Logistic Regression Analyses in the Water Level Study

A. Introduction.

166 students participated in the 'Water level Study'. 70 passed and 96 failed to correctly draw the water level in the glass. There were two main research questions:

1. Why was the passing rate so low? What factors affect passing?
2. There was a major difference in the proportion of females and males who passed? Can some of the variables in the study explain this?

B. Frequency Tables:

1. **Table of y by sex**

Frequency, Percent , Row Pct , Col Pct	female	male	Total	
Fail	75 45.18 78.13 70.09	21 12.65 21.88 35.59	96 57.83	29.91% of females passed (32/107)
Pass	32 19.28 45.71 29.91	38 22.89 54.29 64.41	70 42.17	64.41% of males passed (38/59)
Total	107 64.46	59 35.54	166 100.00	

Statistics for Table of y by sex

Statistic	DF	Value	Prob
Chi-Square	1	18.5617	<.0001
Likelihood Ratio Chi-Square	1	18.6578	<.0001

2. **Table of y by gravity**

Frequency, Percent , Row Pct , Col Pct	Gravity Score 0	1	2	3	4	5	Total
Fail	10 6.02 10.42 100.00	23 13.86 23.96 92.00	18 10.84 18.75 64.29	18 10.84 18.75 58.06	23 13.86 23.96 53.49	4 2.41 4.17 13.79	96 57.83
Pass	0 0.00 0.00 0.00	2 1.20 2.86 8.00	10 6.02 14.29 35.71	13 7.83 18.57 41.94	20 12.05 28.57 46.51	25 15.06 35.71 86.21	70 42.17
Total	10 6.02	25 15.06	28 16.87	31 18.67	43 25.90	29 17.47	166 100.00

Statistics for Table of y by gravity

Statistic	DF	Value	Prob
Chi-Square	5	43.1342	<.0001
Likelihood Ratio Chi-Square	5	50.7636	<.0001

C. Logistic Regression of Pass/Fail in Water Level Study on Sex

Model: $\ln\{\pi(\text{sex})/[1-\pi(\text{sex})] = \beta_0 + \beta_1 * (\text{sex}) =$

$\beta_0 + \beta_1$, **for females**

β_0 , for males

```
options ls=72;                    Note 1: Females are coded '1'
data sex;                                 Males are coded '0'
input sex  r n @@;                Note 2: Frequency counts are used
cards;
1 32 107 0 38 59

;
Proc Logistic ; Model r/n=sex ;
output out=pred p=phat lower=lcl upper=ucl;
proc print;
run;
```

Output: The LOGISTIC Procedure

Model Information

Data Set	WORK.SEX
Response Variable (Events)	r
Response Variable (Trials)	n
Number of Observations	2
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	Binary Outcome	Total Frequency
1	Event	70
2	Nonevent	96

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.036	211.378
SC	231.148	217.602
-2 Log L	226.036	207.378

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	18.6578	1	<.0001
Score	18.5617	1	<.0001
Wald	17.6086	1	<.0001

Test H_0: No sex effect or H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 \neq 0$. $G^2 = 18.6578 = LRT$

Reject H_0: No sex effect and conclude there is a statistically significant difference between females and males in proportion passing the task.

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.5931	0.2719	4.7572	0.0292
sex	1	-1.4446	0.3443	17.6086	<.0001

Fitted Model: fitted logit(females) = 0.5931 – 1.4446 = -0.8515 for females
fitted logit(males) = 0.5931 for males

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
sex	0.236	0.120	0.463

Odds ratio (females vs. males) = $s^{-1.4446} = 0.236$

Obs	sex	r	n	phat	lcl	ucl
1	1	32	107	0.29911	0.22005	0.39229
2	0	38	59	0.64407	0.51503	0.75510

Odds ratio (males vs females):

	Pass	Fail
Males	**38**	**21**
Females	**32**	**75**

Odds Ratio = (38)(75)/(21)(32 = 4.24 = $s^{1.4446}$

D. Logistic Regression of Pass/Fail in Water Level Study on x = 'Gravity'

Model: ln π(x)/[1-π(x)].

