

# Practice test 1 – HS 67

## Chapter 1 Conceptual Review Questions

1. If statistics is *not* merely a compilation of computational techniques, what then *is* it?
2. The first element of data analysis involves exploring and describing the data. The second element involves [complete this sentence].
3. Define “measurement.”
4. What does the acronym GIGO stand for?
5. Define the following terms: observation variable, value.
6. Provide a synonym for *categorical variable*.
7. Provide a synonym for *quantitative variable*.
8. Describe the structure of a data table.
9. This is the word we use to refer to observing things as they are without shaping them to your own preconceived world view. [Multiple choice: accurate, precise, objective].
10. What do we mean by the term *statistical population*?
11. What is a sample?
12. What do we call the ratio of the sample size to population size?
13. The information on a data collection form corresponds to a(n) [M/C]: a) value b) variable c) observation
14. A column of data in a data table corresponds to a(n): [M/C]: a) value b) variable c) observation.
15. Every value in a simple random sample from a population has a(n) \_\_\_\_\_ chance of being selected.
16. I select the first 10 people on an alphabetized list. Explain why this is not an SRS.
17. A famous statistician once called statistics “the servant of all sciences”. What did he mean by this statement?
18. How does sampling with replacement differ from sampling without replacement?

## Chapter 1 Practice Exercises

**1.8 Cerebellar toxicity.** An investigation was prompted when the U. S. Food and Drug Administration received a report of an increased frequency of cerebellar toxicity associated with a form of chemotherapy from the University of Wisconsin Hospital and Clinics (Madison) after the hospital had switched from the product manufactured by the innovator company to a generic product. To address this issue, the FDA sent a team of investigators to complete a chart review. The investigators wished to determine whether there was a greater risk associated with the generic chemotherapeutic agent than the innovator product. Fifty-four ( $n = 54$ ) leukemia and lymphoma patients undergoing chemotherapy in preparation for bone marrow transplantation were studied.

- (A) Describe the population to which the research applies.  
(B) Describe the sample.

Data on patient and treatment characteristics for the following variables were collected:

Variable	Description
AGE	Age at time treatment (years)
SEX	1 = male; 2 = female
GENERIC	1 = generic manufacturer 2 = innovator manufacturer
DIAG	Underling diagnosis: 1 = leukemia; 2 = lymphoma
STAGE	Stage of disease: 1 = relapse; 2 = remission
TOX	Toxicity? 1 = yes; 2 = no
DOSE	Dose of drug (grams /meters <sup>2</sup> )
SCR	Serum creatinine (mg/dl)
WEIGHT	Body weight (kg)

The first 5 observations are shown here:

ID	AGE	SEX	GENERIC	DIAG	STAGE	TOX	DOSE	SCR	WEIGHT
1	50	1	1	1	1	1	36.0	0.8	66
2	21	1	1	1	2	2	29.0	1.1	68
3	35	1	1	2	2	2	16.2	0.7	97
4	49	2	2	1	1	2	29.0	0.8	83
5	38	1	1	2	2	1	16.2	1.4	97

(C) Classify each variable's measurement scale.

(D) What is the value of the AGE variable for observation 4?

(E) What is the diagnosis of observation 2?

## Chapter 2 Conceptual Review Questions

- How many *leaves* do you plot on a stemplot?
- How many stem values should you use when drawing a stemplot?
- Why would you use split stem-values on a stemplot?
- What is the *depth* of an observation?
- What is the purpose of the stem-multiplier on a stemplot?
- Distributions with long tails toward the higher numbers are said to have a \_\_\_\_\_ skew.
- The three elements of a distribution are shape, location, and \_\_\_\_\_.
- What is the "middle point" of the ordered array called?
- The median has a depth of \_\_\_\_\_.
- This is a value that does not fit in with the general pattern of a distribution.
- Define the following terms: frequency, relative frequency, cumulative relative frequency.
- What is a distribution?
- What is the full name for stemplot?
- What do you look for on a stemplot?
- Why must the stem and leaves of a stemplot align vertically and horizontally?
- What do you do if a value in a data set has more significant digits that can fit on a Tukey stemplot?

