The Language of Sampling

Lecture 6
Sections 2.1-2.4

Robb T. Koether

Hampden-Sydney College

Fri, Sep 5, 2008

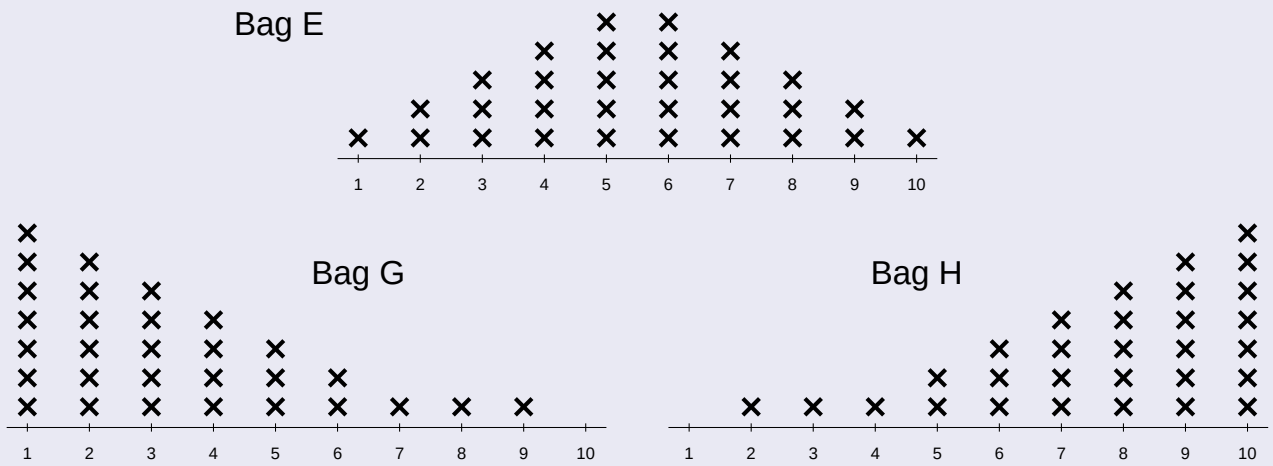
Outline

Homework Review

Exercise 1.16, p. 69.

You are allowed to select just one voucher from the shown bag and must decide whether or not to reject H_0.

- H_0 : The shown bag is ***Bag E***.
- H_1 : The shown bag is ***Bag G*** or ***Bag H***.



Homework Review

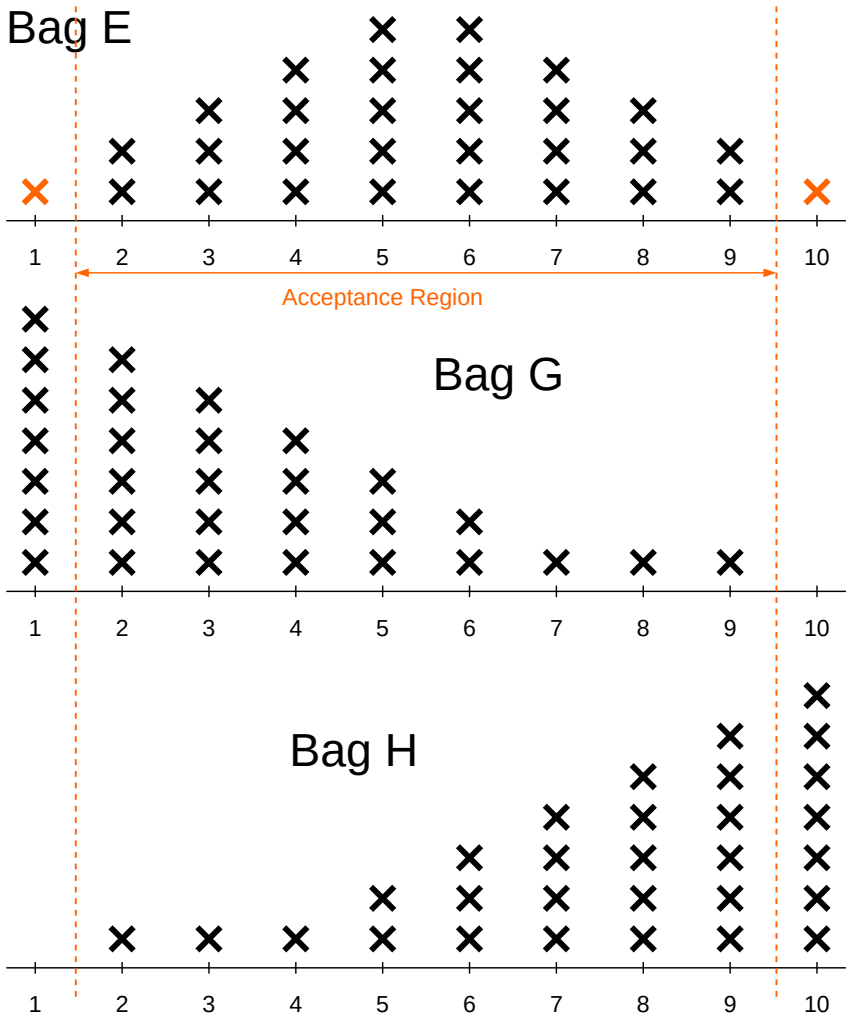
Exercise 1.16, p. 69.

