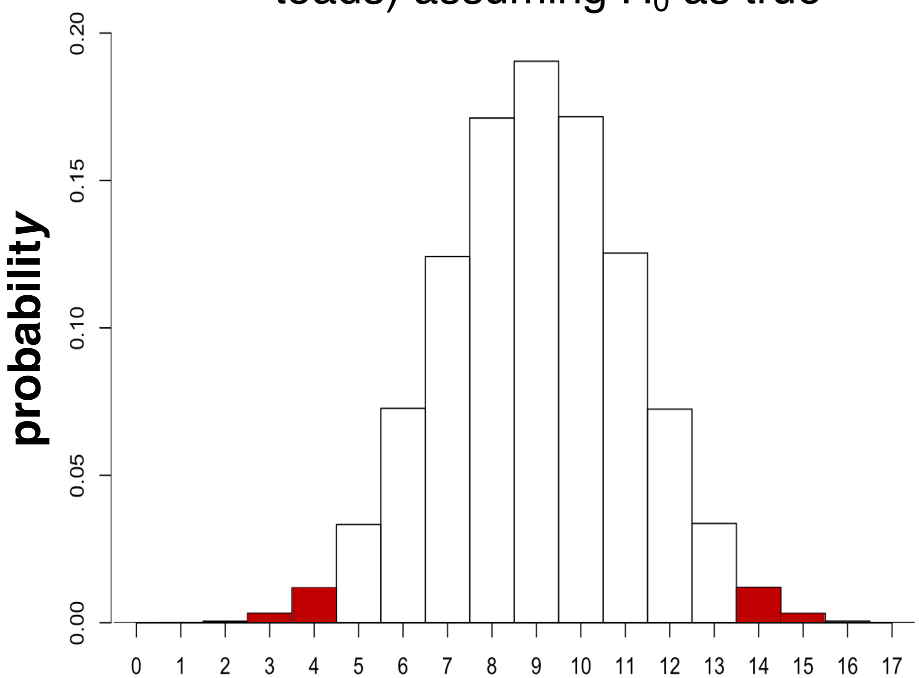


Pr[14 or more right-handed toads] =
Pr[14] + P[15] + P[16] + P[17] + P[18] =
0.0155 x 2 (symmetric distribution) = 0.031

Back to toads

Remember that we assume the null hypothesis to be true to build the sampling distribution under the null hypothesis (H_0). As such, all values in the sampling distribution of the theoretical populations (including the one of your sample) are also possible. So when using an alpha to reject H_0, we can make a mistake because those values below alpha are possible even when H_0 is true.

Sampling distribution of the test statistic (here number of right-handed toads) assuming H_0 as true



Number of right-handed toads (out of 18 toads)

Any value for the test statistic in the null distribution that is equal or more extreme (bigger or smaller) amounts to evidence against the null hypothesis.

The P-value is a way to judge whether we have evidence against or for the statistical null hypothesis.

As such any value of the test statistic that is equal or more extreme than the observed value amounts to that probability (i.e., as improbable as my own observed value).

Pr[14 or more right-handed toads] =
Pr[14] + P[15] + P[16] + P[17] + P[18] =
0.0155 x 2 (symmetric distribution) = 0.031

Statistical errors - two types of errors in statistical testing (using the criminal trial analogy)

	Reality (unknown)	
Conclusion based on sample (evidence)	H_0 true (innocent)	H_0 false (guilty)
Reject H_0 ("guilty")	Type I error	Correct
Do not reject H_0 ("not guilty")	Correct	Type II error

Type I error = FALSE POSITIVE
(its probability represent the probability of rejecting H_0 when is false, i.e., alpha).

Type II error = FALSE NEGATIVE
(its probability represents the probability of rejecting H_A when is true)

Remember that we assume the null hypothesis to be true to build the sampling distribution for a statistical hypothesis testing.

As such, all values in the sampling distribution from the theoretical populations (including the one in your sample) are possible under sampling.

If your sample value differs a lot from the null distribution (sampling distribution), then you have grounds to state that your sample value in improbable under the null hypothesis; and you reject it.

But you could be wrong! Either rejecting OR not rejecting it.

Remember that we assume the null hypothesis to be true to build the sampling distribution for a statistical hypothesis testing.

As such, all values in the sampling distribution from the theoretical populations (including the one in your sample) are possible under sampling.

If your sample value differs a lot from the null distribution (sampling distribution), then you have grounds to state that your sample value in improbable under the null hypothesis; and you reject it.

But you could be wrong! Either rejecting OR not rejecting it.

The protection we have against rejecting when is true (type I error) is the alpha level (significance level).

The protection we have against not rejecting when is false (type II error) is often based on increasing sample size (greater sample sizes lead to smaller p-values, hence smaller type II errors).

Critical definitions

Type I error is rejecting a true null hypothesis (i.e., reject the null hypothesis when you should not have). Its probability is the significance level α set by us. This probability does not change with sample size n.

Type II error is failing to reject a false null hypothesis (i.e., do not reject the null hypothesis when you should not have). Its probability is β and is more complex to estimate (advanced stats). This probability decreases as sample size increases.

The power of a test ($1-\beta$**)** is the probability of rejecting the null hypothesis when is truly false. This probability increases as sample size increases.

Hypothesis testing involving a continuous variable; the toad problem involved a categorial variable

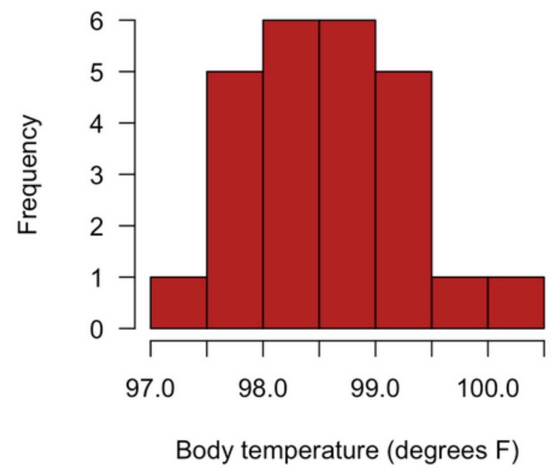
Normal human body temperature, as kids are taught in North America, is 98.6^{o}F (37^{o}C). But how well is this supported by data?

