

**Second MidTerm Exam  
Multiple Regression**

**Question A** (two thought questions; not many points; don't spend too long on them)

1. (3 pts) The following message was posted on a newsgroup for SAS users.

When performing a stepwise regression analysis and you are not sure whether you should consider an intercept or not, you can do the following. Create a new variable INTERC containing the value 1 for each observation and include it in your model statement. This is better than using the default INT-option. The stepwise regression methods will show you whether INTERC is a "variable" to take into account or not.

```
data DATASET;  
  set DATASET;  
  INTERC=1;  
  
proc reg data=DATASET;  
  model DEP = INDEP1 INDEP2 .. INTERC  
    / noint selection=stepwise ...;  
run;
```

What do you think of this advice? Write a **brief** reply to this message.

2. (3 pts) This question is based on

Lyness, S. A. (1993). Predictors of differences between Type A and Type B individuals in heart rate and blood pressure reactivity. *Psychological Bulletin*, 114, 266-295.

Lyness found 156 studies in the literature on this topic. For 99 of these studies, either effect sizes [PREs for us] were presented or there was enough information from which effect sizes could be calculated. The conclusions of the studies about the relationship between personality type (A or B) and blood pressure reactivity varied widely; some concluded there was a relationship, but many found no evidence for such a relationship. Averaging across studies, Lyness estimated that the typical effect size for diastolic blood pressure studies was approximately  $r^2 = .01$ , while for systolic blood pressure it was approximately  $r^2 = .03$ . He also noted that the typical study used about 50 subjects.

Write a **brief** comment using the concept of power to explain why some studies found effects and others did not.

**Question B** (model and data questions)

Data were collected by the US Department of Transportation on characteristics of different types of automobiles for sale during 1993. For each of 93 vehicle

models included in the sample, data were gathered on the following five variables:

MPGH	Highway gas consumption in miles per gallon
WEIGHT	Weight of the vehicle in pounds
WBASE	Length of the vehicle's wheelbase in inches
HP	Maximum horsepower generated
ENGSIZE	Engine size in liters

The Department of Transportation is interested in these data because they want to understand highway gas consumption, the first of these five variables. Notice that the remaining four variables are potential predictors of MPGH, falling into two different classes. WEIGHT and WBASE both have to do with the physical size of the vehicle. HP and ENGSIZE measure characteristics of the vehicle's engine.

On the following pages you will find univariate statistics on these variables and various regressions, estimating models that predict MPGH. Using this output, answer the following questions.

1. (4 pts) A goal of the US Department of Transportation was to have the average highway gas mileage of all cars in the US be better than 28 miles per gallon. Test whether the 1993 sample of vehicles in this data set exceeds this goal. Do this in the context of the simplest model and give the estimated models, the SSE's, PRE, and F\*.
2. (4 pts) Is WBASE a useful predictor of MPGH? Provide the estimated models, the SSE's, PRE, and F\*. Additionally provide a one-sentence interpretation of the estimated slope for WBASE.
3. (5 pts) Given that a vehicle's weight is highly predictive of MPGH, is WBASE a useful additional predictor? Provide the estimated models, the SSE's, PRE, and F\*. Interpret any slopes in Model A that are significantly different from zero.
4. (3 pts) Write a couple of sentences interpreting the difference between the slope for WBASE in the simple regression model (-0.48) and the partial slope for WBASE in the model that also includes WEIGHT as a predictor (.30). What substantively (i.e., five o'clock news summary) does the difference between these two slopes tell us?
5. (3 pts) What is the correlation between WBASE and WEIGHT? Is this correlation significantly different from zero? Provide PRE and F\*.
6. (5 pts) Are the two measures of the vehicles' engine (HP and ENGSIZE) useful predictors of MPG as a set? Provide the estimated models, the SSE's, PRE, and F\*. Interpret any slopes in Model A that are significantly different from zero.

7. (4 pts) Are the two measures of the vehicle's engine (HP and ENGSIZE) useful predictors of MPGH as a set over and above the two predictor variables that assess the car size? Provide the estimated models, the SSE's, PRE, and F\*.
8. (3 pts) Write a couple of sentences that gives a five o'clock news summary of your answers to questions 6 and 7.
9. (2 pts) Describe how you would construct a dependent variable that assesses MPGH over and above (or controlling for or adjusting for) weight, wheel base, horsepower, and engine size.
10. (4 pts) In the context of a model where WEIGHT and WBASE are used as predictors of MPGH, retest the hypothesis in question 1, that the 1993 sample of vehicles in this data set exceeds the goal of 28 miles per gallon in highway driving. Provide Models C and A, SSE's, PRE, and F\*.
11. (6 pts) The final page of the output is from SAS/Insight. It contains the residual plot from the model where MPGH is regressed on WEIGHT and WBASE. Additionally, the two partial regression leverage plots from this analysis are given. One observation in each of these three plots is highlighted. This observation is the Honda Civic. Write a few sentences explaining the meaning or interpretation of the Honda Civic point in each graph. It is obvious that it is an outlier; we are looking for more of a description than that.

