

22S:166, Computing in Statistics  
Fall 2001  
Practice problems for Midterm 2, Cowles

1. See the attached SAS code and output. The code is available for download from the “Datasets” section of the course webpage as `mid2pracexpurg.sas`.

If you are working in the Unix lab, the data step included in the program will access the data file without your having to change anything. If you are working in the PC lab, you will have to download the data file `cigarettes.dat` from the “Datasets” section of the course web page, and will have to modify the `infile` statement to read the datafile from wherever you saved it.

To the SAS program, add the two statements that are needed to make it produce the attached output.

Note that, in order for `proc univariate` to produce plots (shown on half-page 4 of the output) the first line of the procedure step must read:

```
proc univariate plot ;
```

2. The following fields are contained in a flat file kept by a piano teacher to store information about her students, who are children.

- (a) Responsible parent’s name
- (b) Home address
- (c) Phone
- (d) Monthly lesson fee for all children in family
- (e) Child 1 name
- (f) Child 1 birthdate
- (g) Child 1 date started lessons
- (h) Child 1 school
- (i) Child 2 name
- (j) Child 2 birthdate
- (k) Child 2 date started lessons
- (l) Child 2 school
- (m) Child 3 name
- (n) Child 3 birthdate
- (o) Child 3 date started lessons
- (p) Child 3 school
- (q) Child 4 name
- (r) Child 4 birthdate

- (s) Child 4 date started lessons
- (t) Child 4 school

Put this data into first normal form. Write down what tables would be needed, and for each one indicate:

- all fields (including any new ones you add)
- primary key
- foreign key(s) if any

```
options linesize = 72 nodate mprint mlogic ;
```

```
data cigs ;
infile '/group/ftp/pub/kcowles/datasets/cigarettes.dat' ;
input name $16. tar nic wgt co1 ;
run ;
```

```
%macro summ(var= ,plot= ) ;
proc univariate &plot ;
var &var ;
title "Summary Statistics on &var" ;
run ;
%mend ;
```

```
proc reg ;
model co1 = nic ;
where wgt > &lowwgt ;
title "Regression of CO1 on Nicotine, Weight > &lowwgt" ;
run ;

endsas ;
```

Regression of CO1 on Nicotine, Weight > 0.9

1

The REG Procedure  
Model: MODEL1  
Dependent Variable: co1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	321.28604	321.28604	88.34	<.0001
Error	19	69.10539	3.63713		
Corrected Total	20	390.39143			

Root MSE	1.90712	R-Square	0.8230
Dependent Mean	12.94286	Adj R-Sq	0.8137
Coeff Var	14.73496		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	2.41896	1.19456	2.02	0.0572
nic	1	11.64393	1.23889	9.40	<.0001

Summary Statistics on tar

2

5% 4.1  
 1% 1.0  
 0% Min 1.0

The UNIVARIATE Procedure  
 Variable: tar

Moments

N	25	Sum Weights	25
Mean	12.216	Sum Observations	305.4
Std Deviation	5.66580974	Variance	32.1014
Skewness	0.75665781	Kurtosis	2.9515386
Uncorrected SS	4501.2	Corrected SS	770.4336
Coeff Variation	46.3802369	Std Error Mean	1.13316195

Extreme Observations

----Lowest----		----Highest---	
Value	Obs	Value	Obs
1.0	16	15.8	19
4.1	5	16.0	2
4.5	20	16.6	9
7.3	22	17.0	17
7.8	13	29.8	3

Basic Statistical Measures

Location		Variability	
Mean	12.21600	Std Deviation	5.66581
Median	12.80000	Variance	32.10140
Mode	.	Range	28.80000
		Interquartile Range	6.50000

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 10.78045	Pr >  t  <.0001
Sign	M 12.5	Pr >=  M  <.0001
Signed Rank	S 162.5	Pr >=  S  <.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	29.8
99%	29.8
95%	17.0
90%	16.6
75% Q3	15.1
50% Median	12.8
25% Q1	8.6
10%	4.5

Stem Leaf	#	Boxplot
3 0	1	0
2		
2		
1 55556677	8	+-----+
1 1223444	7	*--+-*
0 788999	6	+-----+
0 144	3	
-----+-----+		
Multiply Stem.Leaf by 10***1		

