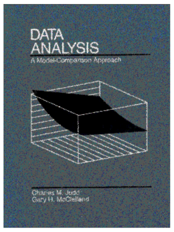
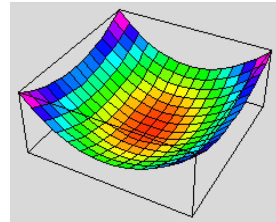


Handout from Psych 5741/5751
Univ of Colorado
used with
Judd, C.M., & McClelland, G.H. (1989). Data Analysis: A Model Comparison Approach. HBJ.



Final Exam: Answers
Spring 1996

Note: The textual and graphical answers are more detailed and complete than was expected on the exam.

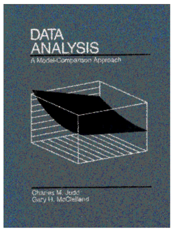
Question 1

A. SAS Code:

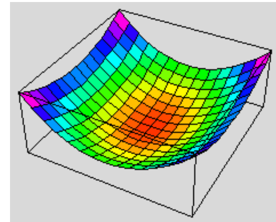
```
data memory;
  infile 'memory.dat';
  input mood$ word$ recall;
  emotvsno = (1/3)*(mood='sad') - (2/3)*(mood='neutral') +
             (1/3)*(mood='pleasant');
  plsvssad = (-1/2)*(mood='sad') + 0*(mood='neutral') +
             (1/2)*(mood='pleasant');
  wordemot = (1/2)*(word='emotional') -
            (1/2)*(word='unemotional');
  inter1 = emotvsno * wordemot;
  inter2 = plsvssad * wordemot;
run;

proc reg;
  title 'Two-Way ANOVA of Recall Data';
  model recall = emotvsno plsvssad wordemot inter1 inter2/
         ss2 pcorr2;
run;
```

B. Source Table Outline

Source		df	
Model	5		
Mood (Main Effect)		2	
Emotional vs Neutral			1
Pleasant vs Sad			1
Word Emotionality (Main Effect)		1	
Interaction: Mood x Word		2	
Inter1: Em vs Neu x Word			1
Inter2: Pl vs Sad x Word			1
Error	12		
Total	17		



Handout from Psych 5741/5751
Univ of Colorado
used with
Judd, C.M., & McClelland, G.H. (1989).
Data Analysis: A Model Comparison Approach. HBJ.



Question 2

A. Source Table

Source	b	df	SS	MS	F*	PRE	p
Between Subjects							
Period		2	13957.8	6978.9	8.82	.60	.0047
Linear **	-51.30	1	13158.4	13158.4	16.63	.58	.0015
Quadratic	10.95	1	799.4	799.4	1.01	.08	.3347
Error		12	9494.0	791.2			
Total Between Subj		14	23451.8				
Within Subjects							
Time **	-11.60	1	1009.2	1009.2	11.99	.50	.0047
Time x Period		2	912.3	456.2	5.42	.48	<.05
Time x Lin **	-7.40	1	68.5	68.5	0.81	.06	.3848
Time x Quad	-22.63	1	843.8	843.8	10.03	.46	.0081
Error		12	1009.6	84.1			
Total Within Subj		15	2931.1				
Total		29	26382.9				

B. On average, clerks sorted 51.3 more dates when using four periods than when using two ($F^*(1,12) = 16.63$, PRE = .58, $p = .0015$) and there was no evidence for a nonlinear effect when comparing the mean of the three-period group to the means of the other two groups ($F^*(1,12) = 1.01$, PRE = .08, $p = .33$). On average, clerks were less efficient in the morning, sorting 11.6 fewer dates than in the afternoon ($F^*(1,12) = 11.99$, PRE = .50, $p = .0047$). However, there was an interaction between time of day and the quadratic trend for period such that there was a quadratic effect for morning but not for afternoon ($F^*(1,12) = 10.03$, PRE = .46, $p = .008$). Or, as can be seen in Figure 1, virtually all of the time-of-day effect is due to the difference for the three-period group.

