Advanced Data Analysis 2, Stat 428/528 Homework 11 Name: Answer Key
Spring 2012 Prof. Erik B. Erhardt

Part I. (80 points) Do all calculations in SAS. Use a word processor of your choice to write a report. Insert computer text output and graphics to support what you are saying, but you need to write something that looks like an academic paper — not a pile of computer output. Turn in a hard copy of your HW in class (i.e., don't email me your HW). Also:

1. Clearly specify parameters and hypotheses when appropriate.
2. Write a coherent conclusion.

(80 pts)

80 pts

1. Alloy fastener failures

The following data are from a study on the compressive strength of an alloy fastener used in the construction of aircraft. Ten pressure loads, increasing in units of 200 psi from 2500 psi to 4300 psi, were used with different numbers of fasteners being tested at each of the loads. The table below gives the number of fasteners failing out of the number tested at each load.

Load	Tested	Failed
2500	50	10
2700	70	17
2900	100	30
3100	60	21
3300	40	18
3500	85	43
3700	90	54
3900	50	33
4100	80	60
4300	65	51

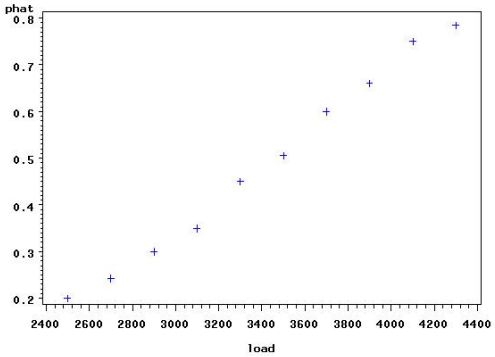
(a) (10 pts) Compute the observed proportion of fasteners failing at each load. Plot the observed proportions against load and comment on how the proportion of failures depends on load.

Solution: The plot indicates an increasing relationship between load and proportion of failures. There is a slight sigmoidal shape to the relationship. The next part will indicate the adequacy of the logit transformation.

Program editor contents:

```
options ls=79 nodate nocenter;

* part (a);
data fasteners1;
  infile "F:\Dropbox\UNM\teach\ADA2_stat528\sas\ADA2_HW_11_fasteners.dat";
  input load ntest nfail;
  phat = nfail/ntest;
  * emperical logits ;
  emplog = log( (phat + 0.5/ntest) / (1 - phat + 0.5/ntest) );
run;

data fasteners2;  * for part (g) ;
  input load ntest nfail;
  datalines;
3400    .     .
;
run;

* put data together, full dataset plus 3400 line with missing values;
data fasteners;
  set fasteners1 fasteners2;
run;

symbol1 v=plus c=blue  i=none;
proc gplot data=fasteners;
  plot phat*load;
run;
```

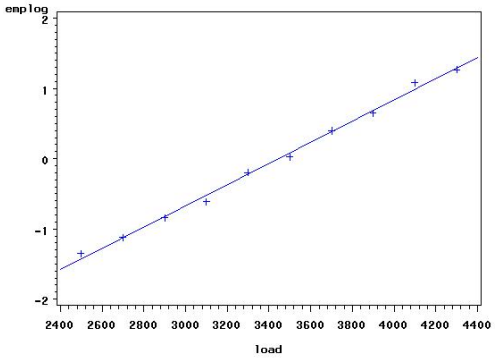
10 pts



(b) (10 pts) Present a graphical summary that provides information on the adequacy of a logistic regression model relating the probability of fastener failure as a function of load. Discuss.

Solution: A plot of the empirical logits against load is roughly linear, which supports a logistic transformation for the response.
Program editor contents:

```
* part (b);
symbol1 v=plus c=blue  i=r;
proc gplot data=fasteners;
  plot emplog*load;
run;
```



(c) (10 pts) *Use the LOGISTIC procedure for the remaining parts.*
Fit a logistic model relating the probability of fastener failure to load. Look at the likelihood ratio goodness-of-fit statistic. Is there evidence of any gross deficiencies with the model?

