Topic 1: Multiple Linear Regression

1. Topics

- Review Simple Linear Regression (SLR) and Multiple Linear Regression (MLR) with two predictors

- More Review of MLR via a detailed example

- Model checking for MLR

  — Keywords: MLR, scatterplot matrix, regression coefficient, 95% confidence interval, t-test, adjustment, adjusted variables plot, residual, dbeta, influence

2. Learning objectives

- Understand and explain what MLR is and how it is used to draw inferences
- Interpret MLR coefficients correctly
- Critically evaluate a multiple linear regression analysis to ensure that substantive findings are appropriate given the data
- Interpret the effects of length of stay and employee salary on per capita health care expenditures

3. What is regression?

- What is the “Regression” of Y on X ?

  — Average Y at each value of X

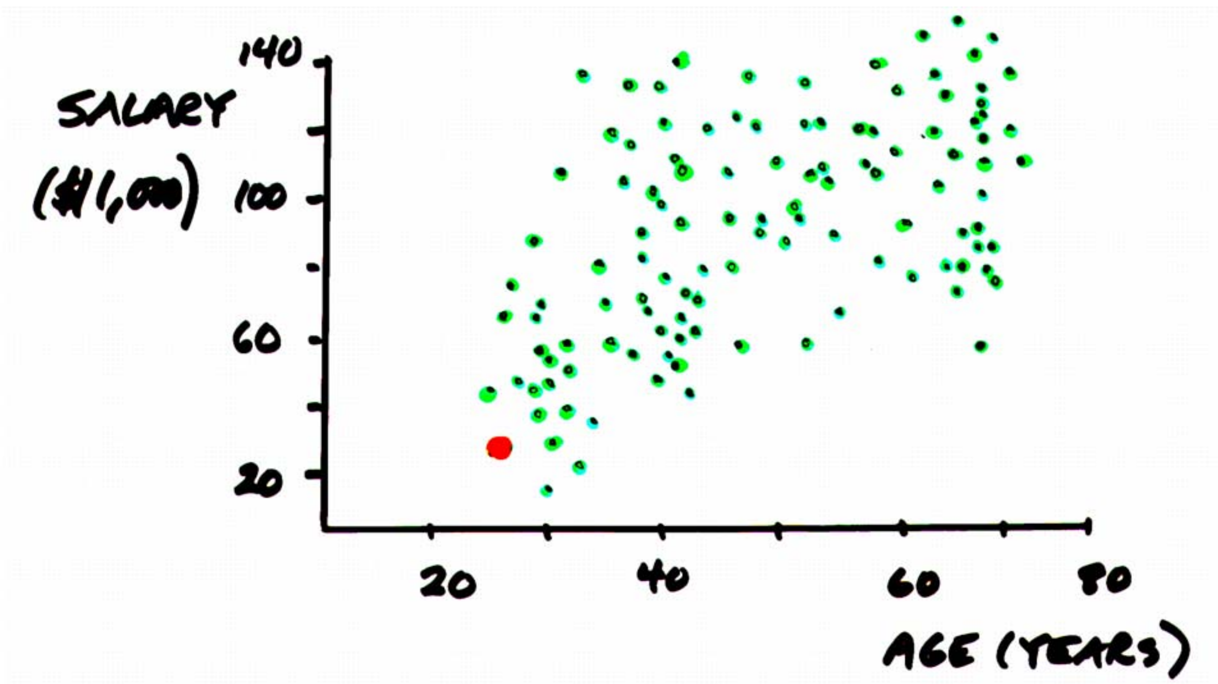
  eg, Regression of salary on age (and, possibly, other Xs)

- Notational convention -- express as either:

  Ave (Y|X), where “ | ” = Given

or,

  E (Y|X), where “ E ” = Expectation

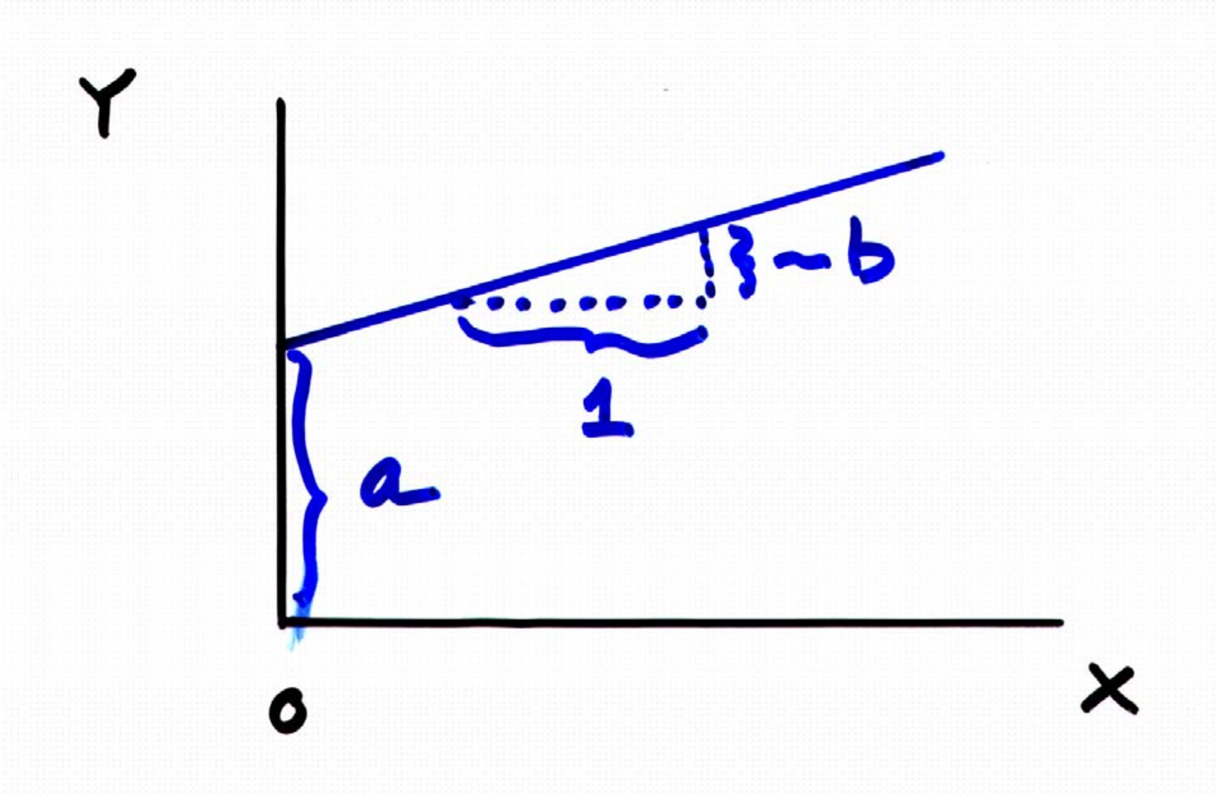


- E (Y|X) is a *linear* function of X

3.1 Simple linear regression (SLR)

- $E(Y|X) = a + bx$

  a = intercept
  b = slope



- $E(Y|X) = \beta_0 + \beta_1 X$

- May use either (a, b) or (β_0, β_1) -- arbitrary notation!

