

Question 1

A national big-box retailer (the kind that builds big discount stores) wants to know whether increasing advertising budgets would increase sales. They don't have enough resources to make large increases in the ad budgets of all their stores so they take a fixed amount of money and then randomly distribute that money as ad budget increases across their stores. As a result, some stores get bigger increases than others, although the specific amount distributed to each store varies randomly. In addition, a randomly selected half the stores are told to spend their increased ad budget on PRINT media (newspapers primarily) and the other half are to spend their increase on BROADCAST media (radio and TV). The following variables are available for analysis:

- ADS the amount of the ad budget increase for each store
- SALES2001 the amount of sales (in thousands of dollars) in January 2001
- SALES2000 the amount of sales (in thousands of dollars) in January 2000
- REGION whether the store is in a RURAL, SUBURBAN, or URBAN location
- MEDIA PRINT or BROADCAST

Use these variables and any other variables you might construct from them (specify how you construct them!) to specify the MODEL A/MODEL C comparisons you would use to answer the following questions.

1. Did increasing ad budgets lead to higher sales in January 2001? (Ignore January 2000 sales.)
2. You've all heard about the dangers of inferring causality from correlations. Can the company make a causal inference based on the model comparison in the previous question? Why or why not? [Note: this is not a Model A/C question.]
3. Within types of advertising media (or holding constant advertising media), did increasing the ad budgets lead to higher sales in January 2001? (Again, ignoring January 2000 sales.)
4. Was the effectiveness of increased advertising budgets for January 2001 sales the same for both types of advertising media? (Again, ignoring January 2000 sales.)
5. Ignoring type of advertising media for the moment, were January 2001 sales over and above January 2000 sales affected by the increase in ad budgets?
6. Were increases in ad budgets more effective on January 2001 sales for stores with higher sales in January 2000?

7. Another manager who is not interested in the amount of the ad budget increase wants to analyze sales data in terms of a store's region and the type of advertising media used. In particular, she wants to know whether the difference in sales due to the type of advertising media is the same for rural stores as it is for the "city" stores (either suburban or urban).
8. Her boss is an old-fashioned fuddy-duddy who took his MBA business statistics course many years ago and tells her that he wants her to test the "main effect" of region.
9. Causal conclusions are often based on the results of ANOVA models. Can we infer cause for the regional affects in the prior question? Why or why not. [Note: this is not a Model A/C question.]
10. Is there any indication that increasing the ad budget is differentially effective in increasing January 2001 sales as a function of a store's region and advertising media? Use an omnibus test and ignore January 2000 sales.

Question 2.

This example is loosely based on the work of Foa, Rothbaum, Riggs & Murdock (1991) as presented by Howell (2000) [although the data have been modified, they are generally consistent with Foa et al's findings]. The authors were interested in whether two new therapies for post-traumatic stress disorder (PTSD) were more effective than the standard existing therapy. Forty-five clients with PTSD were assigned to one of four groups. (Note that assignment was not random; thus there are differences among the groups prior to treatment.) The new therapies of interest were Stress Inoculation Therapy (SIT) in which subjects were taught a variety of coping skills and Prolonged Exposure therapy (PE) in which subjects went over the traumatic event in their minds repeatedly for seven sessions. The effectiveness of these new therapies was compared to that of Supportive Counseling (SC), a standard therapy that served as a therapy control condition, and to a Waiting List (WL) control condition. Measures of the severity of post-traumatic stress symptoms were taken before treatment and after 9 biweekly sessions of therapy (or after an equivalent amount of time for WL subjects).

