

MIDTERM EXAMINATION #1 VERSION A
“Introduction and Statistics Review”
February 7, 2006

INSTRUCTIONS: This exam is closed-book, closed-notes. You may use a calculator on this exam, but not a graphing calculator or a calculator with alphabetical keys. Point values for each question are noted in brackets. A table of the t-distribution is attached.

I. MULTIPLE CHOICE: Circle the one best answer to each question. Feel free to use margins for scratch work [3 pts each—42 pts total]

(1) Which of the following is *not* necessarily true?

- a. $\sum (x_i - \bar{x}) = 0.$
- b. $\sum x_i = n\bar{x}.$
- c. $\sum (x_i / y_i) = (\sum x_i) / (\sum y_i).$
- d. $\sum (\alpha x_i) = \alpha \sum x_i.$
- e. $\sum (x_i - \bar{x})^2 = (\sum x_i^2) - n\bar{x}^2.$

(2) Suppose we wish to fit the equation $y = \beta_1 + \beta_2 x$ to data by the method of least squares. This method minimizes which function of the data?

- a. $f(\beta_1, \beta_2) = \sum (y_i - \beta_1 - \beta_2 x).$
- b. $f(\beta_1, \beta_2) = \sum |y_i - \beta_1 - \beta_2 x|.$
- c. $f(\beta_1, \beta_2) = \sum (y_i - \beta_1 - \beta_2 x)^2.$
- d. $f(\beta_1, \beta_2) = \sum (\beta_1 + \beta_2 x)^2.$
- e. $f(\beta_1, \beta_2) = \sum (y_i^2 - (\beta_1 - \beta_2 x)^2)$

The next two questions assume the following. Suppose X is a Bernoulli random variable, with $\text{Prob}\{X=1\} = 0.4$ and $\text{Prob}\{X=0\} = 0.6.$

(3) The mean or expected value of X is

- a. zero.
- b. 0.24 .
- c. 0.4 .
- d. 0.6 .
- e. one.

(4) The variance of X is

- a. zero.
- b. 0.24 .
- c. 0.4 .
- d. 0.6 .
- e. one.

(5) The correlation of any random variable with itself is necessarily

- a. negative.
- b. zero.
- c. between negative one and one.
- d. exactly one.
- e. infinite.

(6) Which of the following distributions does *not* have a symmetric bell-shaped density function?

- a. normal distribution.
- b. t distribution.
- c. chi-square distribution.
- d. all of the above have bell-shaped density functions.
- e. none of the above have bell-shaped density functions.

The next two questions assume the following. Suppose a random sample of size n is drawn from some population. The population has mean μ and variance σ^2 . Consider the sample mean, defined as

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i .$$

(7) The expected value of \bar{X} is

- a. zero .
- b. μ .
- c. $\mu / (n-1)$.
- d. μ / n .
- e. μ / n^2 .

(8) The variance of \bar{X} is

- a. zero.
- b. σ^2 .
- c. $\sigma^2 / (n-1)$.
- d. σ^2 / n .
- e. σ^2 / n^2 .

(9) An estimator $\hat{\theta}$ of an unknown population parameter θ with the property that $E(\hat{\theta}) = \theta$ is said to be

- a. unbiased.
- b. asymptotically unbiased
- c. consistent.
- d. linear.
- e. efficient.

(10) If the expected value of an estimator approaches the true value of the unknown population parameter, as the sample size grows without bound, that estimator is said to be

- a. unbiased.
- b. asymptotically unbiased.
- c. consistent.
- d. linear.
- e. asymptotically normal.

(11) Which method(s) for finding an estimator requires us to know the formula for the underlying density function of the population?

- a. method of maximum likelihood.
- b. method of moments.
- c. both of the above.
- d. none of the above.

(12) A narrower confidence interval is obtained by

- a. increasing the confidence level.
- b. decreasing the confidence level.
- c. increasing the sample size.
- d. both (b) and (c).
- e. none of the above.

