

Stat 5100 Handout 2.6.1 – SAS: Inference with Multiple Predictors

Example: (Table 7.1) Study seeks to relate (in females) amount of body fat (Y) to triceps skinfold thickness (X_1), thigh circumference (X_2), and midarm circumference (X_3). Amount of body fat is expensive to measure, requiring immersion of person in water. This expense motivates the desire for a predictive model based on these inexpensive predictors.

Q1: Do thigh and midarm both have no effect on body fat when triceps is in the model?

Q2: Do the relationships among the predictors cause any problems in the fitted model?

```
/* Input data */
```

```
data bodyfat;
```

```
    input triceps thigh midarm body @@; cards;
```

19.5	43.1	29.1	11.9	24.7	49.8	28.2	22.8
30.7	51.9	37.0	18.7	29.8	54.3	31.1	20.1
19.1	42.2	30.9	12.9	25.6	53.9	23.7	21.7
31.4	58.5	27.6	27.1	27.9	52.1	30.6	25.4
22.1	49.9	23.2	21.3	25.5	53.5	24.8	19.3
31.1	56.6	30.0	25.4	30.4	56.7	28.3	27.2
18.7	46.5	23.0	11.7	19.7	44.2	28.6	17.8
14.6	42.7	21.3	12.8	29.5	54.4	30.1	23.9
27.7	55.3	25.7	22.6	30.2	58.6	24.6	25.4
22.7	48.2	27.1	14.8	25.2	51.0	27.5	21.1

```
;
```

```
run;
```

```
proc corr data=bodyfat;
```

```
    var body triceps
```

```
        thigh midarm;
```

```
    title1 'Correlation
```

```
matrix';
```

