

Stat 322 (F17) - Assignment #4

(Due Monday, Dec. 4 at 4:00 pm in appropriate STAT 322 slot in assignment box #15 outside the Math Tutorial Centre (MC 4066/4067). Electronic submissions or in-class submissions will not be accepted.

- 1) Consider the *inventory* dataset introduced in Example 1 of Chapter 5 and discussed in class.
- Estimate θ , the mean count error (i.e. the mean difference in actual and stated numbers) **of all items that are in error**.
 - Obtain a 95% confidence interval for θ . Interpret this interval in the context of the study.

You will need R to obtain the standard error of the estimate - in particular, to calculate the

estimated standard deviations of the residuals, $\sqrt{\frac{\sum_{i=1}^n \hat{r}_i^2}{n-1}}$, where

$$\hat{r}_i = y_i - \hat{\theta}x_i$$

$$x_i = \begin{cases} 1 & \text{if } i\text{th item in error} \\ 0 & \text{otherwise} \end{cases}$$

y_i = actual number – stated number of i th item

The command `x<-as.numeric(y!=0)` will give the values of x (0,1) that indicate whether the i th item was in error.

- 2) Consider the farm and residential water quality example in Chapter 7 of the course notes. Suppose information is also available on the farm size, x , (hectares) of all **farms** (but not houses) in the frame. Sample and strata means for the two farm strata are provided below. Assume that scatterplots of sodium concentration vs farm size for these two strata suggest a linear relationship between these two variates.

	<u>Sample mean farm size</u>	<u>Strata mean farm size</u>
Strata #1 (with animals)	309	322
Strata #2 (without animals)	268	258

Suppose the regression model

$$Y_i = \alpha + \beta(x_i - \bar{x}) + R_i \quad R_i \sim G(0, \sigma)$$

is fit to farm size and sodium concentration for both farm stratas, yielding the fitted lines

$$\hat{\mu}_1(y)_{reg} = \hat{\mu}_1(y) + .85(\mu_1(x) - \hat{\mu}_1(x))$$

$$\hat{\mu}_2(y)_{reg} = \hat{\mu}_2(y) + .68(\mu_2(x) - \hat{\mu}_2(x))$$

with respective residual sum of squares, $SS(\text{Res})_1 = 186332$, $SS(\text{Res})_2 = 98878$.

- Use regression estimates of the farm strata means (as well as the SRS estimate of the house strata mean) to obtain an estimate of the mean sodium conc. of all wells in the frame.
- Calculate the standard error of this estimate.
- Provide a 95% confidence interval for the mean sodium concentration of all wells in the frame. Interpret this interval in the context of the study.