

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE AND ENGINEERING
FINAL EXAMINATION, DECEMBER 2001
Third Year Industrial Engineering Program

MIE 337H1F - STATISTICS & EXPERIMENTAL DESIGN

Exam Type: B

Aids Allowed: 1. Aid Sheet supplied with this exam
2. One non-programmable calculator

Examiners - D. R. Edwards

MIE 337F: Statistics & Experimental Design
Final Examination – December 10, 2001

VALUE

4

1. X and Y are two independent continuous random variables with means μ_x and μ_y and variances σ_x^2 and σ_y^2 . Specify the mean and variance of Z where $Z = aX - bY - c$ and a, b and c are constants?

3

2. Find the probability that in successive tosses of a fair die a 3 comes up for the first time on the fifth toss. What distribution models this set of events?

4

3. Houses in a coastal residential area were found to have dry basements in only 62.4% of cases. How many houses would it be necessary to check to estimate the population proportion of dry basements within .03 with 95% confidence?

5

4. Construct the 99% confidence interval for σ^2 from the following sample data: 4, 1, 8, 0, 1, 9, 8, 3, 2, 2. The null hypothesis is that $\sigma^2 = 18$. What decision would you make concerning this null hypothesis?

x	$x - \bar{x}$	$(x - \bar{x})^2$
4	0.2	0.04
1	-2.8	7.84
8	4.2	17.64
0	-3.8	14.44
1	-2.8	7.84
9	5.2	27.04
8	4.2	17.64
3	-0.8	0.64
2	-1.8	3.24
2	-1.8	3.24
Σ		99.60

5. The EPA wants to model the gas consumption ratings, Y, of automobiles as a function of their engine displacement, X . A quadratic model:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$$

is proposed. A sample of 50 engines with displacements from 90 to 212 cubic inches with an average of 150 are selected and the fuel consumption in miles per gallon is measured. X_i is scaled so that $X = 1.00$ when the engine displacement is 100 cubic inches. The least squares model is:

$$\hat{Y}_i = 51.3 - 10.1X_i + 0.15X_i^2$$

and $MSE = 4.72 \quad S_{b_2}^2 = 0.0037 \quad R^2 = 0.93$

4

a) sketch the fitted model between $X = 1$ and $X = 3$ in 0.5 increments.

6

b) We are interested in knowing whether there is evidence that the quadratic term in the model is contributing to the prediction of Y. Conduct a t-test at $\alpha = 0.01$ and appropriate degrees of freedom.

2

c) Assuming that the model is appropriate, use the model to estimate the mean miles per gallon rating for all cars with 195 cubic inch engines (i.e., $X_0 = 1.95$).

4

d) Suppose the 99% confidence interval for the mean of Y quantity estimated in part (c) is (29.27, 35.09). Interpret this interval. With 99% confidence, what is the largest error, ε , that is likely in estimating $\mu_{Y|X}$ from $\hat{Y}_{X=1.95}$? What is \underline{x}_0 for $X_0 = 1.95$?

MIE 337F: Statistics & Experimental Design
Final Examination – December 10, 2001

VALUE
9

e) Suppose you purchase an automobile with a 350 cubic inch engine ($X_0 = 3.50$) and find that the actual miles per gallon rating is 9.7. If you use the fitted model, what fuel consumption is predicted? What is \hat{x}_0 for $X_0 = 3.50$? What is the 99% prediction interval for mileage rating of this engine when Y_0 is estimated from $\hat{Y}_{\text{hat}}|_{x=3.50}$? Is the actual miles per gallon rating in the prediction interval? Discuss your result.

11

6. For each of the items on the left in the list below there is a corresponding definition, partial definition or instance in the list on the right. Write the item number from the left list beside the most appropriate item on the right. Please note that the list on the right contains more items than the list on the left.

A	F_{\max}
B	multicollinearity
C	randomized blocks design
D	squared Pearson product-moment correlation
E	Mallow's C_p
F	hypergeometric distribution
G	polynomial regression
H	heteroskedasticity
I	regression line through origin
J	Type II ANOVA Model
K	probability of rejecting false null hypothesis
L	probability of rejecting null hypothesis when it is true
M	Scheffé test
N	a Gauss-Markov condition for regression
O	Poisson distribution
P	multinomial distributed events
Q	random independent sampling of binary events with replacement
R	Normal Distribution
S	Tukey HSD Test
T	eta-squared (η^2)
U	$F_{1,v,\alpha}$
V	dependent events

	model underfit if this statistic greater than number of predictors
	β
	variance of dependent measure non-constant across range of predictor
	random effects
	α
	power
	$(S_{xy})^2 / (S_{xx} S_{yy})$
	Bernoulli trials
	test of $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$ for k groups
	$Y_i = \beta_0 + \beta_{1,1}X_i^2 + \epsilon_i$
	$t^2_{v,\alpha/2}$
	$P(A \cap B) = P(A) P(B A)$
	pairwise comparisons of means
	correlated regression predictors
	approximates Binomial for very large n and p not extreme
	treatments assigned randomly to like experimental units
	estimated effect size
	$E(\epsilon_i) = 0$
	events with more than 2 outcomes each with probability p_i
	a posteriori comparison of 2 or more means
	limiting form of Binomial for very large n, small p and constant mean
	$Y_i = \beta_{1,1}X_i + \beta_{2,2}X_i + \epsilon_i$
	intercept
	sampling without replacement

