Chapter 11: Quantal Response Data Analysis

11.1 Introduction

- Quantal response data analysis deals with subject response to a stimulus that occurs with greater and greater intensity
- Bioassay: process of determining potency or strength of reagent or stimuli based on elicited response in biological organisms
- Tolerance: amount of stimulus required to produce response
- Researchers interested in:
    - tolerance of subjects to stimulus or drug
    - relative potency of new drug to standard drug
- This chapter concerned with quantal responses (death or survival) analyzed with categorical analysis strategies

11.2 Estimating Tolerance Distributions

	Status	
Bacterial Dose	Dead	Alive
1.2×10^3	0	5
1.2×10^4	0	5
1.2×10^5	2	3
1.2×10^6	4	2
1.2×10^7	5	1
1.2×10^8	5	0

- Assume responses of subjects determined through tolerance distribution
- Assume either logistic or normal distribution for tolerances
  - For the normal distribution, can use logarithms of tolerances
- If tolerances are assumed to follow a normal distribution, we can write the probability of death at level x_i of the drug as

$$p_i = \Phi\left(\frac{x_i - \mu}{\sigma}\right)$$

and

$$\Phi^{-1}(p_i) = -\frac{\mu}{\sigma} + \frac{1}{\sigma}x_i = \alpha + \beta x_i$$

- Φ is the cumulative distribution function for the standard normal distribution and $\Phi^{-1}(p_i)$ is called the probit

- Under the logistic assumption, probability of death at level x_i of drug

$$\Pr\{Y_i \le x_i\} = p_i = \frac{\exp\{\alpha + \beta x_i\}}{1 + \exp\{\alpha + \beta x_i\}},$$

where Y_i represents the tolerance for subject i

$$\log\left\{\frac{p_i}{1 - p_i}\right\} = \alpha + \beta x_i$$

- Often interested in dose at which 50% of subjects produce a response (LD50 or ED50).
- Let x_{50} = log(LD50) and p_{50} = probability of response at median of tolerance distribution, then:

$$\log\left\{\frac{p_{50}}{1-p_{50}}\right\} = \log\left\{\frac{0.5}{0.5}\right\} = 0$$

- Then $\hat{\alpha}+\hat{\beta}x_{50}$ can be set to zero to obtain

$$\hat{x}_{50} = \frac{-\hat{\alpha}}{\hat{\beta}},$$

$$\text{and } \operatorname{var}\{\hat{x}_{50}\} = \{\hat{x}_{50}\}^2 \left\{\frac{V(\hat{\alpha})}{\hat{\alpha}^2} - \frac{2V(\hat{\alpha},\hat{\beta})}{\hat{\alpha}\hat{\beta}} + \frac{V(\hat{\beta})}{\hat{\beta}^2}\right\},$$

where $V(\hat{\alpha}) =$ variance of $\hat{\alpha}$

$V(\hat{\beta})$= variance of $\hat{\beta}$

$V(\hat{\alpha},\hat{\beta})$= covariance of $\hat{\alpha}$ and $\hat{\beta}$

- Then a confidence interval for log(LD50) is:

$$\hat{x}_{50} \pm z_{1-\alpha/2}\sqrt{\text{var}\{\hat{x}_{50}\}}$$

Linear Taylor Series Expansion

Let F and G be consistent and asymptotically normal estimators of respective parameters ξ and η. Let ν_F, ν_G, and ν_{FG} be consistent estimators for the variances of F and G and their covariance. Also, note that ν_F, ν_G, and ν_{FG} are of order $(1/n)$ such that parts of them which do not correspond to the variances and covariance of F and G in large samples tend to 0 faster than $(1/n)$. Then statistical behavior of $R=(F/G)$ as an estimator of (ξ/η) corresponds in large samples to its linear Taylor series counterpart.

$$R_{TL} = (\xi/\eta) + (1/\eta)(F-\xi) - (\xi/\eta^2)(G-\eta)$$
$$= (\xi/\eta)\left\{1 + \frac{(F-\xi)}{\xi} - \frac{(G-\eta)}{\eta}\right\}$$

Note that $R = R_{TL} + \mathbf{O}(1/n)$.

