STA244 **1/22/2001**

Homework 2

Due 1/29/2001

1. (From CB 11.40) Consider the standard simple linear model with normal errors that has been re-parameterized as

$$Y_t = \mu_{\bar{Y}} + \beta t + \epsilon$$

where $t = (X - \bar{X})$ so that $\hat{\mu}_{\bar{Y}} = \bar{Y}$ and $\hat{\beta}$ are independent. Extend Scheffé's procedure to construct simultaneous prediction intervals for predicting future Y^* at t for all t. That is find a constant M_α such that

$$P\left(\frac{\left|Y^* - (\hat{\mu}_{\bar{Y}} + \hat{\beta}t)\right|}{s\sqrt{1 + \frac{1}{n} + \frac{t^2}{S_{XX}}}} \leq M_\alpha \forall\ t\right) = 1 - \alpha$$

or

$$P\left(\max_t \frac{\left(Y^* - (\hat{\mu}_{\bar{Y}} + \hat{\beta}t)\right)^2}{s^2(1 + \frac{1}{n} + \frac{t^2}{S_{XX}})} \leq M_\alpha^2\right) = 1 - \alpha$$

   (a) Show that if a,b,c, and d are constants with $c > 0$ and $d > 0$ that

$$\max_t \frac{(a + bt)^2}{c + dt^2} = \frac{a^2}{c} + \frac{b^2}{d}$$

   (b) Use the result in (a) to find the distribution of

$$\max_t \frac{\left(Y^* - (\hat{\mu}_{\bar{Y}} + \hat{\beta}t)\right)^2}{s^2(1 + \frac{1}{n} + \frac{t^2}{S_{XX}})}$$

   Hint: rewrite $Y^* = \mu_{\bar{Y}} + \beta t + \epsilon$, *where* $\epsilon \sim N(0, \sigma^2)$.

   (c) Use this to find a value of M_α. If the distribution does not exist in closed form, use moment matching to show how to find a value of M_α as an approximation. (See example 7.2.3 in Casella and Berger for moment matching) *There may be an error in the statement/setup of the problem in Casella and Berger 11.40 (c) - be careful in your proof!*

2. The data in the file `oldfaith` (download from the course calendar) gives information about eruptions of the Old Faithful Geyser in Yellowstone National Park during October 1980. Variables are `duration` in seconds of the current eruption and `interval`, the time in minutes to the next eruption. The data were collected by volunteers, and except for the period from midnight and 6 AM, this is a complete record of eruptions for that month. As Old Faithful is an important tourist attraction, the park service would like to use these data to obtain a prediction equation for the time to the next eruption using the length of the current eruption.

(a) Use simple linear regression to obtain a prediction equation for `interval` using `duration`. You may use any software package that you would like. *Briefly* summarize your results in a way that might be useful for the nontechnical personnel who staff the visitors center. Include (with explanation) the following in your typed summary (max of 1 page double spaced with 12 point font):

(b) Construct a 95% confidence interval for E(interval | duration = 250 secs). Construct a 95% prediction interval for interval given duration = 250 secs. Explain to the staff when it would be appropriate to use each of these.

(c) Estimate the 0.90 quantile of the conditional distribution of future intervals between eruptions for durations of 250 secs, assuming a normal population.

(d) Suppose the park service would like to calculate prediction intervals for all possible time points in the range of the observed `duration` and have a computer display that automatically displays a prediction interval for the time until the next eruption. Find the value of M_α using the results for the Scheffé intervals from problem (1). Suppose a person just arrives at the end of an eruption that lasted 250 seconds. Give the two prediction intervals (point-wise, and Scheffé). Which of the two prediction intervals is best from the visitors point of view? the park service's point of view?

3. Problem 7.5 in CW (do not turn in)
4. Problem 7.6 in CW
5. Problem 7.9 in CW
6. Problem 7.10 in CW