Instructor Guide to
Lineups Activity: Building Intuition for Two-Sample Inference

Quick Info

Audience: *Intro undergraduate statistics*

Brief Description: *Students build their intuition about hypothesis testing by answering a research question using a visual test.*

Learning Goals: *Develop intuition for the logic of hypothesis tests using a visual test.*

Topics Covered: *Exploratory data analysis, hypothesis statements, (informal) two-sample hypotheses tests*

Prerequisites: *Descriptive statistics (mean, standard deviation, etc.), basic statistical graphics (histograms, boxplots, etc.), basic experimental design (role of randomization)*

Class Resources: *Students can use R, JMP, SPSS, etc. to explore the data set prior to viewing the lineups, or the graphics could be included on the student handout. Students will need access to the lineup (this can be done using a slide, handout, or a Shiny app).*

Instructor Resources:

- *Student handout and instructor guide*
- *Sample data set*
- *Shiny app (BLINDED URL)*
- *Tutorial on creating lineups in R (BLINDED URL)*

Time: *~50 minutes in class*

Why use this activity in your course?

Students often struggle to make the leap from exploring a data set and describing what they learn in a sample to drawing conclusions about the population. Introducing students to the logic of hypothesis testing is especially difficult if too many technical details are introduced with this new thought process. Using lineups provides the scaffolding to build from an intuitive introduction to testing to the formal statistical ideas. This activity serves as the starting point, where students are only asked to apply "Sesame Street logic" (i.e., "which one of these is not like the others") to draw conclusions. This allows focus to remain on the logic rather than the technical details (i.e., reference distributions and formalized strength of evidence).

When should you use this activity in your course and what are the prerequisites?

This activity is designed to introduce concepts underlying hypothesis tests, so it should be used on the first day you discuss testing. In my course, this comes after a unit on data collection and EDA. Students should already have working knowledge of descriptive statistics (mean, standard deviation, etc.), statistical graphics (boxplots, etc.), and data collection schemes (observational vs. experimental studies, random sampling, randomization, etc.). Instructors could assign a reading or give a mini lecture on crafting hypotheses before this activity.

What is an example lesson plan for the activity?

Note: Italicized text indicates a teaching tip or aside.

- **Introduce logic of testing and data set** (5 minutes, depending on the scope). At the beginning of class, introduce (or review from the reading) the logic behind hypothesis tests—why can't we simply answer our research question using our one sample? Alternatively, the logic behind testing could be introduced via reading or a pre-class video. The key is to make sure students review sampling and randomization variability, and why we can't draw general conclusions based solely on EDA. Students should also be familiar with formulating hypothesis statements, but the notation could be avoided to focus on the intuition. *This is a good time to assign student roles in the groups.*
- **Small group discussion** (10 minutes). Send students into their small groups to discuss questions 1-5. These questions review basic experimental design and EDA and challenge the students to develop appropriate hypotheses.
- **Introduce lineups** (3 minutes)**.** After working through questions 1-5, introduce lineup plots. The important point here is that students know one plot is the observed data and the other plots are decoys from situations where there is "no difference between groups." Don't worry about the technical details and conditions here. *You could also take a few minutes to review the answers to questions #1-5 at this point.*
- **Individual work** (2 minutes). Send students to work individually on questions 6-7 to evaluate the lineup plots and determine which plot is most different from the others. It's important to remind students to work individually because discussion is more interesting and meaningful when each student is an unbiased evaluator. *Saying "take two minutes to evaluate the lineup on your own" is helpful. You could even display a timer.*
- **Small group discussion** (5 minutes). Bring students back into their small groups to discuss questions 8 and 9 and come to a consensus on which plot is most different from the others. Having students unpack their thoughts about the lineups in groups helps them flesh out why they chose the particular plot. In addition, requiring one choice per group makes it easier to hear from "everyone" in a short amount of time in a lower stress environment (it's the group's choice, not the individual's). *I recommend reminding students that they need to arrive at consensus and that the spokesperson should be ready to share the group's thoughts with the class. In addition, you can explain why you are using group discussion if that helps your class "buy in."*
- **Large group discussion** (1 minute per group). Regroup and have each group briefly report what plot they chose and why. A less time-consuming alternative is to have each group put a post-it note on the board under their choice with a brief rationale for their choice. If you choose the post-it approach, then call on one group to provide their rationale. *You could also use some online polling platform (e.g., Google forms, Moodle, etc.) to collect this information. Even if you have each group report their choice and rationale, the post-it note approach provides you with a "progress bar" to gauge how things are going.*
- **Small group discussion** (3-5 minutes). Send students back into their small groups to discuss what their choice of plot suggests about the competing claims. *Encourage students to link directly to the claims as they report their answers, this tends to get more focused answers and guides them toward thinking about evidence.*
- **Debrief** (10 minutes)**.** Regroup and have each group report back, or randomly select a couple groups to share their thoughts if time is tight. Once each group has shared their thoughts, summarize how this activity links to each step of the logic behind hypothesis testing, ending with what the class just concluded and how that differs from using a single plot or test statistic. *You can introduce notation for the hypotheses and point to future topics here.*

