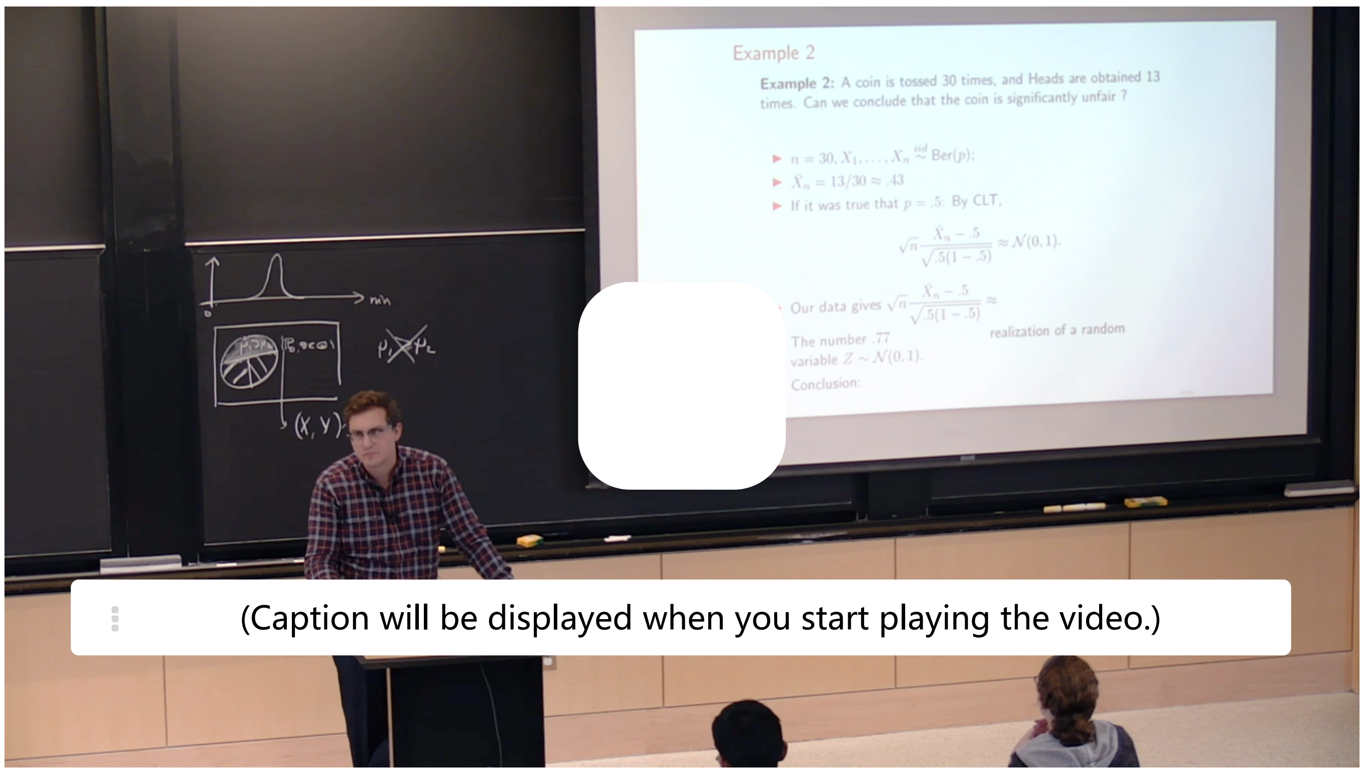
# 9. Example: Is this Coin Fair?

**Video Note:** In the video below, at 3:24, and on the last line of the annotated slide on the conclusion, there is an important misprint: The conclusion should be **It is not unlikely that the coin is fair .**

## Testing Fairness of a Coin




Start of transcript. Skip to the end.

So here is a much more compelling example

that if you work as a data scientist,

you would probably be solving every day, which

is tossing a coin, and deciding whether it's fair.

So this is probably the most important statistical problem,

because that's probably the one that

**视频**
下载视频文件

**字幕**
下载 SubRip (.srt) file
下载 Text (.txt) file

## CLT Concept Check

1/1 point (graded)

In the next few questions, we will flip a coin $200$ times in order to try and answer the hypothesis testing **question of interest**:

"**Is this coin fair?**."