SAS Program:

```
options ls=72;
data gravity;
input gravity r n @@;
cards;
0 0 10 1 2 25 2 10 28 3 13 31 4 20 43 5 25 29
;
Proc Logistic ; Model r/n=gravity ;
output out=pred p=phat lower=lcl upper=ucl;
proc print;
run;
```

The LOGISTIC Procedure

Model Information

Data Set	WORK.GRAVITY
Response Variable (Events)	r
Response Variable (Trials)	n
Number of Observations	6
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	Binary Outcome	Total Frequency
1	Event	70
2	Nonevent	96

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.036	187.859
SC	231.148	194.083
-2 Log L	226.036	183.859

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	42.1765	1	<.0001
Score	37.8303	1	<.0001
Wald	31.1219	1	<.0001

Test H_0: No gravity effect or H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 \neq 0$. $G^2 = 42.1765 = LRT$

Reject H_0: No gravity effect and conclude there is a statistically significant difference between gravity score and proportion passing the task.

The LOGISTIC Procedure **Page 5**

Analysis of Maximum Likelihood Estimates

Parameter	DF	Standard Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.8155	0.5048	31.1055	<.0001
gravity	1	0.7998	0.1434	31.1219	<.0001

Fitted Model: Estimated logit[π(x)] = -2.8156+0.7998x

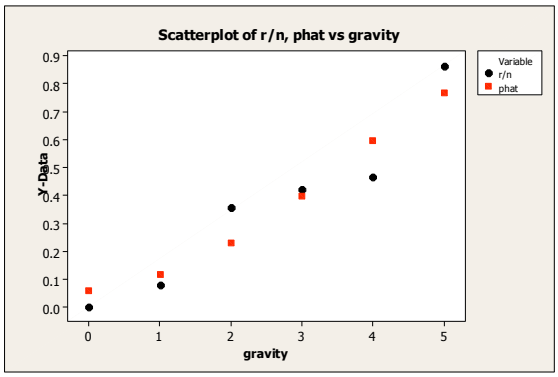
Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits
gravity	2.225	1.680 2.947

Odds of passing the water level task increase by 2.225 for each additional right answer on gravity items.

gravity score	r (pass)	n (fail)	r/n	phat	lcl	ucl
0	0	10	.0000	0.05649	0.02178	0.13871
1	2	25	.0800	0.11756	0.06012	0.21719
2	10	28	.3571	0.22864	0.15170	0.32945
3	13	31	.4194	0 39742	0.31490	0.48622
4	20	43	.4651		0.59473	0.49485 0.68733
5	25	29	.8621		0.76554	0.64295 0.85550

A graph of observed and fitted proportions is Given above, right. How does the 'fit' look?



E. Logistic regression of Pass/Fail on sex and gravity:

```
options ls=72;
data water;
input obs y sex gravity ;
cards;
  1     0      1          4
  2     1      2          5
...     .      .          .
...     .      .          .
166     0      1          4

;
Proc Logistic descending; Model Y=sex gravity;
output out=pred p=phat lower=lcl upper=ucl;
proc print;
```

Note: Females are coded '1'
Males are coded '2'

```
Proc Logistic; Model Y=sex|gravity;     Note: 'descending' not specified
run;
```

Page 6

E1. Logistic Regression of Pass/Fail on Sex and Gravity

Model: $\text{logit}[\pi(\text{sex, gravity})] = \beta_0 + \beta_1 * (\text{sex}) + \beta_2 * \text{gravity}$

$$= \begin{cases} (\beta_0 + \beta_1) + \beta_2 * \text{gravity, for females} \\ (\beta_0 + 2\beta_1) + \beta_2 * \text{gravity, for males} \end{cases}$$

The LOGISTIC Procedure

Model Information

Data Set	WORK.WATER
Response Variable	y
Number of Response Levels	2
Number of Observations	166
Model	binary logit
Optimization Technique	Fisher's scoring

Probability modeled is y=1.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.036	181.059
SC	231.148	190.395
-2 Log L	226.036	175.059

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	50.9766	2	<.0001
Score	45.0940	2	<.0001
Wald	35.2414	2	<.0001

Test H_0: Sex and gravity together do not affect passing the water level task or H_0: $\beta_1 = \beta_2 = 0$ vs. H_a: at least one of the parameters is not 0. $G^2 = 50.9766 = LRT$

Conclude the logistic regression of pass/fail on sex and gravity is statistically significant.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.1676	0.7228	33.2425	<.0001
sex	1	1.1220	0.3824	8.6117	0.0033
gravity	1	0.7404	0.1466	25.4979	<.0001

Estimated logit(sex,gravity) = -4.1676 + 1.1220sex + 0.7404gravity. Note that sex is coded as 1 for females and 2 for males.