Solution: Program editor contents:

```
* part (c);
proc logistic data=fasteners;
  model nfail/ntest = load  /scale=none aggregate;
  output out=outlog p=fitprob l=l95 u=u95 /alpha=0.05;
run;
```

Output window contents: (First general model fit details)

```
The LOGISTIC Procedure

              Model Information
Data Set                          WORK.FASTENERS
Response Variable (Events)        nfail
Response Variable (Trials)        ntest
Model                             binary logit
Optimization Technique            Fisher's scoring

Number of Observations Read          11
Number of Observations Used          10
Sum of Frequencies Read             690
Sum of Frequencies Used             690

          Response Profile
 Ordered     Binary          Total
   Value     Outcome     Frequency
       1     Event             337
       2     Nonevent          353

NOTE: 1 observation was deleted due to missing values for the response or
      explanatory variables.
```

The goodness-of-fit statistic (deviance) gives a p-value=1, thus there is insufficient evidence to conclude the model does not fit the data.

```
                  Model Convergence Status
         Convergence criterion (GCONV=1E-8) satisfied.

       Deviance and Pearson Goodness-of-Fit Statistics
Criterion          Value       DF     Value/DF     Pr > ChiSq
Deviance          0.3719        8       0.0465         1.0000
Pearson           0.3707        8       0.0463         1.0000

Number of unique profiles: 10
```

(d) (10 pts) Does load appear to be a useful predictor of the probability of failure? Do a formal hypothesis test.

Solution: Because there is only one predictor in our model, the Wald and ML parameter estimate give the same test (same Wald Chi-Square statistic and p-value). Both indicate a highly-significant parameter estimate.
Output window contents:

```
        Testing Global Null Hypothesis: BETA=0
Test                 Chi-Square       DF     Pr > ChiSq
Likelihood Ratio       112.4602        1         <.0001
Score                  107.0664        1         <.0001
Wald                    96.6082        1         <.0001  *

             Analysis of Maximum Likelihood Estimates
                              Standard          Wald
Parameter    DF    Estimate       Error    Chi-Square    Pr > ChiSq
Intercept     1      0.5457     95.7497        <.0001
load          1    0.000158     96.6082        <.0001  *
```

(e) (10 pts) Provide an equation relating the fitted probability of fastener failure to the load.

Solution: The MLE of the predicted probabilities satisfy:

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = -5.3397 + 0.00155\text{load}$$

or

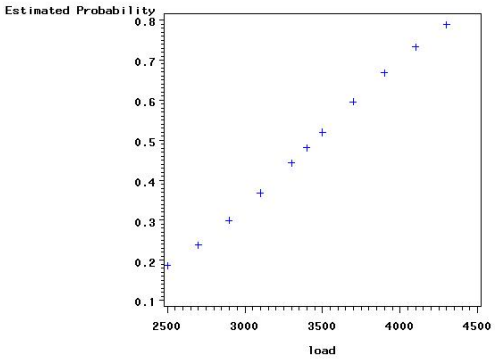
$$\hat{p} = \frac{\exp(-5.3397 + 0.00155\text{load})}{1 + \exp(-5.3397 + 0.00155\text{load})}.$$

(f) (10 pts) Plot the fitted probabilities as a function of load.

Solution: Program editor contents:

```
* part (f);
symbol1 v=plus c=blue  i=none;
proc gplot data=outlog;
  plot fitprob*load;
run;
```

30 pts



(g) (10 pts) Compute the estimated probability of failure when the load is 3400 psi. Provide a 95% CI for this probability.
For the last two parts you can augment the data set with a case

```
3400     .      .
```

having missing numbers tested and failed. Then save and print out the relevant summaries to answer the questions.

Solution: Program editor contents:

```
* part (g);
proc print data=outlog;
run;
```

The predicted probability of failure is 0.48 with a 95% CI of $(0.44, 0.52)$.
Output window contents:

Obs	load	ntest	nfail	phat	emplog	fitprob	l95	u95
1	2500	50	10	0.20000	-1.34993	0.18715	0.14224	0.24225
2	2700	70	17	0.24286	-1.11748	0.23886	0.19229	0.29262
3	2900	100	30	0.30000	-0.83789	0.29959	0.25380	0.34977
4	3100	60	21	0.35000	-0.60825	0.36829	0.32513	0.41366
5	3300	40	18	0.45000	-0.19574	0.44278	0.40210	0.48425
6	3500	85	43	0.50588	0.02326	0.51994	0.47893	0.56068
7	3700	90	54	0.60000	0.40089	0.59616	0.55132	0.63945
8	3900	50	33	0.66000	0.64934	0.66801	0.61776	0.71469
9	4100	80	60	0.75000	1.08222	0.73280	0.67794	0.78133
10	4300	65	51	0.78462	1.26743	0.78894	0.73164	0.83673
11	3400	.	.	.	.	0.48125	0.44087	0.52187

(h) (10 pts) Write a short yet complete summary.

Solution: summary

20 pts