Let’s understand this problem under a statistical hypothesis testing framework

Normal human body temperature, as kids are taught in North America, is 98.6^{o}F. But how well is this supported by data? Researchers obtained body-temperature measurements on randomly chosen healthy people (Schoemaker 1996). The data for the 25 people are as follows:

98.4	98.6	97.8	98.8	97.9
99.0	98.2	98.8	98.8	99.0
98.0	99.2	99.5	99.4	98.4
99.1	98.4	97.6	97.4	97.5
97.5	98.8	98.6	100.0	98.4

The data look relatively symmetric so for now we have a good indication that these data are normally distributed. We'll see later in the course how to test this assumption in a more rigorous way.



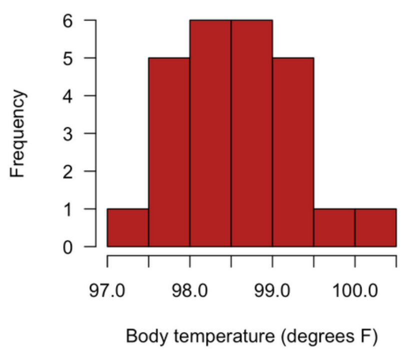
Normal human body temperature, as kids are taught in North America, is 98.6^{o}F. But how well is this supported by data?

Let's understand this problem under a statistical hypothesis testing framework

H_0 (null hypothesis): the mean human body temperature is 98.6^{o}F.

H_A (alternative hypothesis): the true population is different from 98.6^{o}F.

The Probability (or P-value), or estimated probability, is the probability of finding the observed, or more *extreme*, assuming that the null hypothesis (H_0) of a study question (μ = 98.6^{o}F) is true.

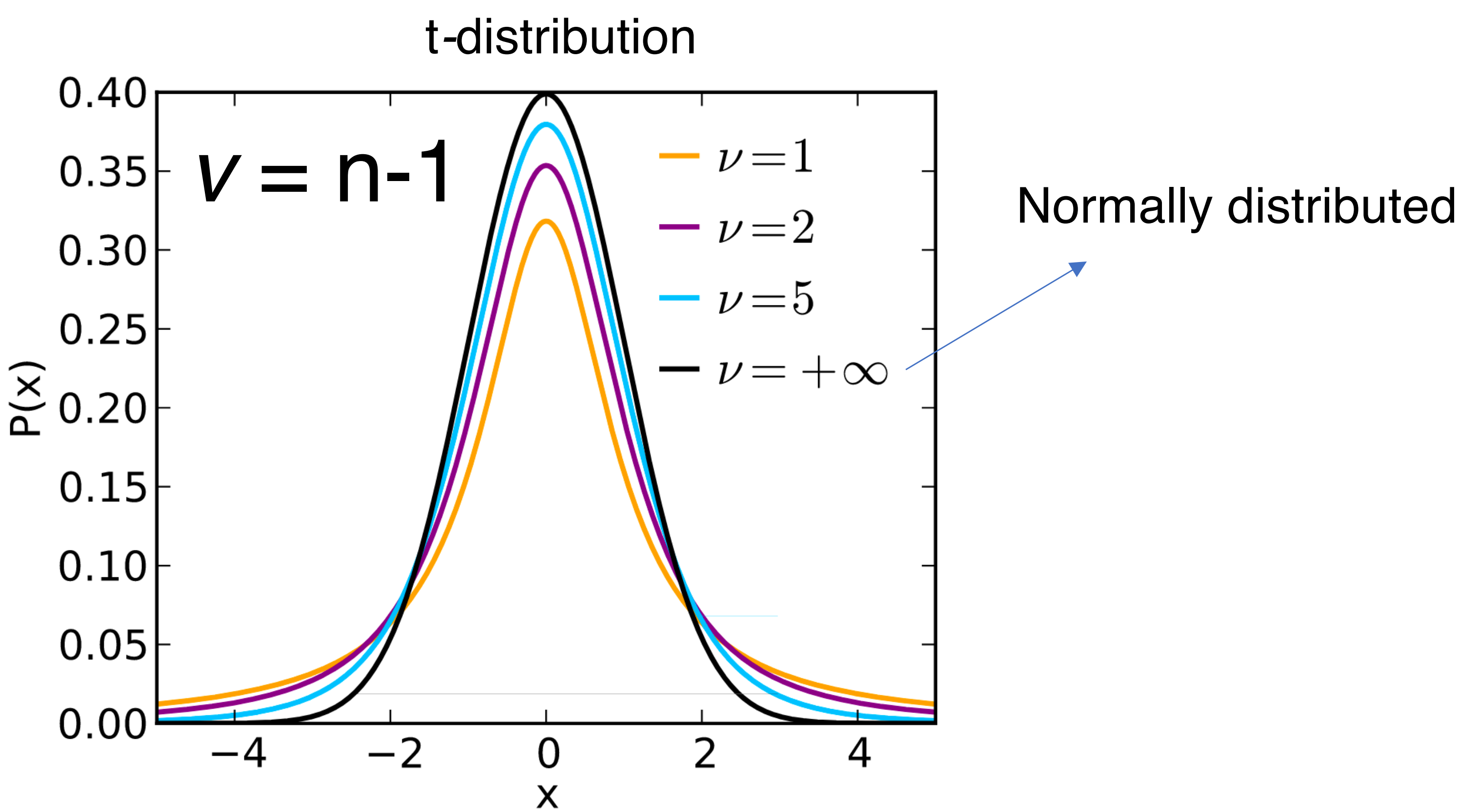


$$\bar{Y} = 98.524$$

H_0 can be stated as "any observed difference between the sample mean (98.524^{o}F) and the theoretical population value (98.6^{o}F) is due to chance alone.

To tackle the human temperature problem, we will use the t distribution (quick review here)

For small sample sizes (n < 60), the sampling distribution of sample means from a normally distributed population are a bit far from normally distributed – the distribution is wider and depends on the sample size (degrees of freedom) – it is called the t-distribution.



