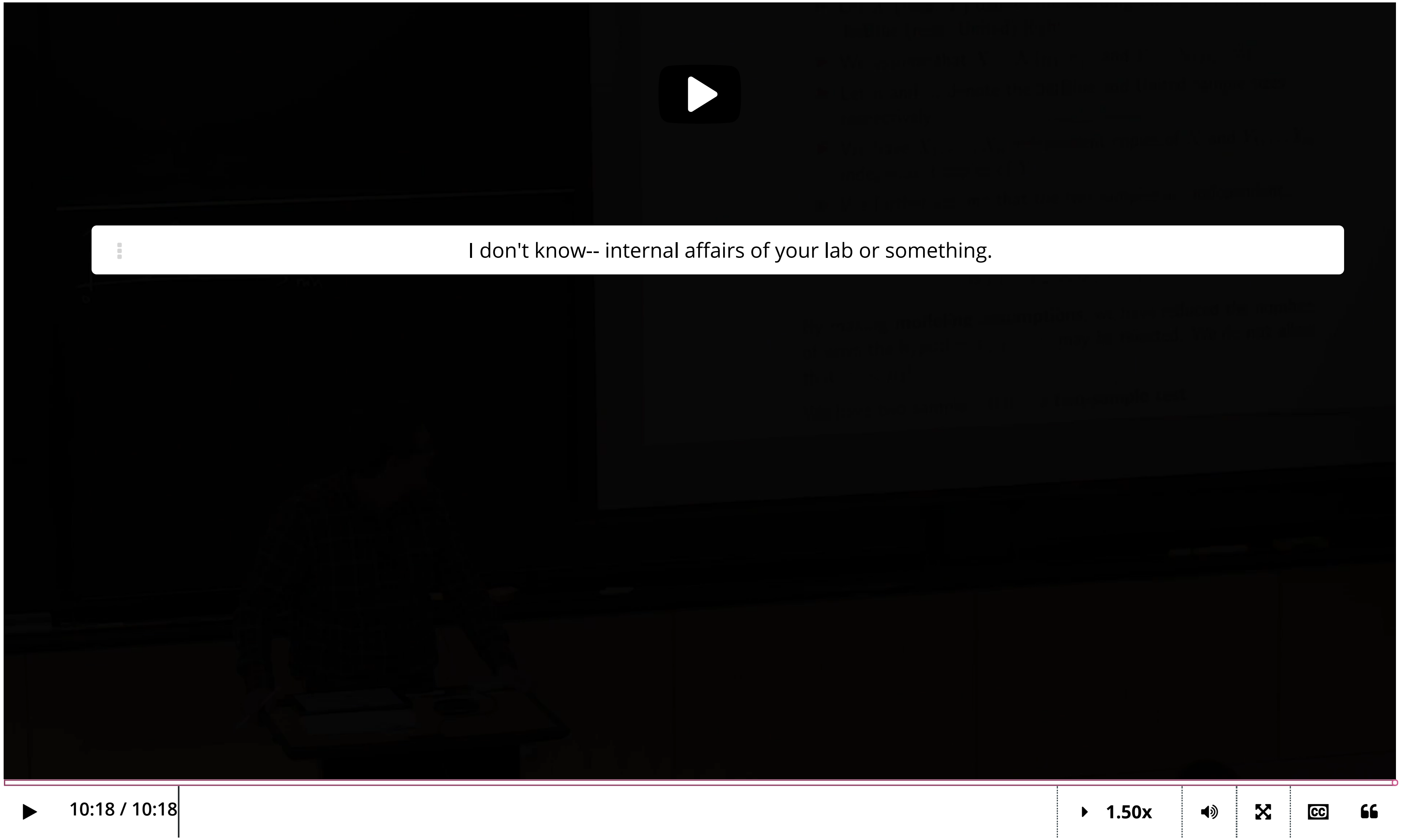
2. Introduction to Hypothesis Testing

Comparing Two Boarding Methods: Modeling Assumptions



Video

[Download video file]

Transcripts

[Download SubRip (.srt) file]

[Download Text (.txt) file]

Modeling Clinical Trials I

1/1 point (graded)

In a clinical trial, a pharmaceutical company wants to determine the efficacy of a cold remedy. To do so, they recruit $2n$ individuals to participate in a study, (randomly) placing n individuals in the **treatment group** and n individuals in the **control group**. Throughout the study, the treatment group will receive the actual drug, while the control group will only receive a placebo (for example, a sugar pill).

