Statistics 406 Problem Set 9

Due in lab, Tuesday December 4th

1. Suppose we have paired continuous measurements X_i, Y_i observed on a sample of n individuals. We are interested in whether these two values are associated (either positively or negatively). One approach is to test the hypothesis that the population Pearson correlation coefficient r_{XY} is nonzero using the Fisher transformed sample correlation coefficient as a test statistic. An alternative approach is to dichotomize the data by setting $A_i = \mathcal{I}(X_i \geq c)$ and $B_i = \mathcal{I}(Y_i \geq c)$, for some constant c, and testing the hypothesis that the population log odds ratio between A_i and B_i is zero.

   (a) Use simulation to assess the powers of the tests based on the continuous and on the dichotomized data when the continuous data are normally distributed. Use a range of values for c and for the population correlation coefficient r. If the log odds ratio is undefined for a particular data set, consider the null hypothesis of no association to be accepted for that data set. Briefly describe your findings.

   (b) For the special case $c = 0$, repeat your simulation analysis, but generating the continuous data from t distributions instead of from normal distributions. Consider both 2 and 1 degrees of freedom. Compare your findings to your results from part (a).

2. Suppose we observe paired continuous data X_i, Y_i, and we are interested in $E(Y_i - X_i)$ (the treatment effect in the case that X_i and Y_i represent pre-treatment and post-treatment measures). For the purposes of this problem, assume that X_i is standard normal, and

$$Y_i = X_i + c + \epsilon_i,$$

   where c is a constant, $E\epsilon_i = 0$ and $\mathrm{var}\epsilon_i = \sigma^2$.

   (a) What are the population values for the mean and variance of $Y_i - X_i$ and of $\bar{Y} - \bar{X}$?

   (b) What is the population Pearson correlation coefficient between X_i and Y_i?

   (c) Suppose we use the unpaired two-sample Z-test to compare $\bar{Y}$ to $\bar{X}$. What do you expect the scaling factor $\hat{\sigma}_X^2/n_x + \hat{\sigma}_Y^2/n_y$ to be?

   (d) Derive expressions for the approximate powers of the two-sample Z-test and the paired Z-test in this case. For purposes of deriving the power, assume that $\hat{\sigma}_X^2 = \sigma_X^2$ and $\hat{\sigma}_Y^2 = \sigma_Y^2$.

   (e) Use R to evaluate your approximate power expressions from part (d). Produce a table showing the relative power (the ratio of the paired test power to the unpaired test power) for a range of values of c and σ^2.