STAT 571A — Advanced Statistical Regression Analysis

Chapter 4 NOTES

Simultaneous Inferences and Other Topics

Simultaneous Inferences

- **(Almost) all of the inferences we've discussed have been Pointwise: they apply to <u>one and only one outcome</u>:**
  - **one conf. interval on one parameter, or**
  - **one hypoth. test on one parameter**
- **To make <u>multiple</u>, <u>simultaneous</u> inferences (hypoth. tests or conf. intervals) on g > 1 parameters or mean responses, we adjust the tests or conf. regions for the multiplicity.**

Multiplicity & FWER

- Why adjust? If we do not correct for the multiple inferences, error rates will be too high (⇔ conf. levels will be too low).
- When $g > 1$ inferences are applied to the same set of data, the Familywise (false positive) Error Rate, or FWER, is

  P[any false positive error(s) among the g inferences]
- Goal: keep the FWER $\leq \alpha$.

Bonferroni's Inequality

- Our mainstay adjustment is the (conservative) Bonferroni correction, based on **Bonferroni's Inequality**:
- As in Equation (4.2), for any events A_k,

$$P[\bar{A}_1 \cap \bar{A}_2 \cap \cdots \cap \bar{A}_g] \geq 1 - \sum_{k=1}^{g} P[A_k]$$

- This can be manipulated to place upper bounds on the FWER, or lower bounds on the FW conf. level.

Bonferroni's Inequality (cont'd)

- For instance, suppose A_1 is 'noncoverage' of a parameter β_0 and A_2 is 'noncoverage' of another parameter β_1.
- Then, the complementary events, $\overline{A}_k$, are correct 'coverage.' Bonferroni tells us that
  P[jointly covering both] ≥
  1 – P[noncover β_0] – P[noncover β_1]
- Say P[noncover β_j] = α. Then
  P[jointly covering both] ≥ 1 – α – α = 1 – 2α.
- (So, divide each orig. α by 2 to get ≥ 1 – α.)

Bonferroni Adjustment

Suppose we study the SLR model with $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$. A pointwise conf. int. on either β_j is

$$b_j \pm t\left(1 - \frac{\alpha}{2}; n-2\right)s\{b_j\}$$

But from Bonferroni, a FW conf. statement on both (g = 2) β_j parameters is

$$b_j \pm t\left(1 - \frac{(\alpha/2)}{2}; n-2\right)s\{b_j\}$$

for j = 0,1.

Bonferroni Adjustment (cont'd)

- Notice what this does: in effect, it simply changes the α-value in any critical point to α/g.
- Above, g = 2, so use $b_j \pm B\, s\{b_j\}$, where

$$B = t\left(1 - \tfrac{\alpha}{4};\, n-2\right)$$

or, more generally

$$B = t\left(1 - \tfrac{(\alpha/g)}{2};\, n-2\right)$$

is the Bonferroni-adjusted critical point

Bonferroni Conf. Intervals on $E\{Y_h\}$

An important example is with multiple conf. intervals for the mean response, $E\{Y_h\}$, at any set of $g > 1$ predictor values X_h ($h = 1,\ldots,g$).

Here, the family of (conservative) simultaneous conf. intervals becomes

$$\hat{Y}_h \pm B\, s\{\hat{Y}_h\}$$

($h = 1,\ldots,g$) for $B = t(1 - ½\{\alpha/g\};\ n-2)$.

Working-Hotelling-Scheffé (WHS) Intervals on $E\{Y_h\}$

- We can alternatively apply the WHS conf. band to build multiple conf. intervals for the mean response, $E\{Y_h\}$, at any set of g >1 predictor values X_h ($h = 1,\ldots,g$).
- The family of (conservative) simultaneous conf. intervals is

$$\hat{Y}_h \pm W\, s\{\hat{Y}_h\}$$

where the **WHS-adjusted critical point** is based on $W^2 = 2F(1-\alpha;\, 2,\, n-2)$.

WHS or Bonferroni?

