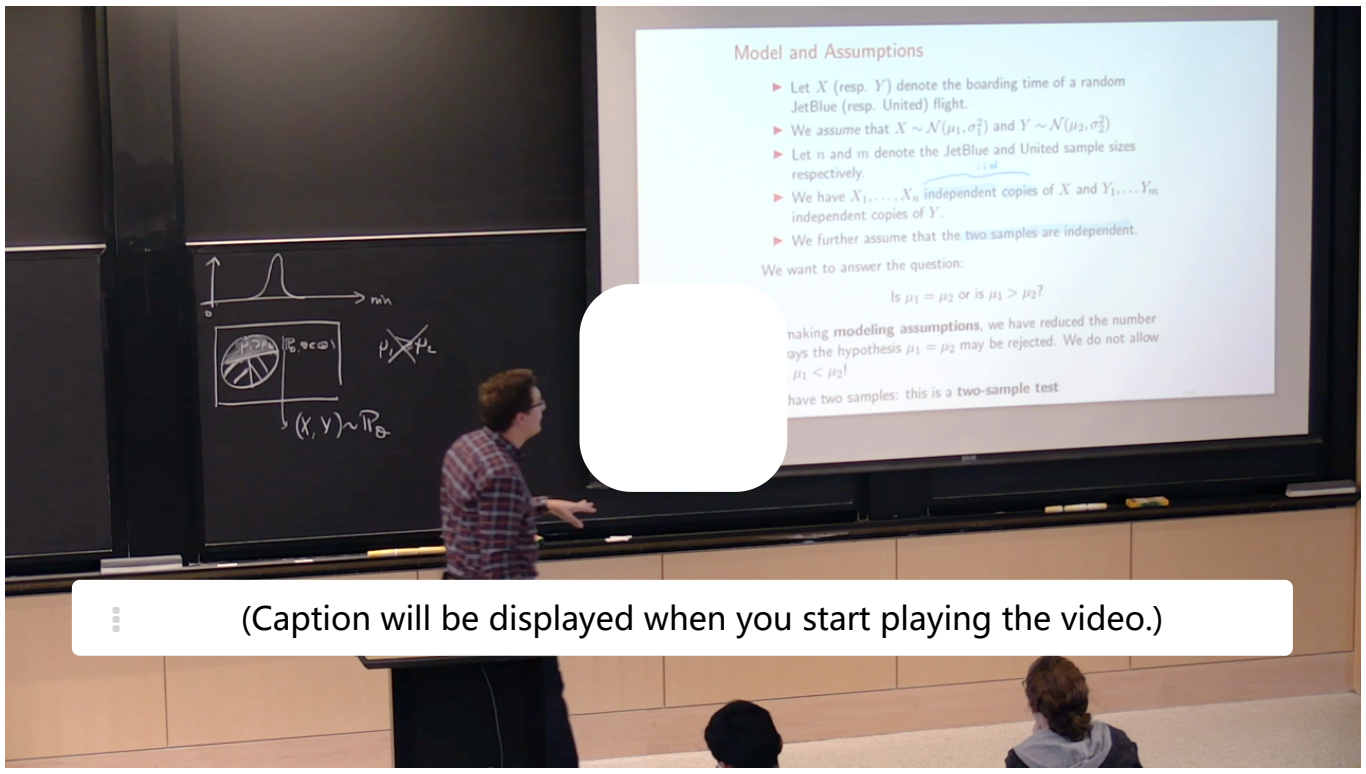
# 5. Heuristics for Two Sample Tests
## Heuristics for Two Sample Tests




Start of transcript. Skip to the end.

By the way, so there's one piece of terminology,

which is actually quite important here,

is we have two samples, right?

One for JetBlue, one for United.

This is the overwhelming majority

of hypothesis testing problems.

You will hear about two sample t-tests, for example.

**视频**
下载视频文件

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

xuetangX.com 学堂在线

## Certainty of a Two-Sample Hypothesis Test

1/1 point (graded)

Let's use the same statistical set-up as above. Recall that $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poiss}(\mu_{\text{drug}})$ and $Y_1, \ldots, Y_n \overset{iid}{\sim} \text{Poiss}(\mu_{\text{control}})$ where $X_i$ denotes the number of coughs per hour of the $i$-th individual in the treatment group and $Y_i$ denotes the number of coughs per hour of the $i$-th individual in the control group. The parameters $\mu_{\text{drug}}$ and $\mu_{\text{control}}$ are unknown. You would like to determine from the two samples if $\mu_{\text{drug}} < \mu_{\text{control}}$.