3.2 Multiple linear regression (MLR)

- In simple linear regression (SLR)

  One X

  $E(Y \mid X) = \beta_0 + \beta_1 X$

- In Multiple linear regression (MLR)

  More than one X

  $$E(Y \mid X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p = \sum_{j=0}^{p} \beta_j X_j$$

- Salary = $\beta_0 + \beta_1$ Age + β_2 Grad Date

- Units are ($1,000) = ($1,000) + ($1,000/Yr) * (Yr) + ($1,000/Yr) * (Yr)

3.3 MLR with two predictors

MLR model

$$Y_i = E(Y_i \mid X_{i1}, X_{i2}) + \epsilon_i$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$$

or,

Response = Prediction using Xs + Error

The errors $\epsilon_1, \epsilon_2, ..., \epsilon_n$ are independent and identically distributed (iid) normal variates, mean 0, variance σ^2 (unknown constant); or, more briefly, ϵ_i iid $N(0, \sigma^2)$, $i=1,...,n$

3.4 Interpretation of MLR coefficient

- β_1 is the expected change in Y when X_1 increases one unit and X_2 remains fixed; or

- β_1 is the difference between average Ys for two populations that differ in X_1 by one unit and have the same X_2

- Example of a regression equation

$$Y = \beta_0 + \beta_1 (\text{Age} - 40) + \beta_2 \text{Gender} + \epsilon$$

$$\text{Salary} = 50 + 1 (\text{Age} - 40) - 3 \text{Gender} + \epsilon$$

Salary in $1,000s,
Age in years and
Gender = 0 if male and 1 if female

What is the average salary for 50 year old males?

$$\text{Ave}(Y) = 50 + 1 (50-40) - 3(0) = \$60K$$

- Interpretations

  — β_0 - average salary ($1,000) for a 40 year old male

  — β_1 - increase in average salary for every year of age for a given gender (M or F)

  — β_2 - difference in average salary for women vs. men of the same age

- Consider observations for two people and subtract expected responses:

  (Y_1, X_1, X_2) and $(Y_2, (X_1 + 1), X_2)$

$$
\begin{array}{lrcl}
 & E\,Y_2 & = & \beta_0 + \beta_1(X_1 + 1) + \beta_2 X_2 \\
- & E\,Y_1 & = & \beta_0 + \beta_1(X_1) \quad + \beta_2 X_2 \\
\hline
 & E\,(Y_2 - Y_1) & = & 0 + \beta_1 \quad + 0 \\
 & & = & \beta_1
\end{array}
$$

- β_1 is the expected difference in Y corresponding to a unit difference in X_1 *given X_2 is the same*

4.1 Data and Question

The data come from the book by Pagano and Gauvreau: *Principles of Biostatistics*

Y - Average expenditure ($s) per admission

X_1 - Average length of stay (days)

X_2 - Average employee salary ($s)

$n = 51$; 50 states + DC

Scientific Question:

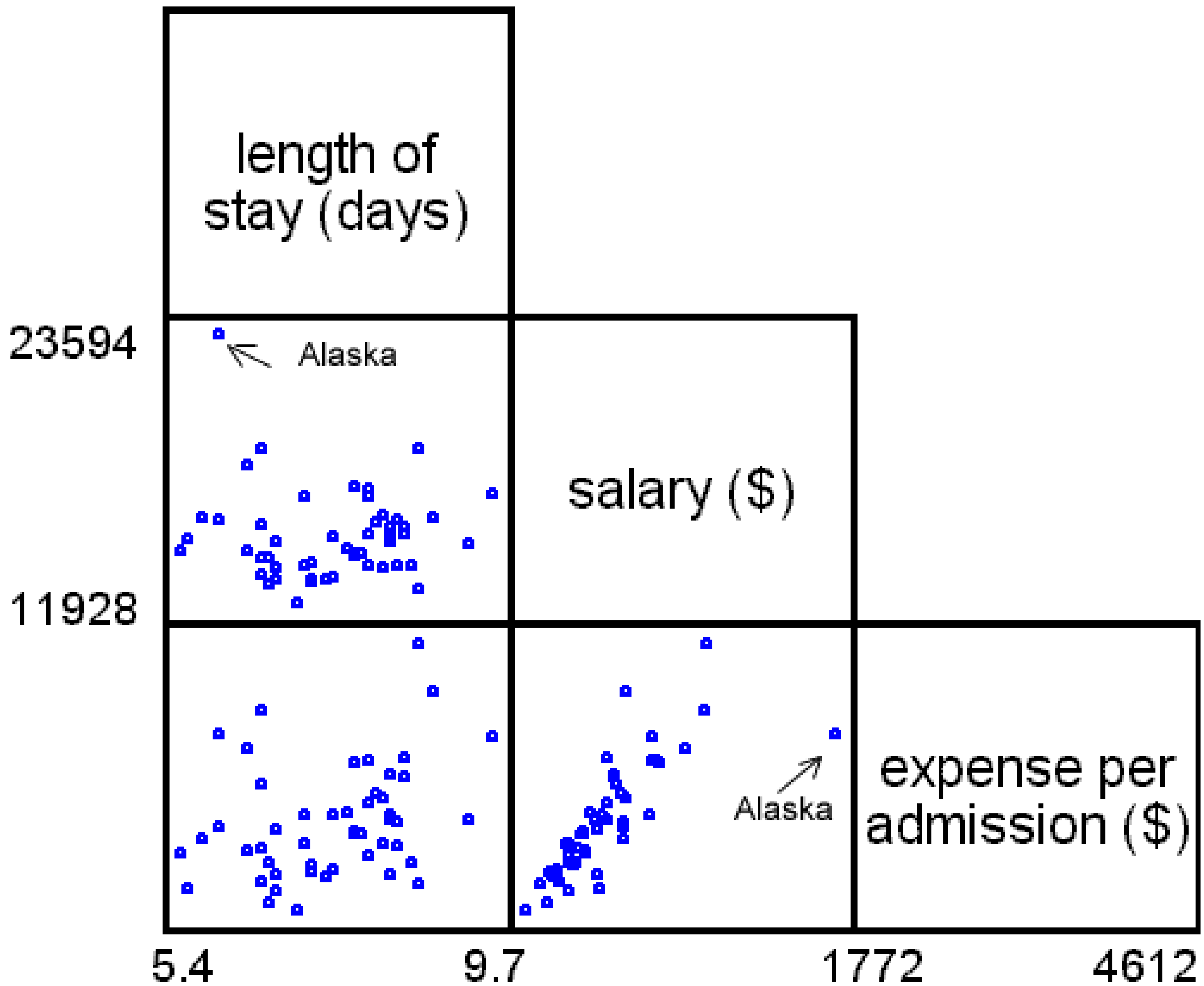
How is per capita expenditure related to:

