2.3.1: SAS - Simple Inference

Dr. Bean – Stat 5100

Example: (The Toluca Company data from Chapter 1 & Chapter 3 Handouts)
We really want to say something about <u>how</u> lotsize affects workhours – does it?

```
/******************************************************/

/* Input Toluca data (recall Ch. 1 example) */
data toluca; input lotsize workhours @@; cards;
   80  399    30  121    50  221    90  376    70  361    60  224
  120  546    80  352   100  353    50  157    40  160    70  252
   90  389    20  113   110  435   100  420    30  212    50  268
   90  377   110  421    30  273    90  468    40  244    80  342
   70  323
  ;
run;

/* Now fit simple linear model with Y=workhours and
   X=lotsize, with residuals and predicted values saved
   in data set tolucaout */
proc reg data=toluca;
  model workhours = lotsize;
  output out=tolucaout r=resid p=pred;
  title1 'Simple linear model';
run;

/* Check assumptions */
/* Define shortcut macro, using line copied from
    www.stat.usu.edu/jrstevens/stat5100/resid_num_diag_1line.sas
 */
%macro resid_num_diag(dataset, ...

%resid_num_diag(dataset=out1, datavar=resid, label='Residual',
  predvar=pred, predlabel='Predicted Value');


/* See output from this on p.5 of Handout #4.
   Only when assumptions are met does inference make sense!
*/
```

```
/* Fit a simple linear model with Y=workhours and X=lotsize;
   output the 95% confidence intervals for the coefficients.
   Get predicted values (call them Predict here) and
   upper and lower 95% prediction and confidence intervals
   for each X value; put all this in a dataset called confidence.
   Also, include prediction for two X-levels not in original
   data set (X=10 and X=130). */
data dummy;  input lotsize @@;  cards;
  10 130
  ;
data trick; set toluca dummy;
run;
proc reg data=trick;
  model workhours = lotsize / clb alpha=.05;
                                      /* 1-alpha is level */
  output out=confidence p=Predict
                  ucl=uPred   /* upper and lower limits for */
                  lcl=lPred   /*      individual prediction */
                  uclm=uConf  /* upper and lower limits for */
                  lclm=lConf; /*      group mean confidence */
  title1 'Regression with 95% interval estimation';
run;
```

Regression with 95% interval estimation							

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	62.36586	26.17743	2.38	0.0259	8.21371	116.51801
lotsize	1	3.57020	0.34697	10.29	<.0001	2.85244	4.28797

Fit Plot for workhours
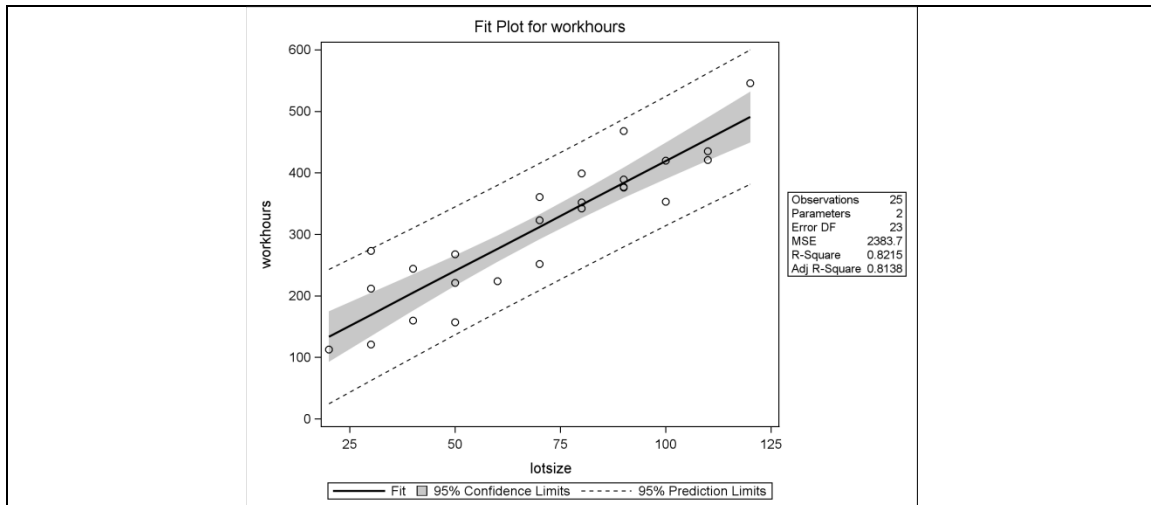
```
/* Look at partial result  */
proc print data=confidence;
  where lotsize < 50;
       /* which observations to use in proc */
  var lotsize workhours Predict lPred uPred lConf uConf;
  title1 'Predicted values and confidence and predicted
intervals';
  title2 'for lotsize < 50; these are 95% intervals.';
run;
```

Predicted values and confidence and predicted intervals
for lotsize < 50; these are 95% intervals.

Obs	lotsize	workhours	Predict	lPred	uPred	lConf	uConf
2	30	121	169.472	62.5464	276.397	134.367	204.577
11	40	160	205.174	99.9483	310.400	175.649	234.698
14	20	113	133.770	24.6977	242.842	92.587	174.952
17	30	212	169.472	62.5464	276.397	134.367	204.577
21	30	273	169.472	62.5464	276.397	134.367	204.577
23	40	244	205.174	99.9483	310.400	175.649	234.698
26	10	.	98.068	-13.5719	209.708	50.500	145.636

```
/************************************************************
  Note: there are other ways to get the CI for Y in SAS, but they
```

3

```
   aren't included here; just know that if you needed to, you
   could get the SE for Yhat using the stdp and stdi options
   in proc reg.
************************************************************/
```

```
/* Look at Reduced model */
proc reg data=toluca;
  model workhours = ;
  title1 'Reduced Model (dropped lotsize predictor)';
run;
```

Reduced Model (dropped lotsize predictor)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	0	0	.	.	.
Error	24	307203	12800		
Corrected Total	24	307203			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	312.28000	22.62753	13.80	<.0001

Fit Diagnostics for workhours

Observations	25
Parameters	1
Error DF	24
MSE	12800
R-Square	0
Adj R-Square	0