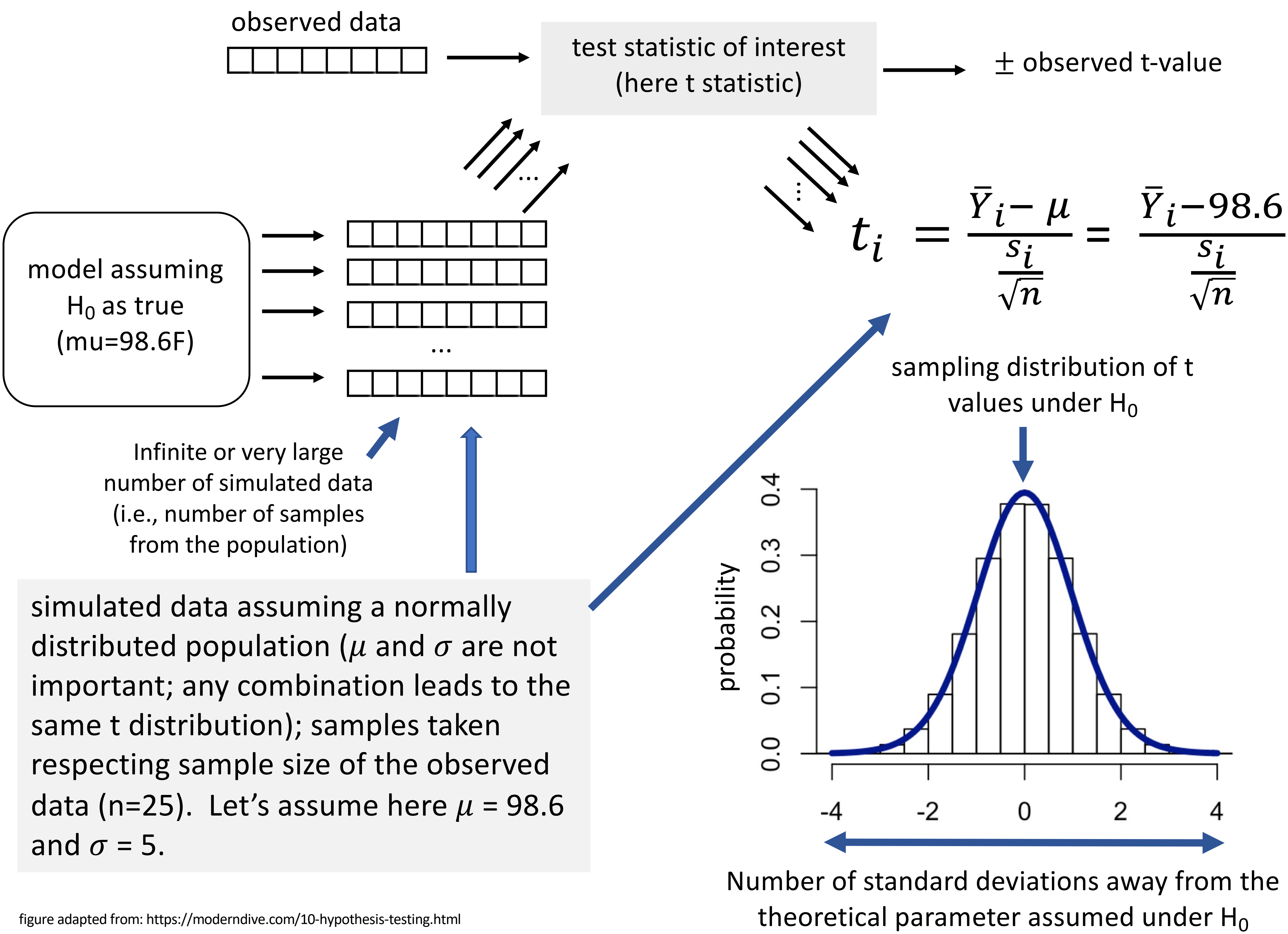
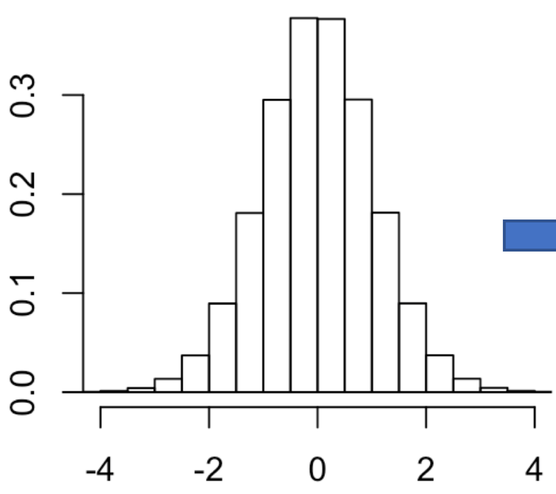
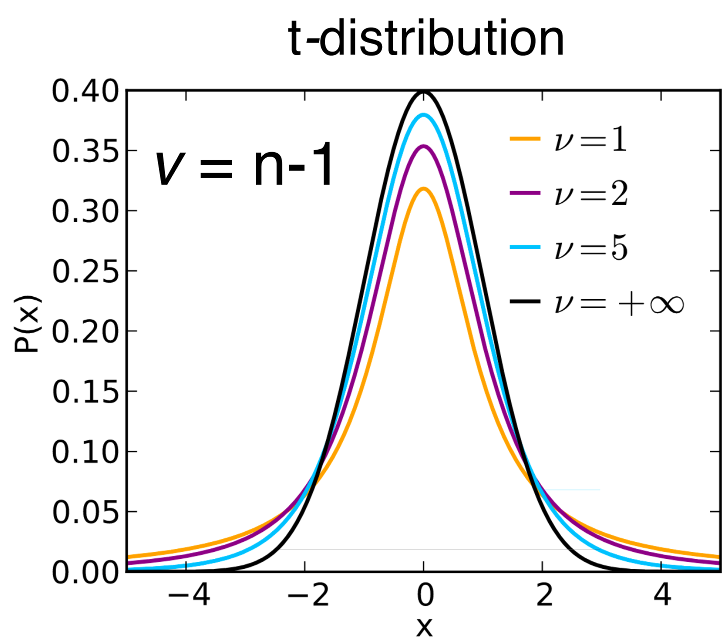


figure adapted from: https://moderndive.com/10-hypothesis-testing.html simulated data assuming a normally distributed population (μ and σ are not important; any combination leads to the same t distribution); samples taken respecting sample size of the observed data (n=25).

```
samples.pop.1 <- replicate(1000000,rnorm(n=25,mean=98.6,sd=5))
sampleMeans.Pop1 <- apply(samples.pop.1,MARGIN=2,FUN=mean)
sampleSDs.Pop1 <- apply(samples.pop.1,MARGIN=2,FUN=sd)
standardized.SampDist.Pop1 <- (sampleMeans.Pop1-98.6)/(sampleSDs.Pop1/sqrt(25))
hist(standardized.SampDist.Pop1)
```