— Length of stay

— Employee salary

4.2 Display the Data

- **Use a scatterplot matrix**



4.3 Model and interpretation of coefficients

- **MLR model**

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

where

Y **Expenditure per admission in $s**

X_1 **LOS in days**

X_2 **Salary in $s**

4.3 Model and interpretation of coefficients

- **Interpretation of coefficients**

β_0 - expected per capita expenditure when LOS = 0 and salary = 0;

Nonsense! But OK- don't need β_0 to answer the questions posed

β_1 - difference in expected per capita expenditure ($s) for two states with same average salary but LOS that differs by one day

β_2 - difference in expected per capita expenditure ($s) for two states with same average LOS but salary that differs by one dollar

---

4.4 Results and interpretation

- **Stata code**

regress expadm los salary

- **Stata log**

```
. * Part b.  MLR:  expense per admission on salary and los
.
. regress expadm los salary

  Source |       SS       df       MS                  Number of obs =      51
---------+------------------------------               F(  2,    48) =   75.55
   Model |  13840987.8     2  6920493.90               Prob > F      =  0.0000
Residual |  4396616.24    48  91596.1716               R-squared     =  0.7589
---------+------------------------------               Adj R-squared =  0.7489
   Total |  18237604.0    50  364752.081               Root MSE      =  302.65

------------------------------------------------------------------------------
  expadm |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     los |   213.7967   42.20769      5.065   0.000       128.9325     298.661
  salary |    .248994   .0217992     11.422   0.000       .2051638    .2928241
   _cons |  -2582.736     464.77     -5.557   0.000      -3517.219   -1648.254
------------------------------------------------------------------------------
```

4.4 Results and interpretation

- **Summary of results**

Variable	$\hat{\beta}$	$SE_{\hat{\beta}}$	*t*	*95% CI*
Intercept	*-2,583*	*465*	*-5.6*	*----*
LOS (days)	*214*	*42.2*	*5.1*	*(129, 299)*
Salary ($s)	*0.249*	*0.022*	*11.4*	*(0.205, 0 .293)*

- **Interpretation**

  — We estimate that expected expenditure per admission will be $214 higher (95% CI: $129-299) in a state whose average LOS is one day longer than another state with the same average employee salary

  — For two states with the same LOS, we estimate an additional average expenditure per admission of $249 (95%: 205-293) by a state whose salary is $1,000 higher