The following variables are available:

Pre Severity of stress symptoms before therapy
Post Severity of stress symptoms after therapy
Treat Treatment group, from which the following contrast-coded predictors were created:

	SIT	PE	SC	WL
TREATvCONT	-1	-1	-1	+3
NEWvSTAND	-1	-1	+2	0
SITvPE	-1	+1	0	0

preTREATvCONT pre*TREATvCONT
preNEWvSTAND pre*NEWvSTAND
preSITvPE pre*SITvPE

On the following pages is SAS output from these data. First, we present means and standard deviation of both Pre and Post for each treatment group. Then results from four Proc Reg Models are presented:

```
Model Post=Pre;  
Model Post= TREATvCONT NEWvSTAND SITvPE;  
Model Post= TREATvCONT NEWvSTAND SITvPE Pre;  
Model Post= TREATvCONT NEWvSTAND SITvPE Pre preTREATvCONT preNEWvSTAND preSITvPE;
```

Based on this output, answer the following questions.

1. Are there differences across the four treatment groups in the severity of their stress symptoms following therapy (i.e., Post)? Ignore Pretest scores in this analysis. Write a paragraph in the style of a journal results section that summarizes the differences and that specifically addresses the authors' hypotheses. (Discuss only contrast effect coded in the analysis.)

2. Another researcher suggests that the most appropriate comparison is between the two new therapies (SIT and PE) and the two control conditions (SC and WL). Provide a test of this additional comparison, including the resulting PRE and F*.

3. Should an adjusted p or critical F be used in answering question 2? If so, what would be an appropriate adjustment?
4. Are there significant differences among the four treatment groups in the severity of their stress symptoms prior to therapy (i.e., Pre)? (You need only provide the PRE and F* for the omnibus test here. Hint: Think about an easy way to get this.)
5. Controlling for Pretest differences in symptom severity, are there differences among the four treatment groups on symptom severity at the time of the Posttest? As you did in response to question 1, write a paragraph that summarizes the differences and that specifically addresses the authors' hypotheses.
6. What are the values of adjusted Post means, adjusting for Pre, for the two treatment conditions using the new therapies (i.e., SIT and PE)? Explain why the coefficient for SITvPE has a different sign in the ANOVA model than in the ANCOVA model, even though neither is significant.
7. Is there any reason to think that the relationship between pretest and posttest scores differs in the four groups? Provide PRE, F* and a substantive conclusion.
8. In Model 3, the slope for NEWvSTAND is significant. In Model 4 it is not. Explain this difference and provide an interpretation for this slope in Model 4.

The MEANS Procedure

treat=SIT

Variable	N	Mean	Std Dev	Minimum	Maximum
post	14	9.1428571	3.6131627	4.0000000	17.0000000
pre	14	11.0714286	3.9509353	3.0000000	18.0000000

treat=PE

Variable	N	Mean	Std Dev	Minimum	Maximum
post	10	10.5900000	9.2767870	2.0000000	30.0000000
pre	10	15.4000000	11.1175537	2.0000000	34.0000000

treat=SC

Variable	N	Mean	Std Dev	Minimum	Maximum
post	11	16.5318182	4.9309367	5.4000000	23.0000000
pre	11	18.0909091	7.1337864	5.0000000	27.0000000

treat=WL

Variable	N	Mean	Std Dev	Minimum	Maximum
post	10	19.2100000	7.1068746	9.0000000	29.0000000
pre	10	19.5000000	7.1063352	12.0000000	30.0000000

The REG Procedure
 Model: MODEL1
 Dependent Variable: post

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1738.13286	1738.13286	109.68	<.0001
Error	43	681.43692	15.84737		
Corrected Total	44	2419.56978			

Root MSE	3.98088	R-Square	0.7184
Dependent Mean	13.50778	Adj R-Sq	0.7118
Coeff Var	29.47099		

Parameter Estimates

Parameter Variable	Standard DF	Estimate	Error	t Value	Pr > t
Intercept	1	1.16967	1.31913	0.89	0.3802
pre	1	0.78978	0.07541	10.47	<.0001

Model: MODEL2
 Dependent Variable: post

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	777.61613	259.20538	6.47	0.0011
Error	41	1641.95365	40.04765		
Corrected Total	44	2419.56978			