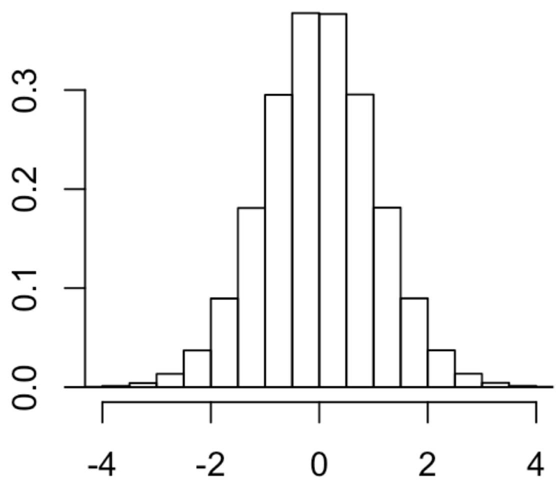


Two populations, two different means, two different standard deviations, but samples have the same size n

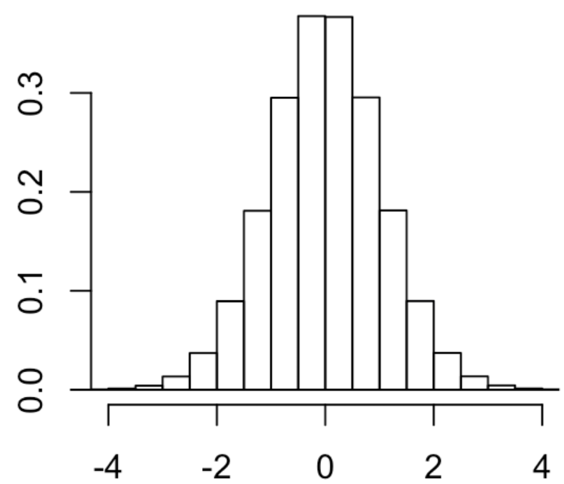
```
samples.pop.1 <- replicate(1000000,rnorm(n=25,mean=98.6,sd=5))
sampleMeans.Pop1 <- apply(samples.pop.1,MARGIN=2,FUN=mean)
sampleSDs.Pop1 <- apply(samples.pop.1,MARGIN=2,FUN=sd)
standardized.SampDist.Pop1 <- (sampleMeans.Pop1-98.6)/(sampleSDs.Pop1/sqrt(25))
hist(standardized.SampDist.Pop1)
```

```
samples.pop.2 <- replicate(1000000,rnorm(n=25,mean=13,sd=15))
sampleMeans.Pop2 <- apply(samples.pop.2,MARGIN=2,FUN=mean)
sampleSDs.Pop2 <- apply(samples.pop.2,MARGIN=2,FUN=sd)
standardized.SampDist.Pop2 <- (sampleMeans.Pop2-13)/(sampleSDs.Pop2/sqrt(25))
hist(standardized.SampDist.Pop2)
```

Because of the standardization process, the standardized t-distribution has always means 0 and the standard deviation only depends on the sample size.



=



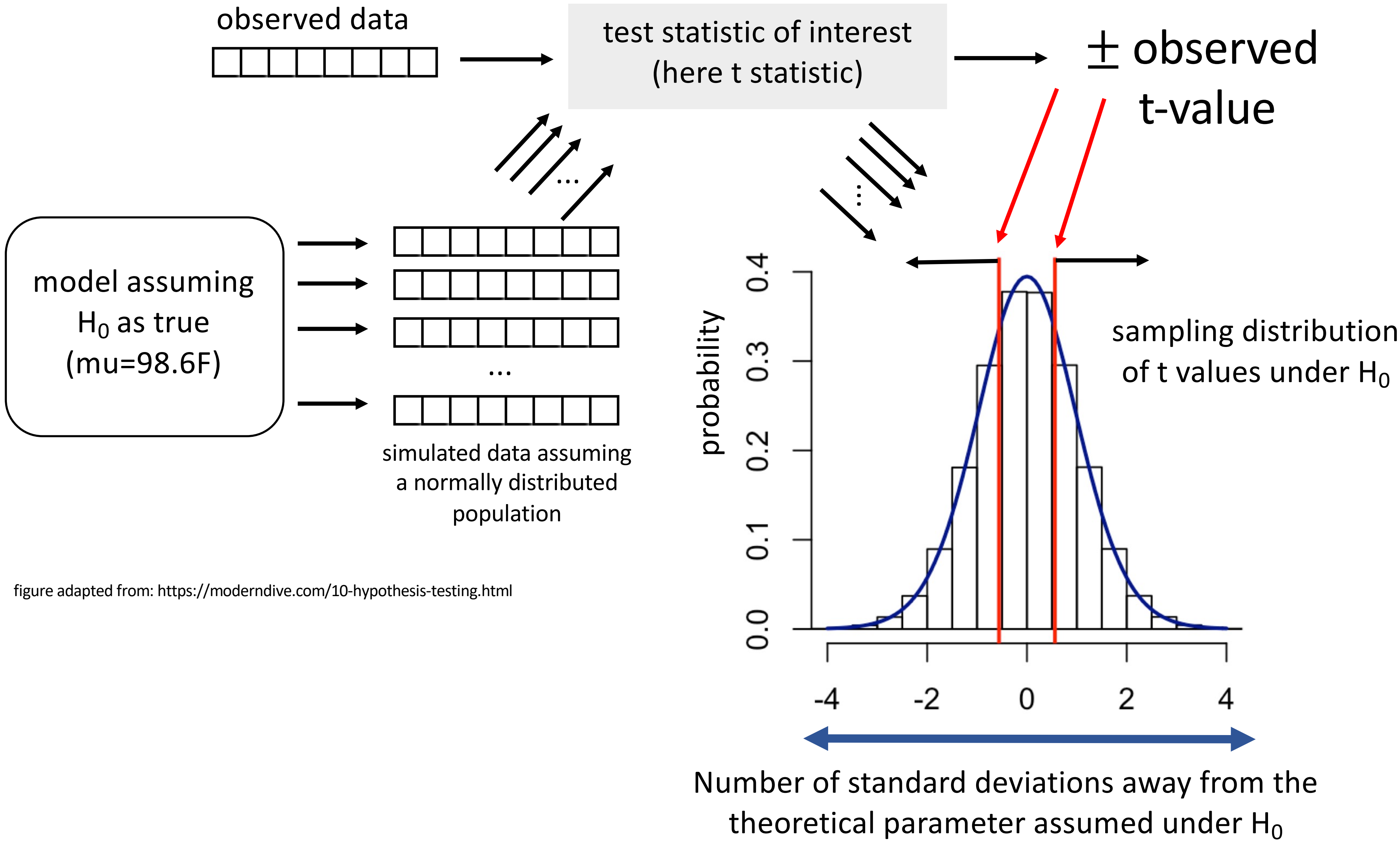


figure adapted from: https://moderndive.com/10-hypothesis-testing.html

Estimating p-value using the t-distribution

P-value is calculated as the number of values equal or greater than the observed and equal or smaller than -observed

The data look relatively symmetric so for now we have a good indication that these data are normally distributed. We'll see later in the course how to test this assumption in a more rigorous way.