Do you have any hints for using lineups?

1. Spend some time introducing the data set. This can be done while you distribute the handouts and as students get situated in their groups.
2. Have students work in small groups, perhaps 3 or 4. This provides students with the opportunity to discuss their understanding in a lower-stakes environment and to learn from each other.
3. Assign roles to each group member to help groups function efficiently and to avoid the situation where one student does all of the work. We recommend assigning roles such as

- Facilitator: makes sure the group stays on task and that each member has room to contribute
- Spokesperson: reports back to the class, reads from the recorder's notes
- Recorder: completes the worksheet for the group, takes coherent notes
- Encourager/Questioner: suggests alternatives if the group gets stuck, asks for clarification, poses questions

Follow-up Activities and Discussion Questions:
You could follow-up this activity with homework questions or warm-up questions for the next class where students evaluate another lineup or two. This would give you the opportunity to include a lineup for a situation where there is not a discernible difference between the two groups and/or using a different plot type. You could design a homework problem and/or additional class activities to grapple with each of the following questions (or some subset of them):

- Do you think plot type influences your ability to detect differences between groups? If so, how? *Plots are test statistics here, so it can be linked to the choice of a test statistic later in the course (e.g., difference in means vs. difference in medians).*
- If you cannot identify the data plot from the lineup, does that mean there is no difference between groups?
- If you can identify the data plot from the lineup, does that mean there is an important difference between groups?
- Consider two situations: (1) you are the only person in class who selected the data plot; (2) twenty people in class select the data plot. What do these two situations suggest about the competing claims?

What else is in this Instructor Guide?
In the next section, we provide a commented version of the student activity. I suggest possible alternative formats you can use, questions that you can ask students to facilitate discussion, and possible issues you may encounter.

References
This activity was adapted from a case study discussed in Chapter 1 of *The Statistical Sleuth*, 3rd ed.

The original study citation is
Amabile, T. M. (1985). Motivation and creativity: Effects of motivational orientation on creative writers. *Journal of Personality and Social Psychology*, 48(2), 393.

The format of this instructor guide was inspired by Shonda Kuiper's Stat2Labs.

Lineups Activity
Building Intuition for Two-Sample Inference

The goal of this activity is to develop your intuition about the logic of hypothesis tests using a visual lineup test.

BACKGROUND

Evidence suggests that reward systems may operate in the opposite way from what is intended (e.g., ranking systems may decrease productivity; grading systems may not stimulate learning). To investigate this phenomenon, Amabile, T. M. (1985) designed a study to explore whether motivation type (intrinsic or extrinsic) impacted creativity scores. In this study, 47 creative writers were randomly assigned to one of two questionnaires where they ranked reasons they write. One questionnaire listed intrinsic motivations and the other listed extrinsic motivations. After completing the questionnaire, all subjects wrote a Haiku about laughter, which was graded for creativity by a panel of poets. The average rating (out of 40 points) for each subject was recorded.

In this activity, you will explore whether there are discernible differences in creativity between the two groups.

