The Language of Sampling

Lecture 6
Sections 2.1-2.4

Robb T. Koether

Hampden-Sydney College

Fri, Jan 27, 2012

Outline

1. Two-Sided Tests
2. Parameters and Statistics
3. Bias
4. Types of Bias
5. Assignment

Review Quiz

Example (Review Quiz)

1. The results of a hypothesis test were found to be statistically significant at the 0.05 significance level. We may conclude that
   (a) The results were also significant at the 0.01 level.
   (b) The results were also significant at the 0.10 level.
   (c) The *p*-value was greater than 0.05.
   (d) There is not enough information to tell.

Review Quiz

Example (Review Quiz)

2. If the significance level of a test is $\alpha = 0.10$ and the *p*-value of the observation is 0.05, then
   (a) H_0 should be accepted.
   (b) H_0 should be rejected.
   (c) The sample size is too small.
   (d) There is not enough information to decide.

Review Quiz Answers

Example (Review Quiz Answers)

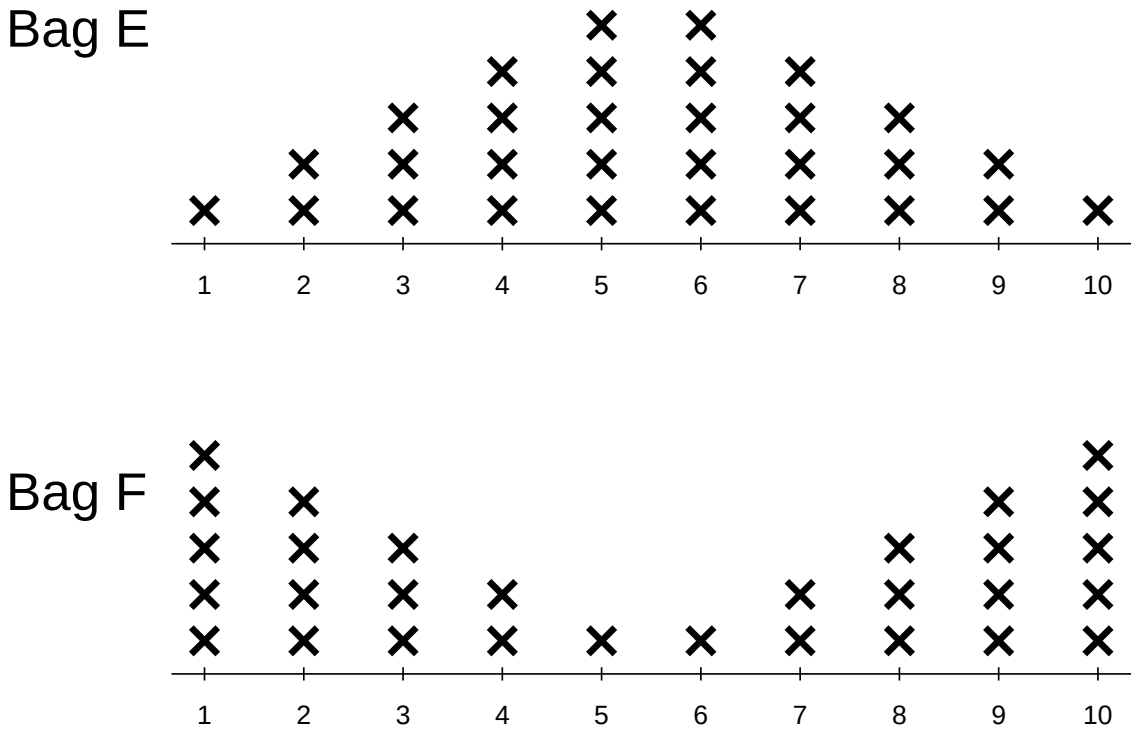
1. (b) The results were also significant at the 0.10 level.
2. (b) H_0 should be rejected.