Remember that we assumed a normally distributed population to generate (either via simulation or infinite sampling calculus) the t distribution. If the original population is not normal, then the standardized sampling distribution of means may not be normal! And the standard error may not be unbiased (as we saw previously).

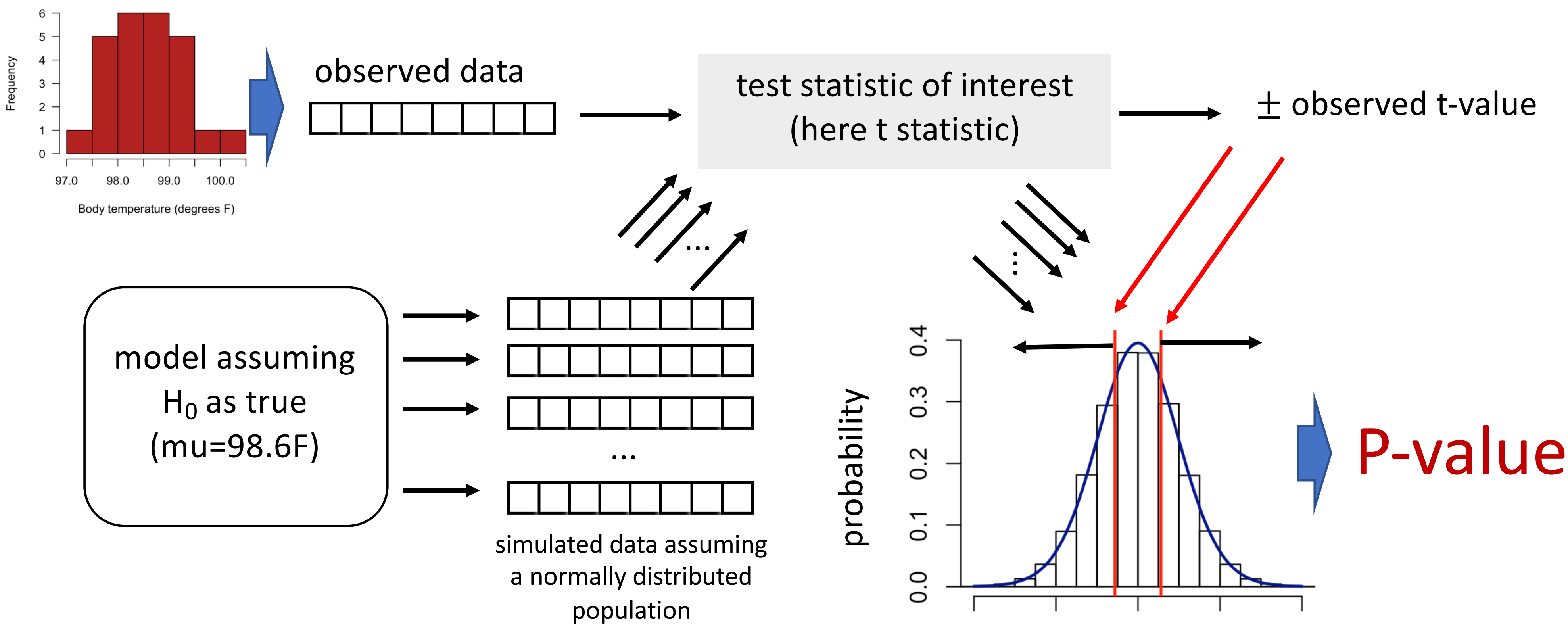


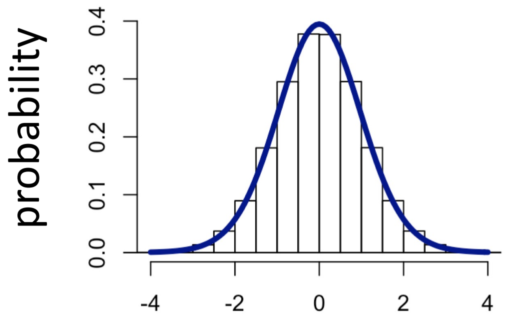
figure adapted from: https://moderndive.com/10-hypothesis-testing.html

The use of probability distribution functions for continuous variables

Unlike discrete values (e.g., handedness in toads), the t-distribution is mostly used for continuous variables (e.g., temperature) and, as such, it needs to be described as a probability density function (pdf). Instead of probabilities of a particular value, which is zero for a pdf, we calculate the probability that a particular value is within a range defined by any two values in the pdf.

Wikipedia has a great intuitive explanation about a pdf:

"Suppose a species of bacteria typically lives 4 to 6 hours. What is the probability that a bacterium lives *exactly* 5 hours? The answer is 0%. A lot of bacteria live for *approximately* 5 hours, but there is no chance that any given bacterium dies at *exactly* 5.0000000000... hours. Instead, we might ask: What is the probability that the bacterium dies between 5 hours and 5.01 hours? Let's say the answer is 0.02 (i.e., 2%)." The same applies for human temperatures and any other continuous variable (impossible to precise its value).



Let’s take a break - 2 minutes

One sample t-test

Normal human body temperature, as kids are taught in North America, is 98.6^{o}F. But how well is this supported by data? Researchers obtained body-temperature measurements on randomly chosen healthy people (Schoemaker 1996).
The data for the 25 people are as follows:

98.4	98.6	97.8	98.8	97.9
99.0	98.2	98.8	98.8	99.0
98.0	99.2	99.5	99.4	98.4
99.1	98.4	97.6	97.4	97.5
97.5	98.8	98.6	100.0	98.4

$\bar{Y} = 98.524$

$s = 0.678$

$\mathrm{SE}_{\bar{Y}} \frac{0.678}{\sqrt{25}} = 0.136$

Normal human body temperature, as kids are taught in North America, is 98.6ºF. But how well is this supported by data?