(a) Consider the following decision rule: Reject H_0 if your selected voucher is either $\leq \$1$ or $\geq \$10$. What is the significance level, α, corresponding to this decision rule?

Solution

- This rule includes only 1 voucher (out of 30) in each direction. Therefore, $\alpha = \frac{2}{30}$.

Homework Review



Homework Review

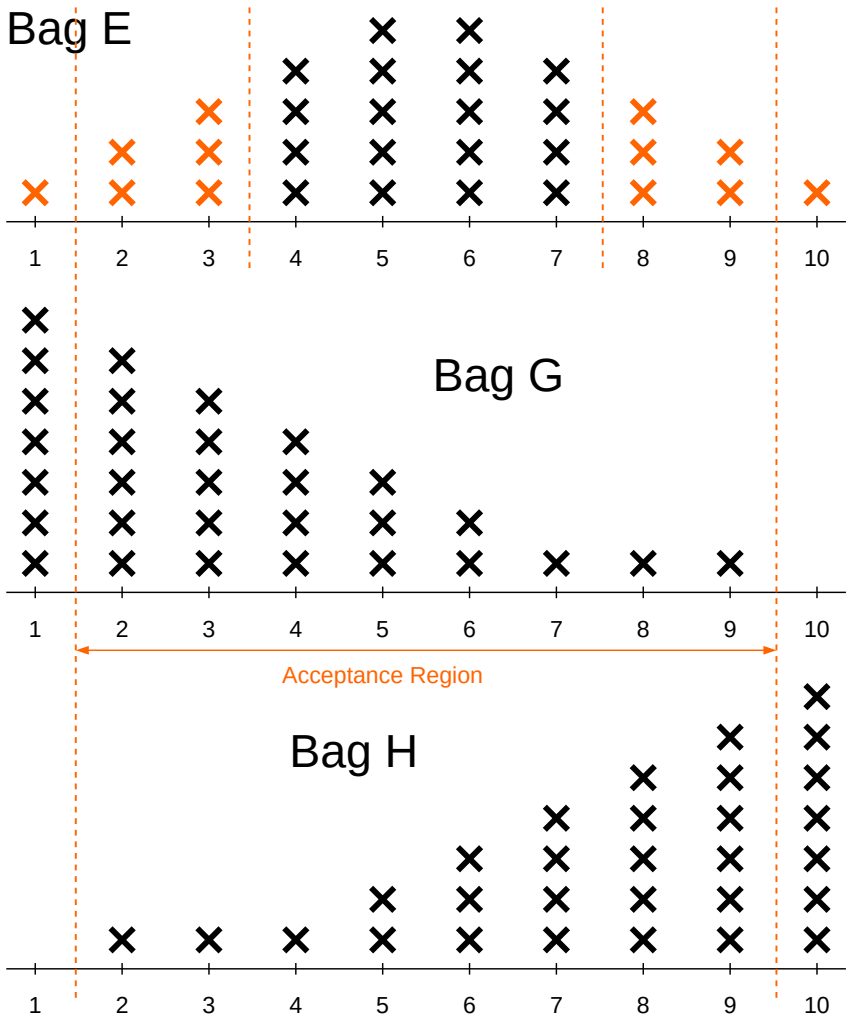
Exercise 1.16, p. 69.

