

Practice Problems

Math 141

Fall 2020

Problem 1

We are interested in exploring whether the presence (or lack thereof) of a protoplanetary disk around a young star depends on the location of that star in its birth cluster. In this dataset, the **Region** variable gives the location of the stars in the cluster as “east”, “west” and “halo”. And the **Type** variable states disk presence as “disk”, “no disk” and “envelope”. You can either include cases of **Type** “envelope” with the “disk” cases or ignore them. For this problem:

- Perform an exploratory data analysis, choosing an appropriate visualization and table.
- Consider two cases, first, just the “east” and “west” categories in **Region** and second, all three categories in **Region**: “east,”west" and “halo”. By both choosing an appropriate probability model and using a computational method with `infer()`, for each case:
 - Calculate an appropriate point estimate and construct a 95%-confidence interval around your point estimate. For the case of more than two categories, discuss the appropriateness of a confidence interval.
 - Formulate null and alternative hypotheses and construct a hypothesis test, calculate a p-value and draw a conclusion.
 - Discuss the relevant criteria for each approach and whether the data conform to the criteria.

Note: After using the `filter()` function, it may be useful to use `fct_drop()` to ensure any categories with a 0 count are dropped.

Problem 2

We are now interested in exploring whether there are differences in the light emitted from each **Type** of star. We will do this by creating a new *color index* composed of the measurements made in certain bands. To do this create a new variable called **Color** made of the difference of two bands (typically a higher band subtracted from a lower band i.e **Band8-Band9**). You can choose any two bands for this problem.

- Perform an exploratory data analysis, choosing an appropriate visualization and table.
- Consider two cases, first, just the “disk” and “no disk” categories in **Type** and second, all three categories in **Type**: “disk,”no disk" and “envelope”. By both choosing an appropriate probability model and using a computational method with `infer()`, for each case:
 - Calculate an appropriate point estimate and construct a 95%-confidence interval around your point estimate.
 - Formulate null and alternative hypotheses and construct a hypothesis test, calculate a p-value and draw a conclusion.
 - Discuss the relevant criteria for each approach and whether the data conform to the criteria.