# Lab #6: Choosing the right statistical approach

An important part of this class is knowing the right approach to use to answer a research question. In this lab you practice seeing a research question, identifying its important aspects, and addressing it using an appropriate statistical method. Below is a summary of the scenarios and statistical approaches we’ve seen so far in the class:

|  |  |  |
| --- | --- | --- |
| Data Scenario | Method(s) | SE |
| One Categorical Variable  (1 sample data) | Confidence interval, one-sample z-test, exact binomial test, randomization test for a single proportion |  |
| One Quantitative Variable  (1 sample data) | Confidence interval, one-sample t-test, randomization test for a single mean |  |
| One Categorical Variable with groups (2 sample data) | Confidence interval, two-sample z-test, randomization test for a difference in proportions |  |
| One Quantitative Variable with groups (2 sample data) | Confidence interval, two-sample t-test, randomization test for a difference in means |  |
| Paired Quantitative data | Confidence interval, paired t-test, randomization test for a single mean |  |
| Quantitative Outcome Variable with a Quantitative Predictor Variable | Regression, correlation coefficient | Coming Soon |

**DIRECTIONS:**

In this lab I would like you will be presented with a research question. I’d like you to decide upon the proper statistical course of action (testing or confidence intervals, and what type of test/interval). Then you can carry out your chosen approach in Minitab, but I’d like you to include the same work that you’d show when doing the method by hand. In some instances this might be as brief as showing that a method’s assumptions aren’t met and then using bootstrapping/randomization in StatKey. In some instances (namely regression and correlation) we’ve only seen how to conduct statistical inference in StatKey, so minimal work can be shown.

There won’t be any step-by-step examples in this lab, but the procedures you’ll need can be found under the “Stat -> Basic Statistics” menu (1-sample t, 2-sample t, paired t, one proportion, two proportions)

**For Questions 1 – 5 use the “GSWarriors2016” data set:**

This dataset contains information for 82 games played during the 2015-16 NBA regular season by the Golden State Warriors. It contains the following variables:

* **Game**: ID number for each game
* **Date**: Date the game was played
* **Location**: Whether the game was home or away
* **Opp**: Opposing team’s name
* **Win**: Whether the game was a win (W) or a loss (L) for Golden State
* **Points**: Number of points scored by Golden State
* **OppPoints**: Number of points scored by the opponent
* **FG**: Number of field goals made by Golden State
* **FGA**: Number of field goals attempted by Golden State
* **FG3**: Number of 3-point shots made by Golden State
* **FG3A**: Number of 3-point shots attempted by Golden State
* **FT**: Number of free throws made by Golden State
* **FTA**: Number of free throws attempted by Golden State
* **Rebounds**: Total number of rebounds by Golden State
* **OffReb**: Number of offensive rebounds by Golden State
* **Assists**: Number of assists by Golden State
* **Steals**: Number of steals by Golden State
* **Blocks**: Number of blocked shots by Golden State
* **Turnovers**: Number of turnovers made by Golden State
* **Fouls**: Number of fouls committed by Golden State
* **OppFG**: Number of field goals made by the opponent
* **OppFGA**: Number of field goals attempted by the opponent
* **OppFG3**: Number of 3-point shots made by the opponent
* **OppFG3A**: Number of 3-point shots attempted by the opponent
* **OppFT**: Number of free throws made by the opponent
* **OppFTA**: Number of free throws attempted by the opponent
* **OppRebounds**: Total number of rebounds by the opponent
* **OppOffReb**: Number of offensive rebounds by the opponent
* **OppAssists**: Number of assists by the opponent
* **OppSteals**: Number of steals by the opponent
* **OppBlocks**: Number of blocked shots by the opponent
* **OppTurnovers**: Number of turnovers made by the opponent
* **OppFouls**: Number of fouls committed by the opponent

## Question #1:

How accurately does the Golden State Warriors team make the free throws they attempt?

|  |  |
| --- | --- |
| Data scenario: | One Categorical Variable |
| Method chosen (include your rationale): | Confidence Interval for p, this is okay because n\*p and n\*(1-p) are large |
| Null and alternative hypotheses (if applicable): | NA |
| Test statistic or Confidence Interval (show your calculation): |  |
| p-value (if applicable): | NA |
| One sentence interpretation: | With 95% confidence we believe the Golden State Warriors team makes between 74.3% and 78.2% of their free throws |

## Question #2:

Looking at the season totals, how much better/worse were Golden State Warriors than their opponents at making free throws during the 2016 season?

|  |  |
| --- | --- |
| Data scenario: | One Categorical Variable with groups |
| Method chosen (include your rationale): | Confidence Interval for difference in proportions, this is okay because n\*p and n\*(1-p) are large for each group |
| Null and alternative hypotheses (if applicable): | NA |
| Test statistic or Confidence Interval (show your calculation): |  |
| p-value (if applicable): | NA |
| One sentence interpretation: | We estimate the Warriors make between 2.3% fewer and 3.3% more of their free throw attempts than their opponents |

## Question #3:

Some basketball fans theorize that referees are more likely to make calls in favor of the home team, did the Golden State Warriors attempt more free throws at home than they did on the road?

|  |  |
| --- | --- |
| Data scenario: | One Quantitative Variable with groups |
| Method chosen (include your rationale): | t-test, this is valid because the data in each group appears to be reasonably symmetric |
| Null and alternative hypotheses (if applicable): |  |
| Test statistic or Confidence Interval (show your calculation): |  |
| p-value (if applicable): | Using a t-distribution with 79 df, p = 0.49 |
| One sentence interpretation: | We do not have evidence to believe that Golden State attempts more free throws at home than they do on the road |

