



Better Business with Basic Statistics: Step by Step from asking the Right Question to Interpreting the Answer

Selecting NHL Teams by Linear Regression Analytics with SAS

Tim Trussell, SAS Canada

**THE
POWER
TO KNOW[®]**

Step By Step

- Step 1: Asking the Right Questions
- Step 2: Know the Right Tools
- Step 3: Check the Assumptions
- Step 4: Point and Click
- Step 5: Interpret the Results

Disclaimer



Step 1: Asking the Right Questions

1. Is there value in knowing if two groups are related?
2. How can I tell if one group is better than another?
3. If I change one attribute, can I expect a change in a separate attribute?
4. How do I use what I know to predict what I want to know?
5. How can I test whether change is for the better?

A fun application



- Is there any relation between where a player is from and excellence at their position?
- Do players from Canada outperform players from around the World?
- If a player gets more penalties are they likely to get better/worse results in other categories?
- What factors can tell me I want a player on my team?
- How can I test an experiment to improve performance?

Step 2: Know the Right Tools

Chi Square

- Is there any relation between where a player is from and excellence at their position?
- Do players from Canada out perform players from around the World?
- If a player gets more penalties are they likely to get better/worse results in other categories?
- What factors can tell me I want a player on my team?
- How can I test an experiment to improve performance?

Linear Regression

Correlation

ANOVA

Odds Ratio

Step 2: Know the Right Tools

Chi Square



- Is there any relation between where a player is from and excellence at their position?
- Do players from Canada out perform players from around the World?
- If a player gets more penalties are they likely to get better/worse results in other categories?
- What factors can tell me I want a player on my team?
- How can I test an experiment to improve performance?

Linear Regression

ANOVA

Correlation

Odds Ratio

Step 2: Know the Right Tools

Chi Square

- Is there any relation between where a player is from and excellence at their position?
- Do players from Canada out perform players from around the World?
- If a player gets more penalties are they likely to get better/worse results in other categories?
- What factors can tell me I want a player on my team?
- How can I test an experiment to improve performance?

Linear Regression

ANOVA

Correlation

Odds Ratio



Step 2: Know the Right Tools

Chi Square

- Is there any relation between where a player is from and excellence at their position?
- Do players from Canada out perform players from around the World?
- If a player gets more penalties are they likely to get better/worse results in other categories?
- What factors can tell me I want a player on my team?
- How can I test an experiment to improve performance?

Linear Regression

Correlation

ANOVA

Odds Ratio

Step 2: Know the Right Tools

Chi Square

- Is there any relation between where a player is from and excellence at their position?
- Do players from Canada out perform players from around the World?
- If a player gets more penalties are they likely to get better/worse results in other categories?
- What factors can tell me I want a player on my team?
- How can I test an experiment to improve performance?

