Grad Stats First Spring Exam 1999

<u>Question 1</u> (all sub-questions worth 3 points except question h which is worth 6.)

Life insurance companies are interested in identifying factors that predict how much life insurance one carries. They collect data from a sample of 100 individuals who hold life insurance policies. For these individuals they have available the following variables:

AMT	Amount of life insurance carrried in thousands of dollars.
INC	Average annual income in thousands of dollars.
RISK	A measure of risk aversion, from 1 to 9, with higher numbers meaning the individual is more reluctant to take risks.
GEN	The respondent's gender (+1 if female; -1 if male).

Assume that you have access to this dataset. For each of the following questions, specify the models A and C that you would compare to provide the answer. You will need to create some additional variables from those listed above. Clearly specify how these additional variables are computed. (Feel free to use the first letters of the above variables (A, I, R, and G) in refering to them.)

a. Do respondents with higher incomes spend more on life insurance than respondents with lower incomes?

b. Is there a difference in the strength of the relationship between income and the amount spent on life insurance for male and female respondents?

c. The industry expectation is that for every $1000 of income, people might be expected to spend $10 on life insurance. Is there reason to doubt this expectation based on this dataset? (Ignore all other potential predictor variables in these models.)

d. Is there evidence to suggest that the relationship between income and amount spent on life insurance is less strong at higher income levels? (Again, ignore all other potential predictor variables in these models.)

e. Assuming there is evidence to suggest that the relationship between income and amount spent on life insurance is less strong at higher income levels, at the average income is there reason to doubt the industry expectation given in question 3 (i.e., that for every $1000 of income, people might be expected to spend $10 on life insurance)?

f. Assuming only a linear relationship of income with amount spent on life insurance, is that relationship stronger among people who are more risk averse than among less risk averse people?

g. Assuming that the linear relationship between income and amount spent depends on how risk averse someone is, is there continuing evidence that the relationship between income and amount spent on life insurance is less strong at higher income levels?

h. The last question is *not* a model comparison question. Assume that RD is computed by deviating each respondent's RISK score from the mean of all respondents. In the following model, provide brief interpretations for parameter estimates b_1, b_2, and b_4 (*not* b_3).

$$I = b_0 + b_1RD + b_2I + b_3(I*RD) + b_4(RD*RD) + e$$

Question 2 (All sub-question worth 3 points.)

Conti and Musty (1985) examined the effects of different amounts of THC, the major active ingredient in marijuana, injected directly into the nucleus accumbens of rats. The dependent variable was locomotor activity, which normally increases with the administration of THC by more traditional routes. Because of the nature of the experimental setting (all animals were observed under baseline conditions and then again after the administration of THC), activity should decrease in all animals as they become familiar with the apparatus. If THC has its effect through the nucleus accumbens, however, the effects of moderate doses of THC should partially compensate for this anticipated decrease, leading to relatively greater activity levels in the moderate-dose groups as compared to the low- or high-dose groups. To avoid technicalities about the actual doses, we will refer to the five drug conditions as indicated in the top row of this table. The other rows define the codes.

	Control	Low	Low-Moderate	Hi-Moderate	High
DRUGCONT	-4	1	1	1	1
LIN	0	-3	-1	1	3
QUAD	0	1	-1	-1	1
CUBIC	0	-1	3	-3	1

The following variables are used in the analysis:

DRUGCONT	Drug vs Control Contrast (-4, 1, 1, 1, 1)
LIN	Linear Trend for Drug (0, -3, -1, 1, 3)
QUAD	Quadratic Trend for Drug (0, 1, -1, -1, 1)
CUBIC	Cubic Trend for Drug (0, -1, 3, -3, 1)
PRE	Pre Injection Activity
POST	Post Injection Activity
POSTMPRE	Diff = Post - Pre
DCPRE	Drugcont x Pre
LINPRE	Linear x Pre
QUADPRE	Quadratic x Pre
CUBPRE	Cubic x Pre

On the attached outputs are available the overall means and the means by drug group as well as the results of the following Proc Reg models:

1. model post = pre;
2. model post = drugcont lin quad cubic;
3. model pre = drugcont lin quad cubic ;
4. model postmpre = drugcont lin quad cubic ;
5. model post = drugcont lin quad cubic pre ;
6. model post = drugcont lin quad cubic pre linpre quadpre cubpre dcpre ;

Use the outputs to answer the following questions. Note that many (but not all!) of these questions ask for omnibus, rather than 1-df, tests.

a. Ignoring the amount of drug (if any) an animal received, is PRE activity a useful predictor of POST activity? Give PRE, F*, p, and a brief interpretation of the coefficient.

b. Ignoring the measure of PRE activity, does the amount of post-injection activity differ across the five treatment conditions? Give PRE, F*, and p.