Outline

1. Two-Sided Tests
2. Parameters and Statistics
3. Bias
4. Types of Bias
5. Assignment

A Two-Sided Test

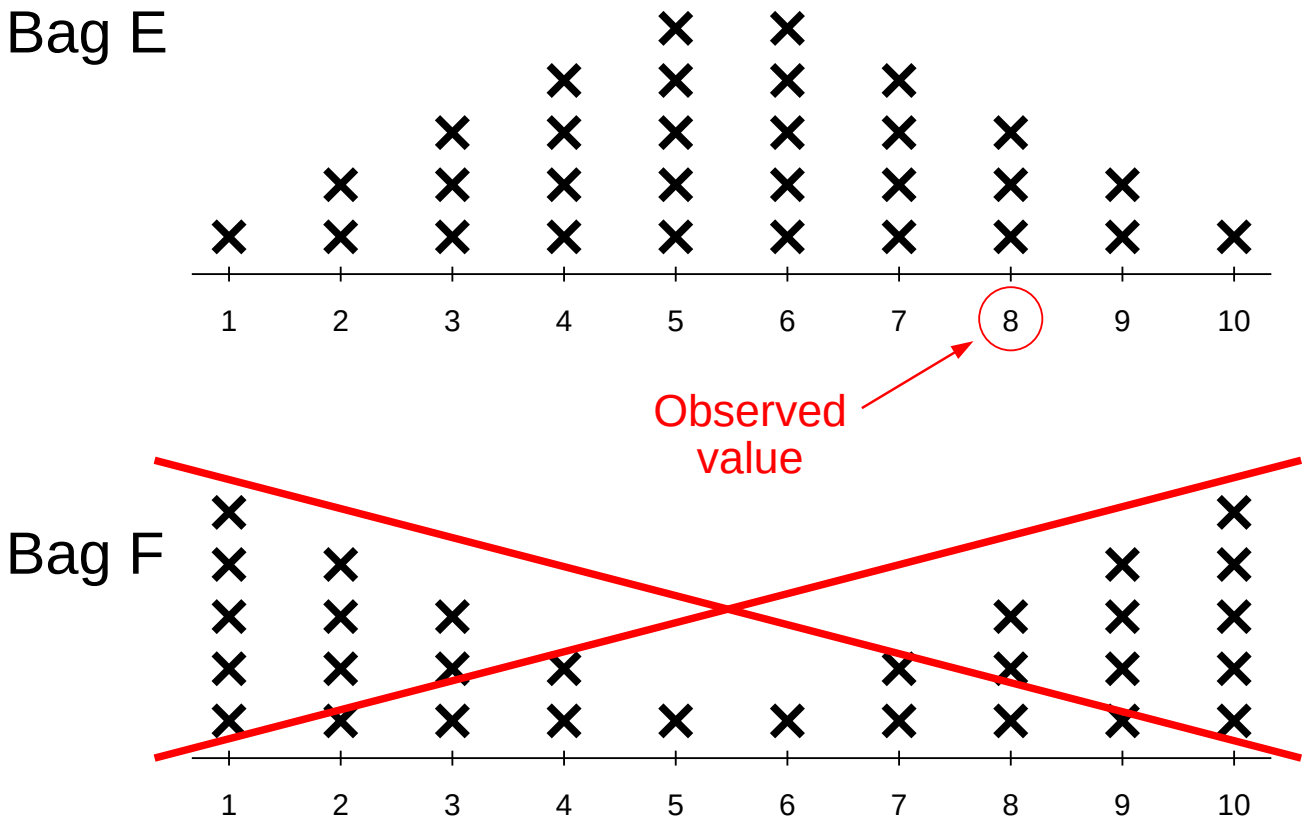
- Now consider Bag E and Bag F.