## Chapter 2 Practice Exercises

**2.4 Outpatient wait time.** The length of time patients wait for attention in doctors' offices is an important consideration in health care. Waiting times (minutes) for 25 patients at a health clinic are:

35      22      63      6      49      19      16      31      24      29  
 23      32      72      13      51      45      77      16      33      55  
 10      42      28      72      13

- (A) Draw a stemplot of the data. Describe the distribution of values.  
 (B) Using 10-unit class intervals, create a frequency table showing frequency counts, relative frequencies, and cumulative relative frequencies.  
 (C) What percentage of wait times were less than 20 minutes?  
 (D) What percentage were *at least* 20 minutes?

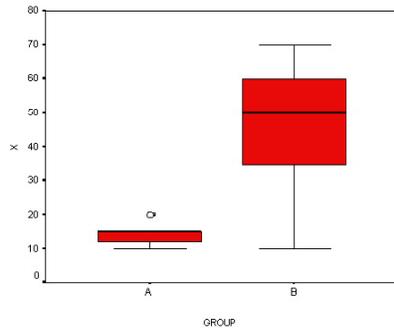
**2.10 Health insurance coverage.** Data for the percentage of people without health insurance by state for the period 2002 through 2004 are shown below. Create an informative stemplot and then describe the distribution. [Note: Plots with too many or too few stem values are not acceptable, as they will fail to show the shape of the distribution.]

STATE	% w/out	STATE	% w/out	STATE	% w/out	STATE	% w/out	STATE	% w/out
Alabama	13.5	Hawaii	09.9	Michigan	11.4	North Carolina	16.6	Utah	13.4
Alaska	18.2	Idaho	17.3	Minnesota	08.5	North Dakota	11.0	Vermont	10.5
Arizona	17.0	Illinois	14.2	Mississippi	17.2	Ohio	11.8	Virginia	13.6
Arkansas	16.7	Indiana	13.7	Missouri	11.7	Oklahoma	19.2	Washington	14.2
California	18.4	Iowa	10.1	Montana	17.9	Oregon	16.1	West Virginia	15.9
Colorado	16.8	Kansas	10.8	Nebraska	11.0	Pennsylvania	11.5	Wisconsin	10.4
Connecticut	10.9	Kentucky	13.9	Nevada	19.1	Rhode Island	10.5	Wyoming	15.9
Delaware	11.8	Louisiana	18.8	New Hamp	10.6	South Carolina	13.8		
DC	13.5	Maine	10.6	New Jersey	14.4	South Dakota	11.9		
Florida	18.5	Maryland	14.0	New Mexico	21.4	Tennessee	12.7		
Georgia	16.6	Massach.	10.8	New York	15.0	Texas	25.1		

## Chapter 3 Conceptual Review Questions

- List three measures of central location.
- The two distinct measures of spread are:
- A study shows a mean of 0.98 and median of 0.56. What does this suggest?
- The line inside a boxplot represents the distribution's \_\_\_\_\_.
- The top of the whisker on a box-and-whiskers plot goes up to the [3 words].
- What visual clues of spread are on a boxplot?
- Besides being the gravitational center of the distribution, the mean tells you what three things?
- This is the reason we don't often use the variance for descriptive purposes.
- The *hinge-spread* is the same as the \_\_\_\_\_.

10. What five points comprise a 5-point summary?
11. Q1 is the median of the \_\_\_\_\_ half of the data set.
12. Do the upper and lower fences appear on a boxplot?
13. Distinguish between  $\mu$  and  $\bar{x}$ . Distinguish between  $\sigma$  and  $s$ .
14. Discuss these side-by-side boxplots.



15. When  $n$  is odd and you are splitting the data into a high group and low group for the purpose of determine Tukey's quartiles, does the median go into the low group, the high group, neither group or both groups?

## Chapter 3 Practice Exercises

**3.10 Air pollution at two sites.** Suspended particulate matter in air is an indicator of pollution. Particulate air pollution levels (micrograms per cubic meter) from two sites are shown as stemplots on a common stem below.

Site 1		Site 2
42	2	
8	2	
2	3	234
86	3	6689
2	4	0
	4	
	5	
	5	
	6	
8	6	
(×10)		

- (A) Transcribe the values for the two groups. Then, using the step-by-step approach used in class (demonstrating deviations and squared deviations), calculate the mean and standard deviation for the values at Site 1.
- (B) Use your calculator to determine the mean and standard deviation for the values at Site 2.
- (C) Compare the means and standard deviations from the two sites. Do the means differ significantly? How do the groups differ?
- (D) Draw a boxplot for the data from site 1. Then, on the same axis, draw the boxplot for site 2. What do you observe?