edX

Currently enrolled in **Audit Track** (expires December 25, 2019)   Upgrade ($300)

10. Statistical Formulation of Hypothesis Testing
Statistical Formulation of Hypothesis Testing

So that's a good analogy to keep in mind.

10:03 / 10:03　　　　　　　　　　1.50x

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

Properties of the Null and Alternative Hypothesis

1/1 point (graded)

Let $X_1, \ldots, X_n \overset{iid}{\sim} P_{\theta^*}$ for some unknown parameter θ^*. The associated statistical model is $(E, \{P_\theta\}_{\theta \in \Theta})$.

Which of the following are true statements regarding the **null hypothesis** $H_0 : \theta^* \in \Theta_0$ and the **alternative hypothesis** $H_1 : \theta^* \in \Theta_1$?
(Choose all that apply.)

- [x] Θ_0 and Θ_1 must be subsets of the parameter space Θ.

- [x] Θ_0 and Θ_1 must be disjoint, i.e. $\Theta_0 \cap \Theta_1 = \emptyset$.

- [ ] Θ_0 and Θ_1 must make up all of the parameter space Θ, i.e. $\Theta_0 \cup \Theta_1 = \Theta$.

- [ ] The null and alternative hypotheses play symmetric roles: *i.e.* if we set $H_0' : \theta \in \Theta_1$ and $H_1' : \theta \in \Theta_0$ then we are still doing the exact same hypothesis test as in the problem statement.

✔

Solution:

We examine the answer choices in order.

- "Θ_0 and Θ_1 must be subsets of the parameter space Θ." is correct. We are trying to decide if θ^* is in a particular region of the parameter space, so Θ_0 and Θ_1 must be subsets of Θ.

- "Θ_0 and Θ_1 must be disjoint, i.e. $\Theta_0 \cap \Theta_1 = \emptyset$" is correct. For hypothesis testing, we are trying to determine, based on observations, whether or not it is likely that $\theta^* \in \Theta_0$. Since we only want to test whether or not the parameter lies in some region (or not), it makes sense to impose that the region Θ_0 determined by the null hypothesis and Θ_1, the region determined by the alternative hypothesis are disjoint.

- "Θ_0 and Θ_1 must make up the entire parameter space Θ" is **not** correct. For example, recall in the two sided test comparing the boarding times of the rear-to-front and the WILMA from the beginning of this lecture, the parameter space is
$\Theta = \{(\mu_1, \mu_2) : \mu_1 \in \mathbb{R}, \mu_2 \in \mathbb{R}\} = \mathbb{R}^2$, while the null and alternative hypotheses are given by

$$
\begin{aligned}
\Theta_0 &= \{(\mu_1, \mu_2) : \mu_1 = \mu_2\} &&\text{a line in } \mathbb{R}^2 \\
\Theta_1 &= \{(\mu_1, \mu_2) : \mu_1 > \mu_2\} &&\text{the region on one side of the line } \mu_1 = \mu_2 \text{ in } \mathbb{R}^2,
\end{aligned}
$$

The region $\Theta_0 \cup \Theta_1$ does not make up the all of $\mathbb{R}^2$ because the region on the other side of the line $\mu_1 = \mu_2$ is not included.
A simpler example is when Θ_0 and Θ_1 both consists of single values of θ when the parameter space Θ is for example $\mathbb{R}^2$.

- "The null and alternative hypotheses play symmetric roles: *i.e.* if we set $H_0' : \theta \in \Theta_1$ and $H_1' : \theta \in \Theta_0$ then we are still doing the exact same hypothesis test as in the problem statement." is incorrect. Actually, H_0 and H_1 play asymmetric roles. Our only goal in hypothesis testing is to use the data to determine whether or not we can **reject** H_0. This is a different statistical objective than using the data to determine whether or not we can reject H_0'.

Remark: Regardless of the data, our conclusion will never be to *accept* the null. On observing the data, we will either **reject** the null in favor of the alternative OR we will **fail to reject** the null. In the latter case, we are not claiming that the null is true, rather we are stating that the data does not provide us with enough evidence to refute the null hypothesis.

<table>
<tr><td>Submit</td><td>You have used 1 of 2 attempts</td></tr>
</table>

ⓘ   Answers are displayed within the problem

## Formulating the Null and Alternative Hypothesis: Is This Coin Fair?

1/1 point (graded)
Refer back to the statistical experiment where you flip the coin 200 times in order to answer the question of interest: "**Is this coin fair?**."

You take as the **status quo** that the coin is fair. Hence, the data must show strong evidence to the contrary in order for this status quo to be rejected. In other words, the coin is considerd fair until "proven" otherwise.

You model the coin flips by $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Ber}\,(p)$ where p is an unknown parameter and rephrase the question as:
"**Does p $= 0.5$ or does p $\neq 0.5$?**".

