Classroom Voting Questions: Statistics

Displaying Distributions with Graphs

1. A data set consists of fifty three-digit numbers ranging from 180 to 510. The best choice for stems in a stem-and-leaf display would be to use _____.

   (a) 1 digit stems (1, 2, ..., 5)

   (b) 2 digit stems (18, 19, ..., 51)

   (c) 3 digit stems (180, 181, ..., 510)

   Answer: (a).

   by Roxy Peck for the textbooks: Roxy Peck and Jay Devore, Statistics: The Exploration and Analysis of Data, 6th Edition, Brooks/Cole Cengage Learning 2008 and Roxy Peck, Chris Olsen and Jay Devore, Introduction to Statistics and Data Analysis, 3rd Edition, Brooks/Cole Cengage Learning 2008.

   STT.01.01.010

   CC HZ MA207 F09: **10**/85/5 time 1:15
   CC KC MA207 F09: **23**/67/9
   AS DH MA3321 Su12: **0**/100/0 time 1:20

2. Which of the following statements is most completely true in comparing an appropriately drawn histogram to a stem-and-leaf display of the same data?

   (a) Both convey the same information about the shape of the distribution.

   (b) Both convey the same information about gaps in the distribution.

   (c) Both convey the same information about outliers.

   (d) Both convey the same amount of information generally.

   (e) Two from (A)-(D) are correct.

   (f) Three from (A)-(D) are correct.

   (g) All from (A)-(D) are correct.

   Answer: (f). (A) This statement is true but so are (B) and (C).

   (B) This statement is true but so are (A) and (C).

   (C) This statement is true but so are (A) and (B).

(D) A stem-and-leaf display also conveys information about at least approximate data values from the data set on which the display is based. A histogram conveys less of this details about the data values.

(E) Three, not two, of the statements are true. That is, (A), (B), and (C) are true.

(F)* correct The statements in (A), (B), and (C) are true.

(G) The statement in (D) is not true.

by Murphy, McKnight, Richman, and Terry

STT.01.01.020

CC HZ MA207 F09: 15/0/0/30/20/**35**/0 time 1:30
CC KC MA207 F09: 6/0/6/3/46/**40**/0
AS DH MA3321 Su12: 0/7/0/7/20/**60**/7 time 2:20
AS DH MA1333 010 F12: 6/0/0/0/35/**24**/35 time 3:00
AS DH MA1333 020 F12: 0/0/0/26/0/**5**/68 time 2:50
AS DH 1333 010 S13: 0/0/0/0/21/**5**/74 time 3:30
AS DH 1333 020 S14: 0/0/0/27/4/**50**/19 time 2:50 ,
AS DH 3321 010 S14: 10/5/0/0/10/**43**/33 time 3:00 ,
AS DH 1333 010 F14: 4/4/0/17/21/**31**/24 time 3:10 ,
AS DH 1333 020 S15: 0/0/0/24/29/**19**/29 time 3:30 ,
AS DH 1333 020 F15: 0/0/0/4/17/**46**/33 time 2:30 ,
AS DH 1342 020 F18: 0/0/7/18/32/**18**/25 time 3:00
AS DH 1342 040 S19: 0/0/0/0/13/**81**/6 time 3:10
AS DH 1342 030 F19: 4/0/0/11/21/**21**/43 time 3:00
AS DH 1342 030 S20: 0/0/0/19/62/**19**/0 time 3:30

3. Suppose you take a random sample of 10 juniors who bought the same model laptop when they were freshmen. You test these 10 laptops to determine how long their batteries last before needing to be recharged, and you obtain the following data (in hours): 1.2, 1.3, 3.8, 3.9, 3.9, 4.0, 4.1, 4.1, 4.2, 4.3. What should be done with the values 1.2 and 1.3? Which of the following is the best course of action?

   (a) Delete them from the data set since they are outliers.

   (b) Keep them in the data set even though they are outliers.

   (c) Determine why these values were so much lower than the rest, then delete them.

   (d) Determine why these values were so much lower than the rest, then keep them in the data set, provided they weren't due to data entry errors.