Let's "transform" this question into a probabilistic statement:

$$t_i = \frac{\bar{Y}_i - 98.6}{\frac{s_i}{\sqrt{n}}}$$

What is the probability of obtaining a sampling *mean as extreme or more extreme* than 98.524ºF given that the theoretical population mean (assumed under H_0) is 98.6ºF?

$$t = \frac{98.524 - 98.6}{0.136} = -0.56$$

The sample mean is -0.56 standard deviations away from the mean of the theoretical population (assumed under H_0)!

H_0 (null hypothesis): the mean human body temperature is 98.6^{o}F.

H_A (alternative hypothesis): the true population is different from 98.6^{o}F.

Should we reject or not reject the H_0?

$$t = \frac{98.524 - 98.6}{0.136} = -0.56$$

The sample mean is -0.56 standard deviations away from the mean of the theoretical population (assumed under H_0)!

The answer lies in the probability of finding a sample value smaller or equal to -0.56 in the sampling distribution of the theoretical population assumed under H_0 ($\mu = 0$)?

We started with: *Normal human body temperature, as kids are taught in North America, is 98.6ºF. But how well is this supported by data?*

Then "translated" the above question into: What is the probability of obtaining a sampling *mean as extreme or more extreme* (i.e., *smaller*) than 98.524ºF given that the population mean is 98.6ºF?

$$t = \frac{98.524 - 98.6}{0.136} = -0.56$$

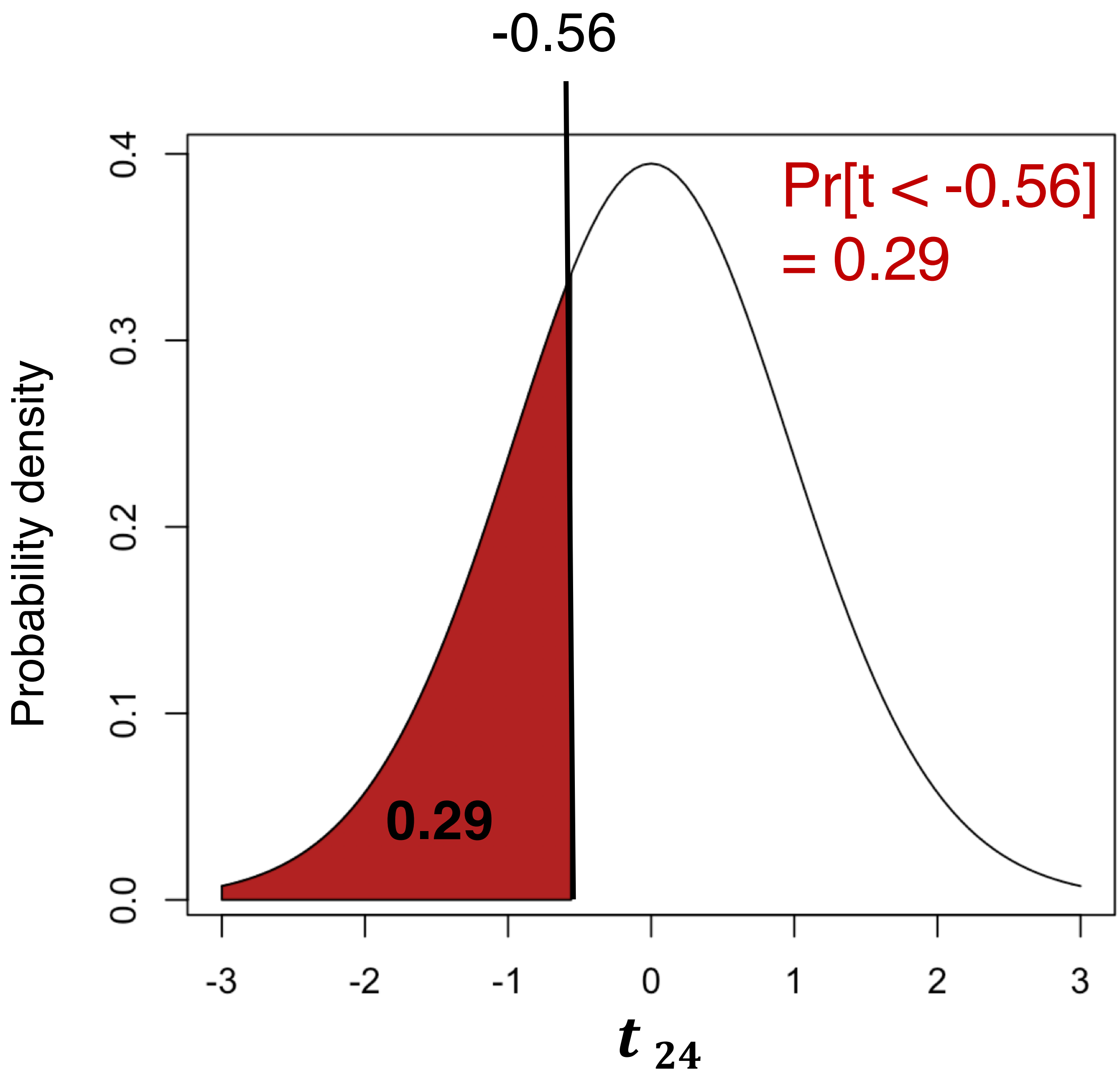
The sample mean is -0.56 standard deviations away from the mean of the theoretical population (assumed under H_0)!

In probabilistic terms, the question then becomes: What is the probability of finding a sample t value equal or smaller than -0.56 in the sampling distribution of the theoretical population (i.e., the t-distribution; where $\mu = 0$)?

```
> 1 - pt(q=abs(-0.56),df=24)
[1] 0.2903347
```

$Pr[t < -0.56] = 0.29$

What is the probability of finding a sample t value equal or smaller than -0.56 in the sampling distribution of the theoretical population (i.e., the t-distribution; where $\mu = 0$)?



We started with: Normal human body temperature, as kids are taught in North America, is 98.6ºF. But how well is this supported by data?

Then "translated" the above question into: What is the probability of obtaining a sampling mean as extreme or more extreme (i.e., smaller) than 98.524ºF given that the population mean is 98.6ºF?

