Analysis of Alberta height data... via SAS

```
PROC sort data=sasuser.alberta; by I_female;
PROC MEANS MEAN CLM ALPHA=0.05 MAXDEC=2;
WHERE (age = 11) ; by I_female;
 var height weight;
run;
```

MEAN gives mean

CLM gives Confidence Limits for mean

95% Confidence: ALPHA=0.05, by default if ALPHA omitted, but safer to spell it out.

I_FEMALE=0

Variable	Mean	Lower 95.0% CLM	Upper 95.0% CLM
HEIGHT	57.21	55.93	58.50
WEIGHT	83.86	79.09	88.62

I_FEMALE=1

Variable	Mean	Lower 95.0% CLM	Upper 95.0% CLM
HEIGHT	57.25	55.43	59.07
WEIGHT	80.69	72.28	89.10

Inference re single mean (μ height, 11 yr old boys)

See notes for annotated analysis using output from SYSTAT.

```
proc reg data=sasuser.alberta;
       where ( I_Female = 0 and age =11);
 model height = ;
run;
```

Dependent Variable: HEIGHT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	0	0.00000	.	.	.
Error	13	64.35714	4.95055		
C Total	13	64.35714			

Root MSE	2.22498	R-square	0.0000
Dep Mean	57.21429	Adj R-sq	0.0000
C.V.	3.88886		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	57.214286	0.59465173	96.215	0.0001

------ how data were read in and dataset set up -------

```
DATA sasuser.alberta;
  INPUT Age Height Weight IDNumber I_Female ;
LINES;
11  56  85  874   1
11  57  69   31   1
11  54  69  613   1
11  62 104  814   1
...
...
16  72 150   28   0
16  65 135  268   0
16  65  98  736   0
;
RUN;
```

Inference re difference of 2 means
(μ height, 11 yr old boys versus girls)

SAS "WHERE" statement is very useful: can restrict analysis, without making a new dataset.
It can be used with any PROCedure.

Chapter 7

```
PROC TTEST DATA=sasuser.alberta;
     where (age =11);
 CLASS I_female;
 VAR height;
run;

PROC REG DATA=sasuser.alberta;
     where (age =11);
 MODEL height = I_Female;
run;
```

TTEST PROCEDURE

Variable: HEIGHT

I_FEMALE	N	Mean	Std Dev	Std Error
0	14	57.21	2.22	0.594
1	16	57.25	3.41	0.853
diff. in means		0.04 (rounded)		

Variances	T	DF	Prob>|T|
Unequal	-0.0343	26.0	0.9729
Equal	-0.0334*	28.0	0.9736

*Procedure subtracts the mean in the first (I_Female=0) group from the mean in the second (I_female=1)

Chapter 10 (regression)

```
PROC REG DATA=sasuser.alberta;
     where (age =11);
 MODEL height = I_Female;
run;
```

Dependent Variable: HEIGHT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	0.00952	0.00952	0.001	0.9736
Error	28	239.35714	8.54847		
C Total	29	239.36667			

Root MSE	2.92	R-square	0.0000
Dep Mean	57.23	Adj R-sq	-0.0357
C.V.	5.10		

Parameter Estimates

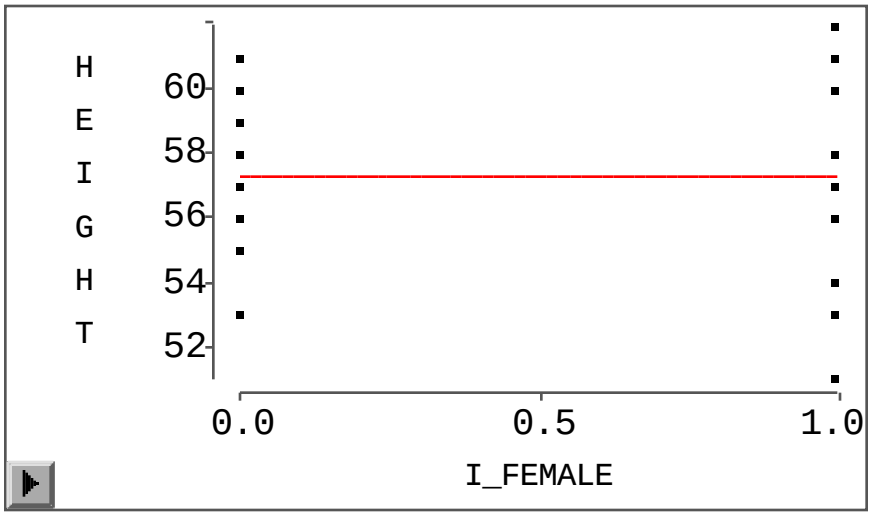
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	57.21	0.781412	73.219	0.0001
I_FEMALE	1	0.04	1.069992	0.033	0.9736

Note the identical P Values from pooled variances t-test of the difference in two means, and the test of whether the regression parameter which represents this difference is zero. See notes for the explanation, and the link between the two ways of assessing the difference in 2 means.

I can't get SAS to directly print CI to accompany each estimate, but could use parameter estimate $\pm Z_{\alpha/2} \times$ Standard Error[parameter estimate] for 100(1-α)% CI in this instance, since df for t are large enough (28) that t can be approximated by Z.

Via SAS INSIGHT rather than SAS PROC REG

HEIGHT = I_FEMALE
Response Distribution: Normal
Link Function: Identity



Parametric Regression Fit

Curve	Degree(Polynomial)	Model DF	Model Mean Square	Error DF
	1.00	1.00	0.01	28.00

Error Mean Square	R-Square	F Stat	Prob > F
8.55	0.00	0.00	0.9736

Summary of Fit

Mean of Response	57.23	R-Square	0.00
Root MSE	2.92	Adj R-Sq	0.00

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
Model	1	0.0095	0.0095	0.0011	0.9736
Error	28	239.3571	8.5485	.	.
C Total	29	239.3667	.	.	.

Parameter Estimates

Variable	DF	Estimate	Std Error	T Stat	Prob >|T|
INTERCEPT	1	57.2143	0.7814	73.2191	0.0001
I_FEMALE	1	0.0357	1.0700	0.0334	0.9736

95% C.I. for Parameters*

Variable	Estimate	Lower	Upper
INTERCEPT	57.2143	55.6136	58.8149
I_FEMALE	0.0357	-2.1561	2.2275

** In INSIGHT: Use FIT under ANALYZE menu. Ask for "CI(Wald) for Parameters" --avaibable as an 'on or off option" under TABLES menu when the FIT output window is active (or specify using the Output button when setting up in the FIT dialog window)*

Can reduce the number of decimal places in Tables

95% C.I. for Parameters

Variable	Estimate	Lower	Upper
INTERCEPT	57.2143	55.6136	58.8149
I_FEMALE	0.0357	-2.1561	2.2275

Click on the triangle at upper left conner of output an choose Format 10.2 say...

95% C.I. for Parameters

Variable	Estimate	Lower	Upper
INTERCEPT	57.21	55.61	58.81
I_FEMALE	0.04	-2.16	2.23

Clicking on the "Save" under this same triangle will write the text of the table out to the Output window in 'regular' SAS -- from where you can cut and paste it as text into a wordprocessor etc..

This same analysis via EXCEL:
see Excel file under Ch 10 Resources.