c. In the analysis for the previous question, is there evidence of a quadratic effect of amount of THC on post activity level? Give PRE, F*, p, and a brief interpretation of the coefficient.

d. The researchers randomly assigned rats to drug conditions. However, with small group sizes, random assignment doesn't always produce equivalent groups. Is there any indication that the five treatment groups differed on their activity levels *before* the injections were given? Give PRE, F*, and p.

e. Believing that the previous result cast doubt on their findings about the effects of THC on activity levels, the researchers controlled for each rat's initial activity level by analyzing the POSTMPRE, the difference score between POST and PRE activity levels. When doing so, did they find that the change in activity level differed across the five treatment groups? Give PRE, F*, and p.

f. After consulting with an expert statistician like yourself, they performed an analysis of covariance of POST scores controlling for PRE scores. Now what is their conclusion about differences among the five treatment groups? Give PRE, F*, and p. Briefly explain why the results of this analysis and the previous analysis are so different.

g. When doing the analysis of covariance in the previous question, is there evidence for a quadratic effect of THC on activity level when controlling for PRE activity levels? Give PRE, F*, and p.

h. The adjusted means for each drug group are 1.7, 2.4, ???, 2.4, and 1.9, respectively. Calculate the missing adjusted mean (i.e., the adjusted mean for the low-moderate drug group). Is the pattern of means consistent with the hypothesis that there would be "relatively greater activity levels in the moderate-dose groups as compared to the low- or high-dose groups"?

i. Is the ANCOVA significantly more powerful than the change score analysis? In other words, can we reject the null hypothesis that the slope for PRE equals 1 when controlling for drug condition? Give PRE, F*, and p.

j. Is there any reason to reject the assumption of homogeneity of regression? Give, PRE, F*, and p.

Across all groups:

Variable	Label	N	Mean	Std Dev
PRE	Pre Injection Activity	47	4.8059574	2.3787836
POST	Post Injection Activity	47	2.2857447	1.4946931

By drug groups:

DRUG=control

Variable	Label	N	Mean	Std Dev
PRE	Pre Injection Activity	10	3.3770000	1.3962891
POST	Post Injection Activity	10	1.0940000	0.5849634

DRUG=low

Variable	Label	N	Mean	Std Dev
PRE	Pre Injection Activity	10	5.3870000	3.4447837
POST	Post Injection Activity	10	2.5860000	1.5331681

DRUG=lowmod

Variable	Label	N	Mean	Std Dev
PRE	Pre Injection Activity	9	6.4944444	2.3781406
POST	Post Injection Activity	9	3.9055556	1.4768134

DRUG=himod

Variable	Label	N	Mean	Std Dev
PRE	Pre Injection Activity	8	5.0450000	1.6876018
POST	Post Injection Activity	8	2.4850000	1.1873861

DRUG=high

Variable	Label	N	Mean	Std Dev
PRE	Pre Injection Activity	10	3.9430000	1.2207470
POST	Post Injection Activity	10	1.5600000	0.8764829

Model: MODEL1
Dependent Variable: POST Post Injection Activity

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	73.41964	73.41964	112.571	0.0001
Error	45	29.34931	0.65221		
C Total	46	102.76895			

Root MSE	0.80759	R-square	0.7144
Dep Mean	2.28574	Adj R-sq	0.7081
C.V.	35.33174		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-0.266676	0.26786180	-0.996	0.3248
PRE	1	0.531095	0.05005631	10.610	0.0001

Model: MODEL2
Dependent Variable: POST   Post Injection Activity

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	44.30285	11.07571	7.956	0.0001
Error	42	58.46610	1.39205		
C Total	46	102.76895			

Root MSE	1.17985	R-square	0.4311
Dep Mean	2.28574	Adj R-sq	0.3769
C.V.	51.61783		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	2.377449	0.17382788	13.677	0.0001
DRUGCONT	1	0.308028	0.084177923	3.659	0.0007
LIN	1	-0.224928	0.08417792	-2.672	0.0107
QUAD	1	-0.561139	0.19478971	-2.881	0.0062
CUBIC	1	0.161783	0.08995160	1.799	0.0793

Variable	DF	Type II SS	Squared Partial Corr Type II	Tolerance
INTERCEP	1	260.398463	.	.
DRUGCONT	1	18.639673	0.24174159	0.99819328
LIN	1	9.939046	0.14529675	0.99733073
QUAD	1	11.552170	0.16498793	0.99672324
CUBIC	1	4.503027	0.07151167	0.99557024

Model: MODEL3
Dependent Variable: PRE   Pre Injection Activity

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	57.35828	14.33957	2.968	0.0302
Error	42	202.93785	4.83185		
C Total	46	260.29613			

