Master's Comprehensive Exam

STATISTICS

September 17, 2005

Instructions:
Do 7 problems out of the 10 problems given below.

1. A biologist considered studying the effects of various environmental pollutants on the health of mice by using a 2^{5-1} fractional factorial design.
   a. How many factors are involved?

   b. How many levels are there for each factor?

   c. How many experimental trials will be required for the single replicate of the experiment?

   d. Choose a defining relation. What treatments are used for the design?

   e. Which effects can be estimated?

   f. What are the aliases for the main effects?

2. A state university developed a training program to teach general computer skills. Researchers conducted an experiment to evaluate the effects of three different incentive methods on achievement during the program. The blocking variables were IQ and age of subject. The data below give the achievement scores for the participants in the experiment.

	Age		
IQ	Young	Middle	Old
High	19(B)	20(A)	25(C)
Normal	24(C)	14(B)	14(A)
Low	10(A)	12(C)	7(B)

   a. Write down the statistical model for the design. Estimate the parameters.

   b. Analyze the data. Which blocking factors are significant?

   c. The program offers morning, day, and night classes.
      I) Does the time of the day matter?
      II) What design should be used to answer this question?
      III) Give an example of how this design might be run.

3. An accounting firm, prior to introducing in the firm a widespread training in statistical sampling for auditing, tested three training methods:

Method 1: study at home with training materials,
Method 2: training sessions at local offices conducted by local staff, and
Method 3: training session in Chicago conducted by national staff.
Thirty auditors were grouped into 10 blocks of three, according to time elapsed since college graduation, and the auditors in each block were randomly assigned to the three training methods. At the end of the training, the subjects took a test. The results were (block 1 consists of the most recent graduates).

	Training Method		
Block	1	2	3
1	73	81	92
2	76	78	89
3	75	76	87
4	74	77	90
5	76	71	88

	Training Method		
Block	1	2	3
6	73	75	86
7	68	72	88
8	64	74	82
9	65	73	81
10	62	69	78

a. Why do you think the blocking variable "time elapsed since college graduation" was employed?

b. Why is it important to randomize subjects within each block?

c. Use the attached SAS output to analyze the data. Give the ANOVA table. What is your conclusion?

d. What design would you recommend if there are not enough subjects to run a complete block design?

4. The Kenton Toy Car Company wishes to test three different production technologies. The company executives would like to know whether parts should be glued, screwed on, or welded to produce a long-lasting toy. Four toy cars from each technological process were randomly selected and given to kids to play. The number of times the cars were thrown on the floor during the play until a piece fell off was recorded. The data are

Method		
Glued	Screwed on	Welded
9	5	12
6	2	10
4	5	9
6	8	11

Use the attached SAS output when appropriate to answer the following questions:

a. Write down a suitable model for this experiment. State the assumptions and estimate the parameters.
b. Are the model assumptions valid? Discuss results of the diagnostic check procedures.
c. What are the two quantities that are displayed by the normal probability plot?
d. Test the hypothesis that the cars last equally long when assembled by the three methods.
e. How would the analysis differ if these methods were randomly selected from a large list of possible methods?

5. Particle-board is made from wood chips and resin. An experiment is conducted to study the effect of using slash chips (waste wood chips) along with standard chips. The researchers make eighteen boards by varying the amount of resin (factor A), the target density (factor B), and the fraction of slash (factor C). The response is the actual density of the boards produced.

	Density=42		Density=48	
	Slash		Slash	
Resin	0%	50%	0%	50%
6%	40.9	41.9	42.0	44.4
	45.4	46.0	46.2	48.4
9%	42.8	43.9	44.8	48.2
	48.6	50.7	46.2	49.9

a. Set up a design to run these observations in two blocks with ABC confounded. You don't need to analyze the data.

b. Set up a design to run these observations in two blocks with ABC confounded in Replicate I and AC confounded in Replicate II. Do not do the analysis.

6. Three states (factor A) participated in a health awareness study. Each state independently devised a health awareness program. Three cities (factor B) within each state were selected for participation and five households within each city were randomly selected to evaluate the effectiveness of the program. All members of the selected households were interviewed before and after participation in the program and a composite index was formed for each household measuring the impact of the health awareness program. The data on health awareness follow (the larger the index, the greater the awareness).

