

Statistics 512 more review problems for the final exam

1. Short answer questions. Each part is unrelated.

- (a) If the design matrix has dimensions 20×4 , what is the dimension of the error vector?
- (b) Using the matrices from least squares estimation, what are the entries of the vector $\mathbf{X}'\mathbf{e}$?
- (c) The true means (μ_1, μ_2, μ_3) in a cell means model have values 40, 50, and 40, respectively. The error variance σ^2 is 20, and the design is balanced with 4 replicates per level. What is the distribution of the following contrast estimate?

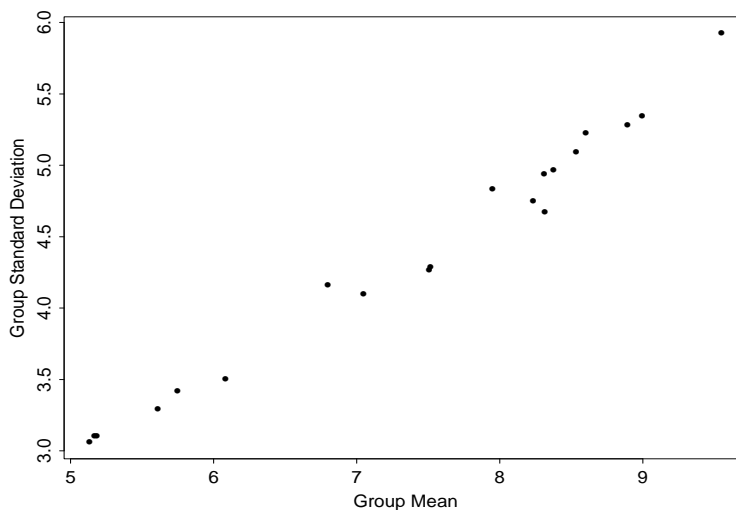
$$\hat{L} = \bar{Y}_1 + \bar{Y}_2 - 2\bar{Y}_3$$

- (d) In a 2×2 ANOVA, the population means are $\mu_{1,1} = 20$, $\mu_{2,1} = 25$, $\mu_{1,2} = 31$, and $\mu_{2,2} = 36$. Is there an interaction in this model? Explain why or why not.
- (e) After performing a least squares regression, you find that the squares of the residuals (r_i^2) are linearly related to the predictor X according to the following equation:

$$E(r_i^2) \approx 1 + X_i.$$

You decide to use weighted linear regression in relating the response Y to the predictor X . Give a formula for the weights w_i in terms of the variables in the model. (Assume all the X_i are positive.)

- (f) Given the following plot of μ_i versus σ_i , what transformation of the response would you recommend to make the variance constant?



- (g) A two-way **additive** ANOVA model has 3 and 4 levels for the two factors. If the data have 25 observations (i.e., $n_T = 25$), what is the error degrees of freedom?
- (h) Show that the criterion $C_p \leq p$ is equivalent to the criterion

$$\frac{MSE_p}{MSE(Full)} \leq 1.$$

- (i) The price per unit increases with lot size until the size is 200. The price per unit stays constant for lot sizes greater than 200. (In other words, the function relating price to lot size is a piecewise linear function. The slope of the line for lot size greater than 200 is zero.)

The output from `proc print` after reading in the data is

Obs	cost	lotsize	cslope
1	128.7	100	
2	107.7	125	
3	85.0	150	
4	70.3	175	
5	48.3	200	
6	46.0	225	
7	47.6	250	
8	47.9	275	
9	48.3	300	

Fill in an additional column of predictor values (under the `cslope` heading) that you would use to fit the previously described piecewise linear model. What `model` statement would you use in `proc reg` for this model?

2. Refer to the SAS output marked OUTPUT FOR PROBLEM 2.
 - (a) Write the factor effects model used for this analysis. Include numerical values for the number of levels being compared and the numbers of observations per treatment. Also state the distributional assumption.
 - (b) Summarize the results of the hypothesis tests for main effects and interactions.
 - (c) Explain the results of the Tukey procedures for the main effects.
 - (d) Estimate the residual variance σ^2 .

3. Refer to the SAS output marked OUTPUT FOR PROBLEM 3. Greenhouse benches were set up as blocks. Within each block, plants of different genetic varieties were grown. The maximum height of each plant was measured.
 - (a) Write the cell means model for this analysis. Include numerical values for the number of levels being compared and the numbers of observations per treatment. Also state the distributional assumption.
 - (b) Explain why no interaction term was included in the model. Describe graphical evidence that would justify this assumption (of no interaction)?
 - (c) Write the factor effects model used for this analysis. Then, give a numerical estimate for each parameter in the model.

OUTPUT FOR PROBLEM 2

The GLM Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	204035.1465	40807.0293	12.91	<.0001
Error	18	56914.1953	3161.8997		
Corrected Total	23	260949.3418			

R-Square	0.781896
Coeff Var	18.56378
Root MSE	56.23077
Y Mean	302.9058

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	2	25361.0880	12680.5440	4.01	0.0363
B	1	174964.1113	174964.1113	55.34	<.0001
A*B	2	3709.9471	1854.9736	0.59	0.5665

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	2	25361.0880	12680.5440	4.01	0.0363
B	1	174964.1113	174964.1113	55.34	<.0001
A*B	2	3709.9471	1854.9735	0.59	0.5665

Tukey's Studentized Range (HSD) Test for Y

Alpha	0.05
Error Degrees of Freedom	18
Error Mean Square	3161.9
Critical Value of Studentized Range	3.60930
Minimum Significant Difference	71.755

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	A
A	346.19	8	1
A			
B A	294.68	8	3
B			
B	267.85	8	2

OUTPUT FOR PROBLEM 3

The GLM Procedure

Dependent Variable: height

Source	DF	Squares	Sum of Mean Square	F Value	Pr > F
Model	8	196.5800000	24.5725000	57.95	<.0001
Error	15	6.3600000	0.4240000		
Corrected Total	23	202.9400000			

R-Square	Coeff Var	Root MSE	height Mean
0.968661	3.807911	0.651153	17.10000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
variety	3	172.9200000	57.6400000	135.94	<.0001
bench	5	23.6600000	4.7320000	11.16	0.0001

Level of variety	N	Mean	Std Dev
1	6	18.0000000	1.83084680
2	6	21.0000000	0.96540147
3	6	15.5000000	0.76681158
4	6	13.9000000	1.06395489

Level of bench	N	Mean	Std Dev
1	4	18.2000000	3.25269119
2	4	16.4000000	2.57811301
3	4	16.3000000	3.10590835
4	4	17.1500000	3.40734501
5	4	18.6000000	3.10805405
6	4	15.9500000	3.40832315