$$\varepsilon\{R_{TL}\} = (\xi/\eta), \varepsilon(R) = (\xi/\eta) + \mathbf{O}(1/n) \to (\xi/\eta)$$

$$Var\{R_{TL}\} = (\xi/\eta)^2 \left\{\frac{Var(F)}{\xi^2} - \frac{2Cov(F,G)}{\xi\eta} + \frac{Var(G)}{\eta^2}\right\}$$

$$Var\{R\} = Var\{R_{TL}\} + \mathbf{O}\left(\frac{1}{n^{1+\delta}}\right) \text{ with } \delta \geq \tfrac{1}{2}$$

A consistent estimator for Var$\{R\}$ is given by

$$\upsilon_R = R^2 \left\{\left(\frac{\upsilon_F}{F^2}\right) - \left(\frac{2\upsilon_{FG}}{FG}\right) + \left(\frac{\upsilon_G}{G^2}\right)\right\}$$

11.2.1 Analyzing the Bacterial Challenge Data

```
data bacteria;
   input dose status $ count @@;
   ldose = log(dose);
   datalines;
1200         dead     0    1200         alive        5
12000        dead     0    12000        alive        5
120000       dead     2    120000       alive        3
1200000      dead     4    1200000      alive        2
12000000     dead     5    12000000     alive        1
120000000    dead     5    120000000    alive        0
;
run;

proc logistic data=bacteria descending;
   freq count;
   model status = ldose ldose*ldose / scale=none aggregate
         selection=forward include=1 details covb;
run;
```

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSQ
Intercept	1	-9.2680	3.1630	8.5857	0.0034
Ldose	1	0.7071	0.2354	9.0223	0.0027

Estimated Covariance Matrix

Parameter	Intercept	ldose
Intercept	10.00458	-0.73338
ldose	-0.73338	0.055418

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
0.2580	1	0.6115

To compute the log LD50, use the estimated values of $\hat{\alpha}$ and $\hat{\beta}$.

$$\log \text{LD50} = \frac{-\hat{\alpha}}{\hat{\beta}} = \frac{9.2680}{0.7071} = 13.1070$$

Using the covariances from above in the formula for var$\{\hat{x}_{50}\}$ yields the value 0.6005. Thus, a confidence interval for the log LD50 is written

$$13.1070 \pm 1.96\sqrt{0.6005}$$

Thus, the confidence interval for log LD50 is (11.588, 14.626).

To determine the LD50 on the actual dose scale, exponentiate the LD50 for the log scale. Exponentiate bounds of the confidence interval to get corresponding CI.

Actual LD50 = $e^{13.1070} = 4.9238 \times 10^5$

Confidence Interval: $(1.0780 \times 10^5, 2.2490 \times 10^6)$

PROC PROBIT vs. PROC LOGISTIC

- PROC LOGISTIC assumes an underlying logistic tolerance distribution, whereas PROC PROBIT can handle either
    - Default is normal distribution; specify logistic with DIST=LOGISTIC option

- PROC PROBIT computes LD50 automatically, as well as computing the estimates for the dose values that yield user-defined response rates and corresponding confidence intervals based on Fieller's Theorem

- PROC LOGISTIC provides residual score tests with the SELECTION=FORWARD option to assist in determining the fit of the model

Data Structures

Bacterial Dose	Status	
	Dead	Alive
1.2×10^3	0	5
1.2×10^4	0	5
1.2×10^5	2	3
1.2×10^6	4	2
1.2×10^7	5	1
1.2×10^8	5	0

For PROC LOGISTIC, you should specify the number dead, the number alive, and differentiate them with a status variable. For PROC PROBIT, you only need to provide the number dead and the total count of alive and dead.