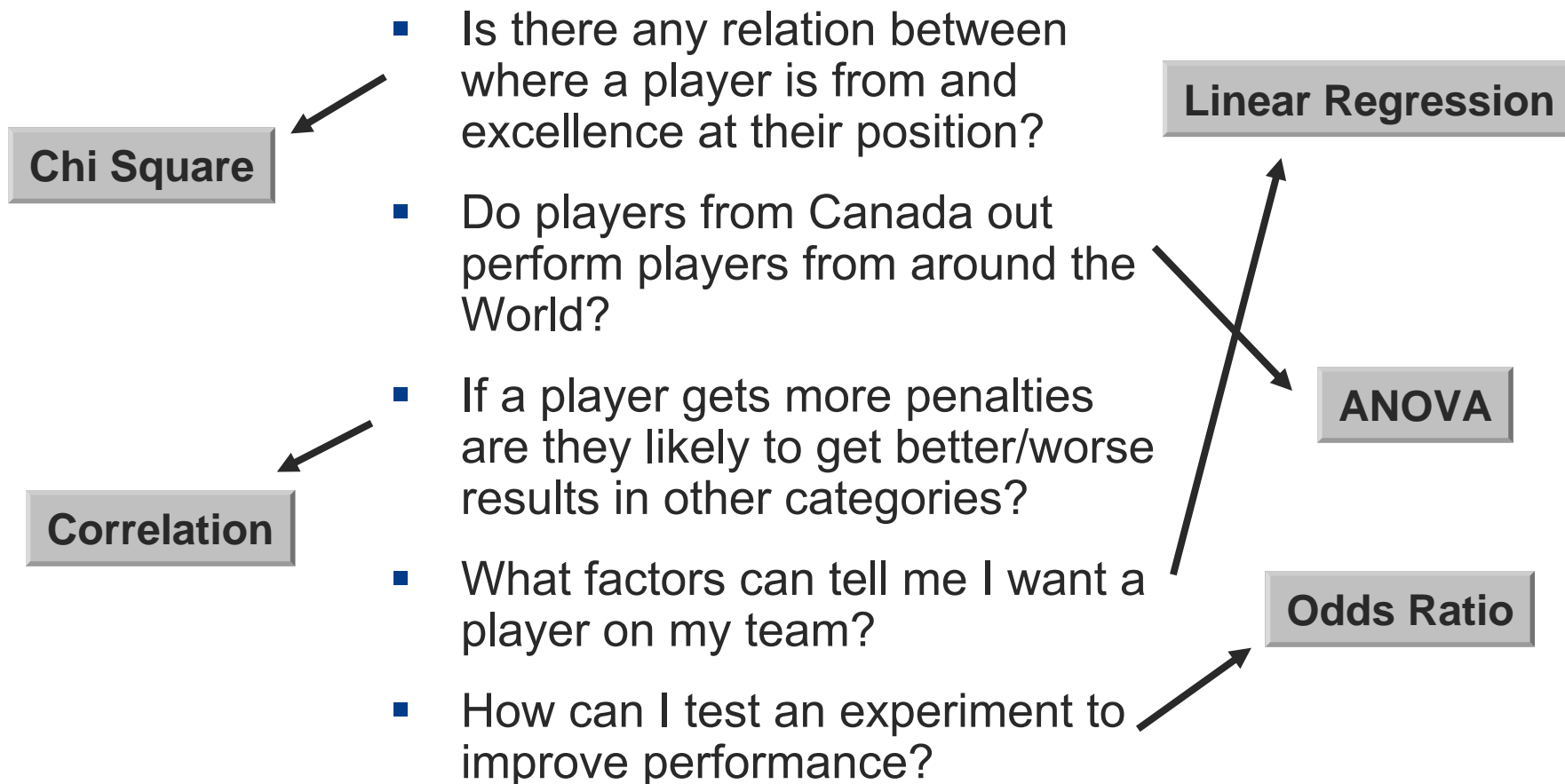
Correlation

Linear Regression

ANOVA

Odds Ratio

Step 2: Know the Right Tools



Chi Square

- When to use: When comparing if there is an association between categorical groups
- H_0 : There is no association between the groups (they are independent)
- Considerations:
 - Does not perform when expected for an individual cell is less than 5.

ANOVA

- When to use: Testing for differences in the outcome depending on group involvement.
- H_0 : There is no significant difference between the groups on the outcome variable.
- Considerations:
 - Even with an significant or insignificant test, a post-hoc test should be performed to look at individual comparisons.

Correlation (Pearson)

- When to use: Descriptive look at the independence/dependence of the target variable
- H_0 : The two variables are independent of one another.
- Considerations:
 - Linear relationship

Regression

- When to use: To determine the effect of predictor variables on the dependant variables while controlling for the effect of other variables
- H_0 : The value of the dependant variable is completely independent of the predictor variables.
- Considerations:
 - Normality of Residuals
 - Too many variables (min 1/10 of respondents)
 - Try variable selection techniques
 - Collinearity
 - Linear Relationships
 - Categorical inputs

Odds Ratio

- When to use: Testing impact of a binary influencer in predicting a binary target
- H_0 : There is equal odds of being in the target group with or without the influencer
- Considerations:
 - 2 x 2
 - Order is important