(13) The probability that a test will mistakenly reject the null hypothesis when it is true is called the

- a. critical point of the test.
- b. power of the test.
- c. size or significance of the test.
- d. test statistic.
- e. standard error.

(14) If the computed p-value for a test statistic is less than the size of the test, we

- a. can reject the null hypothesis.
- b. cannot reject the null hypothesis.
- c. cannot compute the test statistic.
- d. answer cannot be determined from the information given.

II. PROBLEMS: Please write your answers in the boxes on this question sheet. Show your work and circle your final answers.

(1) [Least-squares calculation: 12 pts] Suppose the following three observations on x_i and y_i are given.

<i>Observation (i)</i>	x_i	y_i
1	3	2
2	1	8
3	2	14

- a. Compute $\hat{\beta}_2$, the least-squares estimate of the slope of the line $y = \beta_1 + \beta_2 x$.

- b. Compute $\hat{\beta}_1$, the least-squares estimate of the y-intercept of the same line.

- c. Compute the three fitted values \hat{y}_i of this least-squares estimated regression line.

- d. Compute the three residuals $\hat{\varepsilon}_i$ of this estimated least-squares regression line.

(2) [Moments: 12 pts] Suppose X_1 and X_2 are random variables with the following moments.

$$E(X_1) = 6$$

$$E(X_2) = 7$$

$$\text{Var}(X_1) = 1$$

$$\text{Var}(X_2) = 4$$

$$\text{Cov}(X_1, X_2) = 0.75$$

Now let $Y = 2X_1 + 3X_2$. Compute the following and circle your final answers.

a. Compute $E(Y)$.

b. Compute $\text{Var}(Y)$.

c. Compute $\text{SD}(Y)$.

d. Compute $\text{Corr}(X_1, X_2)$.

(3) [Estimation: 12 pts] Suppose we wish to estimate the mean of a population using the following (peculiar) estimator applied to a random sample of 10 observations.

$$\hat{\mu} = 3 + \frac{1}{12} \sum_{i=1}^{10} x_i$$

Compute the following properties of the estimator under the assumption that the true population mean is $E(X_i) = 6$ and the true population variance is $\text{Var}(X_i) = 18$. Circle your final answers.

a. Compute $E(\hat{\mu})$.

b. Compute $\text{Bias}(\hat{\mu})$.

c. Compute $\text{Var}(\hat{\mu})$.

d. Compute $\text{MSE}(\hat{\mu})$.

(4) [Inference for arbitrary distribution, large sample: 18 pts] Suppose we wish to analyze the distribution of the number of children per family in a population. Let μ denote the unknown true population mean number of children per family. Observations X_i have been collected on 500 families, with the following summary values. Here, \bar{X} is the sample mean.

$$\sum_{i=1}^{500} X_i = 1025 \qquad \sum_{i=1}^{500} (X_i - \bar{X})^2 = 225$$

- a. [3 pts] Is the population distribution discrete or continuous? Justify your answer.

- b. [3 pts] Compute an unbiased estimate of μ .

- c. [3 pts] Compute a 95% asymptotic confidence interval for μ .

- d. [9 pts] Test the null hypothesis that $\mu = 2$ against the one-sided alternative hypothesis that $\mu > 2$, at 5% significance using an asymptotic test. Give

- the *value* of the test statistic
- the *critical point* from the appropriate table
- your conclusion: whether you reject the null hypothesis at 5% significance.

III. CRITICAL THINKING [4 pts] Suppose a researcher proposes to apply the following estimator for the population mean to a sample of n observations.

$$\hat{\mu} = \frac{1}{5} \sum_{i=1}^5 x_i$$

Note that only the first five observations will be used, no matter how large the sample size n . What is wrong with this estimator? Is it biased? Is it asymptotically biased? Is it inconsistent? Justify your answer.

[end of exam]