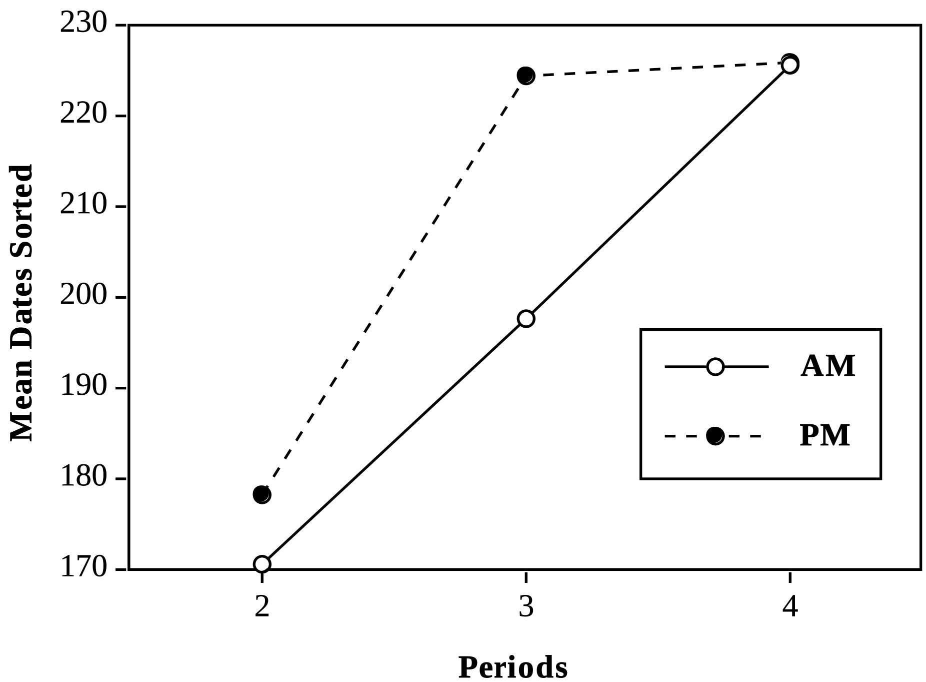
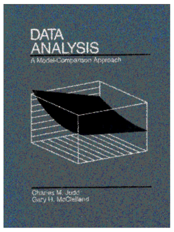
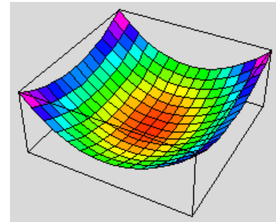


Figure 1.
Mean Number of Dates Sorted by Periods and Time of Day

C. Note that **`lin`** and **`quad`** are contrast codes and with equal numbers of observations they are effectively mean-deviated and not redundant with one another. On the other hand, **`periods`** is not mean-deviated and is necessarily redundant with **`periods*periods`**. However, they provide a complete set of codes so none of the omnibus rows nor the error rows in the source table will be affected. Because of the redundancy, the **`periods`** rows will not be the same as the **`lin`** rows. However, **`periods*periods`** captures the same higher-order (in this case, quadratic) interaction as **`quad`**, so those rows will not change. Because **`periods`** is not mean-deviated, the intercept will change. This is an issue only for the **`Time`** row in the within-subject portion of the source table.





Question 3

A. PRE = .09, F*(1,18) = 1.71, n.s.; there is no reliable evidence showing a difference in the time to adoption for stock and mutual insurance companies.

B. PRE = .64, F*(1,17) = 30.48, p < .0001; mutual companies adopted the innovation approximately 8 months faster than stock companies. The coefficient for type of company is -8.20, which is the difference in the adjusted means (i.e., the means after controlling for size). Hence, the adjusted means equal the grand mean +/- (-8.20)/2 or

Mutual	= 19.4 - 4.1 = 15.3
Stock	= 19.4 + 4.1 = 23.5
Diff	-8.2

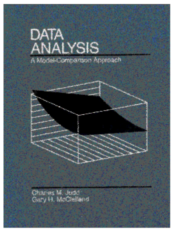
C. The size of the insurance company has an enormous effect on time to adopt the innovation. Among the companies in our study, the model predicts that the largest company ($305 million) will be approximately 27 months slower to adopt the innovation than the smallest company ($31 million). Not taking account of this large effect obscures the differences in adoption times for mutual versus stock companies. Using statistical methods to equate companies in terms of size gives us more power to detect the difference due to type of company. Hence, the model predicts that if a mutual and stock company had the same size, then the mutual company is expected to adopt the innovation approximately eight months sooner.

