

FINAL EXAM STAT 572

Spring 1999

Instructions: You are to provide your own answers to every problem. Direct references to the notes, book, or other source will not be accepted.

1. A study was conducted to investigate the effect of three medications to treat neuralgia (headache). Data was obtained on 60 patients. The variables measured on these patients were:

- Treat: Medication (A, B, or P)
- Sex: F or M
- Age: In years
- Duration: The number of days with pain before medication was administered
- Pain: Whether the patient had pain 3 hours after the medication was administered (Yes or No)

The response was Pain and the results of a logistic regression analysis appear below.

Response: Pain

Iteration History

Converged by Gradient

Whole-Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	16.383737	5	32.76747	<0001
Full	24.367859			
Reduced	40.751596			
RSquare (U)		0.4020		
Observations (or Sum Wgts)		60		

Lack of Fit

Source	DF	-LogLikelihood	ChiSquare	Prob>ChiSq
Lack of Fit	54	24.367859	48.73572	
Pure Error	0	0.000000		Prob>ChiSq
Total Error	54	24.367859		0.6770

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	18.7872197	6.9653219	7.28	0.0070
Treat[A-P]	0.88494522	0.5291471	2.80	0.0944
Treat[B-P]	1.41179715	0.6079184	5.39	0.0202
Sex[F-M]	0.91610035	0.3981303	5.29	0.0214
Age	-0.2620931	0.0970204	7.30	0.0069
Duration	0.00585866	0.0329935	0.03	0.8591

Effect Test

Source	Nparm	DF	Wald ChiSquare	Prob>ChiSq
Treat	2	2	12.530957	0.0019
Sex	1	1	5.294631	0.0214
Age	1	1	7.297681	0.0069
Duration	1	1	0.031531	0.8591

a) Explain in detail the effect of the three medications. Which appear to be the fastest? Slowest?

b) Explain the effect (coefficients) of the covariates (Sex, Age and Duration) on the response.

c) What is the role of the covariates? If the interest is on the effect of the medication, why were these variables included in the model?

d) When the treatment-sex interaction is included in the model the -loglikelihood drops to 20.53. Test the significance of this term.

e) Interpret in **practical terms** what the treatment-sex interaction means in this context.

2. A corporation is assessing the compliance of federal regulations on their compensation practices. For that purpose, records for 106 employees selected at random were used to determine the factors that influence their annual salary. A forward selection was conducted to reduce the number of original regressors (13) and the following variables were included in the final model:

- Salary: (in dollars)
- Level: Job level
- TOP: Time on post
- PAR: Most recent Performance Appraisal Rating
- SLPAR: Second last PAR
- HS: High school diploma (0 = No, 1 = Yes)
- White: 1 if white, 0 if not
- Asian: 1 if asian, 0 if not
- Age: In years

The results appear below

Response: Salary

Summary of Fit	
RSquare	0.972014
RSquare Adj	0.969391
Root Mean Square Error	1942.802
Mean of Response	33953.77
Observations (or Sum Wgts)	106

Whole-Model Test				
Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	9	1.25854e10	1.3984e9	370.4820
Error	96	362350121	3774480	Prob>F
C Total	105	1.29477e10		<0001

Parameter Estimates					
Term	Estimate	Std Error	t Ratio	Prob> t	VIF
Intercept	21708.871	2353.206	9.23	<.0001	0
Level	-998.3489	169.9166	-5.88	<.0001	40.513768
Level*Level	59.321169	3.919555	15.13	<.0001	38.82018
TOP	168.614	97.25627	1.73	0.0862	1.3098044
PAR	229.51453	181.8947	1.26	0.2101	1.2811867
SLPAR	128.65582	93.98944	1.37	0.1742	1.3235513
HS	-979.232	646.7285	-1.51	0.1333	1.0924392
White	980.8439	476.5974	2.06	0.0423	1.5941804
Asian	1251.4682	555.1818	2.25	0.0265	1.5161164
Age	98.995895	25.26084	3.92	0.0002	1.40418

Effect Test					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
Level	1	1	130301742	34.5218	<.0001
Level*Level	1	1	864575742	229.0582	<.0001
TOP	1	1	11345124	3.0057	0.0862
PAR	1	1	6009490	1.5921	0.2101
SLPAR	1	1	7072256	1.8737	0.1742
HS	1	1	8653351	2.2926	0.1333
White	1	1	15986506	4.2354	0.0423
Asian	1	1	19179017	5.0812	0.0265
Age	1	1	57969041	15.3582	0.0002

a) Comment on the collinearity structure. Are you concerned?

b) The inclusion of the quadratic term for Level was suggested by analyzing the model with only linear terms. What specific diagnostic do you think was used and why?

c) The CEO is concerned about the negative sign of the variable Level. Is this a valid concern?
[Note: Level varies from 10 to 39]

d) The CEO is also concerned about the significance of the two race variables. Is this a valid concern?

e) The average length of the confidence intervals (see bottom of p. 66 in the notes) at the observed points was \$2,289.59. Interpret this value.

f) Do you think that this model will predict **future** average salaries as well as the figure in e) suggests? Justify your answer.

3. Determine whether the following statements are True or False. Justify clearly your answers.

T F When fitting a model to data, overfitting is always preferable to underfitting.

T F Adjusted effects are usually more important than direct effects.

T F The results of a robust regression analysis will always be different to those of an OLS analysis

T F Models that include polynomial terms are more prone to having collinearity.

T F The quality of prediction of a model is uniform across a region.

T F Using transformations to linearize a nonlinear model and using OLS to do the analysis is equivalent to fitting the nonlinear model directly.