

BATTING AVERAGES.

(The attached PDF file has better formatting.)

You can choose other hypotheses from web sites providing sports statistics. You might relate batting averages by year to the age of the player.

- ~ Young players improve each year, so the expected batting average increases.
- ~ Old players deteriorate, so the expected batting average decreases.

Collect batting averages by player and year, form a hypothesis, and test it.

Jacob: What types of F tests can we do?

Rachel: Your F test may compare batting averages in two teams (or two leagues) by age of player. Use the regression analysis explained in this project template to get an optimal regression equation for batting averages for the players on one team or in one League.

- ~ Use players on the starting roster. Players who sit on the bench most of the year have high variability in their batting averages. Exclude also pitchers who play every third or fourth game also have high variability. Use judgment to decide which players to include.
- ~ Use players in their third or later years. This gives at least two previous years.
- ~ Include a player's age as a dummy variable in the regression analysis, assuming that batting averages rise as players gain experience and fall as players get old. Use baseball age: the years since being drafted or since first playing in the major leagues.

Many sports sites show batting averages. You may need an hour or two to convert the data to Excel format. If the conversion is tedious, use just one or two teams.

For the F test, compare players on two teams (or two leagues), old vs young players, good vs poor players, or any other dichotomy.

PRINCIPLES OF PROJECT DESIGN

Take heed: Designing a project takes patience. One is tempted to start with an idea, plan all the associated items, and search for data on the web. In practice, you may not find data that matches exactly your hypothesis. Be flexible: if you use players' ages, see what is available on web sites.

- If the exact data you want is not available, use the closest proxy.
- Other variables may obscure the effects you are trying to test. If you can adjust for other variables, do so. If you cannot adjust for distorting variables, explain the distortion in the write-up, and do the project.

Take heed: Some candidates send emails saying: “I have N data points showing ... My null hypothesis is ... I will use an F test (or some other procedures) to examine if Will this satisfy the student project requirement?”

Answer: We can't say.

- Any topic satisfies the requirements if it is done well.
- Even a good topic fails if you don't use the procedures correctly.

Recommendation: You enjoy the student project more if you design it yourself. If you state the potential problems and do the analysis, you receive credit for the student project.

Illustration: You would like to analyze women's college basketball, but you have only four teams with a few years of data. Do the analysis, and explain why the sparse data make it hard to interpret the results.

Take heed: The SOA prefers that candidates design their own projects. We know that it is hard to collect data in useable format. If you design your own project, we do not grade the student project adversely for sparse data or poor data quality.

Illustration: A candidate uses an F test to compare regression equations for batting averages vs pitching records. The candidate

- Compares outfielders and pitchers in National League teams.
- Develops regression equations using two past years for each statistic.
- Normalizes each statistic to a mean of 1.000.
- Shows that the same regression equation should not be used for both statistics.

In truth, we use F tests to compare the same statistic in two samples. But batting averages and pitching records are different statistics. They have different means, different variances, and different relations to other aspects of the team. Players' batting averages are unrelated to each other, but pitchers' won-loss records are correlated with the team's overall won-loss record.

- We would not give a project template for this F test, since it is not proper.
- If a candidate designs the project, we give credit. We check if he/she understands the degrees of freedom, the constraints, and the expression for the F test.