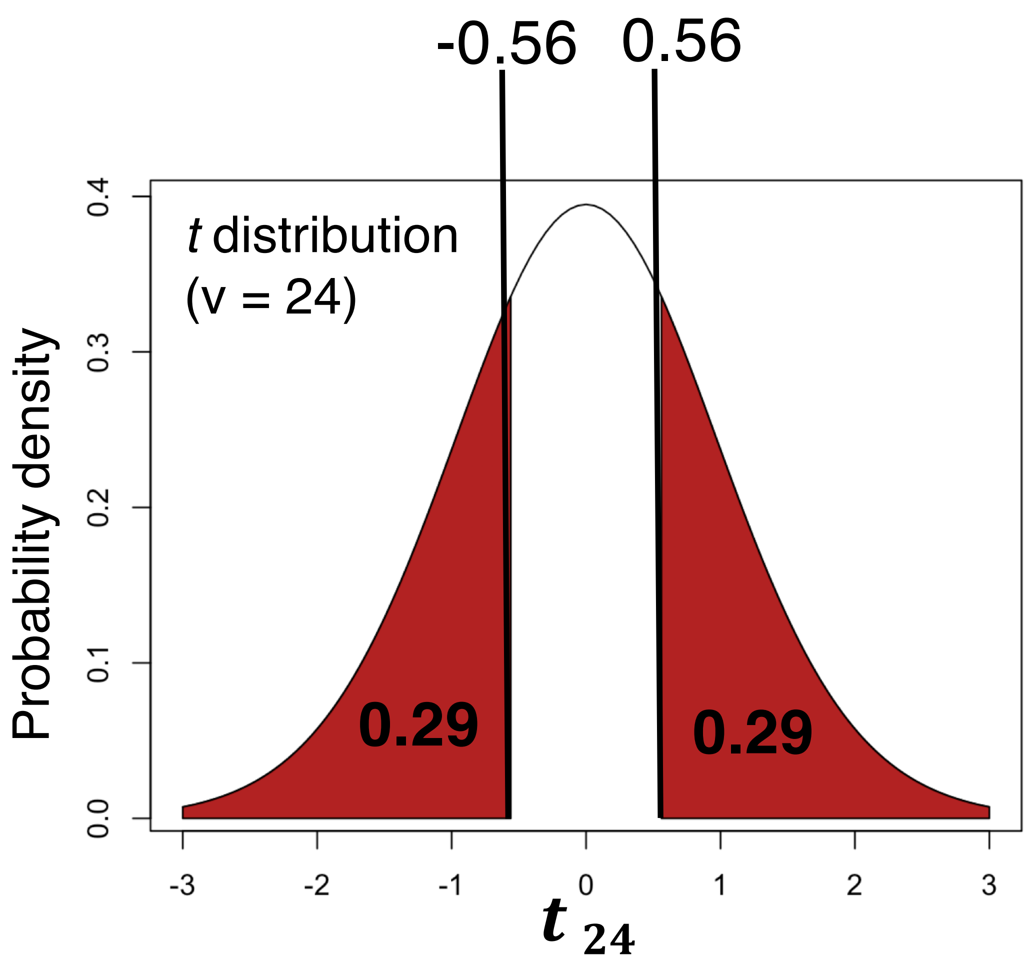
Need to consider here:

1) In principle, we are not interested in knowing if the sample mean we obtained would be smaller or greater than the theoretical population mean assumed under the null hypothesis (98.6ºF).

We are interested here in stating whether we have evidence to say that the sample mean we obtained is *consistent* (i.e., a common sample mean among the potential samples from the theoretical population, i.e., 98.6ºF) or *inconsistent* with the theoretical population assumed true under H_0 (inconsistent = an uncommon sample mean among the potential samples from the theoretical population).

SO, WE NEED TO CONSIDER BOTH SIDES OF THE t DISTRIBUTION.

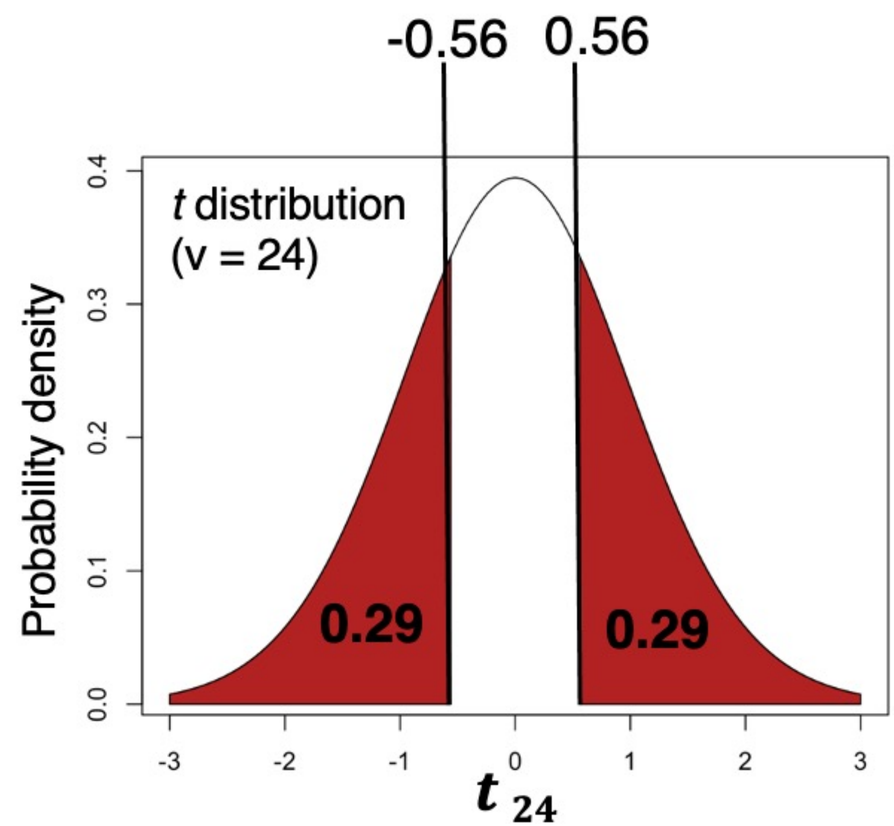
What is the probability of finding a sample t value equal or smaller than -0.56 AND finding a sample t value equal or greater than +0.56 in the sampling distribution of the theoretical population (i.e., the t-distribution; where $\mu = 0$)? P-value=0.58