MIE 337F: Statistics & Experimental Design
Final Examination – December 10, 2001

7. A manufacturing firm is interested in boosting the sales of its three product lines by introducing a commission component to supplement the salaries of their sales staff. Management wishes to examine three commission alternatives as a Treatment Factor in an analysis of variance experiment: 60% of Current Salary + a 10% Commission (t_1), 40% of Current Salary + 20% commission (t_2), and 30% of Current Salary + 30% Commission (t_3). The trial period for the three commission programs (Treatments) is one year, and the dependent measure is the number of sales \$10K units surpassed above the sales rep's normal sales quota. Realizing that SALES TERRITORIES, East (a_1), Central (a_2) and West (a_3), as well as the PRODUCT LINES (b_1, b_2, b_3) to which sales persons are dedicated are important determinants of the sales posted by an individual sales rep, TERRITORY and LINE are included as blocking factors in the experiment. It is also assumed that the two Blocking Factors (Territories and Line) and the Treatment Factor (Commission Program) each contribute independently to the prediction of sales. Within each blocking combination, one commission program was assigned to 4 employees with a total of 36 sales reps participating in the experiment.

Table 1: Territory (A) x Product Line (B) x Commission Program (T)

Design				Observed Data						Σn_{jk} = 36
	b_1	b_3	b_2		b_1	b_3	b_2	Mean $\bar{y}_{.j}$	$S_{.j}$	Grand Mean
a_2	t_3	t_2	t_1	a_2	6, 8, 12, 7	0, 0, 1, 4	0, 2, 2, 5	2.92	3.80	
a_1	t_2	t_1	t_3	a_1	2, 5, 3, 1	2, 2, 4, 6	9, 10, 12, 12	5.67	4.08	
a_3	t_1	t_3	t_2	a_3	0, 1, 1, 4	2, 1, 1, 5	0, 1, 1, 4	1.75	1.66	
Mean $\bar{y}_{.j}$					4.17	2.33	4.83			
$S_{.j}$					3.59	1.97	4.67			

Table 2

	Treatments		
	t_1	t_2	t_3
Mean $\bar{y}_{..k}$	2.42	1.83	7.08
$n_{..k}$	12	12	12
$S_{..k}$	1.93	1.75	4.17

Table 3: Factor B (Product Line) x T (Commission Plan)

Product Line	Commission Plan	Mean Sales Level
1	1	1.50
1	2	2.75
1	3	8.25
2	1	2.25
2	2	1.50
2	3	10.75
3	1	3.50
3	2	1.25
3	3	2.25

