Directions: Thoroughly, clearly and neatly answer the following two problems in the space given, showing all relevant calculations. Unless otherwise noted, use $\alpha = 5\%$ throughout.

1. (1 + 1 + 1.5 + 1 = 4.5 points) Suppose that of the members of a given population equally exposed to a virus infection, a percentage (which can be assumed to contain a fair cross-section of the population as a whole) has been inoculated. After the epidemic has passed, a random sample of people is drawn from the population and the numbers of inoculated and uninoculated that have escaped infection and otherwise are recorded giving the data shown in the following table. These data are analyzed on p.3 of the *Appendix* using Logistic regression.

	Infected	Not infected
Not inoculated	20	130
Inoculated	3	97

(a) This Logistic regression makes several assumptions. *Clearly list these in the context of this study*.

(b) Give and interpret the odds ratio (point estimate) in the context of this study.

(c) Give the 95% confidence interval for the above odds ratio.

(d) Does it appear that inoculation was effective here? Why or why not?

2. (0.5 + 1.5 + 1.5 + 2 = 3.5 or 5.5 points) Data for this exercise, analyzed in Appendix pp. 4-5, comes from a study of mental health for a random sample of 40 adult residents of Alachua County, Florida. It relates mental impairment to two explanatory variables. Mental impairment is an ordinal response, with categories (well, mild symptom formation, moderate symptom formation, impaired). One of the explanatory variables is the life events index (denoted "*levents*"), which is a composite measure of the number and severity of important life events such as birth of child, new job, divorce, or death in family that occurred to the subject within the past 3 years. The other explanatory variable that is used here is socioeconomic status ("*ses*"), where 1 represents "high" and 0 represents "low". Using $\alpha = 10\%$ here, we note that both explanatory variables are significant.

(a) Identify the assumed model and comment on whether you feel a key assumption appears to be met for these data.

- (b) Interpret the effect of the life events index variable in terms of the odds ratio. Please be clear and detailed
- (c) Interpret the effect of the socioeconomic status variable in terms of the odds ratio. Please be clear and detailed.

(d) [Mandatory for G students; EC for UG students] Give the predicted probabilities that a resident will have each of (well, mild symptom formation, moderate symptom formation, impaired) given that her socioeconomic status is "high" and her life events index is 5. Keep at least four decimal places in all calculations. Show all work.

First Exercise - Data, Program and Output

```
data two;
    do innoc='no ','yes'; shotno=(innoc='no ');
        input inyes inno @@; n=inyes+inno; output;
        end; datalines;
20 130 3 97
;
proc logistic;
    model inyes/n=shotno;
run;
```

The LOGISTIC Procedure							
Model Information							
Data Set Posnonso Varia	blo (Evonts)	i nyos					
Response Varia	able (Events)	rnyes					
Medel	able (Intals)	li hinamu lagit					
	Fa ahut au a	Figher's geer					
optim zation	rechnique	FISHER'S SCOR	ing				
Number of	Observations Rea	d 2					
Number of	Observations Use	d 2					
Sum of Fre	Sum of Frequencies Read						
Sum of Fre	Sum of Frequencies Used						
	Response Profi	le					
Ordere	d Binary	Total					
Value	e Outcome	Frequency					
1	Event	23					
2	2 Nonevent	227					
N.		Ch - t					
Convorgence	onvergence	Status	.d				
convergence of	GCONV=	TE- 6) Satisine	eu.				
	Model Fit Statis	stics					
		Intercept					
	Intercept	and					
Criterio	n Only	Covari ates					
AIC	155. 570	148. 751					
SC	159.092	155. 794					
-2 Log L	153. 570	144. 751					
Tosting C	obal Null Hypoth	osis: BETA-O					
Tost	Chi-Square	DF Pr	> Chi Sa				
Likelihood Ratio	8 8196	1	0.0030				
Score	7 6693	1	0.0056				
Wal d	6 4130	1	0.0113				
VICLE CL	0. 4150	I.	0. 0110				
Analysis of	f Maximum Likelih	ood Estimates					
	Standard	Wal d					
Parameter DF Estin	mate Error	Chi-Square	Pr > Chi Sq				
Intercept 1 -3.4	1761 0. 5862	35. 1623	<. 0001				
shotno 1 1.6	0. 6335	6. 4130	0.0113				

data one; do mi='well ','mild ','moderate','impaired'; ni=12; if mi='moderate' then ni=7; if mi='impaired' then ni=9; do j=1 to ni; input ses levents @@; drop ni j; output; end; end; datalines; 1 1 1 9 1 4 1 3 0 2 1 0 0 1 1 3 1 3 1 7 0 1 0 2 1 5 0 6 1 3 0 1 1 8 1 2 0 5 1 5 1 9 0 3 1 3 1 1 0 0 1 4 0 3 0 9 1 6 0 4 0 3 1 8 1 2 1 7 0 5 0 4 0 4 1 8 0 8 0 9 ; proc print; run;