(b) Suppose that the observed voucher value is \$3. Compute the corresponding p-value.

Solution

- The values at least as extreme as 3 are 1, 2, 3 to the left and 8, 9, 10 to the right. The total number of vouchers of those values is 12. Therefore, the p-value is $\frac{12}{30}$.

Homework Review



Homework Review

Exercise 1.16, p. 69.

(c) Recall that β is the chance of failing to reject H_0 when H_1 is true. If H_1 is true, the shown bag is either Bag G or Bag H. To compute the value for β, we need one more piece of information–namely, which of these two alternative bags is the shown bag.

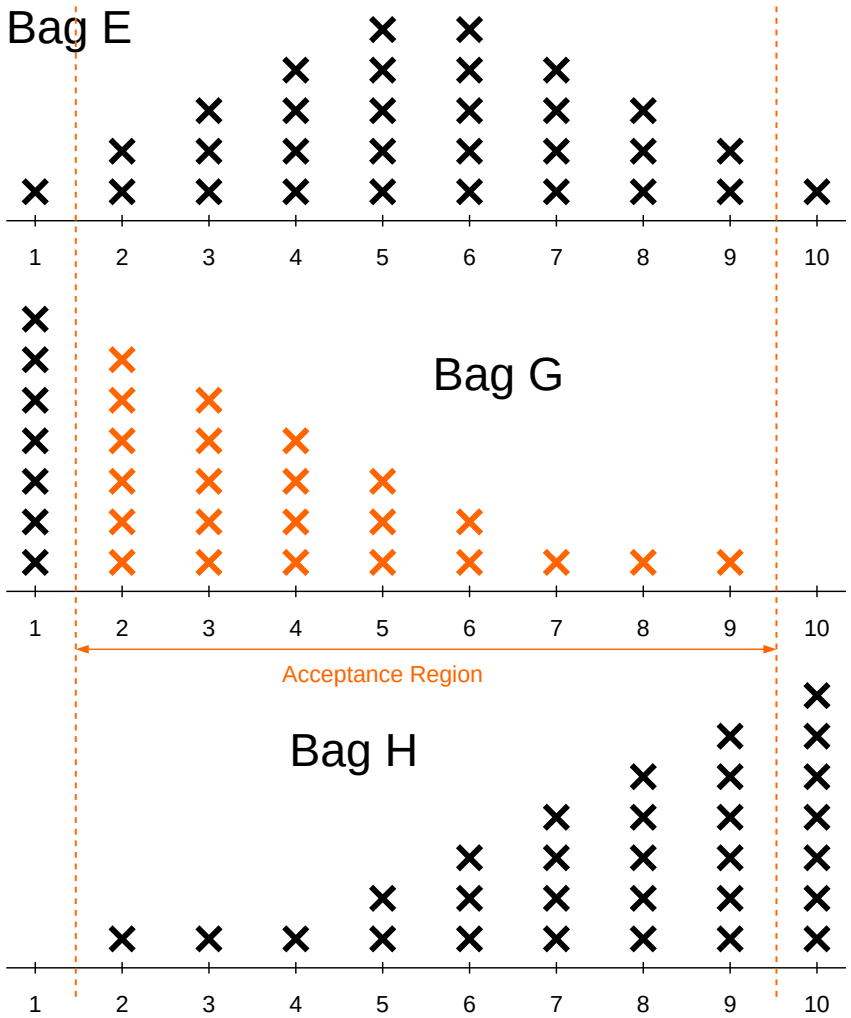
- (i) Compute β assuming that the shown bag is Bag G.
- (ii) Compute β assuming that the shown bag is Bag H.