Page 7

Test the hypothesis that there is no gravity effect, adjusted for 'sex'"

Calculate the change in G^2 for the models with both variables included and with only sex.

G^2 (sex, gravity) - G^2 (sex) = 50.9766 – 42.1765 = 8.801, or calculate the change in -2log likelihood: -2ln (sex) - [-2ln(sex, gravity) = 183.859 – 175.059 = 8.800. compare this value with the Wald chi-square 8.6117.

Test the hypothesis that there is no sex effect, adjusted for gravity score:

Calculate the change in G^2 for the models with both variables included and with only gravity.

G^2 (sex, gravity) - G^2 (sex) = 50.9766 – 18.6568 = 32.319, or calculate the change in -2log likelihood: -2ln (gravity) - [-2ln(sex, gravity) = 207.478 – 175.059 .
Compare this value with the Wald chi-square 25.4979.

Odds Ratio Estimates

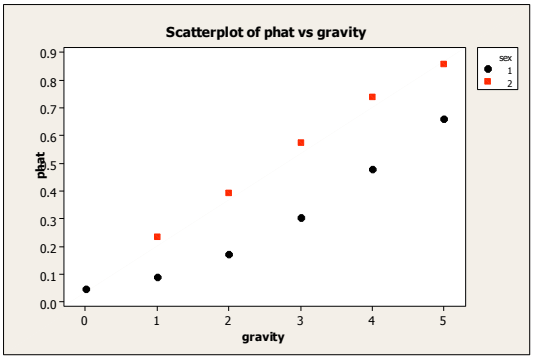
Effect	Point Estimate	95% Wald Confidence Limits	
sex	3.071	1.452	6.498
gravity	2.097	1.573	2.795

Predicted Values and Confidence Limits for Population Proportions:

Obs	obs	y	sex	gravity	_LEVEL_	phat	lcl	ucl
1	1	0	1	4	1	0.47898	0.35601	0.60455
2	2	1	2	5	1	0.85548	0.73431	0.92689
...	...	.	.	.	.	.....	.....	.....
166	166	0	1	4	1	0.47898	0.35601	0.60455

Edited Fitted Values are given below; a plot of phat vs. gravity for females and for males is given in the graph.

Row	sex	gravity	phat	lcl	ucl
1	1	0	0.04541	0.01658	0.11831
2	1	1	0.09069	0.04332	0.18012
3	1	2	0.17295	0.10478	0.27199
4	1	3	0.30481	0.21613	0.41080
5	1	4	0.47898	0.35601	0.60455
6	1	5	0.65841	0.49314	0.79246
7	2	1	0.23448	0.11133	0.42822
8	2	2	0.39107	0.24091	0.56514
9	2	3	0.57384	0.42507	0.71034
10	2	4	0.73844	0.60244	0.84026
11	2	5	0.85548	0.73431	0.92689



E2. Logistic Regression of Pass/Fail on Sex, Gravity and Sex*Gravity (Interaction Model)

Model: logit[π(sex, gravity)] = $\beta_0 + \beta_1$*(sex) + β_2*gravity+β_3*(sex*gravity)

$$(\beta_0 + \beta_1) + (\beta_2 + \beta_3)\text{gravity, for females}$$

=

$(\beta_0 + 2\beta_1) + (\beta_2 + 2\beta_3)$gravity, for males

The LOGISTIC Procedure

Model Information

Data Set	WORK.PRED
Response Variable	y
Number of Response Levels	2
Number of Observations	166
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	y	Total Frequency
1	0	96
2	1	70

Probability modeled is y=0.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	228.036	182.944
SC	231.148	195.392
-2 Log L	226.036	174.944

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	51.0922	3	<.0001
Score	45.1521	3	<.0001
Wald	34.9621	3	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	4.6340	1.5633	8.7873	0.0030
sex	1	-1.4606	1.0646	1.8822	0.1701
gravity	1	-0.8823	0.4452	3.9281	0.0475
sex*gravity	1	0.1026	0.3009	0.1162	0.7332