Unit 7 Project - Teacher Guide

Table of Contents

Project Overview

This unit-long performance task invites students to use statistics as a tool for investigating representation in media. Across the unit, students pose their own statistical questions, collect and analyze data, and make evidence-based claims about similarities and differences between groups. Rather than aiming for definitive answers, the project emphasizes careful reasoning about variability, sampling, and uncertainty, and asks students to consider what conclusions are reasonable and responsible to make.

The project is intentionally student-driven. Students choose the media type, identity-related variable, and comparison groups that matter to them, grounding abstract statistical ideas in contexts connected to culture, identity, and lived experience. By the end of the unit, students synthesize their work into a final presentation that communicates both their mathematical reasoning and the real-world meaning of their findings.

Project Task/Lesson Sequence

Lesson	Focus / Student Work
1 **Ask a Question**	• Brainstorm noticings and wonderings about representation. • Pose a comparative statistical question.
2 **Plan Sampling**	• Identify subgroup populations. • Choose a fair random sampling method.
3 **Collect Data**	• Gather a small random sample. • Record data in an organized table.
4 **Increase Sample Size**	• Collect more data points in each sample to reduce variability. • Repeat trials to increase reliability.
5 **Center & Variability**	• Calculate mean and MAD. • Interpret typical values and spread.
7 **Visualize Data**	• Create dot or box plots. • Compare center and variability visually.
9 **Inferences**	• Make inferences mean difference and MAD comparisons. • Present findings and real-world implications.

Data Set Overview

Link: [Simulated Media Representation Data Set]

Students are provided with a large, simulated dataset designed to reflect realistic patterns found in youth media. While simulated, the data is modeled after real-world trends, allowing students to draw conclusions that are plausible and meaningful without the logistical challenges of true population-level data collection.

Variables included in the dataset:

- ID
- Medium (Movie, TV Show, Book, Video Game)
- Year Bin (Pre-2010, 2010–2020, 2020–Present)
- Format (Animated / Live Action for Movies & TV; N/A otherwise)
- Genre (Action, Fantasy, Drama, Comedy, Family, Horror)
- Intended Audience Age Group (Children, Teen/YA, Adult/MA)
- Lead Character Gender (Male, Female, Nonbinary, Unknown)
- Lead Race/Ethnicity (White, Black, Latinx, Asian, Indigenous, Middle Eastern, Multiracial, Unknown)
- LGBTQIA+ Lead (Yes, No, Unknown)
- Disability Lead (Yes, No, Unknown)

Students may compare any two subgroups that can be defined using these variables, as long as their question is measurable and allows for a fair comparison.

Using the Data in Google Sheets

Students will primarily work in Google Sheets to filter, sort, and sample data. Before beginning independent work, it may be helpful to ensure students are able to use the tools and see how they support strong random sampling.

- Filter columns to isolate a specific medium, year bin, or subgroup
- Sort data within a filtered set, when useful for organization (not for sampling)
- Count category values (e.g., number of "Yes" responses in a sample)
- Use row numbers or IDs alongside a random number generator to select random samples

Common student challenges to anticipate:

- Forgetting to apply the same filters to both comparison groups
- Accidentally using convenience sampling (e.g., first visible rows)
- Feeling unsure about how to handle "Unknown" values

Rather than resolving these issues for students, use questions to press their reasoning:

- *How does this choice affect who is included or excluded?*
- *Would every item in the population have an equal chance of being selected?*
- *How will you apply this decision consistently across groups?*

Mathematical Practices in the Project

This project naturally embeds multiple Mathematical Practices throughout the unit:

- **MP.1: Make sense of problems and persevere in solving them** — Students navigate multi-step, open-ended data investigations, make decisions about how to proceed when data is messy or incomplete, and revise their approaches as new questions or challenges arise.
  MP.2: Reason abstractly and quantitatively — Students interpret proportions, means, and MADs in context, connecting numerical results back to patterns and representations in media.
- **MP.3: Construct viable arguments and critique the reasoning of others** — Students justify sampling choices, defend conclusions, and revise claims based on peer feedback and evidence.
  MP.4: Model with mathematics — Students use samples, visual displays, and summary statistics as models for understanding larger populations.
- **MP.5: Use appropriate tools strategically** — Students select and use tools such as random number generators, calculators, and data spreadsheets (e.g., filtering, sorting, and organizing data) to support efficient analysis and informed decision-making.
- **MP.6: Attend to precision** — Students refine statistical questions, clearly define populations and samples, and use careful, precise language when interpreting results and communicating conclusions.