Solution

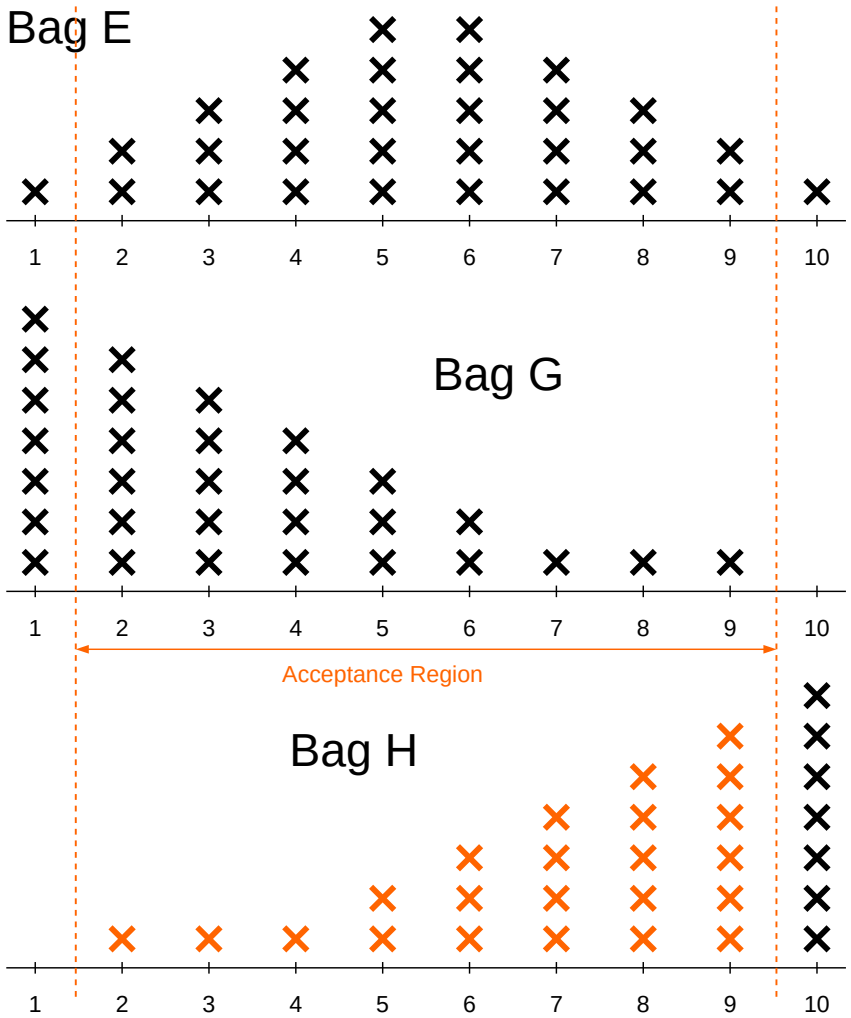
(i) If the shown bag is Bag G, then there are 23 vouchers in the acceptance region. Therefore, $\beta = \frac{23}{30}$.

(ii) If the shown bag is Bag H, then again there are 23 vouchers in the acceptance region. Therefore, $\beta = \frac{23}{30}$.

Homework Review



Homework Review



Why Sample?

- What are the benefits of sampling over studying the entire population?
- What are the disadvantages or risks of sampling over studying the entire population?

The Language of Sampling

Definition (Population size N)

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

Definition (Sample size n)

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 (Variable)

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

Definition (Parameter)

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

Definition (Statistic)

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

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.

Random vs. Representative

Definition (Random sample)

A random sample is a sample that was selected using a procedure, at least one step of which was 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.
- In other words, if a random sample is large enough, then we can be assured that it is representative.

Types of Bias

Definition (Selection bias)

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

Definition (Nonresponse bias)

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

Types of 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.