Pr[t < -0.56] + Pr[t > 0.56] =
2 Pr[t > abs(0.56)] = **0.58**
(*t* is symmetric around μ)

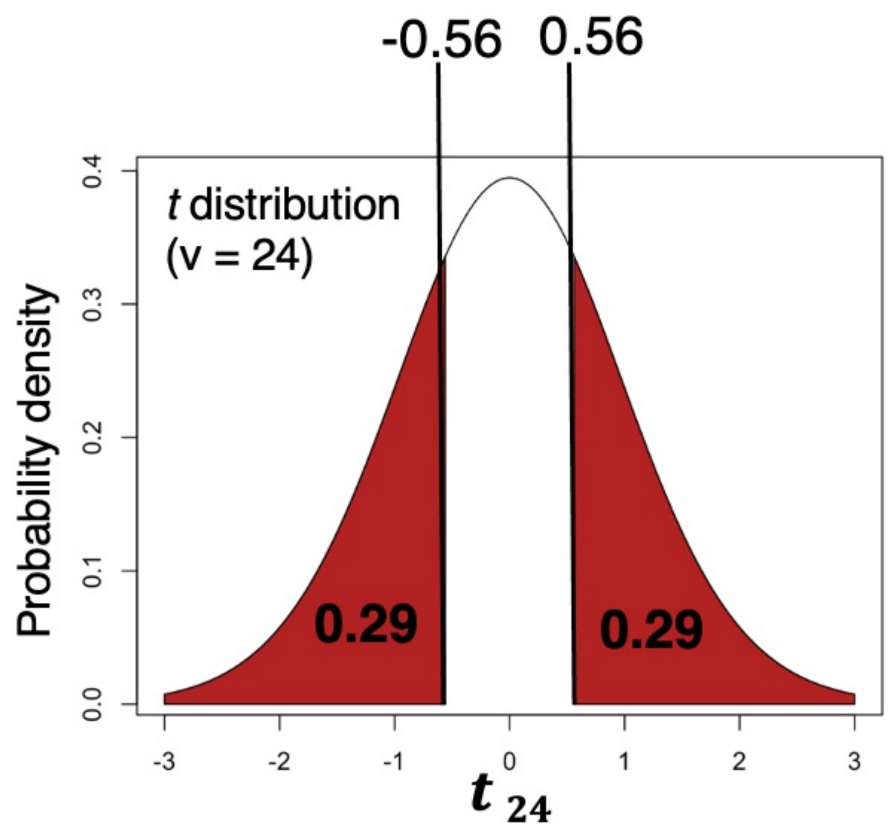
This probability 0.58 is quite large for an alpha = 0.05 (significance level). As such, we lack evidence to suggest that the observed mean does not belong to a population that has a mean (μ) = 98.6°F.

In other words, we DO NOT reject the H0 (null hypothesis) that the mean human body temperature is 98.6°F.



As such, the p-value is the evidence against the null hypothesis. The p-value is relatively large (0.58), and, as such, the evidence against H_0 is weak.

By not rejecting H_0, we cannot state that the true population value is 98.6ºF; all we can state is that we have no evidence to state that it is not (i.e., to the contrary)!



The process for the one sample mean test

1) Establish the theoretical population mean value of interest under H_0. This is the parameter that is used to standardize the sample mean value to generate the t standardized value (called t score, t statistic or t deviate).

2) Take one sample from the population of interest and assume (or verify) that the sample is normally distributed.

3) Standardize the sample mean in relation to the population mean value of interest established in 1 using the t standardization, i.e., calculate the t score.

4) Determine the probability of finding the observed (sample-based) t score (step 3) in the t distribution that is extreme or more extreme (small and large) than the observed. Remember – the t distribution is the standardized sampling distribution of the population of interest (step 1).

5) Based on the probability calculate in step 4 and the established significance value (alpha), reject or do not reject the null hypothesis.

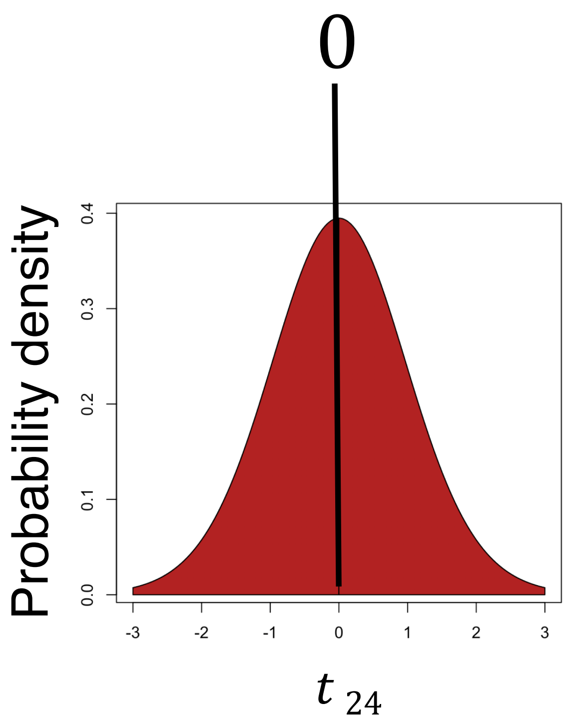
Let's think from another angle:
if the sample had a mean of 98.6ºF, then t = 0

$$t = \frac{98.6 - 98.6}{0.136} = 0$$

$Pr[t < 0] + Pr[t > 0] = 1$

What is the probability of obtaining a sampling mean as extreme or more extreme than 98.6ºF given that the population mean is 98.6ºF? 1.00 (100%).

Even here, by not rejecting H_0, we cannot state that the true population value is 98.6ºF; all we can say is that we have no evidence to state that it does not!



SUMMARY

We started with: Is the normal human body temperature of 98.6^{o}F, as kids are taught in North America, supported by data?

Then "translated" the above question into: What is the probability of obtaining a sampling mean as extreme or more than a sample mean of 98.524^{o}F given that the population mean is 98.6^{o}F?

1) In principle, we were not interested in knowing if the sample mean we obtained would be smaller or greater than the true population mean.

2) As such, all we are interested is to state whether we have evidence to say that the sample mean we obtained is ***consistent*** with H_0 or ***inconsistent*** with H_0.

3) If **consistent** (large P-value), then we can state that we have no evidence to state that the human temperature is different from 98.6^{o}F.

4) If **inconsistent** (small P-value), then we would have stated that we have evidence that the *Normal human body temperature is not 98.6^{o}F.*