As in lecture, we model the $i$'th flip as $X_i$ where $X_i = 1$ for a heads and $X_i = 0$ for a tails. Since the flips should not interact with each other and we always flip the same coin, we make the familiar modeling assumption $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathrm{Ber}(p)$ where $p$ is an unknown parameter. Then our original question of interest can be rephrased:

"**Does $\mathbf{p = 0.5}$ or does $\mathbf{p \neq 0.5}$ ?**".

Note that this is a very specific question. In particular, we do not care so much about the particular value of $p$– we just want to test whether or not it is equal to $0.5$.

To answer this question, we consider the statistic

$$\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5\,(1-0.5)}}$$

where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ denotes the sample mean.

Recall that we do not know the true value of $p$. Assume that $n$ is very large. Can we conclude that the distribution of $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ is very close to the distribution of a standard Gaussian $\mathcal{N}(0,1)$?

Choose the correct response that also has the correct explanation.

- ○ Yes, because $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ is a shift and rescaling of a binomial distribution. We know that for $n$ large enough, the binomial distribution $\mathrm{Bin}(n,p)$ provides a good approximation to the distribution of a standard Gaussian $\mathcal{N}(0,1)$.
- ○ Yes, because the central limit theorem (CLT) guarantees that for $n$ sufficiently large, $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \approx \mathcal{N}(0,1)$ (in distribution).
- ◉ No. Since we do not know for sure that $p = 0.5$, we cannot conclude that $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \to \mathcal{N}(0,1)$ in distribution. (e.g. If $p = 0.6$, then this estimator will not converge to $\mathcal{N}(0,1)$.) ✔
- ○ No. Even if $p = 0.5$, it is not true that $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \to \mathcal{N}(0,1)$ in distribution. Hence, even in the case of a fair coin, we do not expect this estimator be close in distribution to $\mathcal{N}(0,1)$.

**Solution:**

We examine the choices in order.

- "Yes, because $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ is a shift and rescaling of a binomial distribution. We know that for $n$ large enough, the binomial distribution $\mathrm{Bin}(n,p)$ provides a good approximation to the distribution of a standard Gaussian $\mathcal{N}(0,1)$." is incorrect. The explanation is wrong: $\mathrm{Bin}(n,p)$ does **not** provide a good approximation for the distribution $\mathcal{N}(0,1)$.

  **Remark:** However, by the CLT, if $X \sim \mathrm{Bin}(n,p)$, then for as $n \to \infty$,

  $$\sqrt{n}\left(\frac{\frac{X}{n} - p}{\sqrt{p(1-p)}}\right) \to \mathcal{N}(0,1)$$

  in distribution.

- "Yes, because the central limit theorem (CLT) guarantees that for $n$ sufficiently large, $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \approx \mathcal{N}(0,1)$ (in distribution)." is incorrect. We can only apply the CLT to the given estimator if the mean is $0.5$ and the variance is $0.5(1-0.5)$. This is only the case if the coin is fair, *i.e.*, $p = 0.5$.

- "No. Since we do not know for sure that $p = 0.5$, we cannot conclude that $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \to \mathcal{N}(0,1)$ in distribution. (e.g. If $p = 0.6$, then this estimator will not converge to $\mathcal{N}(0,1)$.)" is the correct response. We can only apply the CLT to conclude $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \to \mathcal{N}(0,1)$ in distribution if $p = 0.5$, as discussed in the previous bullet.

- "No. Even if $p = 0.5$, it is not true that $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \to \mathcal{N}(0,1)$ in distribution. Hence, even in the case of a fair coin, we do not expect this estimator be close in distribution to $\mathcal{N}(0,1)$." is incorrect. Though the answer is "No", the explanation is incorrect: the case where $p = 0.5$ is the **only** situation in which we can apply the CLT to say that $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \to \mathcal{N}(0,1)$ in distribution.

提交 你已经尝试了2次（总共可以尝试2次）

Answers are displayed within the problem

In the next two problems, we will illustrate some of the basic steps behind hypothesis testing.