MIE 337F: Statistics & Experimental Design
Final Examination – December 10, 2001

- | VALUE | a) What kind of Design is this? | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---------------------|--|---------------------|----------------|----|---------------------------------|----|---------------------------------|--------------|-------|--|--|--|--|--------------|-------|--|--|--|--|------------|--------|--|--|--|--|----------|-------|----------------|--|--|--|-------------|-------|--------------|--|--|--|--------------|---------------|--|--|--|--|
| 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | b) Write the model equation representing Sale Level as Y_{ijk} , Territory as α_i , Product Line as β_j and Commission Plan as τ_k . | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | c) What assumptions (3) about the parameters are necessary to make them estimable? | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | d) State the assumptions about the relation among the blocking factors and the treatment in terms of the parameters $\alpha\beta\gamma \dots \alpha\beta\tau_{ijk}$? | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | e) Both blocking factors include all levels possible (i.e., there are only three territories, and only three product lines sold by the company). The treatment levels were selected from a population of alternatives, but these are the only commission programs upper management is willing to consider. What type of ANOVA Model Type is appropriate for this experiment? | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8 | f) <u>Copy</u> the partial ANOVA summary Table below <u>to your answer booklet</u> and complete it. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">SOURCE OF VARIATION</th> <th style="text-align: left;">SUM OF SQUARES</th> <th style="text-align: left;">DF</th> <th style="text-align: left;">MEAN SQUARE</th> <th style="text-align: left;">F*</th> <th style="text-align: left;">YOUR DECISION at $\alpha = .05$</th> </tr> </thead> <tbody> <tr> <td>A: Territory</td> <td>92.39</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>B: Prod.Line</td> <td>40.22</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>T: Program</td> <td>198.72</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Residual</td> <td>33.39</td> <td>$(k-1)(k-2) =$</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Within.Cell</td> <td>99.50</td> <td>$k^2(n-1) =$</td> <td></td> <td></td> <td></td> </tr> <tr> <td>TOTAL</td> <td>464.22</td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> | SOURCE OF VARIATION | SUM OF SQUARES | DF | MEAN SQUARE | F* | YOUR DECISION at $\alpha = .05$ | A: Territory | 92.39 | | | | | B: Prod.Line | 40.22 | | | | | T: Program | 198.72 | | | | | Residual | 33.39 | $(k-1)(k-2) =$ | | | | Within.Cell | 99.50 | $k^2(n-1) =$ | | | | TOTAL | 464.22 | | | | |
| SOURCE OF VARIATION | SUM OF SQUARES | DF | MEAN SQUARE | F* | YOUR DECISION at $\alpha = .05$ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| A: Territory | 92.39 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| B: Prod.Line | 40.22 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| T: Program | 198.72 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Residual | 33.39 | $(k-1)(k-2) =$ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Within.Cell | 99.50 | $k^2(n-1) =$ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TOTAL | 464.22 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6 | g) List three assumptions underlying the F-Test of Treatments. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6 | h) Plot the Treatment means by Employee Type using Employee Type as the parameter. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | i) Use the ANOVA results and your plot to discuss whether the design / model is appropriate? | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | j) What is the Treatment factor null hypothesis? Are you in a position to test it? Explain. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | k) Use Cochran's test to assess the homogeneity of variance within Treatments at $\alpha = .05$. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | l) Discuss whether it is appropriate to conduct contrasts on the Treatment main effect. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 11 | m) Making an <u>assumption</u> that contrasts are appropriate, construct and test a contrast at $\alpha = .01$ that assesses whether the 30/30 commission plan (t_3) is significantly more effective than the other two plans in increasing sales. State the null hypothesis for your contrast in terms of the population treatment means and the contrast coefficients. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

MIE 337F: Statistics & Experimental Design
Final Examination – December 10, 2001

VALUE	n) Suppose that that model has been shown to be inappropriate in i). Management is willing to commit additional funds to expand the experiment with less restrictive assumptions, but has asked you to salvage the data already collected. The company has many hundreds of sales representatives who can be included in the experiment, and it is known that the gross revenues of the company in the coming quarter are typically the same as those of the quarter in which the first phase of the experiment was conducted.
6	i. make a table for our your new design in a format similar to Table 1, indicating how treatments are assigned to sales reps within blocks
1	ii. indicate the number of cases per cell
4	iii. in place of the "Observed Data" in Table 1, copy the data salvaged from phase one into the appropriate cells.
2	iv. what kind of fourth treatment level might serve as a control group for the experiment. Add it to the table you created in 7.n.i.
4	v. discuss why your new design overcomes the shortcomings of the original design.and the benefit of including the control group.
6	vi. without data, set up a partial ANOVA summary table for your new design, including the sources of variation, degrees of freedom, mean squares and composition of the F-ratios.
TOTAL MARKS	
130	



TABLE A.6* Critical Values of the F-Distribution

$f_{\alpha}(v_1, v_2)$		v_1									
v_2		1	2	3	4	5	6	7	8	9	
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5		
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38		
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81		
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00		
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77		
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10		
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68		
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39		
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18		
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02		
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90		
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80		
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71		
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65		
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59		
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54		
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49		
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46		
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42		
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39		
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37		
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34		
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32		
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30		
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28		
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27		
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25		
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24		
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22		
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21		
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12		
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04		
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96		
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88		

*Reproduced from Table 18 of *Biometrika Tables for Statisticians*, Vol. I, by permission of E. S. Pearson and the Biometrika Trustees

TABLE A.6 (continued) Critical Values of the F-Distribution

$f_{\alpha}(v_1, v_2)$		v_1									
v_2		1	2	3	4	5	6	7	8	9	
1	-0.52	4999.5	5403	5625	5764	5839	5928	5981	6022		
2	96.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39		
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35		
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66		
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16		
6	13.75	10.92	9.78	9.15	8.75	8.47	8.25	8.10	7.98		
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72		
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91		
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35		
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94		
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63		
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39		
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19		
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03		
15	8.68	6.26	5.42	4.89	4.56	4.32	4.14	4.00	3.89		
16	8.53	6.23	5.39	4.77	4.44	4.20	4.03	3.89	3.78		
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68		
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60		
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52		
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46		
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40		
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35		
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30		
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26		
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22		
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18		
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15		
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12		
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09		
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07		
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89		
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72		
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56		
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41		