University of Houston

High School Math Contest

Spring 2012

Statistics Test

Name _____

School _____

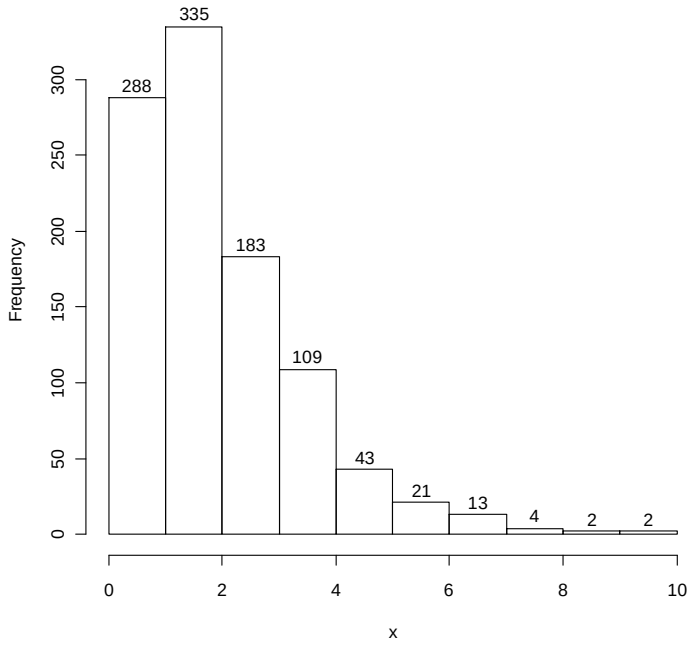
Exam Time: 1 hour

Any type of calculator is allowed. Tables are provided.

The first 15 problems are multiple choice. The last question calls for a written response. It will be used to break ties.

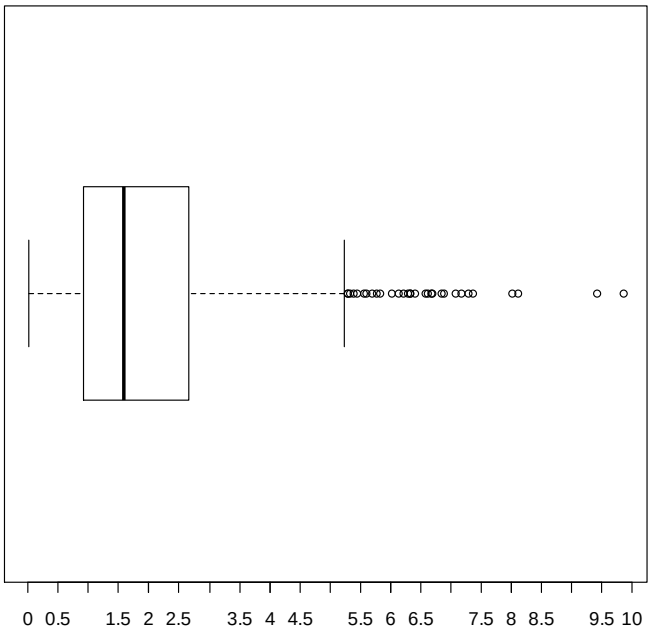
1. The figure below is a histogram of 1000 measurements. The heights of the histogram bars are written above them. The third quartile of the data is closest to

   (a) 2.50
   (b) 3.50
   (c) 4.50
   (d) 1.80
   (e) Since the data is highly skewed, the third quartile cannot be accurately estimated.



2. The figure below is a box and whisker diagram of some data. Which of the following is true?

   (a) A normal distribution would accurately describe the data.
   (b) The mean of the data is more robust (i.e., resistant to outliers) as an indicator of central location than the median.
   (c) The mean of the data is considerably larger than the median.
   (d) Not many of these observations could be considered outliers.
   (e) A five number summary would give more of the important information about this data than the box and whisker diagram does.



3. Refer to the same figure above. The MAD statistic is defined to be the median of all the absolute values of the differences between the observations and their median. For this data the MAD statistic is

   (a) more than the mean of the data.
   (b) more than the standard deviation of the data.
   (c) less than or equal to the distance from the first quartile to the median.
   (d) less than or equal to the distance from the median to the third quartile.
   (e) more than 3.

4. Scores on a national achievement test are normally distributed with a mean of 700 and a standard deviation of 75. The 85^{th} percentile of test scores is closest to

(a) 800
(b) 775
(c) 850
(d) 625
(e) 720

5. Previous studies of levels of a pollutant in Galveston Bay indicated a mean level of 0.035 with a standard deviation of 0.005. Scientists would like to revise their estimate of the mean level with an error no greater than 0.001 with 95% confidence and to be as economical about it as possible. How large a sample of measurements should they take? You may assume that pollutant levels are normally distributed.