Root MSE	2.19815	R-square	0.2204
Dep Mean	4.80596	Adj R-sq	0.1461
C.V.	45.73798		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	4.910634	0.32385373	15.163	0.0001
DRUGCONT	1	0.368072	0.15682945	2.347	0.0237
LIN	1	-0.289072	0.15682945	-1.843	0.0724
QUAD	1	-0.552361	0.36290710	-1.522	0.1355
CUBIC	1	0.145217	0.16758624	0.867	0.3911

Variable	DF	Type II SS	Squared Partial Corr Type II	Tolerance
INTERCEP	1	1110.941397	.	.
DRUGCONT	1	26.614886	0.11594236	0.99819328
LIN	1	16.416147	0.07483860	0.99733073
QUAD	1	11.193580	0.05227434	0.99672324
CUBIC	1	3.628022	0.01756351	0.99557024

Model: MODEL4
Dependent Variable: POSTMPRE   Diff = Post - Pre

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	1.59450	0.39862	0.197	0.9386
Error	42	84.98660	2.02349		
C Total	46	86.58110			

Root MSE	1.42249	R-square	0.0184
Dep Mean	-2.52021	Adj R-sq	-0.0751
C.V.	-56.44343		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-2.533185	0.20957651	-12.087	0.0001
DRUGCONT	1	-0.060044	0.10148955	-0.592	0.5573
LIN	1	0.064144	0.10148955	0.632	0.5308
QUAD	1	-0.008778	0.23484924	-0.037	0.9704
CUBIC	1	0.016567	0.10845063	0.153	0.8793

Variable	DF	Type II SS	Squared Partial Corr Type II	Tolerance
INTERCEP	1	295.630917	.	.
DRUGCONT	1	0.708279	0.00826512	0.99819328
LIN	1	0.808308	0.00942139	0.99733073
QUAD	1	0.002827	0.00003326	0.99672324
CUBIC	1	0.047218	0.00055528	0.99557024

Model: MODEL5
Dependent Variable: POST Post Injection Activity

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	5	82.64350	16.52870	33.673	0.0001
Error	41	20.12544	0.49086		
C Total	46	102.76895			

Root MSE	0.70062	R-square	0.8042
Dep Mean	2.28574	Adj R-sq	0.7803
C.V.	30.65160		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	0.243000	0.26264493	0.925	0.3603
DRUGCONT	1	0.148042	0.05316322	2.785	0.0081
LIN	1	-0.099280	0.05196881	-1.910	0.0631
QUAD	1	-0.321050	0.11881686	-2.702	0.0100
CUBIC	1	0.098664	0.05389023	1.831	0.0744
PRE	1	0.434659	0.04918122	8.838	0.0001

Variable	DF	Type II SS	Squared Partial Corr Type II	Tolerance
INTERCEP	1	0.420180	.	.
DRUGCONT	1	3.806354	0.15905006	0.88246040
LIN	1	1.791431	0.08173752	0.92269189
QUAD	1	3.583867	0.15115863	0.94462019
CUBIC	1	1.645341	0.07557565	0.97808453
PRE	1	38.340658	0.65577586	0.77964221

Model: MODEL6
Dependent Variable: POST Post Injection Activity

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	9	84.66027	9.40670	19.220	0.0001
Error	37	18.10868	0.48942		
C Total	46	102.76895			

Root MSE	0.69959	R-square	0.8238
Dep Mean	2.28574	Adj R-sq	0.7809
C.V.	30.60658		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	0.247220	0.30736953	0.804	0.4264
DRUGCONT	1	-0.012437	0.14024471	-0.089	0.9298
LIN	1	-0.089474	0.14474619	-0.618	0.5403
QUAD	1	-0.350454	0.35309926	-0.993	0.3274
CUBIC	1	-0.156411	0.17005933	-0.920	0.3637
PRE	1	0.412482	0.06460155	6.385	0.0001
LINPRE	1	0.003087	0.03182115	0.097	0.9232
QUADPRE	1	0.021088	0.06912001	0.305	0.7620
CUBPRE	1	0.045769	0.02997407	1.527	0.1353
DCPRE	1	0.044116	0.03614989	1.220	0.2300

Variable	DF	Type II SS	Squared Partial Corr Type II	Tolerance
INTERCEP	1	0.316615	.	.
DRUGCONT	1	0.003849	0.00021249	0.12643530
LIN	1	0.187008	0.01022142	0.11859096
QUAD	1	0.482118	0.02593317	0.10664569
CUBIC	1	0.414016	0.02235180	0.09793075
PRE	1	19.953004	0.52422804	0.45053814
LINPRE	1	0.004607	0.00025434	0.08537947
QUADPRE	1	0.045557	0.00250942	0.08440935
CUBPRE	1	1.141142	0.05928062	0.08554827
DCPRE	1	0.728887	0.03869324	0.11521875