State	1			2			3		
City	1	2	3	1	2	3	1	2	3
	42	26	34	47	56	68	19	18	16
	56	38	51	58	43	51	36	40	28
	35	42	60	39	65	49	24	27	45
	40	35	29	62	70	71	12	31	30
	28	53	44	65	59	57	33	23	21

a. What design is used for these data? Explain why.

b. Do the analysis assuming the fixed factor effects. Use the attached SAS output. What are your findings about the appropriateness of the model?

c. Analyze the data assuming a mixed effects model.

d. Analyze the data assuming a random effects model.

e. When writing a SAS code for this problem, what procedure would you use?

7. An automobile manufacturer wished to study the effect of difference between drivers (factor A) and differences between cars (factor B) on gasoline consumption. Two drivers were selected at random; also three cars of the same model with manual transmission were randomly selected from the assembly line. Each driver drove each car twice over a 40-mile test course and the miles per gallon were recorded. The data follow.

	Car		
Driver	1	2	3
1	25.3	28.9	24.8
	25.2	30.0	25.1
2	29.2	32.4	27.7
	29.3	32.4	28.9

a. State the model. Obtain point estimates of the variances.

b. Are the main effects present? Is there an interaction? Which factor appears to have the greater effect on gasoline consumption?

8. For the data in Problem 4,

a. Conduct all pairwise comparisons using the Fisher LSD method at the 0.05 significance level. Use $\sqrt{MSE}$ =1.993.

b. Compute 95% confidence intervals for the difference in treatment means.

9. In a two-factor study, the treatment means are as follows:

	Factor B			
Factor A	B1	B2	B3	B4
A1	250	265	268	269
A2	288	273	270	269

a. Obtain the factor B main effects. What do your results imply about factor B?

b. Prepare a treatment means plot and determine whether the two factors interact. How can you tell that interactions are present?

10. In a chemical reaction, the reactant concentration is factor A run at two levels, and the catalyst is factor B, with two levels. Only four experiments can be made from a single batch of raw material. Therefore, three batches (blocks) are used to run three replicates. The data are

	Batch		
Treatments	1	2	3
(1)	28	25	27
a	36	32	32
b	18	19	23
ab	31	30	29

a. Write the statistical model for these data. Explain the model components.
b. Compute the analysis of variance. Is it necessary to do the blocking?

Problem 3. SAS OUTPUT

Dependent Variable: score

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	1728.366667	157.124242	25.18	<.0001
Error	18	112.333333	6.240741		
Corrected Total	29	1840.700000			

R-Square	Coeff Var	Root MSE	score Mean
0.938972	3.240139	2.498147	77.10000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
block	9	433.366667	48.151852	7.72	0.0001
method	2	1295.000000	647.500000	103.75	<.0001

Problem 4. SAS OUTPUT

Dependent Variable: falls

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	66.5000000	33.2500000	8.37	0.0088
Error	9	35.7500000	3.9722222		
Corrected Total	11	102.2500000			

R-Square	Coeff Var	Root MSE	falls Mean
0.650367	27.49025	1.993043	7.250000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
method	2	66.50000000	33.25000000	8.37	0.0088

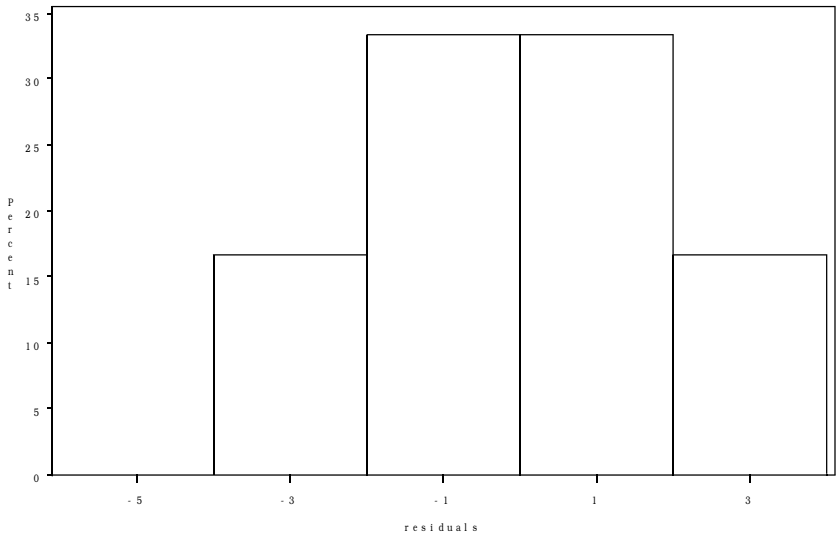
Bartlett's Test for Homogeneity of falls Variance