To statistically model this scenario, we let

- $X_1, \ldots, X_n$ be random variables that denote the number of coughs per hour of individuals $1, \ldots, n$, respectively in the treatment group, and
- $Y_1, \ldots, Y_n$ be random variables that denote the number of coughs per hour of individuals $1, \ldots, n$, respectively, in the control group.

Let's assume that the individuals participating in the trial are separated throughout the trial, so that it's reasonable to expect the coughs per hour of one individual in the study will not affect the coughs per hour of some other individual in the study. Moreover, we expect the drug administered to induce the same distribution (on coughs) on each individual in the treatment group. We will also assume that the number of coughs per hour for individuals in the control group have the same distribution as each other.

What collection of mathematical assumption(s) below would capture exactly all of the assumptions stated in the previous paragraph, but nothing more? (Choose all that apply.)

- [ ] $X_1, \ldots, X_n$ are independent, but may not all have the same distribution. The same holds for $Y_1, \ldots, Y_n$.
- [ ] $X_1, \ldots, X_n$ all have the same distribution, but some of them are correlated. The same holds for $Y_1, \ldots, Y_n$.
- [x] The random variables $X_1, \ldots, X_n$ are iid and the random variables $Y_1, \ldots, Y_n$ are iid (though perhaps from a different distribution from $X_1, \ldots, X_n$).
- [ ] The random variables $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ are all iid (in particular, the X_i's and Y_i's are sampled from the *same* distribution).
- [x] The random variable X_i for any i is independent of Y_j for any j.

✔

Solution:

The third choice "The random variables $X_1, \ldots, X_n$ are iid and the random variables $Y_1, \ldots, Y_n$ are iid (though perhaps from a different distribution from $X_1, \ldots, X_n$)." " and the last choice "The random variable X_i for any i is independent of Y_j for any j." together captures all assumptions we need. Since, intuitively speaking, we do not expect individuals in the study will affect one another, this translates to imposing that all random variables $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ are independent of one another. We also assumed that $X_1, \ldots, X_n$ will have the same distribution induced by the drug and that the treatment group $Y_1, \ldots, Y_n$ will have a common distribution on coughs. Thus, the assumption $X_1, \ldots, X_n$ are iid and $Y_1, \ldots, Y_n$ are iid captures all of the information described. It is important to note, however, that X_i and Y_i may be sample from **different** distributions.

We now look at the incorrect choices in order.

- The first and second choices, "$X_1, \ldots, X_n$ are independent, but may not all have the same distribution. The same holds for $Y_1, \ldots, Y_n$." and "$X_1, \ldots, X_n$ all have the same distribution, but some of them are correlated. The same holds for $Y_1, \ldots, Y_n$.", respectively, are incorrect because each directly contradicts the iid assumption.

- The fourth choice "The random variables $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ are all iid (in particular, the X_i's and Y_i's are sampled from the *same* distribution.)" is incorrect. The paragraph mentioned does not assume anywhere that the X_i's should have the same distribution as the Y_i's. Since we are mainly interested in deciding, based on the data, whether or not the X_i's and Y_i's have the same (or differing) distribution, for the purpose of modeling, it would not make sense to impose that they have the same distribution.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

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

Add a Post

Show all posts ▼ by recent activity ▼

same distribution

It's unclear from the description of the problem of the samples are selected from the same distribution or if they perhaps are not from the same. We only read "they are selec... 8

Nice magic trick.
I like how not only did he seamlessly change clothes in the middle of lecture but also grew a couple inches of hair in a matter of seconds!
5