Culturally Responsive Math Teaching (CRMT) Connections

The project is designed to center students' identities, interests, and community knowledge without positioning students as representatives of any group. Opportunities for CRMT include:

- Allowing students to choose questions connected to identities and forms of media that matter to them.
- Valuing multiple interpretations and emphasizing that data describes patterns, not people.
- Encouraging careful language that avoids stereotypes or overgeneralizations.

Teachers can support this work by:

- Framing discussions around curiosity and impact rather than judgment.
- Asking whose stories may be missing from a dataset or a sampling method.
- Validating student insights that connect data to lived experiences.

Throughout the unit, students are encouraged to see statistics as a tool for understanding systems and patterns, not for labeling or ranking groups.

ELA Connections and Extensions

The final presentation aligns closely with ELA skills related to argumentation, evidence, and audience. Students must:

- Clearly state a claim
- Support it with multiple forms of evidence (graphs and numerical summaries)
- Explain limitations and uncertainty
- Consider why their findings matter

Possible extensions include:

- Writing a short reflection or commentary explaining how their thinking changed.
- Revising conclusions for a specific audience (e.g., media creators, consumers, or peers).
- Connecting findings to themes of identity, belonging, or representation explored in ELA texts.

Project Assessment Checklist (Sample)

Student projects will vary. Look for evidence of the following:

Data Representation

- ☐ Dot plots are accurate, clearly labeled, and appropriate for the data.
- ☐ Graphs make it possible to compare the distributions visually.
- ☐ Axes, titles, and group labels support interpretation.

Quantitative Reasoning

- ☐ Means and MADs are calculated correctly for each group.
- ☐ The difference between means is clearly identified.
- ☐ Students explicitly compare the difference in means to the MAD.
- ☐ Reasoning reflects an understanding of variability, not just averages.

Informal Inference

- ☐ Conclusions are supported by multiple pieces of evidence (visual + numerical).
- ☐ Claims use careful, data-informed language (e.g., tends to, more likely, suggests).
- ☐ Students acknowledge overlap or variability when appropriate.

Impact, Importance, and Recommendations

- ☐ Groups explain why their question mattered to them or to a broader community.
- ☐ Conclusions connect data findings to real-world implications or lived experiences.
- ☐ Groups make recommendations, raise new questions, or describe how the findings could be used.
- ☐ Students reflect on how data can inform understanding, challenge assumptions, or support more equitable decisions.

Communication & Clarity

- ☐ Presentation is organized and easy to follow.
- ☐ Mathematical decisions and conclusions are clearly explained.
- ☐ Visuals and written or spoken explanations work together to communicate ideas.

Sample Project

This sample project reflects the Student Responses from all relevant problems in Unit 7 with a focus on the Unit Project. This is intended to show what an on-track group's work might look like across the unit. Projects will show variation in questions, samples, and conclusions.

Lesson 1: Statistical Question

My group noticed that animated movies often feel more diverse in lead characters than live-action movies.

Our statistical question is: What proportion of lead characters are Asian in live-action movies compared to animated movies?

Lesson 2: Populations and Sampling Plan

Our statistical question that we are investigating is, "What proportion of lead characters are Asian in live-action vs. animated movies?"

a. The population for our two subgroups would be all live-action movies and all animated movies.
b. We would use a sample statistic to determine the typical number of lead characters who are Asian for each of the two groups. We think that thousands or even millions of movies have probably been produced, so collecting data on every single one of them would be impossible.
c. One biased way to collect data would be to look only at the movies that are currently trending on streaming platforms. Those movies are picked by the app, not randomly, and they usually push certain types of movies, so that wouldn't represent all movies in the population.
   Another unrepresentative way would be to select 20 action movies. If we only had one specific genre of movie included in our data set, we would not be able to generalize our claims to include all genres.
d. One unbiased method would be to create a full list of all live-action movies, assign each a number, and use a random number generator to select the movies we will analyze. We could repeat this with all of the animated movies. This gives every movie an equal chance of being selected, which helps us get a representative sample of all animated or live action movies.