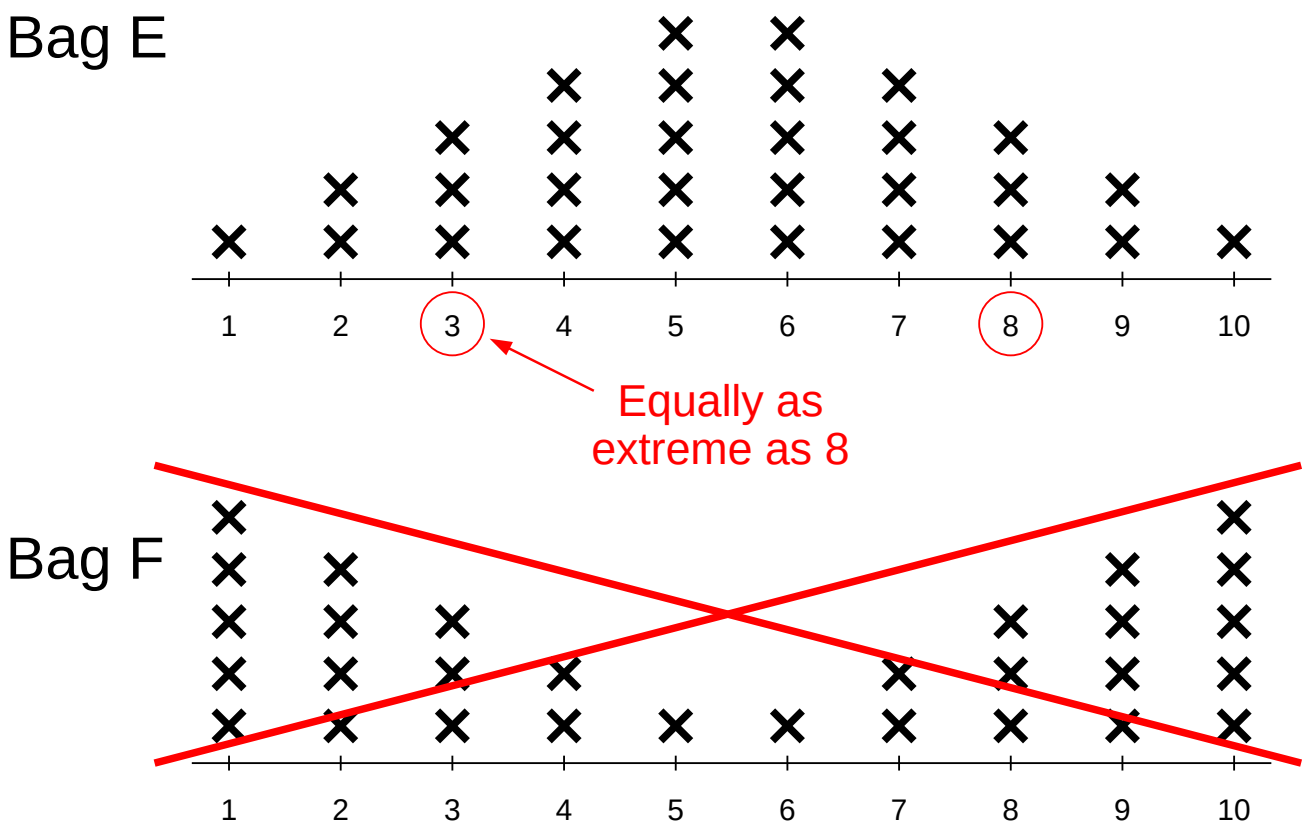
A Two-Sided Test

- If the selected token is worth \$8, what is the *p*-value?
- First, what is the direction of extreme?
- Next, which values are at least as extreme as 8?

