BIOS 145 Lab #4 9/17/97

- 1st hour: KKM Ch.13: #2a, c(1), d(1), e, maybe #3
- 2nd hour: questions
- Other:

    1. Extra 1 hr. for Thurs/Fri?
    2. Problems downloading datasets?
    3. Using word processors to edit ASCII file
       -- Make sure to save file as “Text only”.
    4. Web page -- problems, suggestions, etc.
    5. Lab overheads will be on web page, notebooks in HSL

Lack of Fit (LOF) Tests:

-- Refer to KKM pp.238-240, LEC notes pp.54-65

-- Will try to use LEC notation as closely as possible

-- Can only be applied if have *replicate* observations, i.e. at least 2 experimental units have the same X value.

Setup:

- n total observations in the dataset
- D distinct values of X
- $r = n - D$ replicates

Example:

Subject	X	Y
1	10	130
*2	10	145
*3	10	128
4	15	140
5	20	135
*6	20	145

Here, we have $n = 6$ observations, $D = 3$ distinct X values, and thus $r = 6 - 3 = 3$ replicates (marked by *)

- From algebra, we know that if we have D distinct points, we can fit a polynomial of (maximum) order $D-1$.

- Classical LOF test compares fit of model of order $D-1$ with fit of model under consideration

  e.g. in dataset above, $D-1=2$ so we might compare

  $$Y = \beta_0 + \beta_1 X + e$$

  with

  $$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + e$$

- In random variable notation (LEC),

  • $X_1, X_2, \ldots, X_D$ are the distinct values of X

  • $Y_{11}, Y_{12}, \ldots, Y_{1n_1}$ observed at X_1
  $Y_{21}, Y_{22}, \ldots, Y_{2n_2}$ observed at X_2
  $\vdots \quad \vdots \quad \vdots$
  $Y_{D1}, Y_{D2}, \ldots, Y_{Dn_D}$ observed at X_D

-

- MSPE $=$ SSPE $/(n-D) =$ pure-error estimate of σ^2
- SSE $=$ SSPE $+$ SSLF
- MSLF $=$ SSLF$/(D-2)$
- Test statistic: (H_0: model fits)

KKM Ch13 #2:

a. Complete the ANOVA table for straight-line regression of TIME (Y) on INC (X) using the given printout:

Source	df	SS	MS	F
Regression	1	442.914	442.914	4.16
	5	15.201	3.040	
Total	24	2891.05		

- Here,
  - SSR = MSR = SS(Degree 1)
  - SSPE given, MSPE = SSPE / 5
  - SSLF = SS(Lack of Fit) + SS(Degree 2) + SS(Degree 3), so MSLF = SSLF / 18
  - Overall $F = \frac{\text{MSR}}{\text{MSE}} = \frac{\text{MSR}}{(\text{SSLF+SSPE})/23}$
  - LOF $F = \frac{\text{MSLF}}{\text{MSPE}}$

c) Plot scatter diagram and fitted regression equation
-- See plot

d) Calculate R^2

For the straight-line model,

$$R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{442.914}{2891.045} = .1532$$

e) Carry out F-tests

- For the hypothesis H_0: "no significant straight-line regression", we use the overall $F = 4.16$.
  - Under H_0, overall $F \sim F_{1,23}$. The p-value is .053 so we don't reject H_0 at the α = .05 level.
- For the hypothesis H_0: "straight-line model fits", we use we use the LOF $F = 44.5$
  - Under H_0, LOF $F \sim F_{18,5}$. The p-value is clearly much smaller than .05 so we reject H_0 at the α = .05 level.

g) Do you think straight-line model "fits"?