Lesson 3: Initial Data Collection

Our group studied the proportion of lead characters who are Asian in all animated and live-action movies.
When we filtered the data for live-action movies, there were 753 movies, so we used a random number generator to select 10 rows at random and recorded the data below. Then we repeated the process with the animated movies, which had 247 movies.

Live Action Movies		
Row # (RNG)	ID	Lead Character Race
496	ID0661	White
45	ID0066	White
666	ID0889	White
474	ID0260	Asian
47	ID0069	White
463	ID0617	Middle Eastern
653	ID0874	Multiracial
402	ID0525	Asian
751	ID0995	Indigenous
111	ID0149	Latinx

Animated Movies		
Row # (RNG)	ID	Lead Character Race
196	ID0779	White
160	ID0632	White
5	ID0013	Black
9	ID0028	White
197	ID0787	White
81	ID0321	Black
108	ID0436	White
100	ID0402	Unknown
15	ID0050	Latinx
135	ID0544	White

Based on this data, there were 2 Asian lead characters out of 10 for live-action movies, and 0 Asian lead characters out of 10 for animated movies.

Lesson 5: Larger and Repeated Samples

a. We increased our sample size to 20 randomly selected movies instead of 10, for each subgroup.

Live Action Movies		
Row # (RNG)	ID	Lead Character Race
496	ID0661	White
45	ID0066	White
666	ID0889	White
474	ID0260	Asian
47	ID0069	White
463	ID0617	Middle Eastern
653	ID0874	Multiracial
402	ID0525	Asian
751	ID0995	Indigenous
111	ID0149	Latinx
448	ID0595	White
624	ID0832	Latinx
630	ID0840	Black
125	ID1067	Multiracial
82	ID0113	White
124	ID0165	White
557	ID0741	White
238	ID315	White
455	ID0604	Black
229	ID304	White

Animated Movies		
Row # (RNG)	ID	Lead Character Race
196	ID0779	White
160	ID0632	White
5	ID0013	Black
9	ID0028	White
197	ID0787	White
81	ID0321	Black
108	ID0436	White
100	ID0402	Unknown
15	ID0050	Latinx
135	ID0544	White
40	ID0148	Latinx
106	ID0425	White
120	ID0496	Latinx
152	ID0609	Latinx
72	ID262	Black
193	ID0774	Middle Eastern
149	ID0592	White
35	ID0117	White
85	ID334	White
182	ID0714	White

Larger sample sizes give us more information about the population, which reduces the impact of unusual or extreme data points. With more data, the averages and percentages we calculate are closer to the true population values, so our predictions are more accurate.

b. Each member of my group repeated the data collection methods to get 5 samples for each subgroup, and each sample had 20 movies. We counted the number of Asian lead characters, shown in the data table below:

Live Action Movies	
Sample	Count of Asian Lead Characters
1	0 / 20
2	4 / 20
3	1 / 20
4	1 / 20
5	0 / 20
6	0 / 20
7	2 / 20
8	3 / 20
9	3 / 20
10	1 / 20
11	1 / 20
12	2 / 20
13	1 / 20
14	3 / 20
15	0 / 20

Animated Movies	
Sample	Count of Asian Lead Characters
1	2 / 20
2	1 / 20
3	1 / 20
4	0 / 20
5	3 / 20
6	0 / 20
7	0 / 20
8	0 / 20
9	0 / 20
10	1 / 20
11	0 / 20
12	2 / 20
13	2 / 20
14	2 / 20
15	1 / 20

Repeated trials show us how much our results change from sample to sample. By looking at multiple samples, we can see patterns and ranges of variation, which helps us understand how confident we should be in our conclusions.