The set up is the same as in the problem above:

Let $X_1, \ldots, X_{200} \stackrel{iid}{\sim} \mathrm{Ber}(p)$, and we are interested in determining from the sample whether or not $p = 0.5$. The hypothesis testing question of interest is then

$$\textbf{Does p} = \textbf{0.5} \qquad \textbf{or} \qquad \textbf{does p} \neq \textbf{0.5 ?}$$

To answer this question, we introduced the statistic, which is also an estimator:

$$\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5\,(1-0.5)}}.$$

The reason for considering this estimator is that, **if $\mathbf{p = 0.5}$, then the CLT applies** (check this!), so that for $n$ very large we may assume

$$\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5\,(1-0.5)}} \approx \mathcal{N}\,(0,1)\,.$$

In other words, **if $\mathbf{p = 0.5}$**, then the above estimator distributed approximately as a standard Gaussian when $n$ is large enough.

Our strategy will be to evaluate this estimator on the data set. Supposing that $p = 0.5$, then the value of our statistic should resemble the typical value of a single observation of a standard Gaussian random variable. Hence, if the value $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ lies deep in the tails of the standard normal distribution, we would logically conclude that it is **unlikely** that $p = 0.5$. Otherwise, we will not be able to refute that $p = 0.5$.

## Hypothesis Testing: A Sample Data Set of Coin Flips I

3/3 points (graded)

We use the statistical set-up from the previous problem. Consider a statistical experiment where you flip the coin $200$ times. In one run of this experiment, you observe **80 heads**. We will use this data and the estimator $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ (as in the previous problem) to provide an answer to the hypothesis testing **question of interest**: "**Does $\mathbf{p = 0.5}$ or does $\mathbf{p \neq 0.5}$?**".

Let $D_1$ denote the value of the realization of the statistic $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ on the given data set. (Here $n = 200$, the number of flips.) What is $D_1$?

$D_1 =$ -2.828 ✔ **Answer:** -2.82842

Let $Z \sim \mathcal{N}\,(0,1)$. What is $\mathbf{P}\,(Z < D_1)$?

(You are welcome to use table or any computational tools e.g. R, or this online normal table calculator.)

$P\,(Z < D_1) =$ 0.0023 ✔ **Answer:** 0.00234

Since $n = 200$ is fairly large, we may assume that if $p = 0.5$ that $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \sim \mathcal{N}\,(0,1)\,.$

Suppose that $p = 0.5$ and you ran the experiment above (consisting of $200$ coin flips) a total of $1000$ times (i.e. a total $200 \times 1000$ coin flips). What is the expected number of experiments such that the estimator $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ is smaller than the value $D_1$ attained in the first experiment? (Round your answer to the nearest integer.)

2 ✔ Answer: 2

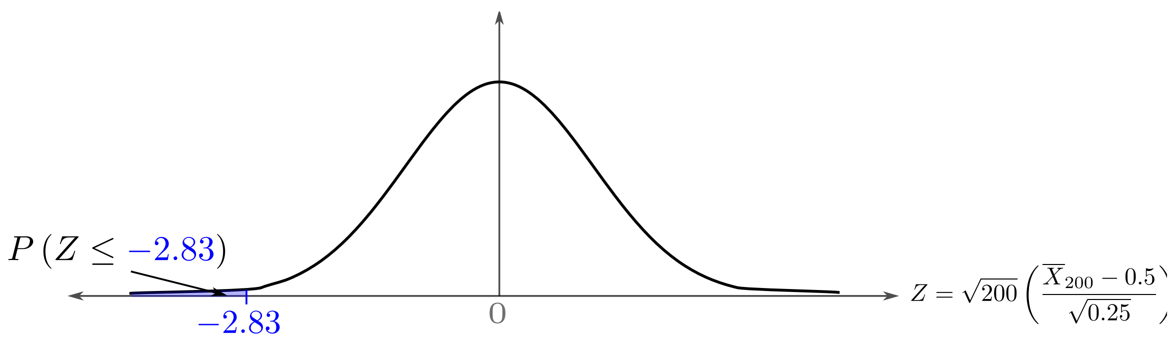
**Solution:**

First,

$$D_1 = \sqrt{200}\left(\frac{\frac{80}{200} - 0.5}{\sqrt{0.25}}\right) \approx -2.82842.$$

Using a table or computational software, we can also compute that if $Z \sim \mathcal{N}(0,1)$,

$$P(Z < D_1) = \int_{-\infty}^{-2.82842} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \approx .00234$$




Hence, for a single experiment, if $p = 0.5$, then there is (approximately) a $0.23\%$ chance of seeing an observation smaller than $D_1 \approx -2.82842$. Thus if we run $1000$ experiments, we would expect to see

$$1000 * (.00234) \approx 2.33907$$

experiments where $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ is smaller than $D_1 \approx -2.82842$.