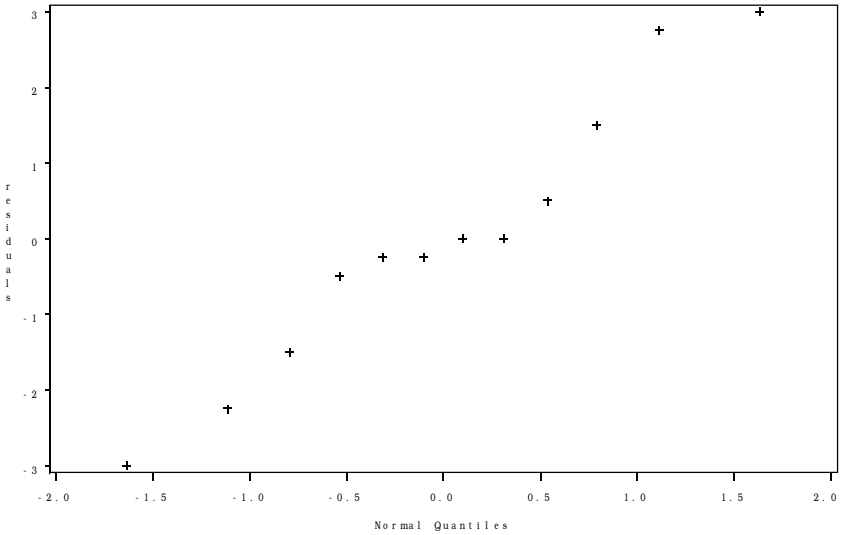
Source	DF	Chi-Square	Pr > ChiSq
method	2	1.0150	0.6020

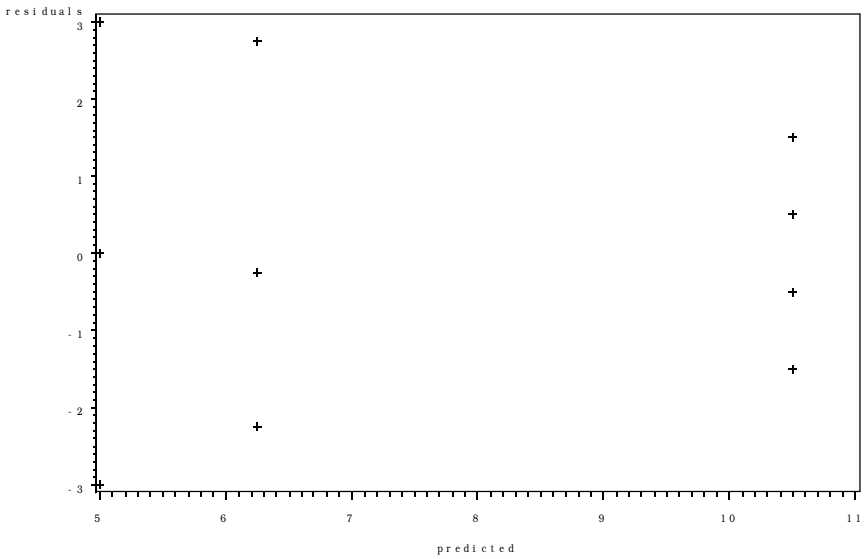
Tests for Normality

Test	--Statistic---		-----p Value-----	
Shapiro-Wilk	W	0.954223	Pr < W	0.699
Kolmogorov-Smirnov	D	0.166667	Pr > D	>0.150
Cramer-von Mises	W-Sq	0.053283	Pr > W-Sq	>0.250
Anderson-Darling	A-Sq	0.299158	Pr > A-Sq	>0.250

Problem 4. SAS OUTPUT (cont.)







Problem 6. SAS OUTPUT

Nested Random Effects Analysis of Variance for Variable index

Variance Source	DF	Sum of Squares	F Value	Pr > F	Error Term	Mean Square	Variance Component	Percent of Total
Total	44	11038				250.855556	338.843704	100.0000
state	2	6976.844444	124.88	<.0001	city	3488.422222	230.699259	68.0843
city	6	167.600000	0.26	0.9526	Error	27.933333	-16.042222	0.0000
Error	36	3893.200000				108.144444	108.144444	31.9157