LOGISTIC:

```
data bacteria;
   input dose status $ count @@;
   ldose = log(dose);
   datalines;
1200        dead     0    1200        alive        5
12000       dead     0    12000       alive        5
120000      dead     2    120000      alive        3
1200000     dead     4    1200000     alive        2
12000000    dead     5    12000000    alive        1
120000000    dead    5    120000000    alive       0
;
```

PROBIT:

```
data bacteria2;
   input dose dead total @@;
   ldose = log(dose);
   datalines;
1200            0       5
12000           0       5
120000          2       5
1200000         4       6
12000000        5       6
120000000       5       5
;
run;
```

Probit Analysis

```
data bacteria2;
   input dose dead total @@;
   ldose = log(dose);
   datalines;
1200          0      5
12000         0      5
120000        2      5
1200000       4      6
12000000      5      6
120000000     5      5
;
run;

ods graphics on;
proc probit data=bacteria2 log plot=ippplot;
   model dead/total = dose / dist=logistic lackfit
            inversecl (prob=.25 .50 .75);
run;
ods graphics off;
```

Goodness-of-Fit Tests

Statistic		Value	DF	Value/DF	Pr> ChiSq
Pearson	Chi-Square	1.3379	4	0.3345	0.8549
L.R.	Chi-Square	1.7508	4	0.4377	0.7815

Analysis of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-9.2680	3.1630	-15.4674	-3.0687	8.59	0.0034
Ln(dose)	1	0.7071	0.2354	0.2457	1.1685	9.02	0.0027

Probit Model in Terms of Tolerance Distribution

MU	SIGMA
13.1070197	1.41421791

Log Dose Analysis

Probit Analysis on Ln(dose)

Probability	Ln(dose)	95% Fiducial Limits	
0.25	11.5533	7.5414	13.0735
0.50	13.1070	11.0067	15.0196
0.75	14.6607	13.1430	18.2947

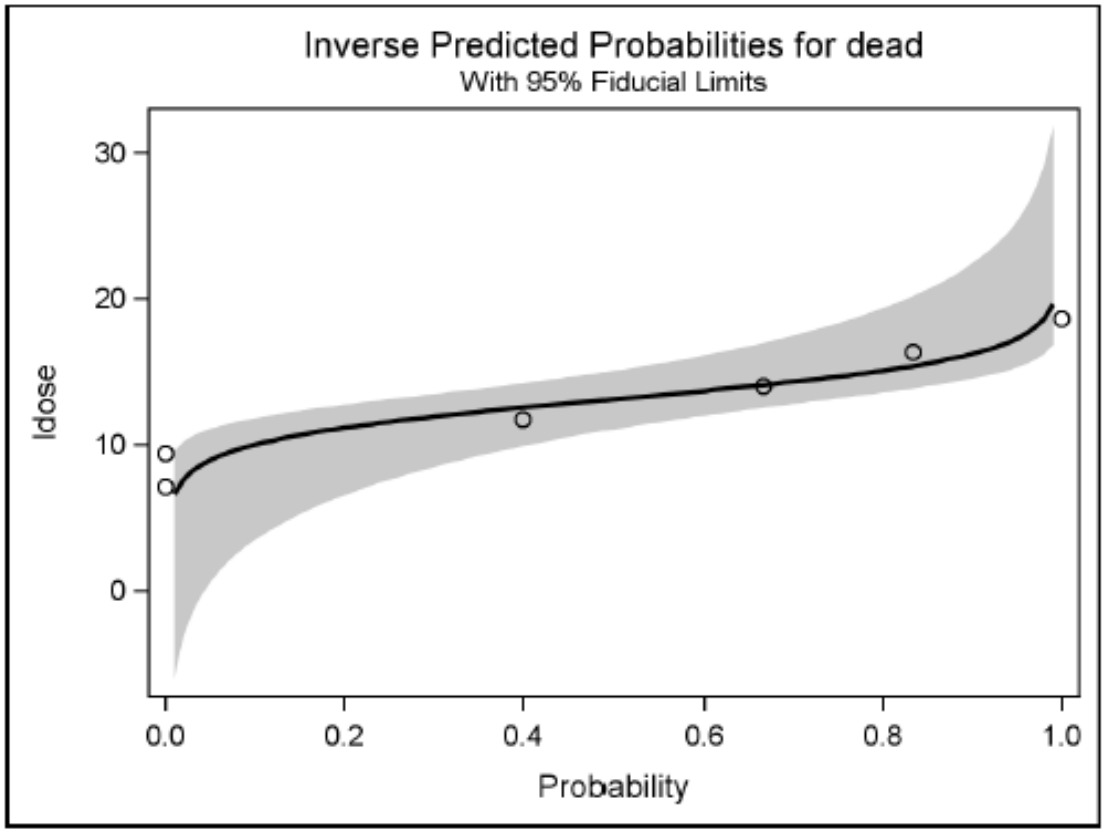
Actual Dose Analysis

Probit Analysis on dose

Probability	dose	95% Fiducial Limits	
0.25	104124	1885	476140
0.50	492387	60276	3333723
0.75	2328412	510434	88159975