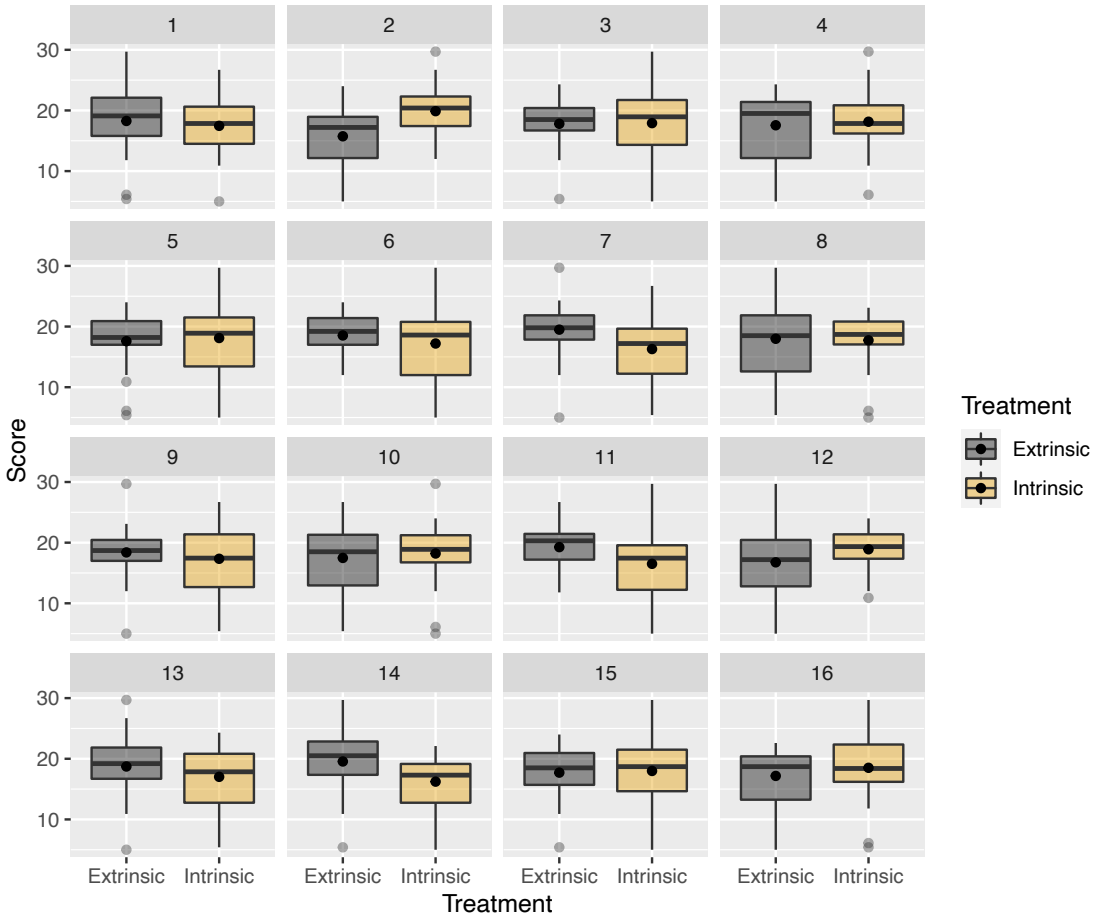
GROUP QUESTIONS

1. Is this an experimental or observational study? How do you know?
2. What research question is being investigated? That is, what is the objective of this study?
3. Write down the competing claims (i.e., hypotheses) being investigated.
4. Identify the response and explanatory variables.
5. What type of plot could you use to compare the two treatment groups? Why did you choose this plot type? How does this plot help you investigate the competing claims?

INDIVIDUAL QUESTIONS *Please do not discuss your answers with your group until you start question #8.*

Below is a lineup of plots where one plot displays the observed data and the other 19 plots are "decoy" plots where there is no difference between the treatment groups.

6. Which plot do you think is the most different from the others?
7. What feature(s) of the plot led you to this choice?



GROUP QUESTIONS

8. Which plot does your group think is the most different from the others?
9. What feature(s) of the plot led you to this choice?

STOP HERE! *We will have a large group discussion sharing the results and then the plot that displays the observed data will be revealed.*

The observed data are in plot # _____.

10. Did your group choose the observed (data) plot?
11. Based on your answer to question 10, what does this suggest about your competing claims?

We will regroup shortly to discuss your thoughts. If you have any questions, please ask!