**Remark**: By the previous result, it seems reasonable to conclude, for our first experiment, that it is **unlikely** that $p = 0.5$. Indeed, if $p = 0.5$, observing the value $D_1 \approx -2.82842$ would be a very rare event, intuitively speaking. In practice, one has to set the threshold of what determines a "very rare" event, and this will be studied later in this lecture.

提交 你已经尝试了1次（总共可以尝试3次）

Answers are displayed within the problem

## Hypothesis Testing: Another Sample Data Set of Coin Flips

3/3 points (graded)
We repeat the above exercise with a different data set.

As above, consider a **statistical experiment** where you flip the coin $200$ times. However, in this run of the experiment, you observe $\mathbf{106}$ **heads**. We will use this data and the statistic $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ from the previous problem to provide an answer to the hypothesis testing question of interest:

"**Does $\mathbf{p = 0.5}$ or does $\mathbf{p \neq 0.5}$?**"

Let $D_2$ denote the value of the realization of the estimator $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ on the given data set. (Here $n = 200$, the number of flips.) What is $D_2$?

$D_2 =$ 0.8485 ✔ **Answer:** 0.8485

Let $Z \sim \mathcal{N}(0,1)$. What is $\mathbf{P}(Z > D_2)$?
(You are welcome to use any tables or any computational tools e.g. R or this online normal table calculator.)

$\mathbf{P}(Z > D_2) =$ 0.1981 ✔ **Answer:** 0.19808

Since $n = 200$ is fairly large, we may assume if $p = 0.5$ that $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \sim \mathcal{N}(0,1)$.

Suppose that $p = 0.5$ and you ran the experiment above (consisting of $200$ coin flips) a total of $1000$ times. What is the expected number of experiments such that the estimator $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ is larger than the value $D_2$ attained in the first experiment? (Round your answer to the nearest integer.)

198 ✔ **Answer:** 198

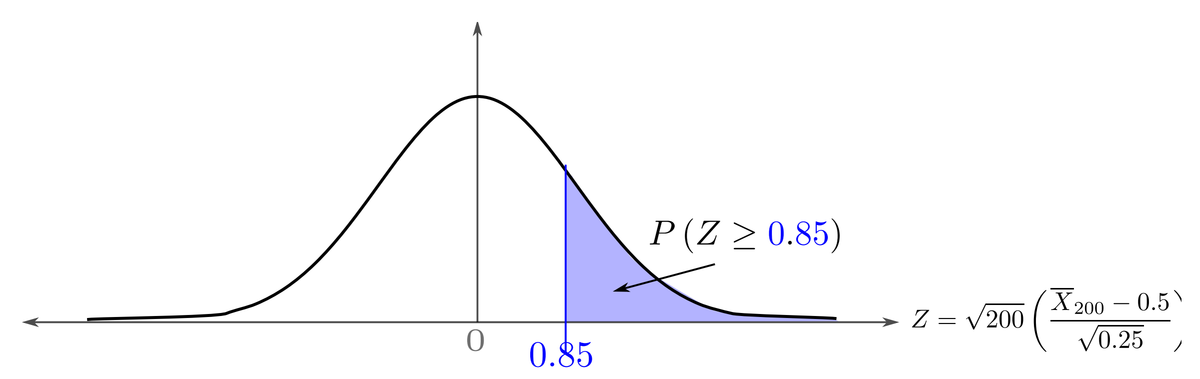
**Solution:**

First,

$$D_2 = \sqrt{200}\left(\frac{\frac{106}{200} - 0.5}{\sqrt{0.25}}\right) \approx 0.8485.$$

Using a table or computational software, we can also compute that if $Z \sim \mathcal{N}(0,1)$,

$$\mathbf{P}(Z > D_2) = \int_{0.8485}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \approx 0.19808.$$




Hence, for a single experiment, if $p = 0.5$, then there is (approximately) a $19.8\%$ chance of seeing an observation larger than $D \approx 0.8485$. Thus if we run $1000$ experiments, we would expect to see

$$1000 * (.00234) \approx 198.08$$

experiments where $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ is larger than $D_2 \approx 0.8485$.