- **For any set of g > 1 predictor values X_h (h = 1,...,g), both the B and W crit. points are valid, if conservative.**
- **So, use the WHS value if W ≤ B, and use Bonferroni if B < W.**
- **Notice: The WHS is exact for *all* X-values. So, it can be used for post hoc intervals on $E\{Y_h\}$ at any finite collection of X_h's. (It is the only valid conf. int. for post hoc "data snooping.")**

Toluca Example (cont'd)

- In the Toluca Data example (CH01TA01), suppose we want g = 3 (three) 90% conf. int's at X_h = 30, 65, 100. Here, $df_E = 25 - 2 = 23$.
- The Bonfer. point is $t(1 - \frac{1}{2}\{0.10/3\}; 23)$:

```
> qt( 1-(.10/6), 23 )
  2.263728
```

- The WHS point is $\{2F(1 - 0.10; 2, 23)\}^{1/2}$:

```
> sqrt( 2*qf(.90,2,23) )
  2.258003
```

- Since W ≤ B, use the WHS adjusted crit. point.

§4.4: Regression Thru the Origin

- If $\beta_0 = 0$, the SLR model simplifies to $E\{Y_i\} = \beta_1 X_i$ $(i = 1,\ldots,n)$.
- The LS estimate of β_1 is $b_1 = \sum Y_i X_i / \sum X_i^2$
- The corresp. std. error is

$$s\{b_1\} = \sqrt{MSE/\sum X_i^2}$$

  where the MSE now has **n–1** df.
- A $1 - \alpha$ conf. interval for β_1 is

$$b_1 \pm t(1 - \alpha/2;\, n-1)s\{b_1\}$$

Regression Thru the Origin (cont'd)

For inferences on $E\{Y_h\}$, use:

- LS estimator: $\hat{Y}_h = b_1 X_h$
- std. error: $s\{\hat{Y}_h\} = |X_h| \sqrt{\dfrac{MSE}{\sum X_i^2}}$
- (pointwise) conf. int.:

$$\hat{Y}_h \pm t\left(1 - \tfrac{\alpha}{2};\ n-1\right) s\{\hat{Y}_h\}$$

Regression Thru the Origin (cont'd)

For prediction of a future $Y_{h(new)}$ at $X_{h(new)}$, use:

- LS estimator: $\hat{Y}_{h(new)} = b_1 X_{h(new)}$
- prediction error:

$$s\{pred\} = \sqrt{MSE\left(1 + \frac{X_{h(new)}^2}{\sum X_i^2}\right)}$$

- (pointwise) prediction int.:

$$\hat{Y}_{h(new)} \pm t\left(1 - \tfrac{\alpha}{2};\ n-1\right)s\{pred\}$$

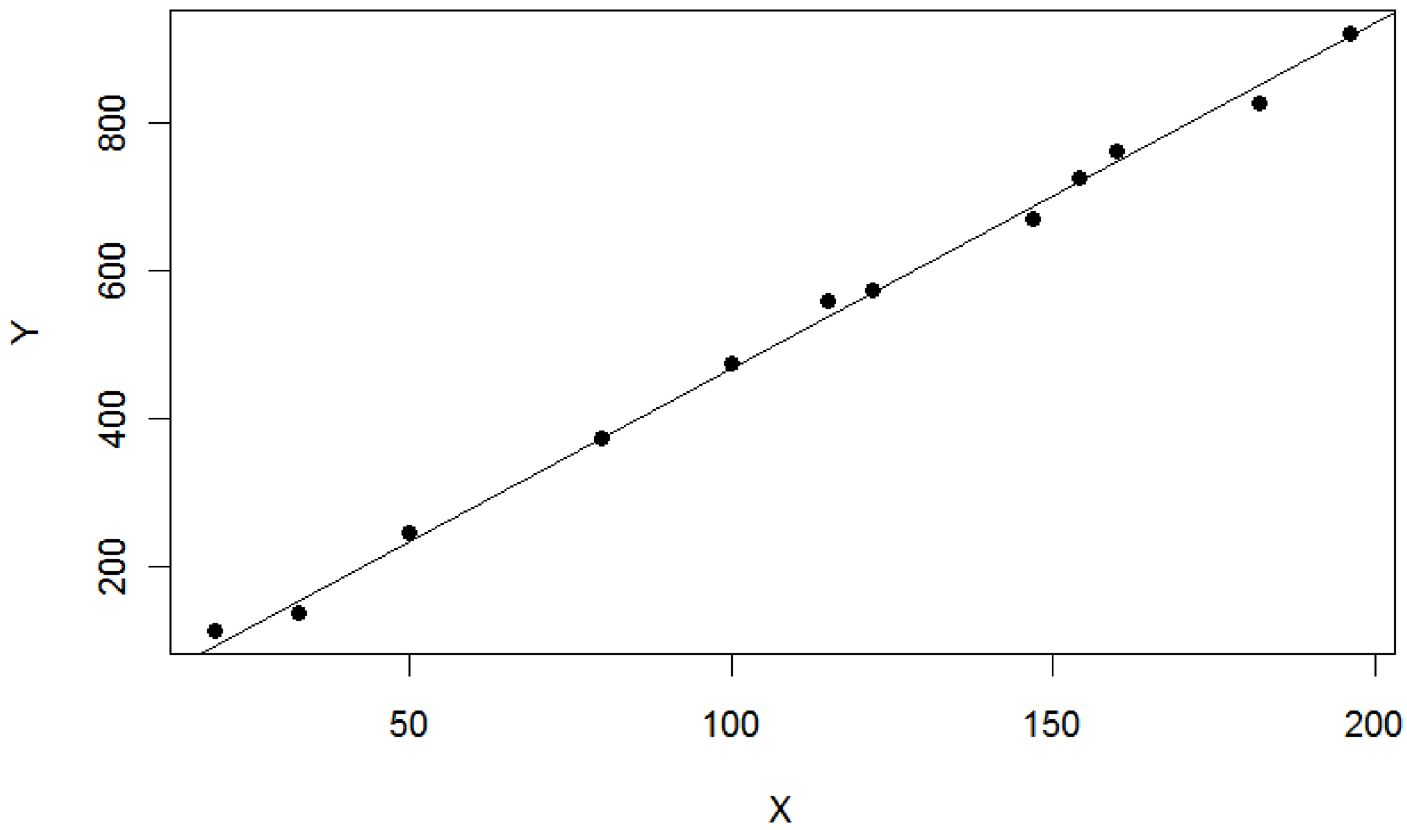
Warehouse Data (CH04TA02)

- X = work units, Y = variable labor costs.
- Expect E{Y} = 0 when X = 0, so fix $\beta_0 = 0$:

```
> CH04TA02.lm = lm(Y ~ X -1 )
> summary( CH04TA02.lm )
 Call:
 lm(formula = Y ~ X - 1)
 Coefficients:
   Estimate Std. Error  t value  Pr(>|t|)
 X  4.68527    0.03421      137   <2e-16
> confint(CH04TA02.lm)
        2.5 %    97.5 %
 X  4.609989  4.760559
```

