WSABI Summer Research Lab **Summer 2025**

Lab 7: Significance and P-Values

Instructor: Jonathan Pipping *Authors: RB, JP*

7.1 Permutation Test of Independence

In today's lecture, you learned how to set up a permutation test to establish a null distribution for a test statistic. This resampling method involves shuffling the labels of the data and recomputing the test statistic for each permutation. This is a powerful way to establish the null distribution of a test statistic without making any distributional assumptions.

7.1.1 Data

The Olympic diving data is stored in `07_diving.csv` and contains the following variables:

- `Event`: the event that the diver competed in.
- `Diver`: the diver's name
- `Country`: the nationality of the diver
- `Rank`: the diver's rank in the event
- `DiveNo`: the dive number
- `Difficulty`: the difficulty of the dive
- `Judge`: the judge's name
- `JudgeCountry`: the nationality of the judge
- `JScore`: the judge's score for the dive

You will use this data to test whether judges' scores are independent of the diver's country of origin (i.e. whether judges are biased towards divers from their own country).

7.1.2 Your Task

Replicate the permutation tests from slide 19 of Lecture 7. Which judges exhibit evidence of a nationality bias?

7.2 Parametric Inference

7.2.1 Time Through the Order Penalty

Recall the models from our Time Through the Order analysis in Lecture 6.

Model 1:

$$y_i = \beta_1 + \beta_2 \cdot \mathbb{1}\{t_i \geq 2TTO\} + \beta_3 \cdot \mathbb{1}\{t_i \geq 3TTO\} + \beta_{BQ} \cdot BQ_i + \beta_{PQ} \cdot PQ_i + \beta_{\text{hand}} \cdot \text{hand}_i + \beta_{\text{home}} \cdot \text{home}_i + \epsilon_i$$

When we fit this model in `R`, we observed $\widehat{\beta}_2 = 0.013$ and $\widehat{\beta}_3 = 0.0054$, and plotted our predicted wOBA values against batter sequence number in Figure 7.1.

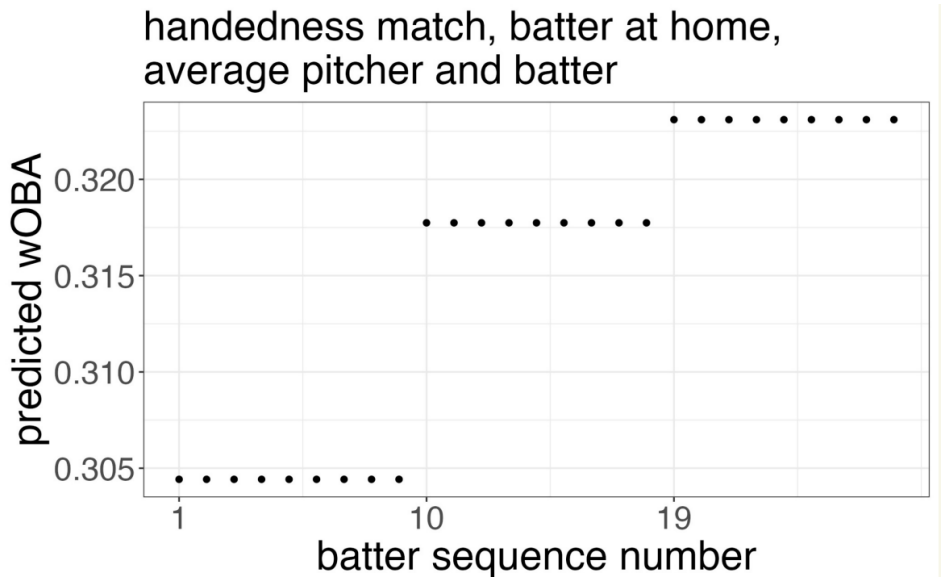

Figure 7.1: Predicted wOBA values against batter sequence number for Model 1.

Model 2:

$$y_i = \beta_0 + \beta_1 t_i + \beta_2 \cdot \mathbb{1}\{t_i \geq 2TTO\} + \beta_3 \cdot \mathbb{1}\{t_i \geq 3TTO\} + \beta_{BQ} \cdot BQ_i + \beta_{PQ} \cdot PQ_i + \beta_{\text{hand}} \cdot \text{hand}_i + \beta_{\text{home}} \cdot \text{home}_i + \epsilon_i$$

When we fit this model in `R`, we observed $\widehat{\beta}_1 = 0.0016$, $\widehat{\beta}_2 = -0.0015$, $\widehat{\beta}_3 = -0.0083$, and plotted our predicted wOBA values against batter sequence number in Figure 7.2.