**Remark 1**: By the previous result, from a heuristic perspective, we would be unable to refute the hypothesis that $p = 0.5$ (Note that this is a **different** conclusion than saying "We may conclude that $p = 0.5$"). Indeed, if $p = 0.5$, observing a value larger than $D_2 \approx 0.8485$ would be **not** be a rare event, intuitively speaking. In practice, one has to set the threshold of what determines a "rare" event, and this will be studied later in this lecture.

**Remark 2**: Though we are considering a very specific example and applying a very specific test, the steps taken in this problem and the previous one are illustrative of the general principles of hypothesis testing. In general, we will transform our data into a given statistic whose distribution we know well that does **not** depend on the true parameter (e.g., as in this problem, the standard Gaussian). Such a distribution is known as **pivotal** . Then we can reduce our hypothesis testing question to a problem of deciding whether or not a given observation is likely (or not) for this pivotal distribution.

提交 你已经尝试了3次（总共可以尝试3次）

Answers are displayed within the problem

## 讨论

隐藏讨论

**主题：** Unit 2 Foundation of Inference:Lecture 6: Introduction to Hypothesis Testing, and Type 1 and Type 2 Errors / 9. Example: Is this Coin Fair?

**Add a Post**

**All Posts**

### Conclusion at end of video

discussion posted 8 days ago by **mrBB** (Community TA)

At the end of the video we're presented the conclusion "It is unlikely that the coin is unfair." The professor contrast that with "The coin is likely to be fair." To me these mean exactly the same, and I think we can't conclude neither of these. I think the conclusion should actually have been "It is *not* unlikely that the coin is fair." Would the logicians among us agree?

此帖对所有人可见。

**添加回复**

4 responses

**karenechu** (Staff)
7 days ago

You're totally right! and none of us caught it. (May the students in class did but just didn't say anything.) I am adding a note. Thanks a lot MrBB.

添加评论

**SergK** (Community TA)
7 days ago

There is no such thing as a fair coin, but a particular test may not detect that a coin is unfair. This is confusing at this point, but the later videos make clear what is meant. The hypotheses "the coin is fair" and "the coin is unfair" are not symmetric, the first is null hypothesis, the second is alternative hypothesis. The purpose of test is to reject the null hypothesis, and it is quite possible that the test can't do it; then we say "it is unlikely that the coin is unfair", using natural language. This statement makes clear sense in context of hypotheses testing, while the statements "it is likely that the coin is fair" or "it is not unlikely that the coin is fair" do not, IMO.

I agree with everything you say, Serg, up to "then we say ...". Perhaps this is a language thingy, but to me "it is unlikely that the coin is unfair" is a *strong* statement saying it is unlikely $H_1$ is true, and by $\neg H_1 \implies H_0$, therefore that it is likely that $H_0$ is true (again strong statement).

On the other hand "it is not unlikely that the coin is fair" to me is a *weak* statement, only saying that $H_0$ is not unlikely ($\neq$ likely) and therefore can't be rejected. And it is exactly this weak conclusion that we want to express IMO.

**mrBB** (Community TA) 在7 days ago前发表

@mrBB, I can't agree with your logic. $H_0$ is the "status quo" and if the data is not against it then it is against the alternative hypothesis. So, it makes sense when rejecting $H_1$ to say that "It is unlikely that ... *whatever the alternative hypothesis is*".

**KRKirov** 在3 days ago前发表

That's just not how hypothesis testing works. We don't reject $H_1$, but we accept or reject $H_0$. Rejecting $H_0$ is only done when we have strong evidence that $H_0$ is unlikely given the data. Accepting $H_0$ is the default and is a weak statement. It only means that we have not enough evidence to reject $H_0$. This doesn't imply the data provides strong evidence $H_0$ is true (or equivalently that $H_1 is false).

**mrBB** (Community TA) 在3 days ago前发表

@mrBB, $H_0$ is the status quo, if the data is not against it it doesn't make sense to say that the status quo is likely. I think SergK's answer above encapsulates the matter very well and am bewildered by the correction note at the top of this page.

KRKirov 在2 days ago前发表

*"if the data is not against it it doesn't make sense to say that the status quo is likely"*

I don't understand. Is this really what you intended to write?