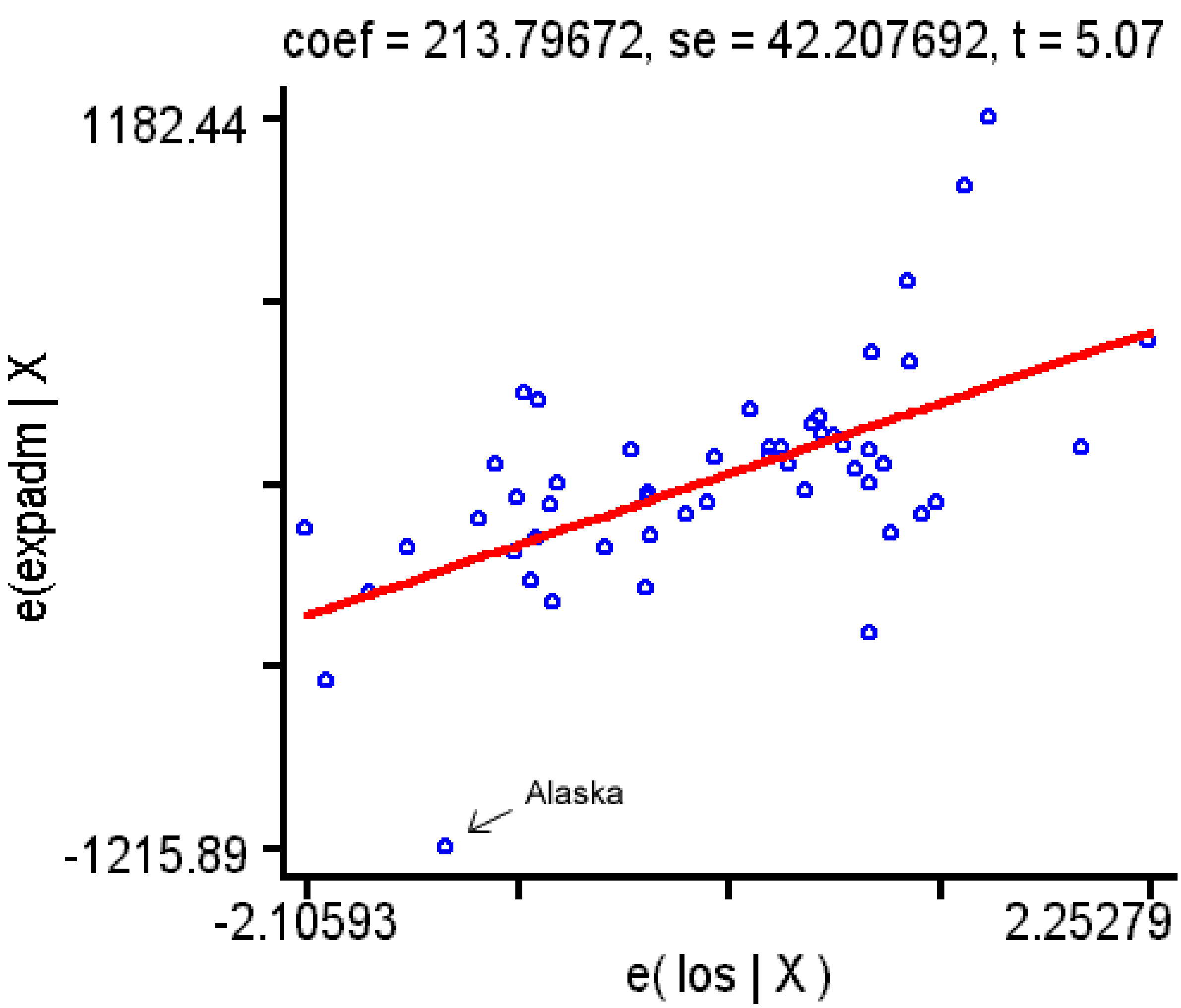
5.1 Added variables plots

- **Check for curvature or other unusual patterns or unusual points**

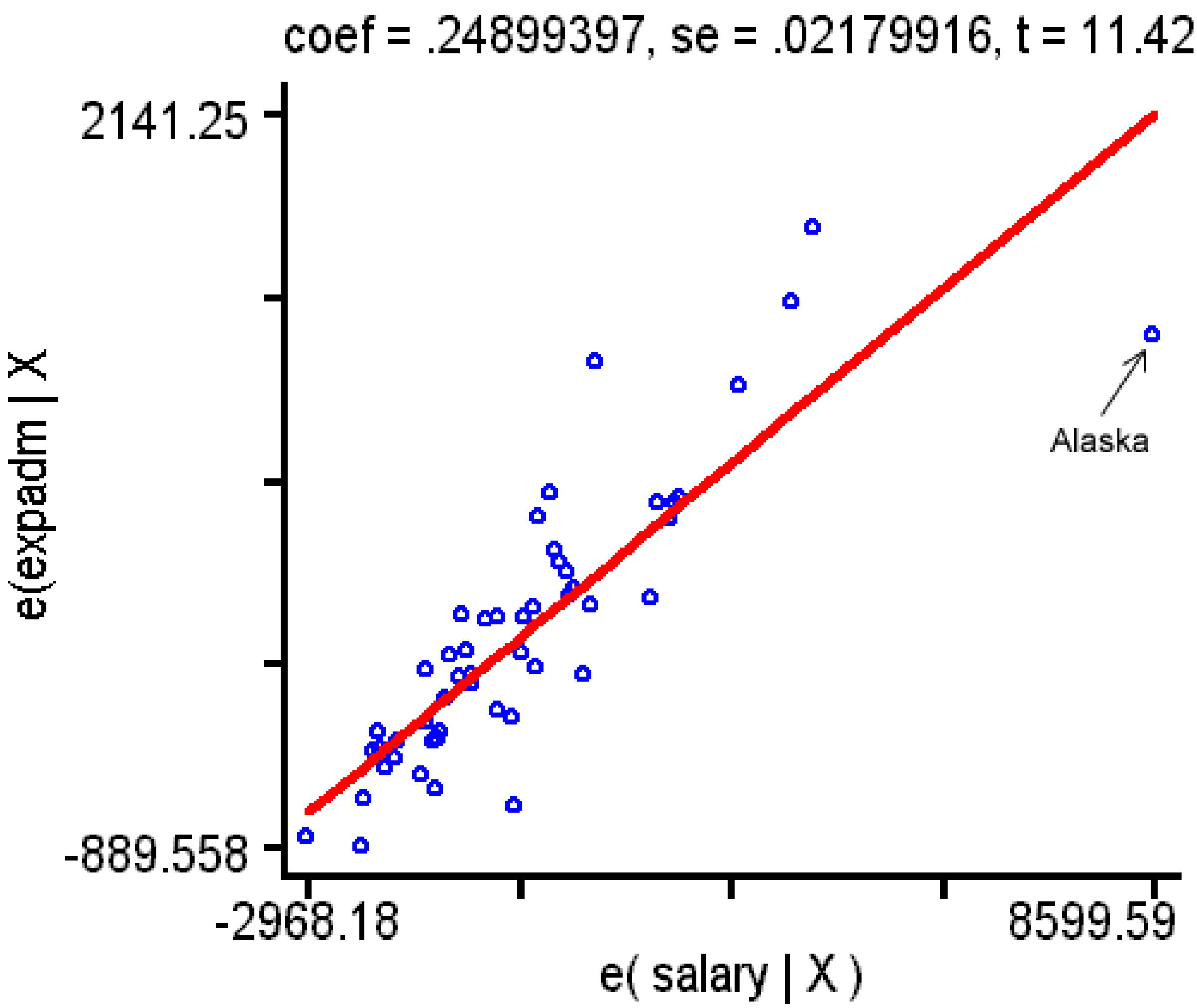
- **Stata commands**

```
regress expadm los salary
avplot los
avplot salary
```

- **Stata AVP graphs**



5.1 Added variables plots



- **Clearly, one point is unusual with respect to the others – it is data for the state of “Alaska”**

5.2 Residual Plots

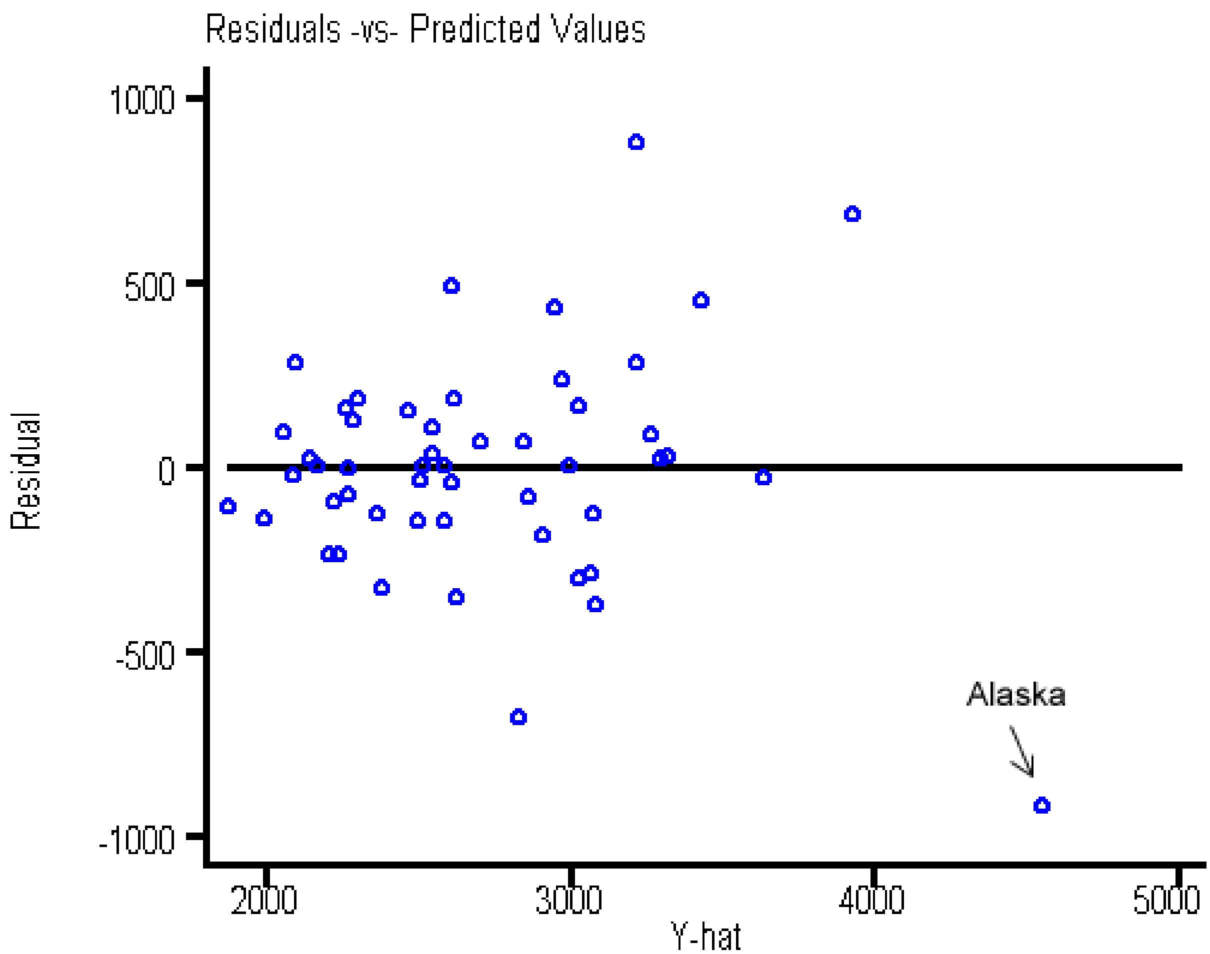
- **Plot residuals -vs- $\hat{Y}$**
- **Plot residuals -vs- each *X* or, perhaps, -vs- a new *X*, not in the current model**
- **Look for departures from random patterns within each "bin" of *X*s or $\hat{Y}$**
- **Look for curvature, unequal variance, points that are outliers, or other unusual patterns or points**
- **Stata Code**