(a) 46
(b) 1350
(c) 210
(d) 4088
(e) 97

6. Which of the following random variables could be described by a binomial distribution?

(a) The number of lightning strikes in a certain area in a 7 day period is counted. X is the number of strikes.
(b) The number of cars counted in one hour by a traffic counter at a certain place on a road is recorded. X is the number of cars.
(c) The sample of size 20 of students in a class of 60 is taken without replacement. X is the number of sampled students who made C or better in the prerequisite course.
(d) A sample of 100 students is taken without replacement from the population of all graduating students in the Houston area. X is the number in the sample who graduated with honors.
(e) A fair die is rolled until two sixes in a row are observed. X is the number of rolls required.

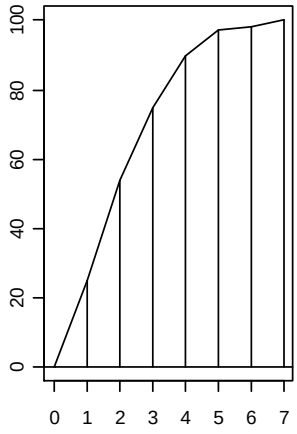
7. Sacks of grain are normally distributed with mean 100 lb. and standard deviation 2 lb. A wagon carries 9 sacks of grain. What is the probability that the wagon load weighs more than 910 lb?

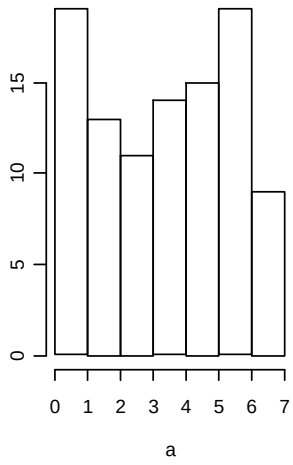
   (a) between 0.045 and 0.049
   (b) between 0.101 and 0.105
   (c) between 0.287 and 0.291
   (d) between 0.019 and 0.023
   (e) between 0.151 and 0.155

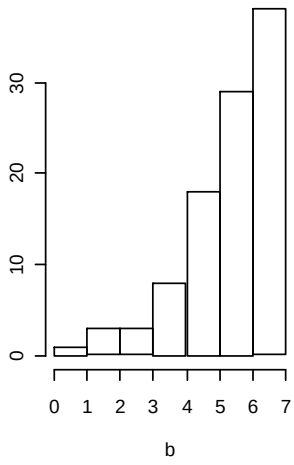
8. A political polling firm conducts a survey by the following method. Random samples of 5% of the registered voters in each of three categories: Democrat, Republican and Independent are selected. The selected voters are polled about which mayoral candidate they will vote for. This is an example of

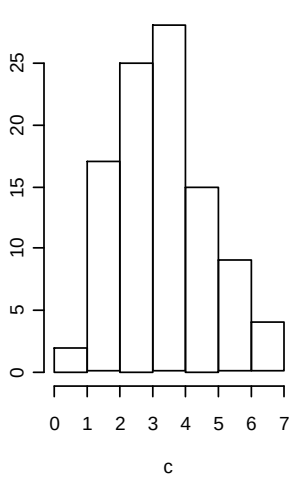
   (a) a randomized complete block design
   (b) stratified sampling with proportional representation
   (c) an observational study
   (d) interval data
   (e) cluster sampling

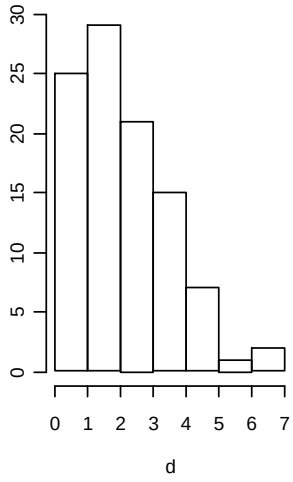
9. The upper left figure on the next page is a frequency polygon of 100 measurements. Which of the histograms represents the same data?

