

STAT 554: Bonus HW

due 7:45 PM on May 12, 2008 (only a 15 minute grace period — cannot be turned in after the final exam period)

Each of the ten parts is worth 1 point. These are fairly straightforward exercises. In order to figure out how to do them you just need to read over pertinent parts of Units 5, 7 and 8. Please present easy to grade solutions and clearly indicate your final answers (but show a little work also (but don't give me a huge amount of paper to go through)).

These exercises are relatively straightforward. Part of the purpose of this assignment is to determine whether you've gained the ability to read a description of a statistical procedure and carry it out. Other than what I may mention while lecturing, I don't intend to tell you step-by-step how to do the necessary computations using software. Rather, I intend to let you figure out how to do things on your own.

The data sets, with some descriptive information, was included at the end of the HW #6 handout distributed in class. The data values (but not the descriptive information) are also available on the course web site.

1) Consider the ball bearing data (supplied with this assignment). Letting σ_1^2 and σ_2^2 be the variances for the distributions underlying the samples from the first and second production lines, test $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 \neq \sigma_2^2$ using each of the procedures indicated below, and report a p-value in each case.

- (a) Use an F test. (See p. 261 of Miller's *Beyond ANOVA*.) (*Note:* Although the F test is perfect if you have exact normality, it is very nonrobust. If you're dealing with heavy-tailed distributions it can produce p-values which are misleadingly small, and for light-tailed distributions (like the ones you're considering in this problem, perhaps) it can be appreciably conservative.)
- (b) Use the version of Levene's procedure based on the $(x_{ij} - \bar{x}_{i.})^2$ values. (The bottom half of p. 269 of Miller's *Beyond ANOVA* gives alternate versions.)
- (c) Use the Box-Andersen test (aka the *APF* test). (See p. 271 of Miller's *Beyond ANOVA*.)

2) Consider the data concerning cork deposits (supplied with this assignment). Let x_i be the total weight of the deposits in the east and west directions for the i th tree and let y_i be the total weight of the deposits in the north and south directions for the i th tree. Give a point estimate and a 99% confidence interval for $E(Y_i/X_i)$.

3) Consider the data concerning bricks (supplied with this assignment).

- (a) Give the value of Pearson's correlation coefficient and also give a scatter plot of mortar dry density (y) versus mortar air content (x).
- (b) Give the value of Spearman's rank correlation coefficient.
- (c) Give the estimate of Kendall's tau.

4) Consider the data concerning corn (supplied with this assignment). Let x_i be the plant density for the i th plot and let y_i be the mean cob weight for the i th plot.

- (a) Do a least squares fit of the model $y_i = \alpha + \beta x_i + e_i$ and give a point estimate of β . Also, assume approximate normality and give a 90% confidence interval for β .
- (b) Give the value of the studentized residual associated with the first data point, (137, 212), and give a probit plot of the studentized residuals.
- (c) Give a residual plot, plotting studentized residuals against plant density.