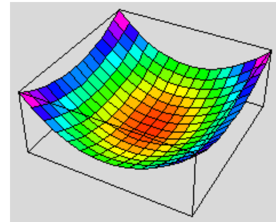
[Technical Note: Although stock companies tend to be somewhat larger than mutual companies ($194.9 million versus $168.8 million), this difference is not statistically significant (Tol = .93, PRE = .03, n.s.) and is not large enough to have produced the dramatic change in F* when Size was controlled. Instead, the dramatic change in F* is due mostly to the reduction in MSE (1535 without controlling for Size, but only 176.4 when controlling for Size) rather than a large change in the coefficient for type of firm (-5.4 versus -8.1). However, the larger difference in the adjusted means did help in the detection of the difference.]

D. PRE = .00, F*(1,16) = .001, n.s.; there is no reliable evidence showing a violation of the ANCOVA assumption of homogeneity of regression.

E. same as **D.** There is no reliable evidence showing a need for separate regression equations relating size to time to adopt for the two types of firms.

F. All outlier indices suggest that Observation #3 is not like the other data values. Its size ($575 million) is almost twice as large as the second largest company ($305 million), resulting in a very high lever of .58 (i.e., one-half of a parameter is allocated in the model to just this observation). It also has the largest residual value and the test of whether it would be useful to add an additional parameter to the model just for this observation is significant ($F^*(1,15) = 96.7$, $p < .0001$). For no other observation would it be useful to add an additional parameter to the model. The combined unusualness of both the predictor (Size) and the dependent variable (Time, with respect to a model of Time) result in an enormous value of Cook's D (6.67, with the next largest being only .15). This outlier does not appear as a long tail in the squished normal-normal quantile plot produced by proc univariate, but does appear in the better plot produced by SAS/INSIGHT. The unusual observation also stands out in the plot of residuals against predicted values. Except for that point, there is an apparent trend in the residuals, but it would probably disappear in an analysis in which Observation #3 is deleted. The obvious next step is to redo the analysis either deleting #3 or, if one's advisor is cranky about deleting observations, by adding a separate parameter just for observation #3. A power transformation would not be in order until heteroscedasticity could be checked in the new analysis. However, given that the data are essentially reaction times or counts, a priori transformations using either the log or the square root would not be inappropriate. Also, Observation #4 has a lever that is a bit unusual (it is the smallest company), but it might not be so unusual once the extremely large company (#3) is removed. Observation #6 has the second most unusual adoption time relative to the model and it is for the company that adopted the innovation first. It too probably won't be so unusual in the new analysis; if it is, then it might be worth considering what special conditions might facilitate the earliest adoption; i.e., what distinguishes the first adopter from subsequent adopters?





Question 4

A. Price reduction percentage is significantly related to the non-redemption rate (Chi-squre(1) = 151, p < .0001), such that higher price reduction percentages were associated with lower non-redemption rates. The logit of non-redemption decreased by -.1 for each 1-pt increase in the percentage price reduction. Or, the odds of non-redemption decreased by a factor of .897 for each 1-pt increase in the percentage price reduction.

B. There is no suggestion of a non-linear effect of price reduction percentage; the quadratic trend is not significant (Chi-square(1) = .04, p = .84).

C.

$$\text{lo}\hat{\text{g}}\text{it} = 2.19 - 0.11 \text{ Reduce}$$

$$\text{lo}\hat{\text{g}}\text{it} = 2.19 - 0.11\,(25) = -0.56$$

$$\hat{p} = \frac{e^{\text{lo}\hat{\text{g}}\text{it}}}{1 + e^{\text{lo}\hat{\text{g}}\text{it}}} = \frac{e^{-.56}}{1 + e^{-.56}} = .36$$

In other words, the predicted rate of non-redemption is .36, so the predicted proportion of coupons redeemed for a 25% price reduction is 1 - .36 = .64.