We are talking about the evidence that the hypothesis test provides. That we might have established $H_0$ firmly previously is not relevant. Moreover, $H_0$ doesn't have to be firmly established status quo. I.e. when we test a new medicine and the null hypothesis is that the new medicine doesn't work better than a placebo, we can hardly call that "status quo". In any case, if we would reject $H_0$ that means we have strong evidence that the medicine works better. If don't reject $H_0$ that does *not* mean we have strong evidence that the new medicine doesn't work better. Stating that "it is unlikely that the new medicine works better" would then be a too strong statement. We have only established with the test that is isn't unlikely that the new medicine doesn't work better.

Also see the lecture example on "not guilty" $\neq$ "innocent" .

mrBB (Community TA) 在a day ago前发表

添加评论

## Ron_is_learning
6 days ago

I think the wording in the "correction" now posted at the top of the page as an intro to the video is actually wrong. (a "not" was missed in MrBB's wording)

That said, the "not unlikely", is, at least for me, a little mental manipulation because it sounds like a double negative, which feels like it should imply "likely".

Perhaps it's easier just to say, "there is insufficient evidence to conclude that the coin is unfair".

I just reread and realized it's definitely wrong.

karenechu (Staff) 在6 days ago前发表

@Ron "Not unlikely" and "likely" to me have very different connotations. When I say "$A$ is likely" I mean $P(A) > 80\%$ (the actual number is not important of course, but only that it is fairly close to $1$): I have quite some confidence $A$ actually is the case. When I say "$A$ is unlikely" I mean $P(A) < 20\%$ (I'm quite sure $A$ is *not* the case). Therefore when I say "$A$ is not unlikely" it to me means $P(A) \nless 20\%$ or equivalently $P(A) > 20\%$ ($A$ might be the case but I'm not sure whatsoever).

mrBB (Community TA) 在6 days ago前发表

@mrBB. Sure, you can use the formal logic route. But if my only option are "likely" and "unlikely", then I would infer "likely" as more often than not, or p > 50%. And unlikely as p < 50%. Because I am not assuming just how likely, that likely has to be... just along the lines of a civil case: more likely than not. So if it's "not unlikly" than it is almost surely "likely" UNLESS it is equally likely: 50/50, but... the probability of any exact number is zero on our continuous p. Hahaha. So equally likely effectively doesn't exist.

All of that aside, my comment was more along the lines of everyday language usage and how people "hear it" to understand what is intended, versus if anybody is going to bother breaking out truth tables to parse sentences as a general habit. ;)

Ron_is_learning 在6 days ago前发表

添加评论

## ptressel
3 days ago

Perhaps the correction note has been changed...but here's what it reads currently:

"**It is not unlikely that the coin is fair.**"

Here is the hand-written statement on the slide for "Example 2":

"**It is unlikely that the coin is unfair.**"

These have identical logical meaning (and, IMO, even the same connotation, and very close to the same valence) in English. The correction appears unnecessary. Just apply negation to both terms "likely" and "fair" in the first statement, and you'll get the second.

To be pedantic 😏 let's use "not" for all negations so that unlikely becomes not likely, unfair becomes not fair. Start with the first form:

**not not likely that coin is fair**

Apply negation to both terms:

**not not not likely that coin is not fair**

not not is the identity (i.e. $\neg\neg A \equiv A$) so remove the double not:

**not likely that coin is not fair**

For speakers of languages that have double negation (which is very common), just treat an English statement containing negation as though it were formal logic -- each negation applied to the same object flips the sense.

I like how Ron_is_learning says it: "There is insufficient evidence to conclude that the coin is unfair".

What we'll likely soon learn to say is that we "fail to reject the null hypothesis". 😁

"not unlikely" to me has a very different connotation than "likely". These two would only be equal if everything is either likely or unlikely. But many things are neither. As I mentioned somewhere above, to me "likely" means that my measure of belief is $> 80\%$. And when my measure of belief is $< 20\%$ I deem something unlikely. Everything in between for me is neither likely nor unlikely. So to me "not unlikely" means that it can fall in either of the categories "likely" or "neither likely nor unlikely".

Do you really feel "it is not unlikely I will visit Paris this year" and "it is likely I will visit Paris this year" express the same level of certainty about visiting Paris later this year?

mrBB (Community TA) 在a day ago前发表

I am wondering how a linguist does statistics.

dxander 在a day ago前发表

添加评论

显示所有的回复

## 添加一条回复：

Preview

提交

认证证书是什么?