   Answer: (d). That's the best course of action listed. However, what you'll probably want to do is to determine why these values were so low, then decide whether or not to keep them. It may be that they are the result of some process that you wish to exclude from your analysis–perhaps they are a result of students who greatly modified

their laptops. It may be, however, that some number of these laptop batteries just don't hold up well over time. If these values were produced by the same process that produced the other values, then they should be kept.

by Derek Bruff

STT.01.01.030

CC HZ MA207 F09: 0/5/0/**95** time 0:55
AS DH MA3321 Su12: 0/0/7/**93** time 1:30
AS DH MA1333 010 F12: 0/79/0/**21** time 2:30
AS DH MA1333 020 F12: 0/95/0/**5** time 2:20
AS DH 1333 010 S13: 0/58/4/**38** time 2:30
AS DH 1333 020 S14: 0/58/0/**42** time 2:00 ,
AS DH 1333 010 F14: 0/21/0/**79** time 2:30 ,
AS DH 1333 020 S15: 0/32/0/**68** time 2:40 ,
AS DH 1333 020 F15: 0/32/0/**68** time 2:00 ,
CC KC MA207 F15: 0/0/0/**100**
CC KC MA315 F15: 0/0/0/**100**
AS DH 1342 020 F18: 0/23/5/**72** time 2:30
AS DH 1342 040 S19: 0/73/0/**27** time 2:10
AS DH 1342 030 F19: 0/32/0/**68** time 2:50
AS DH 1342 030 S20: 0/31/6/**63** time 2:50

4. You consider all of the adult patients in a large hospital. Which of the following variables would you expect to have a distribution that is left-skewed as revealed by a dot plot of the data?

   (a) height

   (b) annual income

   (c) eye color

   (d) age

   Answer: (d).

   (a) Height has a bell-shaped distribution.

   (b) Annual income has a right-skewed distribution.

   (c) Eye color is a qualitative or categorical variable. Both producing a dot plot and discussing skewness make sense only with a quantitative variable.

   (d) In this context, age is left-skewed. (Young adults tend to be healthier. We expect the largest number of hospital patients to be in, perhaps, their 70s and 80s.) [Follow-up question: Is the distribution J-shaped? Answer: No. Although people in their late 90s and 100s are more likely to be in the hospital than people in their 70s and 80s, there are a lot fewer such people overall.]

by David A. Huckaby

STT.01.01.040

AS DH 3321 010 F16: 0/10/0/**90** time 2:00
AS DH 1342 010 F17: 0/11/3/**86** time 2:30
AS DH 1342 020 F18: 0/5/0/**95** time 2:20
AS DH 1342 040 S19: 0/31/0/**69** time 2:00
AS DH 1342 030 F19: 10/14/3/**72** time 3:00
AS DH 1342 030 S20: 0/0/0/**100** time 3:10

5. You consider all of the adult patients in a large hospital. Which of the following variables is continuous?

   (a) height

   (b) weight

   (c) number of past surgeries

   (d) more than one of the above

   Answer: (d). This question can be used when students are first learning about data and variables. For questions on discrete versus continuous in the context of random variables, see the STT.04.03 section of this question collection.

   (a) A person's height can be any number on the real number line in the range of possible human heights, so height is a continuous variable.

   (b) Like height, weight is a continuous variable.

   (c) There are gaps on the real number line between the possible value for the variable "number of surgeries," so the variable is discrete.

   (d) Because both (a) and (b) are valid, (d) is the correct answer.

   by David A. Huckaby

   STT.01.01.050

   AS DH 3321 010 F16: 0/0/0/**100** time 1:30
   AS DH 1342 010 F17: 0/0/0/**100** time 1:40
   AS DH 1342 020 F18: 0/0/0/**100** time 1:30
   AS DH 1342 040 S19: 0/0/0/**100** time 1:30
   AS DH 1342 030 F19: 0/0/0/**100** time 1:40
   AS DH 1342 030 S20: 0/0/0/**100** time 1:50

6. You consider all of the adult patients in a large hospital. Which of the following variables is discrete?

(a) height

(b) eye color

(c) number of siblings

(d) more than one of the above

Answer: (c). This question can be used when students are first learning about data and variables. For questions on discrete versus continuous in the context of random variables, see the STT.04.03 section of this question collection.

(a) A person's height can be any number on the real number line in the range of possible human heights, so height is a continuous variable.

(b) Eye color is a qualitative, or categorical, variable. Only variables that are quantitative can be further classified as either discrete or continuous.

(c) There are gaps on the real number line between the possible values for the variable "number of siblings," so the variable is discrete. If the issue of half-siblings comes up, this provides an opportunity for discussing the fact that the discrete-versus-continuous distinction is not an issue of "decimals-versus-not-decimals" but is more subtle.

(d) Only (c) is correct.

by David A. Huckaby

STT.01.01.060

AS DH 3321 010 F16: 0/3/**27**/70 time 1:00
AS DH 1342 010 F17: 0/0/**61**/39 time 1:00
AS DH 1342 020 F18: 3/0/**35**/63 time 2:00
AS DH 1342 040 S19: 0/0/**63**/38 time 1:50
AS DH 1342 030 F19: 0/11/**64**/25 time 1:50
AS DH 1342 030 S20: 0/5/**67**/29 time 2:10