Inverse Probability Plot

```
ods graphics on;
proc probit data=bacteria2 plot=ippplot;
   model dead/total = ldose / dist=logistic;
run;
ods graphics off;
```



11.3 Comparing Two Drugs

Dose	Drug	Dead	Alive	Total
0.01	N	0	30	30
0.03	N	1	29	30
0.10	N	1	9	10
0.30	N	1	9	10
0.30	S	0	10	10
1.00	N	4	6	10
1.00	S	0	10	10
3.00	N	4	6	10
3.00	S	1	9	10
10.00	N	5	5	10
10.00	S	4	6	10
30.00	S	5	5	10
30.00	N	7	3	10
100.00	S	8	2	10

- Dilution assumption for doses z_s of somatostatin (S) and z_n of neurotensin: $z_s = \rho z_n$
  - ρ = relative potency
- If x_n and x_s represent log (doses), then

$$x_s = log\,\rho + x_n$$

- Assume logistic model structure for somatostatin is

$$p_s(x_{si}) = \{1+ \exp(-\alpha_s - \beta x_{si})\}^{-1},$$

- Then implied structure for log(dose) levels x_{ni} of neurotensin is $p_n(x_{ni}) = p_s(\log\rho + x_{ni})$

$$= \{1+\exp(-\alpha_s - \beta\log\rho - \beta x_{ni})\}^{-1}$$

$$= \{1+\exp(-\alpha_n - \beta x_{ni})\}^{-1},$$

where $\alpha_n = \alpha_s + \beta\log\rho$

- By forming logit ($p_n(x_{ni})$), the following equation results

$$\log\left\{\frac{p_n(x_{ni})}{1-p_n(x_{ni})}\right\} = \{\alpha_s + \beta \log \rho\} + \beta x_{ni}$$

$$= \alpha_n + \beta x_{ni}$$

and

$$\log\left\{\frac{p_s(x_{si})}{1-p_s(x_{si})}\right\} = \alpha_s + \beta x_{si}$$

- Therefore: dilution assumption can be tested by fitting model with separate intercepts and slopes, and testing for common slope.

- $\rho =$ relative potency $= \exp\left\{\dfrac{\alpha_n - \alpha_s}{\beta}\right\}$

- Fieller's theorem can be used to produce confidence intervals for relative potency, as well as LD50

- See example 11.3.1, pages 362-368

<u>Analysis of the Peptide Data</u>

The estimated log LD50s from this model are

$$\log \text{LD50}_n = \frac{-\hat{\alpha}_n}{\hat{\beta}} = \frac{1.1931}{0.7234} = 1.65$$

$$\log \text{LD50}_s = \frac{-\hat{\alpha}_s}{\hat{\beta}} = \frac{2.4445}{0.7234} = 3.38$$

The log relative potency is estimated as

$$\log\{\hat{\rho}\} = \frac{\hat{\alpha}_n - \hat{\alpha}_s}{\hat{\beta}} = \frac{-1.1931 - (-2.4476)}{0.7234} = 1.73$$

```
proc logistic data=assay descending;
   freq count;
   model status = int_n int_s
                  ldose*int_n ldose*int_s
                  ldose*int_n*ldose*int_n
                  ldose*int_s*ldose_int_s
                  / noint scale=none aggregate
                 include=4 selection=forward details;
   contrast 'equal slopes' int_n*ldose 1 int_s*ldose -1;
run;
```

Tests for Quadratic Terms

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
1.4817	2	0.4767

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
int_n	1	-1.1301	0.2948	14.6983	0.0001
int_s	1	-3.3782	0.8797	14.7479	0.0001
int_n*ldose	1	0.6199	0.1240	24.9907	<.0001
int_s*ldose	1	1.0615	0.2798	14.3914	0.0001