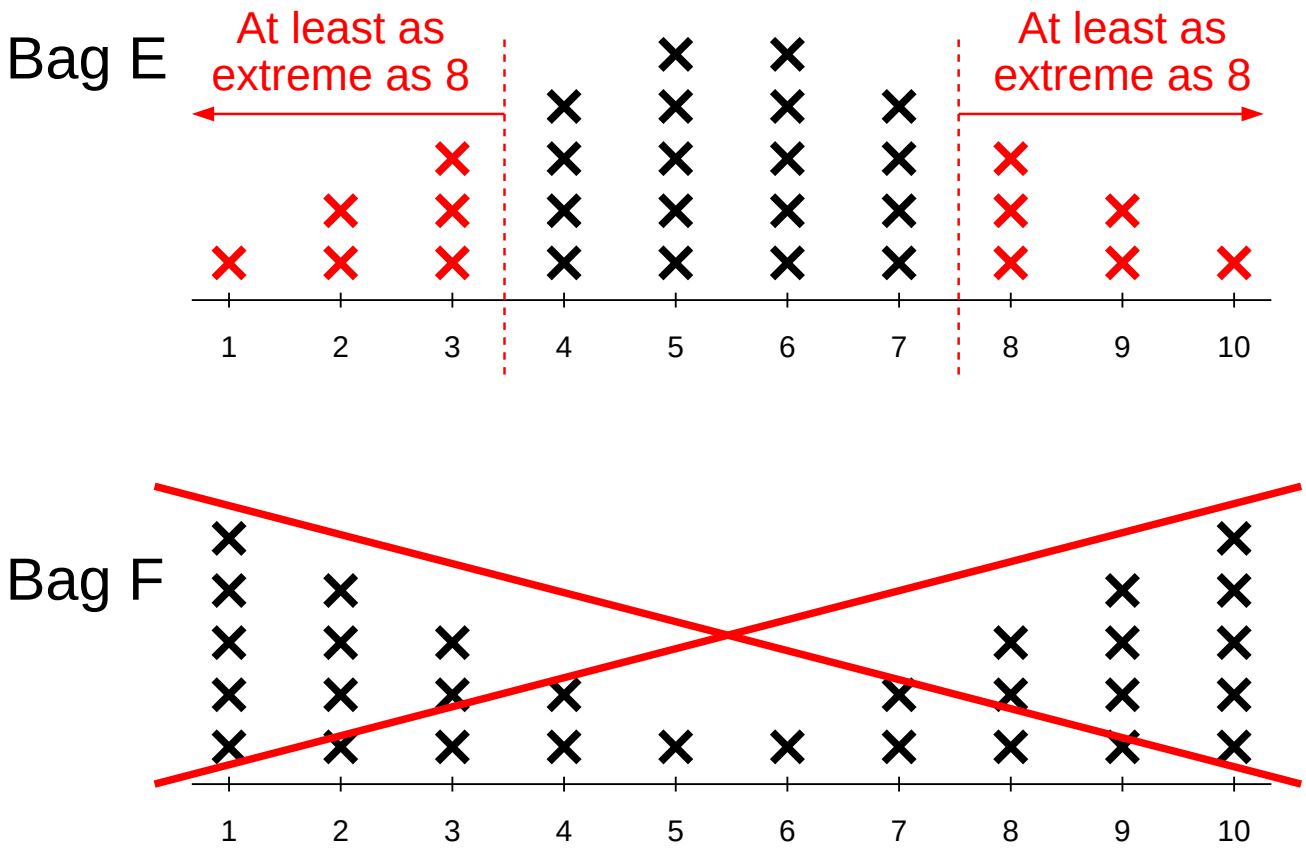
A Two-Sided Test



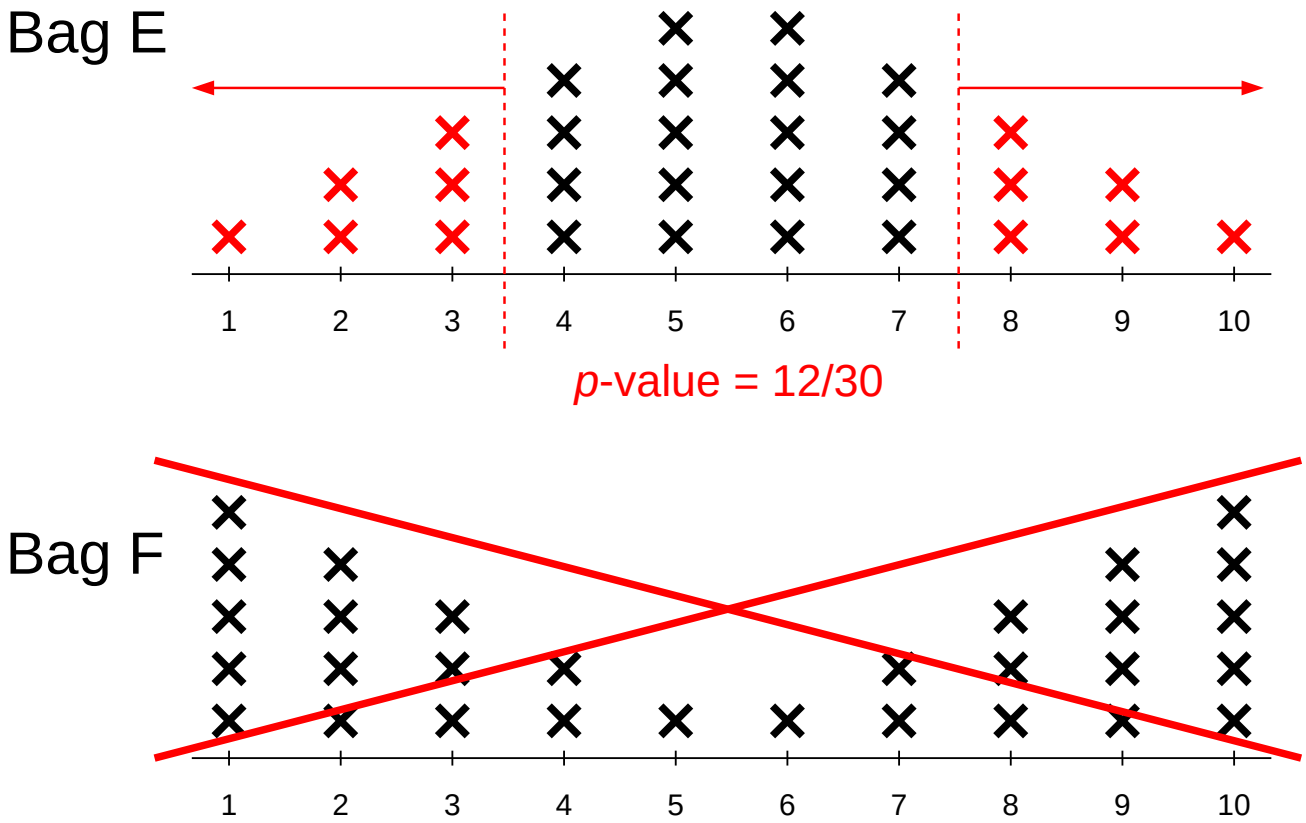
A Two-Sided Test



A Two-Sided Test



A Two-Sided Test



The *p*-value

- If the selected token is worth \$1, what is the *p*-value?
- If the selected token is worth \$5, what is the *p*-value?

Shortcut

- A shortcut:
  - In a two-sided test, if the null distribution is symmetric, then you can compute the probability in one direction, and then double it to get the *p*-value.

Outline

1. Two-Sided Tests
2. Parameters and Statistics
3. Bias
4. Types of Bias
5. Assignment

The Language of Sampling

Definition (Population size)

The population size is the number of members in the population. It is denoted by the symbol N.

Definition (Sample size)

The sample size is the number of members in the sample. It is denoted by the symbol n.

Definition (Unit or subject)

A unit or subject is a single member of the population.

The Language of Sampling

Definition (Parameter)

A parameter is a numerical characteristic of the *population*.

Definition (Statistic)

A statistic is a numerical characteristic of a *sample*.