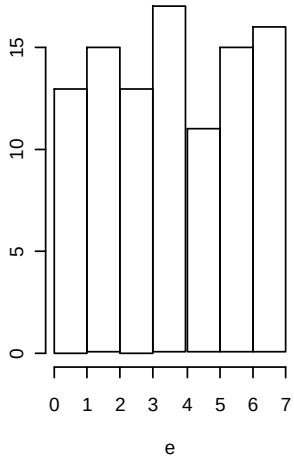

0
20
40
60
80
100
0 1 2 3 4 5 6 7


0
5
10
15
0 1 2 3 4 5 6 7
a


0
10
20
30
0 1 2 3 4 5 6 7
b


0
5
10
15
20
25
0 1 2 3 4 5 6 7
c


0
5
10
15
20
25
30
0 1 2 3 4 5 6 7
d


0
5
10
15
0 1 2 3 4 5 6 7
e

10. A six-sided die is tossed 180 times with the following results.

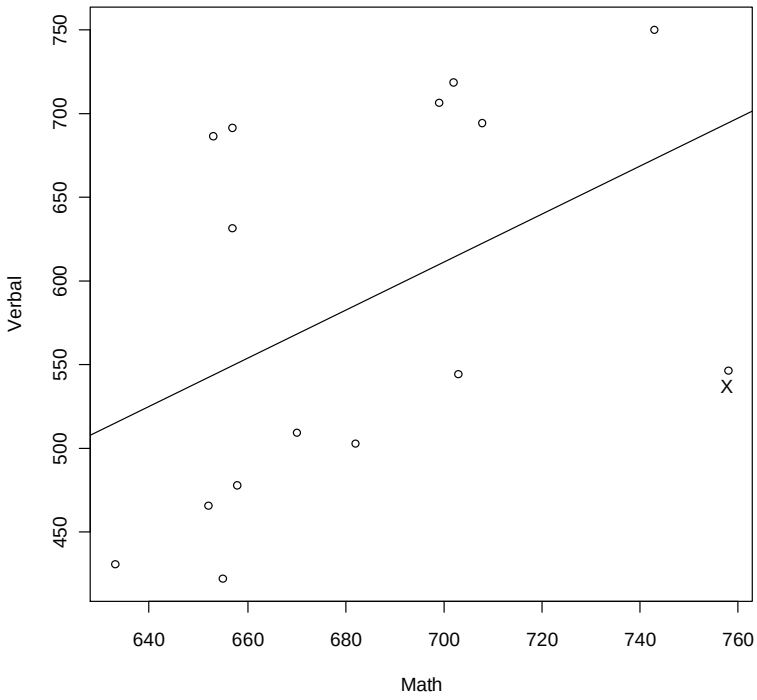
Face	1	2	3	4	5	6
Frequency	20	21	31	35	33	40

We would like to test the null hypothesis that the die is fair against the alternative hypothesis that it is not fair by using a chi-square test. The null hypothesis would be rejected at a chosen significance level (probability of type 1 error) of

(a) 20% but not 10%
(b) 10% but not 5%
(c) 5% but not 2%
(d) 2% but not 1%
(e) 1%

11. The figure below shows math and verbal scores for 15 students. Seven of them are native speakers of the language and eight are not. Which of the following is not true.

(a) The correlation between math and verbal scores is positive and moderately strong.
(b) Correlations calculated separately for each group are even stronger than the overall correlation.
(c) If this is a sample from the general population of students, an increase of 20 in the math score corresponds to an expected increase of about 30 in the verbal score.
(d) The least squares line passes through the point whose coordinates are the average math and average verbal scores.
(e) The point labeled with an X is an outlier but it has little influence on the fitted line.



12. A random sample of 225 measurements of a numeric variable were recorded. The sample average was 2.58 and the sample standard deviation was 1.50. A 99% confidence interval for the population mean is

    (a) from 2.56 to 2.60
    (b) from 1.08 to 4.18
    (c) from 1.44 to 3.72
    (d) from -0.01 to 5.17
    (e) from 2.32 to 2.84.

13. 100 cards are dealt one at a time from a standard deck, with replacement and with reshuffling after each card is dealt. The deck contains 13 cards in each of four suits: hearts, diamonds, clubs and spades. The probability of getting more than 30 hearts is closest to

    (a) 0.2917
    (b) 0.0594
    (c) 0.1587
    (d) 0.0250
    (e) 0.1020

14. An experiment on the effect of a new drug for migraine headaches is conducted using a "crossover" experimental design. Each of twenty adult women is given a placebo for one migraine episode and the experimental drug for another episode. The degree of alleviation of the symptoms is measured after each. Half of the experimental patients are randomly chosen to receive the placebo first and the experimental drug second. The order is reversed for the other half of the patients. This is done because it is possible that the order of administration influences the effectiveness of each treatment. Which of the following analyses would be best for investigating the effectiveness of the drug?

    (a) A student-t test should be performed on the Drug - Placebo differences to test the null hypothesis that the mean difference is zero.
    (b) The Placebo measurements and the Drug measurements should be treated as independent samples from two populations and a t-test performed for the null hypothesis that the population means are equal.
    (c) A straight line should be fitted by least squares expressing the Drug measurement as a linear function of the Placebo measurement. Then the t-test should be performed for the null hypothesis that the slope is zero.
    (d) The mean differences between the first treatment administered and the second treatment administered should be compared for the two groups "Placebo first" and "Drug first". A two-sample t-test should be performed for the null hypothesis that the population means are equal.
    (e) A chi-square test for the hypothesis of independence of the order of administration and the difference between treatments should be performed.

15. Scores on a national achievement test are normally distributed with a mean of 650 and a standard deviation of 35. A student scores 690. Which of the following is closest to her percentile score?

    (a) 87%
    (b) 83%
    (c) 94%
    (d) 70%
    (e) 99%

16. A space probe is aimed at a certain point on the Moon. The square of the distance from the aiming point to the landing point is kW, where k is a constant and W has a chi-square distribution with 2 degrees of freedom. The probability that the probe lands within 1 kilometer of the aiming point is 0.99. Find the distance d such that the probability that the probe lands within d kilometers of its target is 0.90.