

PRACTICE TEST FOR TEST 1 (STAT-461/561)

- (1) Determine if each of the following statements is true or false: (a) In a statistical study, all individuals in the population are usually included; (b) Statistics has two major branches: First we draw conclusions from the data (inferential statistics) and then organize and summarize the data (descriptive statistics); (c) Simple random sampling with or without replacement can get rid of all kinds of bias in a statistical study; (d) In stratified random sampling, all individuals in each stratum is included in the sample; (e) In cluster sampling (where the population is divided into many clusters), you randomly choose the 1st cluster and then follow a system (like choosing every 5th or 10th cluster); (f) Qualitative ordinal data can be classified into interval-type and ratio-type, (g) Quantitative ratio-type data can be classified into discrete and continuous; (h) In a statistical study, the word 'population' always means a group of people; (i) A class frequency distribution table is appropriate for discrete ratio-type data without many repetitions or for continuous ratio-type data; (j) From the histogram of a dataset, you can compute its mean, median and standard deviation; (k) From the stem-and-leaf plot of a dataset, you can compute its mean, median and standard deviation; (l) IQR and median are more robust (or resistant) to extreme data-values than standard deviation and mean; (m) MAD is easily affected by extreme data-values, so it is a worse measure of variability than range; (n) The median of a dataset is always one of the data-values.
- (2) Identify each of the following data-types (for example, the number of buttons on your shirt is quantitative ratio-type discrete data): (a) The amount of liquid hand-soap in a bottle; (b) The year of first child-birth for a working mother; (c) Your level of satisfaction with your current cable/internet provider (on a scale of 1 to 5); (d) The college or university degree that you hold; (e) The name of your best friend; (f) The number of petals of a flower; (g) The start-time of a nationally live televised football game.
- (3) Identify each of the following sampling techniques: (a) In order to get a sample of 20 jack-rabbits, the first 20 burrows that were found in a meadow were invaded and the rabbit therein were caught; (b) In order to get a sample of 9 students from a class of 45 (sitting in a circle along the walls of a classroom), you randomly choose the first student and then select every 5th student starting from him/her; (c) To draw a sample of 30 saw-blades from batches of saw-blades coming out of an assembly line (10 blades in each batch), three batches are randomly selected and all the blades in them were included; (d) To get a sample of 100 Medicare-eligible elderly people in a state, 20 elderly people were randomly selected from each of the races Caucasian white, African-American, Latino, Asian and others.
- (4) 20 patients with severe epilepsy were observed for 8 weeks for seizures. The following are the numbers of major seizures suffered by each patient during that period: 5,0,9,6,0,0,5,0,6,1,5,0,0,0,0,7,0,0,4,7. (a) What is the mean (or average) number of seizures? (b) What is the median number of seizures? (c) What is the IQR? (c) Construct a histogram.

- (5) A botanist grew 15 pepper plants on the same greenhouse bench. After 21 days, she measured the total stem-length (in cm) of each plant. Here are the numbers: 12.4, 12.2, 13.4, 10.9, 12.2, 12.1, 11.8, 13.5, 12.0, 14.1, 12.7, 13.2, 12.6, 11.9, 23.1. (a) Construct a stem-and-leaf plot for this dataset (with only the decimal digit as the 'leaf'); (b) Compute the '5-point summary' of this dataset; (c) Draw a boxplot (with the fences clearly marked on the baseline) and do outlier detection.
- (6) If in the above dataset (in #5), it was later found that the last data-point should have been 13.1 instead of 23.1, which of the following quantities, do you think, will change significantly? (a) Mean, (b) Median, (c) Range, (d) Standard deviation, (e) IQR, (f) MAD. Compute any two of the quantities that you expect to change significantly --- both for the original dataset and the corrected dataset.
- (7) Compute the standard deviation and the MAD of the dataset in #5
- (8) The weights (in ounces) and the postages needed for 8 international letters were as follows: (0.9, 0.98), (1.30, 1.21), (0.68, 0.98), (2.56, 1.88), (1.01, 1.21), (3.90, 2.78), (0.88, 0.98), (2.05, 2.33). Compute the correlation between these two variables. What is the equation of the best-fitting line? What is the ANOVA equation? How well does the best-fitting line fit this dataset?