

4-29-2008

MATH 375 — Final Exam Spring 07

0 Homework problems and midterm exam.

I (12 points) True/False questions. Circle T for true and F for false.

1. (T F) The SD of the scores of the students in a class gets larger if each student in the class gets a three percent increase in his or her score.
2. (T F) If outcomes A and B are independent and $P(A) = P(B) = 0.3$ then $P(A \text{ and } B) = 0.09$.
3. (T F) If outcomes A and B cannot occur simultaneously and $P(A) = P(B) = 0.3$ then $P(A \text{ or } B) = 0.6$.
4. (T F) If $P(A) = 0.3, P(B) = 0.6$ then B is twice as likely as A .
5. (T F) When tossing a coin three times, getting exactly two heads is as likely as getting three heads.
6. (T F) If two events are independent then they cannot occur at the same time.
7. (T F) A sample correlation coefficient between variables x and y measures the linear relationship between x and y , not the relationship in general.
8. (T, F) If a histogram has a long left tail, then the mean is smaller than the median.
9. (T, F) In the U.S. in 1990, there were 2.1 million deaths from all causes, compared to 1.7 in 1960— nearly 25 percent increase. The data show that the public's health got worse over the period 1960–1990.

10. (T F) In an observational study, the subjects assign themselves to the control and treatment groups.
11. (T F) In a series of repeated measurements, bias does not change from measurement to measurement.
12. (T F) In an observational study, the effects of treatment may be confounded with the effects of factors that got the subjects into treatment or control in the first place.

II (12 points) Fill in blank.

1. If A, B are independent, $P(A) = 0.4$ and $P(B) = 0.7$ then $P(A \text{ and } B) = \underline{\hspace{2cm}}$.
2. If A and B cannot occur simultaneously and are independent, then $P(A)P(B) = \underline{\hspace{2cm}}$.
3. A discrete numerical variable is one whose possible values are along the number line.
4. A continuous numerical variable is one whose possible values form along the number line. characteristic to be estimated.
5. Pearson's sample correlation coefficient is a number between and .
6. By Empirical Rule, there are approximately , , of observations within 1, 2, 3 standard deviations of the mean when the histogram is well approximated by a normal curve.
7. By Chebyshev Rule, there are approximately , , of observations within 1, 2, 3 standard deviations of the mean.
8. Coefficient of determination is a measure of the proportion of variation in the observed y 's that can be explained by .

9. A point estimate represents _____ of a population characteristic.
10. A statistic is _____ if it has a sampling distribution with a mean equal to the value of the population
11. Residual sum of squares is a measure of y variation that cannot be attributed to _____.
12. The correlation coefficient between points scored and points missed on an exam is _____. (options: -1, -0.5, 0, 0.5, 1).
13. A density curve serves as a model for the population distribution of _____.
14. Blocking uses extraneous factors to create experimental groups that are _____ with respect to those factors, thereby filtering out their effect.
15. In a double blind experiment the subjects who do not receive the treatment will instead receive a _____.
16. The selection bias is the tendency of a sample to differ from the population because of systematic _____ of some part of the the population.
17. A normal distribution is standard when its mean is _____ and standard deviation is _____. characteristic to be estimated.
18. An experiment is a procedure for investigating the effect of experimental conditions on _____.
19. A normal distribution has a density curve whose shape is _____.
20. A statistic is _____ if it has a sampling distribution with a mean equal to the value of the population

III (14 points) Circle the answer which best completes/answers the sentence/question.

1. Which of the following is a categorical measurement?
 - (a) GPA.
 - (b) Hair color.
 - (c) Weekly income.
 - (d) Both (i) and (iii).
 - (e) None of the above.
2. For a straight line plotted on an x-y graph
 - (a) the slope equals run/rise.
 - (b) the slope equals rise/run.
 - (c) the intercept is where the line crosses the x-axis.
 - (d) both (a) and (c) are true.
 - (e) both (b) and (c) are true.
3. The correlation coefficient between two variables
 - (a) is not affected by outliers.
 - (b) is always between 0 and 1.
 - (c) is a measure of linear association between the two variables.
 - (d) is 0 when there is a tight clustering of the points around the regression line.
 - (e) satisfies all of the above.
4. When summarizing the relationship between variables x and y,
 - (a) the mean of y is needed.
 - (b) the standard deviation of y is needed.
 - (c) the mean of x is needed.
 - (d) the standard deviation of x is needed.
 - (e) all the above are true.
5. For two outcomes which cannot occur simultaneously,