## Question #4:

At times the Golden State Warriors have been criticized as being overly reliant on 3-pt shooting. Is the number of 3 point attempts taken by the team predictive of the number of points the team ended up scoring?

|  |  |
| --- | --- |
| Data scenario: | Correlation and Regression |
| Method chosen (include your rationale): | Regression (we are focused more on prediction than the relationship) |
| Null and alternative hypotheses (if applicable): | H0: B = 0  HA: B =\= 0 |
| Test statistic or Confidence Interval (show your calculation): | b = 0 |
| p-value (if applicable): | Approx. 0.0028 |
| One sentence interpretation: | The number of three point attempts taken by Golden State is predictive of how many points they will score. More 3-pt attempts is predictive of higher scoring. |

## Question #5:

Do the Golden State Warriors attempt more 3 point shots than their opponent in each game?

|  |  |
| --- | --- |
| Data scenario: | Paired Quantitative Data |
| Method chosen (include your rationale): | Paired t-test, this is because the data are paired and each game represents a case |
| Null and alternative hypotheses (if applicable): |  |
| Test statistic or Confidence Interval (show your calculation): |  |
| p-value (if applicable): | Approx. 0 |
| One sentence interpretation: | We have very strong evidence that Golden State takes more three point attempts than their opponents |

**For Questions 6 – 10 use the “HollywoodMovies” data set:**

This dataset contains information for 970 movies released by Hollywood between 2007 and 2013. It contains the following variables:

* **Movie**: Title of the movie
* **LeadStudio**: Studio that released the movie
* **RottenTomatoes**: Rotten Tomatoes rating (from critics)
* **AudienceScore**: Rotten Tomatoes rating (from the audience)
* **Story**: Category of the movies general theme
* **Genre**: One of 14 possible genres
* **TheatersOpenWeek**: Number of theaters the movie was in screened in on opening weekend
* **OpeningWeekend**: Gross revenue on opening weekend
* **BOAverageOpenWeek**: Average box office income per theater on opening weekend
* **DomesticGross**: Gross income for domestic viewers (in millions)
* **ForeignGross**: Gross income for foreign viewers (in millions)
* **WorldGross**: Gross income for all viewers (in millions)
* **Budget**: Production budget (in millions)
* **Profitability**: WorldGross as a percentage of Budget
* **OpenProfit**: Percentage of budget earned on opening weekend
* **Year**: Year the movie was released

## Question #6:

Do movies in the Action genre have larger budgets than the movies in all of the other genres?

|  |  |
| --- | --- |
| Data scenario: | One Quantitative Variable with groups |
| Method chosen (include your rationale): | Two-sample t-test, while the data are highly right skewed the sample size is large in each group |
| Null and alternative hypotheses (if applicable): |  |
| Test statistic or Confidence Interval (show your calculation): | T = 6.85 with 207 df |
| p-value (if applicable): | p-value of nearly 0 |
| One sentence interpretation: | With a p-value of nearly 0, there is extremely strong evidence that Action movies have larger budgets than the other genres. |

## Question #7:

On average, how much of a movie’s profit is expected to be made on opening weekend?

|  |  |
| --- | --- |
| Data scenario: | One Quantitative Variable |
| Method chosen (include your rationale): | Confidence interval using the t-distribution because n is large |
| Null and alternative hypotheses (if applicable): | NA |
| Test statistic or Confidence Interval (show your calculation): | (51.75, 72.70) |
| p-value (if applicable): | NA |
| One sentence interpretation: | We are 95% confidence that movies will on average make between 51.75% and 72.70% of their profit on opening weekend. |

## Question #8:

Do Rotten Tomatoes critics tend to rate movies lower than the audience does?

|  |  |
| --- | --- |
| Data scenario: | Paired Quantitative Data |
| Method chosen (include your rationale): | Paired t-test |
| Null and alternative hypotheses (if applicable): |  |
| Test statistic or Confidence Interval (show your calculation): |  |
| p-value (if applicable): | Nearly 0 |
| One sentence interpretation: | There is overwhelming evidence that critics tend to rate movies lower than the audience does. |

## Question #9:

Which Studio is more focused on Animation genre films, Disney or Paramount? (Hint: You might want to use the “Subset Worksheet” option under the “Data” menu to subset the data to contain only Disney and Paramount films)

|  |  |
| --- | --- |
| Data scenario: | One Categorical Variable with groups |
| Method chosen (include your rationale): | Two-sample difference in proportions test |
| Null and alternative hypotheses (if applicable): |  |
| Test statistic or Confidence Interval (show your calculation): |  |
| p-value (if applicable): | Approx. 0.008 |
| One sentence interpretation: | There is strong evidence that a larger proportion of Disney films are Animation |

## Question #10:

Is the number of theaters that a movie is in screened in during opening week associated with that movie’s overall profitability?

|  |  |
| --- | --- |
| Data scenario: | Correlation and Regression |
| Method chosen (include your rationale): | Correlation (we are more focused on the strength of association than prediction) |
| Null and alternative hypotheses (if applicable): |  |
| Test statistic or Confidence Interval (show your calculation): | The observed correlation is very near zero |
| p-value (if applicable): | The p-value is approximately 0.65 |
| One sentence interpretation: | We conclude there is no association between the number of theaters screening a movie and its overall profitability. |