(Assumes fit already done)

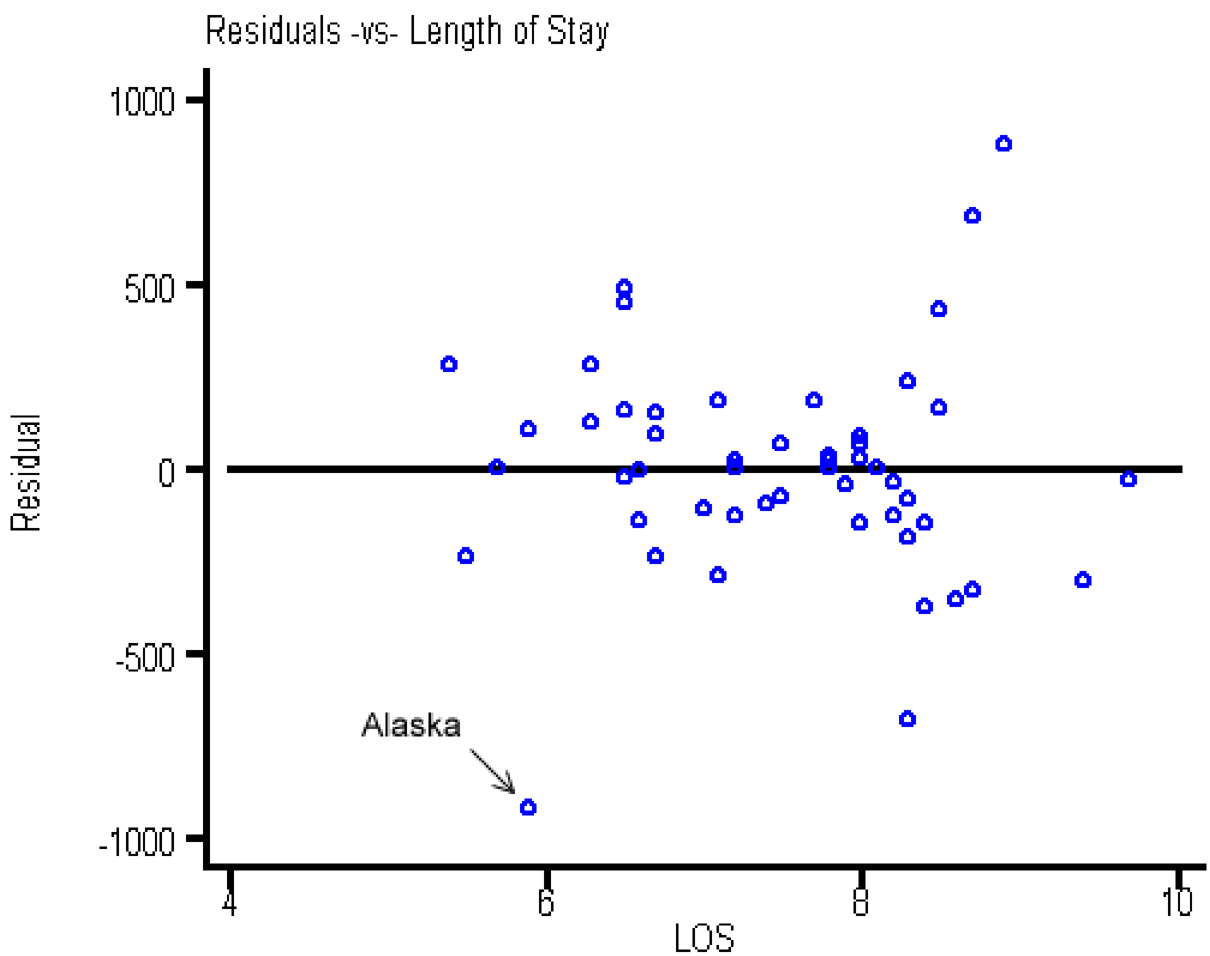
```
predict yhat
predict e, residuals

graph e yhat
graph e los
graph e salary
```