Lesson 7: Representations and Summary Statistics

a. Our group studied the proportion of movies with Asian lead characters across two subgroups: Live Action movies and Animated movies. Each sample included 20 movies, so we set up the count per 20 to find the proportions:

Animated movie samples (15 total):

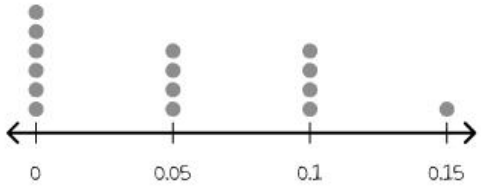
0.00	0.10	0.05	0.00	0.10
0.05	0.05	0.10	0.00	0.00
0.15	0.10	0.00	0.00	0.05

Live Action samples (15 total):

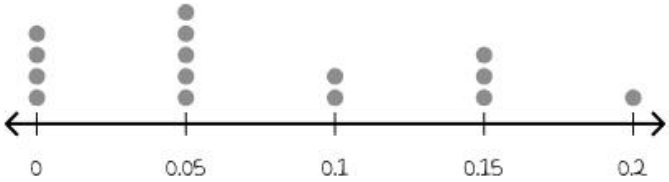
0.10	0.05	0.05	0.05	0.00
0.00	0.15	0.10	0.00	0.15
0.00	0.15	0.05	0.05	0.20

b.

Dot plot for proportion of Animated movies featuring Asian lead characters:



Dot plot for proportion of Live Action movies featuring Asian lead characters:



c. Describe each distribution by determining and recording the following for each dot plot:

i. The mean for Animated movies was 0.05 and the median was 0.05

The mean for Live-Action movies was 0.073 and the median was 0.05

ii. The IQR for Animated movies was 0.1 and the MAD was 0.04.

The IQR for Live-Action movies was 0.1 and the MAD was 0.055.

iii. For Animated movies, the mean is about 0.05 and the median is 0.05, meaning a typical sample has about 5% of movies with Asian leads. Most values are close to 0, and the MAD of 0.04 shows the samples don't vary much.

For Live-Action movies, the mean is about 0.073 and the median is still 0.05, but there are more samples above 0.10. The MAD is larger (about 0.055), which shows more variation between samples.

d. Both groups have the same median, but the Live-Action samples tend to be slightly higher overall, which is shown by the higher mean and more values above 0.10. The variability is also greater for Live Action, meaning the proportion of movies with Asian leads seems to change more from sample to sample. This suggests that Asian leads appear somewhat more often in Live Action movies, even though the difference between the groups is small and there is still quite a lot of overlap.

Lesson 9: Final Presentation and Informal Inference

Our question is: What proportion of lead characters are Asian in live-action movies compared to animated movies? This question is important because movies shape how people see the world and who they see represented in it. Looking at the percentage of movies with Asian leads helps us understand whether different types of movies provide equal representation. Representation matters because when people see characters who look like them in the media, it can affect their sense of belonging and how others view their communities. Comparing animated and live-action movies helps us see whether some types of media are more inclusive than others.

The mean percentage of movies with Asian leads is about 0.05 for animated movies and about 0.073 for live-action movies. The difference in means is approximately 0.023.

The MAD for animated movies is about 0.04, and the MAD for live-action movies is about 0.055. Since the difference in means (0.023) is smaller than both MAD values, the difference between the groups is less than one MAD apart. This suggests that the difference in the sample means is small compared to the typical variability within each group. Based on this, it seems there is not a meaningful difference between the means.

Although live-action movies have a slightly higher mean percentage of Asian leads than animated movies, the difference is small compared to the variability in the data. Because the difference in means is less than one MAD, it is likely that the observed difference could be due to sampling variability rather than a meaningful difference between the populations.

This suggests that families and media consumers should not assume one type of movie is consistently more representative than the other. For producers and decision-makers, the results highlight that representation remains low overall, pointing to an opportunity to intentionally include more Asian leads across both animated and live-action films to better reflect real communities.