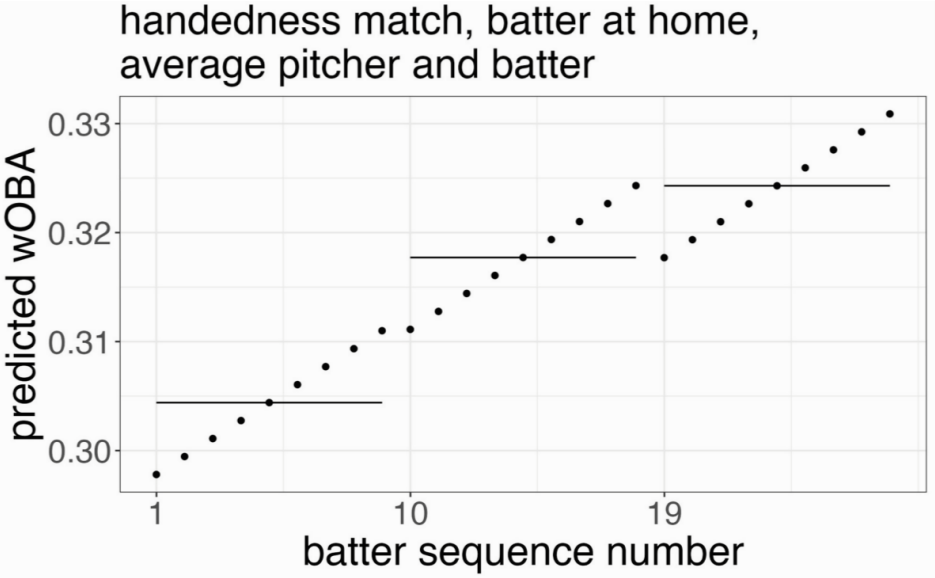

Figure 7.2: Predicted wOBA values against batter sequence number for Model 2.

7.2.2 Testing Significance in Regression Coefficients

In each of these models for the MLB data, we got **point estimates** for our coefficients, which represented a single "best guess" of the parameter values according to the model. The thing is, **ANY** model fit to the same data would also give us a set of estimated coefficients. How do we know if our model is any good?

A more refined way to ask this question is to ask "are the values of the estimated coefficients due to a **real** trend in the data, or just random chance (or **noise**)?" . Since we're interested if these coefficients are zero or not, we can specify our question by asking "is there enough evidence in the data to conclude that the **true** coefficients are very-likely non-zero?" . These questions underly **parametric inference**, the process of making inference about parameters of a model based on some distributional assumptions on the data.

In `R`, regression models assume that the error terms are normally distributed with mean 0 and unknown variance σ^2. Mathematically, we write this as:

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Under this assumption, we have the following theorem:

Theorem 7.1. *Assume that $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ for all i. Let β_j be the true value of the j^{th} parameter in the model and $\widehat{\beta}_j$ be our estimate for that parameter. Letting $\widehat{\sigma}_j$ be the estimated standard error of $\widehat{\beta}_j$, the test statistic*

$$TS = \frac{\widehat{\beta}_j - \beta_j}{\widehat{\sigma}_j}$$

follows a t-distribution with $n - p$ degrees of freedom, where n is the number of observations and p is the number of parameters in the model. Mathematically, we would say that

$$TS \sim t_{n-p}$$

When R fits a regression model, it will also estimate the standard error of each $\widehat{\beta}_j$ and compute the test statistic for the test that the coefficient is zero. Symbolically, this is written as

$$H_0 : \beta_j = 0 \quad \text{vs} \quad H_1 : \beta_j \neq 0$$

and the associated test statistic is

$$TS = \frac{\widehat{\beta}_j}{\widehat{\sigma}_j}$$

From Theorem 7.1, we know that $TS \sim t_{n-p}$. Then the p-value for the test that the coefficient is zero as

$$\begin{aligned} p &= \mathbb{P}_{H_0}(|t_{n-p}| \geq |TS|) \\ &= 2 \cdot \mathbb{P}_{H_0}(t_{n-p} \geq |TS|) \end{aligned}$$

where t_{n-p} is the t-distribution with $n - p$ degrees of freedom. We can get this output for a model fit using the R code below.

```
# fit model
model <- lm(y ~ x, data = data)
# get test statistics and p-values
summary(model)
```

7.2.3 Your Task

1. Fit Models 1 and 2 from Section 7.2.1 in R using the `lm()` function.
2. Get the model summaries using the `summary()` function.
3. Interpret the p-values for the estimated coefficients in each model.
4. Is pitcher decline from one time through the order to the next significant? Explain.