The SAS System

Variable	Mean	Variance	CSS	USS	N
MPGH	29.0860215	28.4273025	2615.31	81293.00	93
WEIGHT	3072.90	347977.89	32013966.13	910188250	93
WBASE	103.9462366	46.5079476	4278.73	1009127.00	93
HP	143.8279570	2743.08	252363.25	2176206.00	93
ENGSIZE	2.6677419	1.0761220	99.0032258	760.8700000	93

Model: MODEL1  
Dependent Variable: MPGH

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	990.41277	990.41277	55.467	0.0001
Error	91	1624.89906	17.85603		
C Total	92	2615.31183			
Root MSE	4.22564	R-square	0.3787		
Dep Mean	29.08602	Adj R-sq	0.3719		
C.V.	14.52808				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	79.096313	6.72924647	11.754	0.0001
WBASE	1	-0.481117	0.06460037	-7.448	0.0001

Model: MODEL2  
Dependent Variable: MPG

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	1718.69528	1718.69528	174.435	0.0001
Error	91	896.61655	9.85293		
C Total	92	2615.31183			
Root MSE		3.13894	R-square	0.6572	
Dep Mean		29.08602	Adj R-sq	0.6534	
C.V.		10.79191			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	51.601365	1.73554978	29.732	0.0001
WEIGHT	1	-0.007327	0.00055477	-13.207	0.0001

Model: MODEL3  
Dependent Variable: MPG

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	1809.85524	904.92762	101.115	0.0001
Error	90	805.45659	8.94952		
C Total	92	2615.31183			
Root MSE		2.99157	R-square	0.6920	
Dep Mean		29.08602	Adj R-sq	0.6852	
C.V.		10.28527			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	29.850459	7.01299911	4.256	0.0001
WBASE	1	0.298072	0.09339401	3.192	0.0019
WEIGHT	1	-0.010332	0.00107971	-9.569	0.0001

Variable	DF	Type II SS	Squared Partial Corr Type II	Tolerance
INTERCEP	1	162.141469	.	.
WBASE	1	91.159954	0.10167106	0.23979856
WEIGHT	1	819.442467	0.50430361	0.23979856

Model: MODEL4  
Dependent Variable: MPGH

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	1172.05385	586.02692	36.544	0.0001
Error	90	1443.25798	16.03620		
C Total	92	2615.31183			
Root MSE		4.00452	R-square	0.4482	
Dep Mean		29.08602	Adj R-sq	0.4359	
C.V.		13.76786			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	39.269132	1.26697532	30.994	0.0001
HP	1	-0.035137	0.01170249	-3.003	0.0035
ENGSIZE	1	-1.922734	0.59083571	-3.254	0.0016

Variable	DF	Type II SS	Squared Partial Corr Type II	Tolerance
INTERCEP	1	15405	.	.
HP	1	144.572598	0.09105039	0.46400070
ENGSIZE	1	169.826907	0.10528082	0.46400070

Model: MODEL5  
Dependent Variable: MPGH

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	1845.38146	461.34536	52.730	0.0001
Error	88	769.93037	8.74921		
C Total	92	2615.31183			
Root MSE		2.95791	R-square	0.7056	
Dep Mean		29.08602	Adj R-sq	0.6922	
C.V.		10.16951			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	29.697332	7.86783921	3.775	0.0003
WBASE	1	0.330011	0.10624879	3.106	0.0026
WEIGHT	1	-0.012476	0.00164731	-7.574	0.0001
HP	1	0.005811	0.01053353	0.552	0.5826
ENGSIZE	1	0.970042	0.58939265	1.646	0.1034

Variable	DF	Type II SS	Squared Partial Corr Type II	Tolerance
INTERCEP	1	124.650115	.	.
WBASE	1	84.406942	0.09879815	0.18113643
WEIGHT	1	501.863916	0.39461092	0.10071142
HP	1	2.662421	0.00344609	0.31246035
ENGSIZE	1	23.699553	0.02986222	0.25439584

Graphical Output from SAS/Insight  
Corresponding to MODEL 3 above