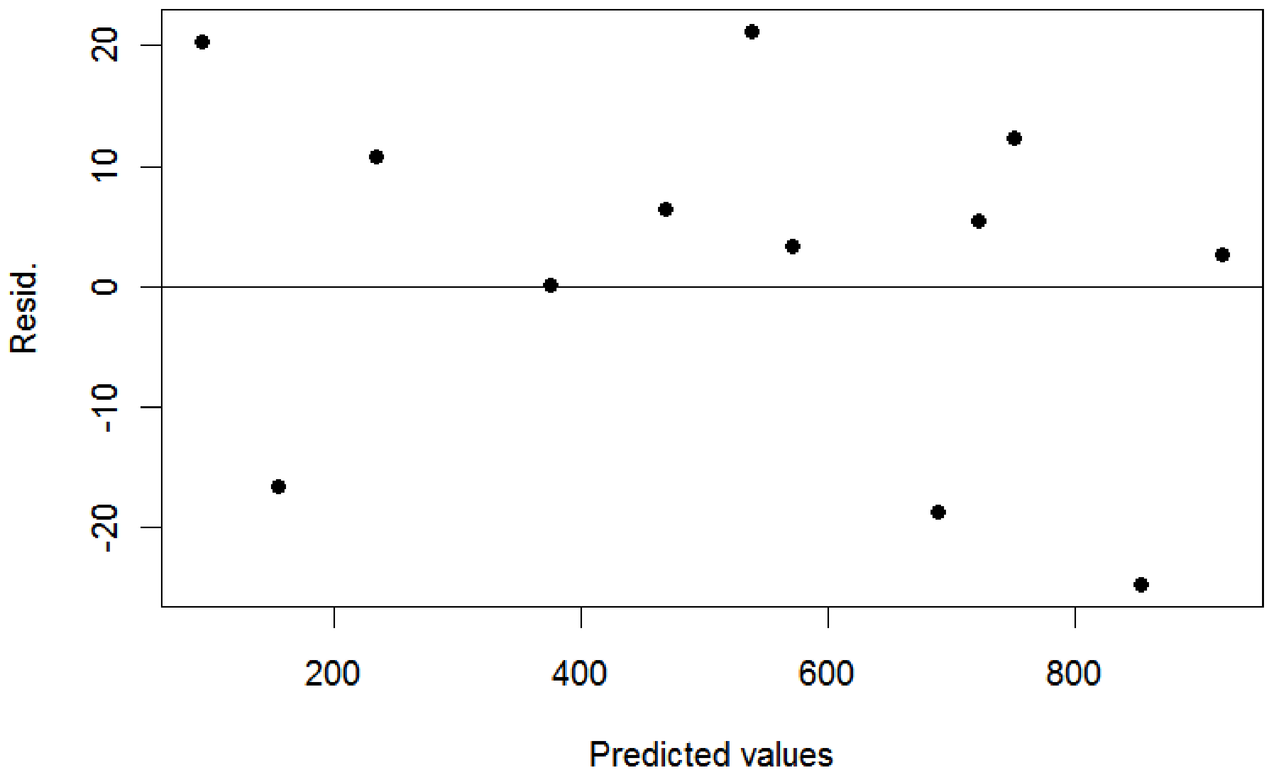
Warehouse Data (CH04TA02) (cont'd)

```
> plot( Y~X, pch=19 )
> abline( lm(Y ~ X-1) )
```



Warehouse Data (CH04TA02) (cont'd)

```
> plot( resid(lm(Y ~ X-1)) ~
                predict(lm(Y ~ X-1)) )
> abline( h=0 )
```



§4.6: Inverse Prediction

- **We can <u>reverse</u> the prediction effort and ask, what value of X produces a given Y? This is an inverse prediction problem.**
  - **also called: "inverse regression" or "calibration"**
- **Assume the SLR model: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, with $\varepsilon_i \sim$ i.i.d. $N(0,\sigma^2)$.**
- **Given $Y_{h(new)}$, we want to find the $X_{h(new)}$ that yields this $Y_{h(new)}$.**

Inverse Prediction (cont'd)

Clearly, if $\hat{Y}_h = b_0 + b_1 X_h$, we can invert this into $\hat{X}_{h(new)} = \dfrac{Y_{h(new)} - b_0}{b_1}$ for $b_1 \neq 0$.

The prediction error can be found as

$$s\{\hat{X}_{h(new)}\} = \sqrt{\frac{MSE}{b_1{}^2}\left(1 + \frac{1}{n} + \frac{(\hat{X}_{h(new)} - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right)}$$

from which a $1-\alpha$ prediction int. is simply

$$\hat{X}_{h(new)} \pm t\left(1 - \tfrac{\alpha}{2};\ n-2\right) s\{\hat{X}_{h(new)}\}$$

§4.7: Optimal Design

- **Notice that terms containing the X_i's, such as $\bar{X}$ and $\sum(X_i - \bar{X})^2$, appear throughout these various expressions. <u>If</u> the X_i's are under control of the investigator, we can manipulate these quantities.**
- **Why? We might be able to make a std. error smaller and hence tighten a conf. int. or increase power in an hypoth. test.**

Optimal Design (cont'd)

Consider the margin of error on the conf. int. for β_1:

$$\text{MOE} = \pm\, t\left(1 - \tfrac{\alpha}{2};\, n-2\right) s\{b_1\}$$

$$= \pm\, t\left(1 - \tfrac{\alpha}{2};\, n-2\right) \sqrt{\frac{\text{MSE}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

Clearly, as $\sum_{i=1}^{n}(X_i - \bar{X})^2 \uparrow$ the MOE $\downarrow$ and the conf. int. will get tighter.

Optimal Design (cont'd)

- So, wherever possible, always try to select the spacing and number (including replicates) of the X_i's to optimize the conf. int's and hypoth. tests.
- See the exposition on pp. 170-172.