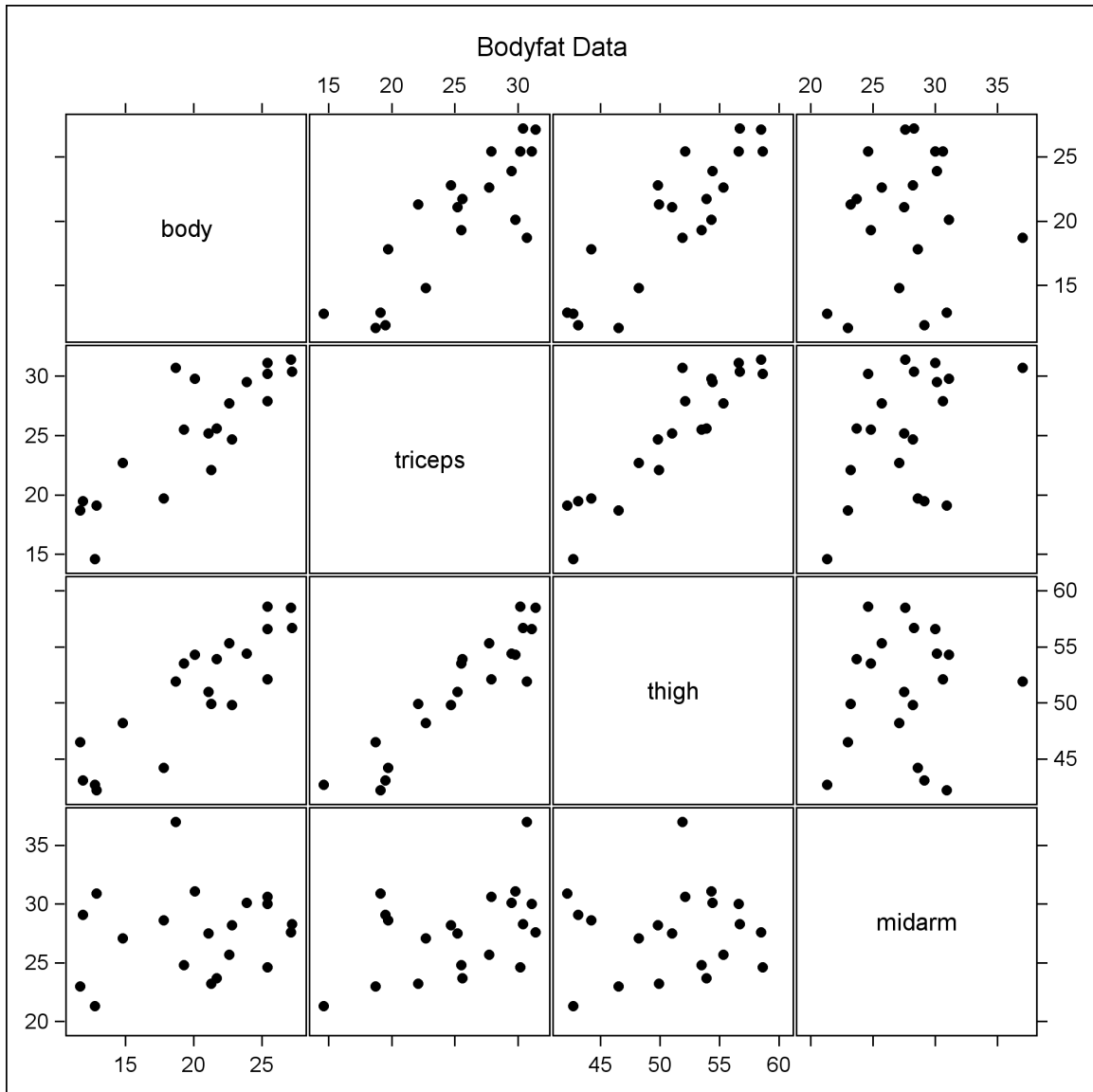
```
run;
```

Pearson Correlation Coefficients, N = 20 Prob > r under H0: Rho=0				
	body	triceps	thigh	midarm
body	1.00000	0.84327 <.0001	0.87809 <.0001	0.14244 0.5491
triceps	0.84327 <.0001	1.00000	0.92384 <.0001	0.45778 0.0424
thigh	0.87809 <.0001	0.92384 <.0001	1.00000	0.08467 0.7227
midarm	0.14244 0.5491	0.45778 0.0424	0.08467 0.7227	1.00000

```

proc sgscatter data=bodyfat;
  matrix body triceps thigh midarm/
    markerattrs=(symbol=CIRCLEFILLED size=2pt);
  title 'Bodyfat Data';
run;

```



```
/* Q1: Test whether thigh and midarm BOTH have
   no effect on body when triceps is in the model */
```

```
proc reg data=bodyfat;
  model body = triceps thigh midarm;
  title1 'Bodyfat Regression';
  title2 '(full model)';
run;
```

<i>Bodyfat Regression (full model)</i>					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	396.98461	132.32820	21.52	<.0001
Error	16	98.40489	6.15031		
Corrected Total	19	495.38950			

```
proc reg data=bodyfat;
  model body = triceps;
  title1 'Bodyfat Regression';
  title2 '(reduced model)';
run;
```

<i>Bodyfat Regression (reduced model)</i>					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	352.26980	352.26980	44.30	<.0001
Error	18	143.11970	7.95109		
Corrected Total	19	495.38950			

```

data temp;
  F = ( (143.11970-98.40489)/2 ) / ( 6.15031 );
  p = 1-probf(F,2,16);
proc print data=temp;
  title1 'Subset F-test, by hand';
run;

```

Subset F-test, by hand		
Obs	F	p
1	3.63517	0.049950

```

/* Do this subset F-test, automatically.
  Also look at related quantities:
  See all sequential sums of squares and
  coefficients of partial determination */
proc reg data=bodyfat;
  model body = triceps thigh midarm / ss1 pcorrl;
  subsetcheck: test thigh=midarm=0;
  title1 'Subset F-test, automatically';
run;

```

Subset F-test, automatically							
Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Type I SS	Squared Partial Corr Type I
Intercept	1	117.08469	99.78240	1.17	0.2578	8156.76050	.
triceps	1	4.33409	3.01551	1.44	0.1699	352.26980	0.71110
thigh	1	-2.85685	2.58202	-1.11	0.2849	33.16891	0.23176
midarm	1	-2.18606	1.59550	-1.37	0.1896	11.54590	0.10501

Test subsetcheck Results for Dependent Variable body				
Source	DF	Mean Square	F Value	Pr > F
Numerator	2	22.35741	3.64	0.0500
Denominator	16	6.15031		

```
/* Q2: Investigate effect of relationships among
predictors. */
```

```
/* Standardizing all variables */
proc reg data=bodyfat;
  model body = triceps thigh midarm / stb;
  title1 'Standardized regression coefficients';
  title2 '(note extra column in output)';
run;
```

Standardized regression coefficients (note extra column in output)						
Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	117.08469	99.78240	1.17	0.2578	0
triceps	1	4.33409	3.01551	1.44	0.1699	4.26370
thigh	1	-2.85685	2.58202	-1.11	0.2849	-2.92870
midarm	1	-2.18606	1.59550	-1.37	0.1896	-1.56142

```

/* Test for multicollinearity */
proc reg data=bodyfat;
  model body = triceps thigh midarm / vif collin;
  title1 'Test for multicollinearity';
run;

```

Test for multicollinearity

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	396.98461	132.32820	21.52	<.0001
Error	16	98.40489	6.15031		
Corrected Total	19	495.38950			

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	1	117.08469	99.78240	1.17	0.2578	0
triceps	1	4.33409	3.01551	1.44	0.1699	708.84291
thigh	1	-2.85685	2.58202	-1.11	0.2849	564.34339
midarm	1	-2.18606	1.59550	-1.37	0.1896	104.60601

Collinearity Diagnostics						
Number	Eigenvalue	Condition Index	Proportion of Variation			
			Intercept	triceps	thigh	midarm
1	3.96796	1.00000	0.00000195	0.00000320	0.00000110	0.00000980
2	0.02052	13.90482	0.00037152	0.00132	0.00003262	0.00139
3	0.01151	18.56570	0.00059915	0.00021875	0.00032550	0.00693
4	0.00000865	677.37207	0.99903	0.99846	0.99964	0.99167