Definition (Experimenter bias)

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

Example

Example (Telephone survey)

- A telephone survey is done using random-digit dialing.
- Which of the following types of bias are *inherent* in this method?
  - Selection bias
  - Non-response bias
  - Response bias
  - Experimenter bias

Example

Example (Mailed survey)

- Survey forms are mailed out (including e-mail).
- Respondents are asked to send their responses back.
- Which of the following types of bias are *inherent* in this method?
  - Selection bias
  - Non-response bias
  - Response bias
  - Experimenter bias

Example

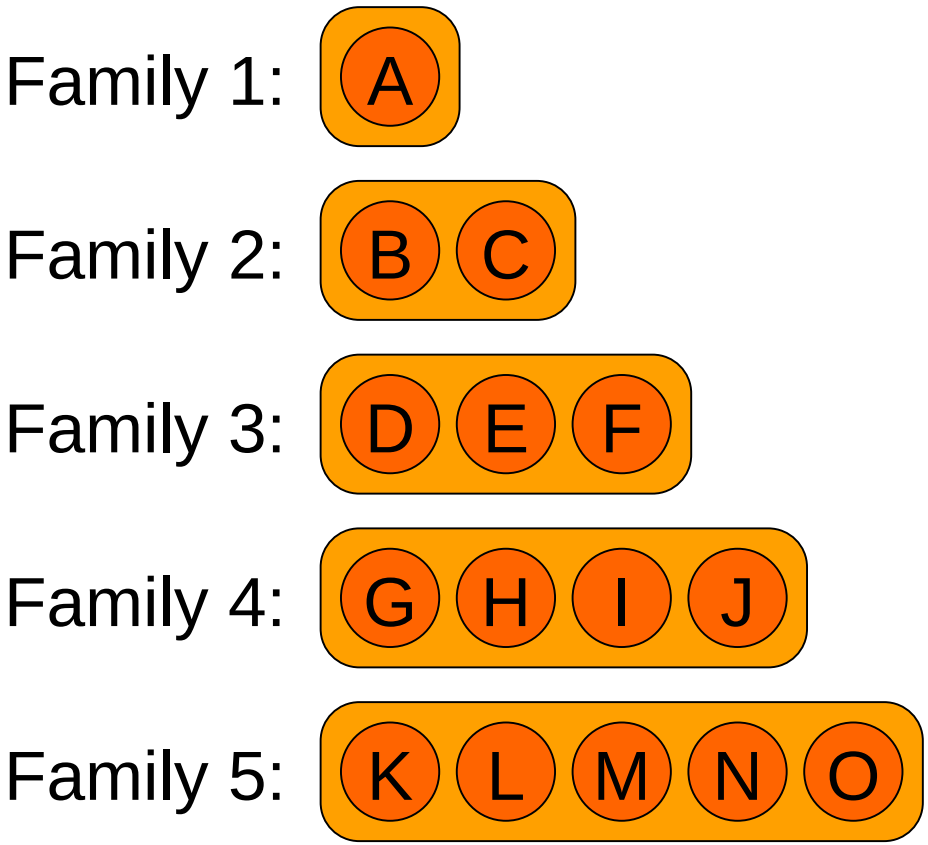
Example (Website survey)

- A survey question is posted on an internet website.
- Visitors to that website are asked to give their response.
- Which of the following types of bias are *inherent* in this method?
  - Selection bias
  - Non-response bias
  - Response bias
  - Experimenter bias

Example

- Suppose that we wish to estimate average family size.
- We randomly select individuals and ask them how many siblings they have.
- Is this method biased?

Example



15 Children - 5 Familes

Example

- If we take a sample of all 15 *children*, then our data are

$$\{1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5\}.$$

- The average is $\frac{55}{15} = 3.667$ children per family.

Example

- If we take a sample of all 5 *families*, then our data are

$$\{1, 2, 3, 4, 5\}.$$

- The average is $\frac{15}{5} = 3$ children per family.
- Which method is correct?

Example

- To estimate average family size, we should take a random sample of *families*, not *individuals*.

Assignment

Homework

- Read Sections 2.1 - 2.4, pages 28 - 47.
- Let's Do It! 2.1, 2.2, 2.3.
- Page 97, exercises 1 - 12.