0bs	mi	ses	levents
1	well	1	1
2	well	1	9
3	well	1	4
4	well	1	3
5	well	0	2
6	well	1	0
7	well	0	1
8	well	1	3
9	well	1	3
10	well	1	7
11	well	0	1
12	well	0	2
13	mild	1	5
14	mild	0	6
15	mild	1	3
16	mild	0	1
17	mild	1	8
18	mild	1	2
19	mild	0	5
20	mild	1	5
21	mild	1	9
22	mild	0	3
23	mild	1	3
24	mild	1	1
25	moderate	0	0
26	moderate	1	4
27	moderate	0	3
28	moderate	0	9
29	moderate	1	6
30	moderate	0	4
31	moderate	0	3
32	i mpai red	1	8
33	i mpai red	1	2
34	i mpai red	1	7
35	i mpai red	0	5
36	i mpai red	0	4
37	i mpai red	0	4
38	i mpai red	1	8
39	i mpai red	0	8
40	i mpai red	0	9

The LOGISTIC Procedure			
Model Informat	ci on		
Data Set	WORK. ONE		
Response Variable	mi		
Number of Response Levels	4		
Model	cumulative logit		

	Optin	mization T	echni que	Fi sher	's scoring		
Number of Observations Read				40			
Number of Observations Used 40							
			Response	Profile			
		Ordered		Т	otal		
		Val ue	mi	Frequ	ency		
		1	well		12		
		2	mild		12		
		3	modera	te	7		
		4	impair	ed	9		
Pr	obabilities m	odeled ar	e cumulate	d over the l	ower Ordered	Val ues.	
		Mode	el Converg	ence Status			
	Conve	rgence cr	iterion (G	CONV=1E-8) s	ati sfi ed.		
		U					
	Score	e Test for	the Propo	ortional Odds	Assumpti on		
		Chi - Squa	re D	$\mathbf{F} = \mathbf{Pr} > \mathbf{C}$	hi Sq		
		2.32	55	4 0.	6761		
		М	bdel Fit S	stati sti cs			
				Int	ercept		
			Interc	ept	and		
	C	ri teri on	0	nly Cova	riates		
	А	10	115.	042 1	09. 098		
	S	C	120.	109 1	17. 542		
	-	2 Log L	109.	042	99. 098		
		-0					
	Τe	sting Glo	bal Null H	vpothesis: B	ETA=0		
	Test	0	Chi - Squa	re DF	Pr > ChiS	a	
	Li kel i hood	Ratio	9, 94	42 2	0.006	i9	
	Score		9.14	31 2	0. 010	3	
	Wald		8. 50	18 2	0. 014	3	
	Ar	alvsis of	Maxi mum L	i kel i hood Es	timates		
		5		Standard	Wald		
Paramet	er	DF E	stimate	Error	Chi-Square	Pr > Chi Sa	
Interce	pt well	1	-0.2818	0. 6231	0. 2045	0.6511	
Interce	pt mild	1	1.2129	0.6511	3, 4700	0, 0625	
Interce	pt moderate	1	2, 2095	0. 7171	9, 4932	0, 0021	
Ses	r =	1	1. 1111	0, 6143	3. 2719	0. 0705	
levents	4	1	-0.3189	0. 1194	7, 1294	0, 0076	
i e i entes		-					
		0d	ds Ratio E	stimates			
Point 95% Wald							
	Effe	ct Es	timate	Confi dence	Limits		
	Ses		3. 038	0. 911	10, 126		
	leve	nts	0. 727	0.575	0.919		
Association of Predicted Probabilities and Observed Responses							
	Percent	Concorda	nt 68.	7 Somers'	D 0.409		
	Percent	Discorda	nt 27.	7 Gamma	0. 425		
	Percent	Tied	3.	6 Tau-a	0. 310		
	Pairs		59	1 c	0. 705		
			50	-	505		