Introduction to Statistical Measurement and Modeling

Homework 2

Due: Wednesday, July 13

1. Osteoporosis analysis, continued:

Please continue our analysis to study how well osteoporosis can be detected by non-invasive measurements by analyzing the dual photon absorptiometry (DPA) data. Using statistical testing and estimation techniques with 95% confidence interval as shown in notes for Lecture 2, please address:

a) Are DPA measurements associated with osteoporosis status?

b) How strong is the evidence of association?

c) Compute and interpret a p-value for the significance of the association.

d) Estimate by how much mean DPA values differ between older women with and without osteoporosis? Describe how precisely we can determine this difference.

Complete a)-d) by hand except that you may obtain the needed rejection region critical points = $z_{1-\alpha/2}$ as well as the p-value using R. The following command obtains $Q\{1-\alpha/2\}$ of the normal distribution:

qnorm(1-α/2) (for example: if α=.1, type 'qnorm(.95)')

To obtain it for the appropriate t distribution:

qt(1-α/2,40) (here, '40' is the 'degrees of freedom' for the t)

The following command obtains the probability that a normal random variable with mean = 0 and variance = 1 is less than or equal to q:

pnorm(q)

And to obtain the same for a t with 40 degrees of freedom:

pt(q,40)

2. Using R and the osteoporosis data: Plot DPA score (Y axis) versus age (X axis), and compute the correlation between DPA score and age. Comment on the association between DPA and age in the sample of women providing our data.

3. Multiple choice: Which of the following choices best completes the statement: Two random variables (X, Y) **are statistically independent** if...? (Circle the one best answer.)

a) ... they are uncorrelated.
b) ... their covariance equals 0.
c) ... their joint density or mass function equals the product of their individual densities or mass functions
d) ... the conditional probability density or mass function of X given Y=y does not equal the individual density or mass function of X.
e) both (a) and (b)

4. Suppose you are developing a drug test which employers will administer to their workers. Suppose that 5% of the population are drug users. You determine that the accuracy of the test is as follows: 95% of drug users test positive, and 85% of non-users test negative.

(a) What is the probability that an individual who tests positive uses drugs?
(b) What is the probability that an individual who tests negative doesn't use drugs?
(c) Comment on where efforts to improve the test should be concentrated.

5. Multiple choice: Which of the following gives the most correct interpretation of a 95% confidence interval for the mean of a measurement, X? (Circle the one best answer.)

a) It is a method producing an interval that includes the mean measurement in the population with probability of no less than 0.95.
b) To write '95% CI = (-202,-79) means that there is probability = 0.95 that the population mean is between -202 and -79.
c) It is a method producing an interval that includes the mean measurement in the sample with probability of no less than 0.95.
d) To write '95% CI = (-202,-79) means that there is probability = 0.95 that the sample mean is between -202 and -79.
e) It includes measurement values X within approximately 2 standard deviations of the mean.

6. Multiple choice: A hypothesis test has probability of Type I error=0.10. What does this mean? (Circle the one best answer.)

a) There is 10% chance of not rejecting the null hypothesis.
b) There is 10% chance of not rejecting the null hypothesis when the null hypothesis is true.
c) The null hypothesis has 10% chance of being rejected.
d) The null hypothesis has 10% chance of being rejected if the alternative hypothesis is true.
e) The null hypothesis has 10% chance of being rejected if the null hypothesis is true.