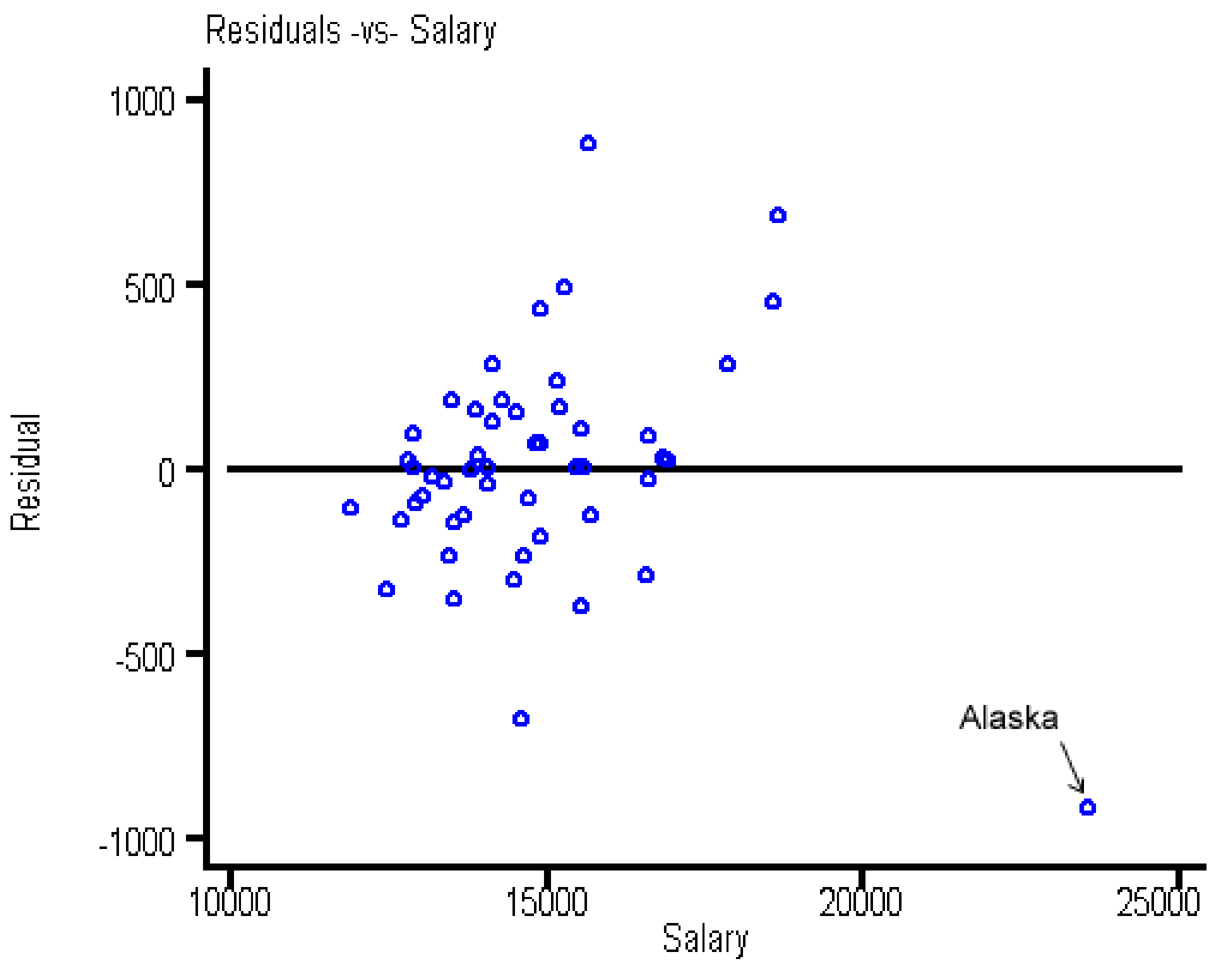
5.2 Residual Plots



5.2 Residual Plots



5.2 Residual Plots



- **CONCLUDE:**

  Exclude "Alaska," an unusual point (outlier) and re-fit the model

5.3 Sensitivity to outliers

- **Determine sensitivity of results to the unusual point discovered above by comparing models *with* and *without* "Alaska"**

- **Stata log**

```
* Since fit will not accept strings in if statements (UGH),
*      must convert state(string variable) to stateno (numeric variable)

encode state , gen(stateno)

* Check
list state stateno , nolabel

regress expadm los salary if stateno ~= 2

  Source |       SS       df       MS                  Number of obs =      50
---------+------------------------------               F(  2,    47) =  119.84
   Model |  14531751.2     2  7265875.59               Prob > F      =  0.0000
Residual |  2849649.30    47  60630.8362               R-squared     =  0.8361
---------+------------------------------               Adj R-squared =  0.8291
   Total |  17381400.5    49  354722.459               Root MSE      =  246.23

------------------------------------------------------------------------------
  expadm |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     los |   168.6216   35.48546      4.752   0.000       97.23405    240.0091
  salary |   .3239785   .0231285     14.008   0.000         .27745    .3705069
   _cons |  -3325.111   405.6919     -8.196   0.000      -4141.258   -2508.965
------------------------------------------------------------------------------
```

5.3 Sensitivity to outliers

Sensitivity of results to including/excluding the extreme point (Alaska)

	with Alaska			w/o Alaska		
Var	$\hat{\beta}$	$se_{\hat{\beta}}$	*t*	$\hat{\beta}$	$se_{\hat{\beta}}$	*t*
Intercpt	*-2583*	*465*	*-5.6*	*-3325*	*406*	*-8.2*
LOS	*214*	*42*	*5.1*	*168*	*35*	*4.7*
Salary	*0.249*	*.022*	*11.4*	*0.323*	*.023*	*14.0*

Var	*DBETA*	*DBETAS*
LOS	*214-168 = 46*	*(214-168) / 42 = 1.1*
Salary	*.249-.323 = -.074*	*(249-.323) / .022 = -3.4*

- **DBETA , calculated above, is the difference between $\hat{\beta}$ (with the outlier) minus $\hat{\beta}$ (without the outlier) – it directly indicates the "influence" of the unusual point on the effect estimate**

5.3 Sensitivity to outliers

- **DBETAS. calculated above, is a standardized version DBETA -- divide DBETA by $se_{\hat{\beta}}$ (from the model including all points)**

  These DBETAS will, approximately, have mean 0 and SD of 1

  Values of | DBETAS | greater than 2 only occur by chance with probability about 0.05; values greater than 3 occur by chance with probability about 0.003

- **Stata calculates DBETAS using a more accurate standardizing formula, so results will not exactly agree with those done by hand**

5.3 Sensitivity to outliers

• Stata log

(Note Stata's DBETAS are more extreme than hand calculated values shown above)

```
* Determine and list DBETA for the outlier

* Fit model with all the data points

regress expadm los salary

dfbeta los salary

DFlos:     DFbeta(los)
DFsalary:  DFbeta(salary)


list state DF*  if stateno == 2

                    state      DFlos   DFsalary
 48.               Alaska   1.315527  -4.227888
```

5.4 Non-linearity

Check for non-linearity in the dependence of expenditures on LOS and salary

- **Use the "broken-arrow" (linear spline) MLR to model the non-linearity**

- **Drop Alaska from the dataset for this analysis**

- **While the AVP and residual plots, showed no obvious non-linearity, fit a spline model with two slopes for LOS (breaking at 7 days) and two slopes for salary (breaking at $15,000) to check, more formally, whether the fit can be improved by extending the model with non-linear terms**