- (a) the probability both outcomes occurring is the product of the individual probabilities.
 - (b) the probability both outcomes occurring is the sum of the individual probabilities..
 - (c) the probability both outcomes occurring is 1.
 - (d) the probability both outcomes occurring is 0.
 - (e) none of the above is true.
6. If the correlation coefficient between two variabels is zero ($r = 0$), then
- (a) none of the points on the scatter plot line within 1 standard deviation of the regression line.
 - (b) there is no linear association between the variables.
 - (c) the scatter plot of the two varables shows no pattern whatsoever.
 - (d) there is tight clustering about the regression line.
 - (e) all of the above statements are true.
7. The outcome A has a chance of 30 percent happening. The outcome B has a chance of 60 percent happening. If the chance that both outcomes happen together is 18 percent then
- (a) the two outcomes are dependent.
 - (b) the two outcomes are independent.
 - (c) the chance that either A or B happens is 90 percent.
 - (d) both (a) and (c).
 - (e) both (b) and (c).
8. If there is a strong association between two variables then
- (a) there is a cause and effect relationship between the variables.
 - (b) knowing one variable helps a lot in predicting the other.

- (c) the correlation coefficient must be close to one.
 - (d) both (ii) and (iii).
 - (e) all of the above are true.
9. A confounding factor
- (a) can never be controlled for in an observational experiment.
 - (b) always increases the standard deviation of the data.
 - (c) may cause the results of an experiment to be misinterpreted.
 - (d) satisfies both (a) and (b).
 - (e) satisfies both (a) and (c).
10. In a completely randomized experiment,
- (a) direct controls may sometimes be used if the investigators have good reason for doing so.
 - (b) the subjects assign themselves to the treatment and control groups.
 - (c) the investigators use their best judgement to assign the subjects to the treatment and control groups
 - (d) there is a serious possibility of confounding with unknown extraneous factors.
 - (e) none of the above is true.
11. If the correlation coefficient between two variables is zero ($r = 0$), then
- (a) none of the points on the scatter plot lie within 1 standard deviation of the regression line.
 - (b) there is no linear association between the variables.
 - (c) the scatter plot of the two variables shows no pattern whatsoever.
 - (d) all of the above statements are true.

12. In an observational study,
- (a) a placebo must always be used.
 - (b) confounding is a potential problem.
 - (c) the investigators use their best judgement to assign the subjects to the treatment and control groups.
 - (d) the subjects are assigned at random by the investigators to the treatment and control groups.
 - (e) confounding is not a potential problem.
13. If there is a strong association between two variables then
- (a) the correlation coefficient must be close to one.
 - (b) there is a cause and effect relationship between the variables.
 - (c) knowing one variable helps a lot in predicting the other.
 - (d) both (ii) and (iii).
 - (e) all of the above are true.

IV (15 points) According to a study, high school GPA (x) and first-year college GPA (y) are mildly correlated. The data, which were collected from a Southeastern public research university, can be summarized as follows:

$$n = 2600, \quad \sum x = 9060, \quad \sum y = 7036$$
$$\sum xy = 27,918, \quad \sum x^2 = 36,168, \quad \sum y^2 = 23,145.$$

1. Find the equation of the least-square regression line.

6. Calculate the coefficient r^2 of determination and interpret the value of it.

V (5 points) Use the attached normal table. Let z denote a variable that has a standard normal distribution. *Find the values and give the R codes.*

(a) $P(z \leq -2.2)$

(b) $P(z > 1.6)$.

(c) $P(-1.2 \leq z \leq 0.8)$.

(d) $z^* = ?$ such that $P(z \leq z^*) = 0.06$?

(e) $z^* = ?$ such that $P(z > z^* \text{ or } z < -z^*) = 0.04$?

VI (5 points) Use the attached normal table. Let x denote a variable that has a normal distribution with mean $\mu = 2.0$ and standard deviation $\sigma = 0.8$. *Find the values and give the R codes.*

(a) $P(x \leq 2.1)$

(b) $P(x > 2.85)$.

(c) $P(0.32 \leq x \leq 3.5)$.

(d) $x^* = ?$ such that $P(x \leq x^*) = 0.02$?

VII (8 points) The amount of aluminum contamination (in parts per million) in plastic of a certain type was determined for a sample of 26 plastic specimens resulting the following data:

30 30 60 63 70 79 87 90 101 102 115 118 119 119

120 125 140 145 172 182 183 191 222 244 291 511

(a) Construct a boxplot that shows outliers, and comment on the interesting features of this plot.

(a) Construct a stem-and-leaf plot and comment on the interesting features of this plot.