The Language of Sampling

- The purpose of a statistic is to estimate the corresponding parameter.
- For example,
  - The average salary of a sample is used to estimate the average salary of the population.
  - The proportion of a sample who approve of President Obama's performance is used to estimate the proportion of the population who approve of his performance.

The Language of Sampling

Definition (Variable)

A variable is a characteristic of *individual members* of the population.

- The data consist of the observed values of the variable.
- Those values are combined as either an average or a proportion to produce a statistic.

The Language of Sampling

Definition (Sampling Error)

The sampling error is the difference between the value of a statistic and the value of the parameter it was meant to estimate.

Parameters and Statistics

Example (Averages and Proportions)

- In the coin-tossing experiment, what was the variable?
- What was the data?
- What was the statistic?
- What was the population?
- Was there sampling error?

Parameters and Statistics

- For numerical data, we usually use the average of the values in the sample.
- For non-numerical data, we usually use the proportion of observations in a specific category.

Outline

1. Two-Sided Tests
2. Parameters and Statistics
3. Bias
4. Types of Bias
5. Assignment

Random vs. Representative

Definition (Random sample)

A random sample is a sample that is selected using a procedure at least one step of which is left to chance.

Definition (Representative sample)

A representative sample is a sample that has all the same characteristics (except size) as the population from which it was taken.

Biased Samples

Definition (Biased sampling method)

A sampling method is biased if it *systematically* produces a sample whose characteristics differ from those of the population.

Definition (Biased sample)

A sample is biased if was selected by a biased sampling method.

Why Sample?

- The purpose of randomness is to prevent bias.
- For random samples, the sampling error gets smaller as the sample size gets larger.
- (We will study that phenomenon in Chapter 8.)
- In other words, if a random sample is large enough, then we can be reasonably assured that it is representative.

Outline

1. Two-Sided Tests
2. Parameters and Statistics
3. Bias
4. Types of Bias
5. Assignment

Selection Bias

Definition (Selection bias)

A sampling method exhibits selection bias if some units in the population are more likely to be selected than others.

Selection Bias

Example (Telephone Survey)

- A telephone survey is done using random-digit dialing.
- This could be an example of selection bias. How?

Nonresponse Bias

Definition (Nonresponse bias)

A sampling method exhibits nonresponse bias if some subjects that were selected to be in the sample choose not to participate.

Nonresponse Bias

Example (E-mailed Survey)

- A list of survey questions is e-mailed to all students on campus, asking them to reply with their responses.
- This could be an example of nonresponse bias. How?

Response Bias

Definition (Response bias)

A sampling method exhibits response bias if the subjects give what they perceive to be the desired response rather than the true response.

Response Bias

Example (Illegal-Drug Survey)

- A group of students is selected at random.
- They are asked how many times a week they use illegal drugs.
- This could be an example of response bias. How?

Experimenter Bias

Definition (Experimenter Bias)

A sampling method exhibits experimenter bias if the observer records the desired values rather than the true values. (Sec. 3.5, p. 176)]

Experimenter Bias

Example (Website survey)

- A professor teaches two sections of statistics.
- He teaches one section by the traditional method.
- In the other section, he tries a new teaching method.
- At the end of the semester, he compares the average grades in the two sections.
- This could be an example of experimenter bias. How?

Example

Example (Medical Study)

- Researchers select all the patients that are suffering from depression at a large hospital for their study.
- Patients agree to be a part of a study to test the effectiveness of a new treatment.
- Patients join either a treatment group that receives the new treatment and a control group that does not receive the treatment.
- After two weeks, the researchers check each patient to see how well they are doing.
- Which of the following types of bias could be present in this method?
  - Selection bias
  - Non-response bias
  - Response bias
  - Experimenter bias

Outline

1. Two-Sided Tests
2. Parameters and Statistics
3. Bias
4. Types of Bias
5. Assignment

Assignment

Homework

- Read Sections 2.1 - 2.4, pages 83 - 96.
- Let's Do It! 2.1, 2.2, 2.3.
- Page 97, exercises 1 - 12.