Formulate the null and alternate hypothesis for this test. Which of the following depicts Θ_0 and Θ_1?
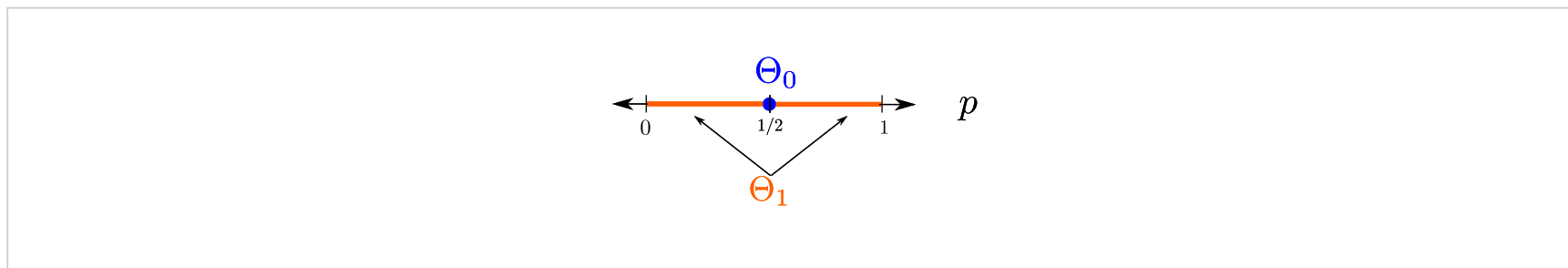
✔
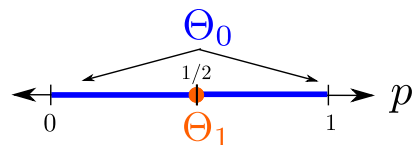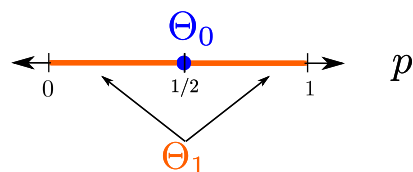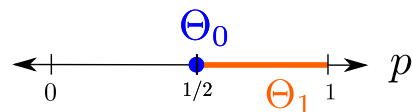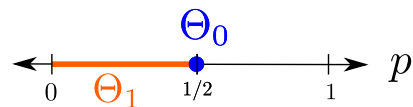
Solution:

In this test, we are looking for evidence in the data to show that the coin is **not** fair. Hence, the null hypothesis is

$$H_0 : p \in \Theta_0 = \{1/2\},$$

and the alternate hypothesis is

$$H_1 : p \in \Theta_1 = (0, 1/2) \cup (1/2, 1),$$

depicted by the figure

Remark: This is called a **two sided test** since Θ_1 lies on both sides of Θ_0.

Submit	You have used 1 of 2 attempts

ⓘ   Answers are displayed within the problem

Formulating the null and Alternative Hypothesis: Are US people taller in 2018 than in 1920?

1/1 point (graded)

As in a previous problem, you try to answer the question "**Were people in the U.S. taller in 2018 than in 1920, when the average height was 5.5 feet?**" by sampling 1 million individuals labeled $1, 2, \ldots, 10^6$ randomly from the 2018 U.S. population.

You take as the **status quo** that the people are **not** taller in 2018, and look for evidence in the data to reject this status quo. In other words, you assume people are **not** taller in 2018, until "proven" otherwise.

You model the height of the i-th individual as a random variable X_i and make the assumption, based on 1920 data, that

$X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1.3)$ where μ is an unknown parameter.

You rephrase the question of interest as: **Is $\mu > 5.5$? or is $\mu \leq 5.5$?**

Formulate the null and alternate hypothesis for this test. Which of the following depicts Θ_0 and Θ_1?
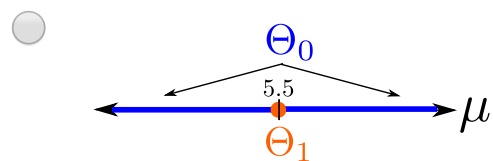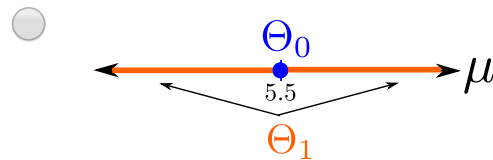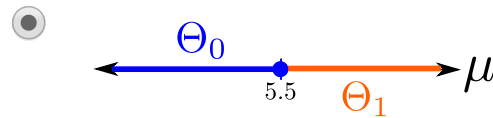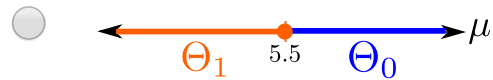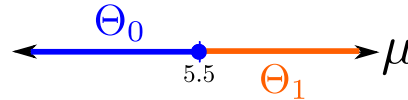

Solution:

In this test, we are looking for evidence in the data to show that $\mu > 5.5$. Hence, the null hypothesis is

$$H_0 : \mu \in \Theta_0 = (-\infty, 5.5],$$

and the alternate hypothesis is

$$H_1 : p \in \Theta_1 = (5.5, \infty),$$

depicted by the figure

Remark: This is called a **one sided test** since Θ_1 lies on only one side of Θ_0.

Submit	You have used 1 of 2 attempts

ⓘ   Answers are displayed within the problem

Discussion

<div style="text-align:right">**Hide Discussion**</div>

Topic: Unit 2 Foundation of Inference:Lecture 6: Introduction to Hypothesis Testing, and Type 1 and Type 2
Errors / 10. Statistical Formulation of Hypothesis Testing

Add a Post

Show all posts ▼	by recent activity ▼

💬 Question: Formulating the Null and Alternative Hypothesis: Is This Coin Fair?
In original example, the question asked is "Can we conclude that the coin is significantly unfair ?". But in this exercise, question has been rephrased as "Is this coin fair?" Well, ... — 1

💬 LOVE the visuals
I understand the different people learn different ways. I happen to be VERY visual, so all of these math concepts can be very dry, but these visuals about the NULL and Alterna... — 3

Audit Access Expires Dec 24, 2019

You lose all access to this course, including your progress, on Dec 24, 2019.

Upgrade by Nov 4, 2019 to get unlimited access to the course as long as it exists on the site. **Upgrade now**

Learn About Verified Certificates