Define

$$L1 = (LOS - 7)^{+}$$

$$= \begin{cases} 0 & \text{if } LOS \leq 7 \\ LOS - 7 & \text{if } LOS > 7 \end{cases}$$

$$SAL1 = (Salary - 15{,}000)^{+}$$

$$= \begin{cases} 0 & \text{if } Salary \leq \$15K \\ Salary - 15{,}000 & \text{if } Salary > \$15K \end{cases}$$

- **Add L1, SAL1 in two separate models**

5.4 Non-linearity

1. $Y = \beta_0 + \beta_1(LOS\text{-}7) + \beta_2(Salary\text{-}15{,}000) + \beta_3(LOS\text{-}7)^+ + \epsilon$

Interpretation of coefficients :

β_0 – Average expenditure when the LOS is 7 days and salary is \$15,000 -- intercept for the LOS line before the break point

β_1 – Rate of change in expenditures with LOS day 7 or before for a given salary

$\beta_1+\beta_3$ – Rate of change in expenditures with LOS after day 7 for a given salary

β_2 – Rate of change in expenditures per \$1000 of salary for a fixed LOS

$\beta_0 - 7\beta_1 - 7\beta_3$ -- Intercept for the LOS line after the 7 day break point if the salary is \$15,000

5.4 Non-linearity

2. $Y = \beta_0 + \beta_1(LOS\text{-}7) + \beta_2(Salary\text{-}15{,}000)$

$+ \beta_3(Salary\text{-}15{,}000)^+ + \epsilon$

Interpretation of coefficients -- you do!:

β_0 --

β_1 --

$\beta_2 + \beta_3$ –

β_2 --

β_0 - $15000\beta_2$ $-15000\beta_3$ –

5.4 Non-linearity

● Stata log

```
* Fit linear splines for LOS and Salary -- in separate models,
*       w/o Alaska

* Center los and salary at their means (excluding Alaska)
egen    losc = mean(los)     if stateno~=2
replace losc = los - losc

egen    salaryc = mean(salary)      if stateno~=2
replace salaryc = salary - salaryc

* Generate non-linear terms

gen l1 = (los-7) * (los>7)
gen sal1 = (salary - 15000) * (salary>15000)

* Fit models

regress expadm losc salaryc l1     if stateno ~= 2

  Source |       SS       df       MS                  Number of obs =      50
---------+------------------------------               F(  3,    46) =   78.48
   Model |  14540359.3     3  4846786.43               Prob > F      =  0.0000
Residual |  2841041.19    46   61761.765               R-squared     =  0.8365
---------+------------------------------               Adj R-squared =  0.8259
   Total |  17381400.5    49  354722.459               Root MSE      =  248.52

------------------------------------------------------------------------------
  expadm |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    losc |   206.6644   108.0117      1.913   0.062      -10.75183    424.0806
 salaryc |   .3267898   .0245278     13.323   0.000        .277418    .3761616
      l1 |  -57.01222   152.7122     -0.373   0.711       -364.406    250.3815
   _cons |   2740.327   117.4716     23.328   0.000       2503.869    2976.785
------------------------------------------------------------------------------
regress expadm losc salaryc sal1   if stateno ~= 2

  Source |       SS       df       MS                  Number of obs =      50
---------+------------------------------               F(  3,    46) =   78.63
   Model |  14544925.1     3  4848308.36               Prob > F      =  0.0000
Residual |  2836475.39    46  61662.5084               R-squared     =  0.8368
---------+------------------------------               Adj R-squared =  0.8262
   Total |  17381400.5    49  354722.459               Root MSE      =  248.32

------------------------------------------------------------------------------
  expadm |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    losc |   170.4847   36.01239      4.734   0.000       97.99555    242.9739
 salaryc |   .3048594   .0474867      6.420   0.000       .2092736    .4004452
    sal1 |   .0360975   .0780963      0.462   0.646      -.1211021    .1932971
   _cons |   2681.589   50.68275     52.909   0.000       2579.569    2783.608
------------------------------------------------------------------------------
```

● Summary of results

Variable	$\hat{\beta}$	$se_{\hat{\beta}}$	t

5.4 Non-linearity

Intercept	*2740*	*74*	*37*
LOS-7	*206.7*	*108.0*	*1.91*
L1	*-57.0*	*152.7*	*-0.37*
Salary-15K	*0.33*	*.024*	*13*

Variable	$\hat{\beta}$	$se_{\hat{\beta}}$	*t*
Intercept	*2682*	*66*	*41*
LOS-7	*170.5*	*36*	*4.7*
Salary-15K	*0.30*	*0.047*	*6.4*
SAL1	*0.036*	*0.078*	*0.46*

What have we learned about non-linearity?

- **Neither L1 nor SAL1 improves model substantially**
- **Therefore, the dependence of expenditures on either LOS or salary is adequately approximated by a linear function**

5.5 Interaction

- **Whether LOS is an "effect modifier" for salary or, equivalently, is there an "interaction" between salary and LOS depends on the answer to either of the following two questions:**

  Is the effect of salary on expenditures different at different levels of LOS?

  or, equivalently,

  Is the effect of LOS on expenditures different at different levels of salary?

- **Once again, leave "Alaska" out of the analysis**

- **Extend the model with an interaction term that is the product of LOS and salary (*LOS*Salary*)**

$$Y = \beta_0 + \beta_1(LOS-7) + \beta_2(Salary-15,000)/1000$$

$$+ \beta_3(LOS-7)*(Salary-15,000)/1000 + \epsilon$$

5.5 Interaction

● Stata log

```
* Fit model with interaction for LOS*Salary
*       w/o  Alaska

* Scale salary, divide by 1000

replace salaryc = salaryc/1000

* Generate interaction term

gen losXsal = losc * salaryc

* Fit model

regress expadm losc salaryc losXsal   if stateno~=2

  Source |       SS       df       MS                  Number of obs =      50
---------+------------------------------               F(  3,    46) =   83.27
   Model |  14678583.7     3  4892861.22               Prob > F      =  0.0000
Residual |  2702816.82    46  58756.8873               R-squared     =  0.8445
---------+------------------------------               Adj R-squared =  0.8344
   Total |  17381400.5    49  354722.459               Root MSE      =  242.40

------------------------------------------------------------------------------
  expadm |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    losc |   153.6725   36.19012      4.246   0.000       80.82559    226.5195
 salaryc |   326.8532   22.84075     14.310   0.000       280.8772    372.8293
 losXsal |    38.3008   24.22849      1.581   0.121      -10.46861    87.07022
   _cons |   2691.143   34.59304     77.794   0.000       2621.511    2760.775
------------------------------------------------------------------------------
```

● Summary of results

Variable	$\hat{\beta}$	$se_{\hat{\beta}}$	t
Intercept	***2691.1***	***34.5***	***77.8***
(LOS-7)	***153.7***	***36.2***	***4.2***
(Salary-15K)/1000	***326.9***	***22.8***	***14.3***
(LOS-7)* (Salary-15K) /1000	***38.3***	***24.2***	***1.58***

