

STA 8250 Exam 1 – SAMPLE EXAM
SHOW ALL WORK

Name: _____

1. Two vectors \mathbf{x} and \mathbf{y} are orthogonal to each other. What is the length of the projection of \mathbf{x} onto \mathbf{y} ? Briefly explain your answer.

2. The following scatterplot corresponds to a sample of bivariate observations of the random vector $\mathbf{x} = (x_1, x_2)'$.

A plot belongs here that looks exactly like figure 1.14 on p.29 of the text except that it should be rotated 90 degrees.

If we are interested in determining how extreme a particular observation $\tilde{\mathbf{x}}$ in comparison with a typical observation,

a. explain why Euclidean distance is inappropriate for this task.

b. Which would be better for this situation, Karl Pearson distance or statistical distance? Briefly explain.

3. Suppose that $\mathbf{x} \sim N_p(\boldsymbol{\mu}_x, \Sigma_x)$ and $\mathbf{y} \sim N_p(\boldsymbol{\mu}_y, \Sigma_y)$ and \mathbf{x} and \mathbf{y} have covariance matrix equal to $\mathbf{0}_{p \times p}$. What is the distribution of $\mathbf{x} - \mathbf{C}\mathbf{y}$ where \mathbf{C} is a $p \times p$ matrix of constants?

4. An investigator is interested in the relationship between height and intelligence. He obtains a sample of 500 healthy adult men and measures their intelligence quotient (IQ) and height. After collecting the data, he forms 95% confidence intervals for the population mean IQ (μ_1) and the population mean height (μ_2) using the sample mean vector $\bar{\mathbf{x}}$ and sample variance-covariance matrix \mathbf{S} and the following formulas:

$$\bar{x}_1 \pm t_{.05/2}(499) \sqrt{s_{11}/n}$$

$$\bar{x}_2 \pm t_{.05/2}(499) \sqrt{s_{22}/n}$$

- a. The investigator hypothesizes that height does not depend on IQ. Without assuming this hypothesis to be true, what is the most precise statement that can be made about the probability that the joint confidence region obtained from the univariate intervals contains $\boldsymbol{\mu} = (\mu_1, \mu_2)'$?

- b. How would the formulas above for the endpoints of the univariate intervals change if we employed the Bonferroni method for simultaneous 95% confidence intervals? How would using the Bonferroni method here affect your probability statement from part (a)?

- c. Suppose that the investigator computed a sample correlation coefficient between height and IQ of 0.048. Compute a test statistic for the hypothesis that height and IQ are independent. With what distribution should you compare your test statistic?

- c. Suppose that a histogram for the lengths of the male turtles was obtained and the histogram looked skewed (unsymmetric), with more large extreme values than small extreme values. The biologist is concerned that this problem may invalidate the statistical analysis of her data. Should she be concerned about this issue? If so, briefly describe how she might deal with the problem to obtain statistically valid conclusions.