To do so, you compute the sample mean corresponding to each group:

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \overline{Y}_n := \frac{1}{n}\sum_{i=1}^{n} Y_i$$

and observe that $\overline{X}_n < \overline{Y}_n$.

Can you conclude with 100% certainty that $\mu_{\text{drug}} < \mu_{\text{control}}$?

Choose the correct answer that also has a correct explanation.

- Yes, because we do not expect the placebo effect to factor in to this trial.
- Yes, because we have carefully chose the treatment and control group so their sample means match the true means: $\overline{X}_n = \mu_{\text{drug}}$ and $\overline{Y}_n = \mu_{\text{control}}$.

- (●) No, we cannot conclude $\mu_{\text{drug}} < \mu_{\text{control}}$. Since there are possible fluctuations in $\overline{X}_n$ and $\overline{Y}_n$ about their respective means $\mu_{\text{drug}}$ and $\mu_{\text{control}}$, there is some positive probability that $\overline{X}_n < \overline{Y}_n$ while at the same time $\mu_{\text{drug}} > \mu_{\text{control}}$. ✔

- ( ) No, because the sample means $\overline{X}_n$ and $\overline{Y}_n$ are biased estimators of their true means, $\mu_{\text{drug}}$ and $\mu_{\text{control}}$, respectively.

**Solution:**

First we examine the correct choice and then look at the incorrect responses in order.

- The third response "No, we cannot conclude $\overline{X}_n < \overline{Y}_n$. Since there are significant fluctuations in $\overline{X}_n$ and $\overline{Y}_n$ about their respective means $\mu_{\text{drug}}$ and $\mu_{\text{control}}$, there is some positive probability that $\overline{X}_n < \overline{Y}_n$ while at the same time $\mu_{\text{drug}} > \mu_{\text{control}}$." is the correct response. Using $\overline{X}_n < \overline{Y}_n$ to predict $\mu_{\text{drug}} < \mu_{\text{control}}$ is only a heuristic, and this may fail at times. For example, perhaps by chance we chose a treatment group that responds extremely well to the drug (*i.e.* $X_1, \ldots, X_n$ are outliers), but the vast majority of the population will not see a significant effect. In this case it is possible that $\overline{X}_n < \overline{Y}_n$ while $\mu_{\text{drug}} > \mu_{\text{control}}$.

- The first response "Yes, because we do not expect the placebo effect to factor in to this trial." is incorrect. On the contrary, the goal of this trial is to compare the placebo effect to the effect of the treatment so we can determine if the drug is useful for treating the cold.

- The second response "Yes, because we have carefully chose the treatment and control group so their sample means match the true means: $\overline{X}_n = \mu_{\text{drug}}$ and $\overline{Y}_n = \mu_{\text{control}}$." is incorrect. We have no way of selecting the participants of the trial so that the sample means match the true means, and this is the case for a couple reasons. First, we do not know the true means, so even if we were given the observations $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ in advance, this would be not possible. And moreover, we do not even have access to the observations $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ until after the clinical trial has completed. Hence, we have no way of controlling the sample mean in advance, and doing so would actually defeat the purpose of running a clinical trial.

- The fourth response "No, because the sample means $\overline{X}_n$ and $\overline{Y}_n$ are biased estimators of their true means, $\mu_{\text{drug}}$ and $\mu_{\text{control}}$, respectively" is incorrect: this choice gives the right answer "No" but for a reason which is false.
  Both $\overline{X}_n$ and $\overline{Y}_n$ are unbiased estimators of $\mu_{\text{drug}}$ and $\mu_{\text{control}}$ respectively.
  In addition, it is possible for a simple comparison test to yield an incorrect answer even if the estimators are unbiased (again, highlighting why the third choice is correct).

提交 你已经尝试了1次（总共可以尝试2次）

Answers are displayed within the problem

## 讨论

显示讨论

**主题：** Unit 2 Foundation of Inference:Lecture 6: Introduction to Hypothesis Testing, and Type 1 and Type 2 Errors / 5. Heuristics for Two Sample Tests

认证证书是什么?