Root MSE	6.32832	R-Square	0.3214
Dependent Mean	13.50778	Adj R-Sq	0.2717
Coeff Var	46.84946		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Type II SS
Intercept	1	13.86867	0.95233	14.56	<.0001	8493.29209
TREATvCONT	1	1.78044	0.56856	3.13	0.0032	392.72331
NEWvSTAND	1	2.22180	0.77151	2.88	0.0063	332.12730
SITvPE	1	0.72357	1.31009	0.55	0.5837	12.21630

Squared Partial

Variable	DF	Corr	Type II	Tolerance	95% Confidence Limits	
Intercept	1	.	.	.	11.94541	15.79193
TREATvCONT	1	0.19302		0.99553	0.63222	2.92867
NEWvSTAND	1	0.16824		0.99073	0.66370	3.77989
SITvPE	1	0.00739		0.98684	-1.92220	3.36935

Model: MODEL3
 Dependent Variable: post

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	1919.32594	479.83148	38.37	<.0001
Error	40	500.24384	12.50610		
Corrected Total	44	2419.56978			

Root MSE	3.53640	R-Square	0.7933
Dependent Mean	13.50778	Adj R-Sq	0.7726
Coeff Var	26.18044		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Type II SS
Intercept	1	2.53231	1.30036	1.95	0.0585	47.42773
TREATvCONT	1	0.95832	0.32917	2.91	0.0059	106.00085
NEWvSTAND	1	1.07624	0.44749	2.41	0.0209	72.33769
SITvPE	1	-0.80838	0.74946	-1.08	0.2872	14.54999
pre	1	0.70783	0.07408	9.55	<.0001	1141.70981

Variable	DF	Squared Partial Corr Type II	Tolerance	95% Confidence Limits	
Intercept	1	.	.	-0.09580	5.16043
TREATvCONT	1	0.17485	0.92751	0.29305	1.62358
NEWvSTAND	1	0.12634	0.91961	0.17182	1.98066
SITvPE	1	0.02826	0.94168	-2.32309	0.70632
pre	1	0.69534	0.81775	0.55811	0.85756

Model: MODEL4
 Dependent Variable: post

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	1950.33621	278.61946	21.97	<.0001
Error	37	469.23357	12.68199		
Corrected Total	44	2419.56978			

Root MSE	3.56118	R-Square	0.8061
Dependent Mean	13.50778	Adj R-Sq	0.7694
Coeff Var	26.36391		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Type II SS
Intercept	1	2.19255	1.45184	1.51	0.1395	28.92342
TREATvCONT	1	-0.42064	0.94558	-0.44	0.6590	2.50967
NEWvSTAND	1	1.36723	1.17583	1.16	0.2524	17.14651
SITvPE	1	-0.97579	1.77047	-0.55	0.5848	3.85234
pre	1	0.71272	0.08899	8.01	<.0001	813.39908
preTREATvCONT	1	0.07489	0.04930	1.52	0.1372	29.27231
preNEWvSTAND	1	-0.00980	0.06944	-0.14	0.8885	0.25269
preSITvPE	1	0.02249	0.13592	0.17	0.8695	0.34733

Variable	DF	Corr	Squared Partial Type II	Tolerance	95% Confidence	Limits
Intercept	1		.	.	-0.74915	5.13425
TREATvCONT	1		0.00532	0.11398	-2.33656	1.49528
NEWvSTAND	1		0.03525	0.13507	-1.01524	3.74969
SITvPE	1		0.00814	0.17111	-4.56310	2.61152
pre	1		0.63416	0.57464	0.53240	0.89304
preTREATvCONT	1		0.05872	0.10942	-0.02499	0.17478
preNEWvSTAND	1	0.00053823		0.12142	-0.15049	0.13089
preSITvPE	1	0.00073965		0.12706	-0.25290	0.29789