5.5 Interaction

- **Interpretation of $\hat{\beta}_3 = 38.3$?**

LOS = 8 (or $X_1 + 1$)

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1(8-7) + \hat{\beta}_2 X_2 + \hat{\beta}_3(8-7)X_2$$
$$= \hat{\beta}_0 + \hat{\beta}_1 + (\hat{\beta}_2 + \hat{\beta}_3)X_2$$

LOS = 7 (or X_1)

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1(7-7) + \hat{\beta}_2 X_2 + \hat{\beta}_3(7-7)X_2$$
$$= \hat{\beta}_0 + \hat{\beta}_2 X_2$$

If *LOS*=7:

Average expenditures per capita are

$326 higher per $1,000 additional salary

but if *LOS = 8,*

Average expenditures per capita are

$326 + $38.3 = $364.3 higher per $1,000 additional salary

- **But, only weak/moderate evidence that this interaction is not 0**

---

6. Scientific Summary

- **Our multiple linear regression analysis shows that a state's expenditures per admission is statistically significantly associated with both LOS and employee salary (Model with two covariates)**

- **LOS being equal, a state whose average employee salary was $1,000 higher averaged $323 more per admission(95% CI: $277, $371) than another state**

- **With equal average salaries, a state with 1 day shorter LOS averaged $168 less per admission (95% CI: $97, $240)**

- **The variation in expenditures per admission among states with comparable LOS and salary was ± $\hat{\sigma}$ = $246**

  (from root MSE on Stata output)

- **The above statements relate to 49 states + DC, excluding Alaska, using the model in Section 4.3**

6. Scientific Summary

- **There is little evidence that the dependence of expenditures on LOS and salary is non-linear or that these two predictors interact with one another**

- **The reason Alaska was excluded is due to the unusually high employee salaries and low expenditures per admission caused by the use of state oil profits to subsidize hospital costs.**

  If the state Alaska is included in the data, the estimates and 95% CI are somewhat different – you do!

- **The do-file and Stata dataset for the examples in this lecture are posted on the course website on the "Classes" page:**

  cl4ex1.do

  hospital.dta

- **Listing of do-file script:**

```
version 7.0

* MLR.DO  MLR

*  Example:  Hospital Cost Data from Pagano and Gauvreau

*  Stata dataset:  hospital.dta


* Assumes files are in folder:  [path]\bio623


* To run this program,  use the following Stata commands:

*          cd [path]\bio623    ... change directory to folder bio623

*          do MLR

* OUTLINE:

* Part a.  Scatterplot matrix

* Part b.  MLR:  expense per admission on salary and los

* Part c.  Added Variable plots -- uses regress above

* Part d.  MLR:  Plot residuals vs Xs and predicteds for model checking

* Part e.  Re-fit without the outlier "Alaska"

* Part f.  Determine and list DBETA for the outlier

* Part g.  Fit linear splines for LOS and Salary -- in separate models,
*       w/o Alaska

* Part h.  Fit model with interaction for LOS*Salary
*       w/o  Alaska
```

```
* Housekeeping

* Clear work space
clear


* Turn off -more- pause
set more off



* Set directory for do and log files or use "cd" to get to it
* cd [path]\regress



* Save log file on disk, use .txt so Notepad will open it

capture log close
log using cl4ex1.log, replace


* Make sub-folder to store graph images
shell md cl4ex1


* Extend linesize for log

set linesize 100


* Access Stata dataset

use hospital,clear




* Dataset contents

describe



* Get variables, codes, descriptive stats

codebook los salary expadm


* Get stats, excluding Alaska

summarize los salary expadm  if state~="Alaska"



* List 5 records for checking

list  in 1/5
```

```
* Part a.  Scatterplot matrix
*   (Increase textsize)
set textsize 150
graph los salary expadm, matrix half
set textsize 100

*  Save the graph on disk in ps1-1 folder as Windows metafile fig1.wmf
gphprint , saving(cl4ex1\fig1.wmf,replace)



* Part b.  MLR:  expense per admission on salary and los
regress expadm los salary



* Part c.  Added Variable plots -- uses regress above
set textsize 150
avplot los , l1(" ") l2("e(expadm | X")
gphprint , saving(cl4ex1\fig2a.wmf,replace)

avplot salary , l1(" ") l2("e(expadm | X")
gphprint , saving(cl4ex1\fig2b.wmf,replace)
set textsize 100



* Part d.  MLR:  Plot residuals vs Xs and predicteds for model checking
*
*
```

```
predict yhat

predict e, residuals


* Make the plots

graph e yhat , yline(0) xlab ylab t1("Residuals -vs- Predicted Values") l1("
") l2("Residual") b1(" ") b2("Y-hat")

gphprint , saving(cl4ex1\fig3a.wmf,replace)


graph e los , yline(0) xlab ylab t1("Residuals -vs- Length of Stay") l1(" ")
l2("Residual") b1(" ") b2("LOS")

gphprint , saving(cl4ex1\fig3b.wmf,replace)


graph e salary , yline(0) xlab ylab t1("Residuals -vs- Salary") l1(" ")
l2("Residual") b1(" ") b2("Salary")

gphprint , saving(cl4ex1\fig3b.wmf,replace)


* Part e.  Re-fit without the outlier "Alaska"

* Since fit will not accept strings in if statements (UGH),

*      must convert state(string variable) to stateno (numeric variable)

encode state , gen(stateno)

* Check

list state stateno , nolabel

regress expadm los salary if stateno ~= 2


* Part f.  Determine and list DBETA for the outlier

* Fit model with all the data points

regress expadm los salary
```

```
dfbeta los salary

list state DF*  if stateno == 2




* Part g.  Fit linear splines for LOS and Salary -- in separate models,
*       w/o Alaska


* Center los and salary at their means (excluding Alaska)

egen    losc = mean(los)    if stateno~=2
replace losc = los - losc

egen    salaryc = mean(salary)      if stateno~=2
replace salaryc = salary - salaryc


* Generate non-linear terms

gen l1 = (los-7) * (los>7)

gen sal1 = (salary - 15000) * (salary>15000)

* Fit models

regress expadm losc salaryc l1     if stateno ~= 2


regress expadm losc salaryc sal1   if stateno ~= 2




* Part h.  Fit model with interaction for LOS*Salary
*       w/o  Alaska


* Scale salary, divide by 1000

replace salaryc = salaryc/1000


* Generate interaction term

gen losXsal = losc * salaryc


* Fit model
```

```
regress expadm losc salaryc losXsal   if stateno~=2


* Close log file -- Only once all errors have been fixed

*log close
```