Contrast Test Results

Label	DF	Wald Chi-Square	Pr > ChiSq
Equal slope	1	2.0820	0.1490

A parallel lines model is fit.

```
proc logistic data=assay descending outest=estimate
                    (drop= intercept _link_ _lnlike_) covout;
   freq count;
   model status = int_n int_s ldose /
                noint scale=none aggregate covb;
run;
```

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
int_n	1	-1.1931	0.3158	14.2781	0.0002
int_s	1	-2.4476	0.4532	29.1632	<.0001
ldose	1	0.7234	0.1177	37.7681	<.0001

Estimated Covariance Matrix

Variable	int_n	int_s	ldose
int_n	0.099702	0.025907	-0.00984
int_s	0.025907	0.20542	-0.03648
ldose	-0.00984	-0.03648	0.013856

You can use Fieller's formula to calculate confidence intervals. SAS macro available in Section 11.6 of text.

Estimate	Value	95% CI	Exponentiated Value	Exponentiated CI
log (Potency)	1.73	(0.4262, 2.9993)	5.64	(1.53, 20.07)
log ($LD50_n$)	1.65	(0.8237, 2.6875)	5.21	(2.28, 14.69)
log ($LD50_s$)	3.38	(2.4863, 4.4506)	29.37	(12.02, 85.68)

Fieller's Formula

When F and G are significantly different from 0 at two-sided level α, a $100(1-\alpha)\%$ two-sided confidence interval can be formed for (ξ / η) without the direct use of the ratio estimator R. It is done by solving

$$\frac{(F-\theta G)^2}{(\upsilon_F - 2\theta\upsilon_{FG} + \theta^2\upsilon_G)} = \chi^2_{(1-\alpha)}(df = 1)$$

as a quadratic equation. The roots of the quadratic equation, θ_L and θ_U are the $100(1-\alpha)\%$ confidence limits for $\theta = (\xi/\eta)$. Here, $\chi^2_{(1-\alpha)}$ (df = 1) is the $100(1-\alpha)$ percentile of the chi-squared distribution with df = 1. This method is called Fieller's formula. With $\alpha = 0.05$, $\chi^2_{(1-\alpha)}$ (df = 1) = 3.84. The quadratic equation is

$$0 = \left(G^2 - 3.84\upsilon_G\right)\theta^2 - 2(FG - 3.84\upsilon_{FG})\theta + \left(F^2 - 3.84\upsilon_F\right)$$

$$\theta_L, \theta_U = \frac{(FG - 3.84\upsilon_{FG}) \pm \sqrt{(FG - 3.84\upsilon_{FG})^2 - (F^2 - 3.84\upsilon_F)(G^2 - 3.84\upsilon_G)}}{(G^2 - 3.84\upsilon_G)}$$

$$= \left(\frac{F}{G}\right)\left\{\frac{\left(1 - 3.84\,\frac{\upsilon_{FG}}{FG}\right) \pm \sqrt{\left(1 - \frac{3.84\upsilon_{FG}}{FG}\right)^2 - \left(1 - \frac{3.84\upsilon_F}{F^2}\right)\left(1 - \frac{3.84\upsilon_G}{G^2}\right)}}{\left(1 - 3.84\upsilon_G / G^2\right)}\right\}$$

$$\rightarrow \frac{F}{G}\left\{1 \pm \sqrt{\frac{3.84\upsilon_F}{F^2} + \frac{3.84\upsilon_G}{G^2} - \frac{2(3.84)\upsilon_{FG}}{FG}}\right\} \quad \text{as } n \rightarrow \infty$$

11.4 Analysis of Pain Study

Dose	Diagnosis I		Diagnosis II	
	Adverse	Not	Adverse	Not
1	3	26	6	26
5	7	26	20	12
10	10	22	26	6
12	14	18	28	4
15	18	14	31	1

- Interest lies in investigating the association of adverse effects with dose and diagnosis

- Unlike the previous bioassay analysis, this does not compare the tolerance distributions of two drugs and is not strictly concerned with estimating the tolerance distribution for either drug. Can still use bioassay methods, however.

See page 362 for data input code.

```
proc logistic data=adverse;
   freq count;
   model status = i_diagI i_diagII
                  dose*i_diagI dose*i_diagII /
               noint scale=none aggregate;
   contrast 'equal slope' dose*i_diagI 1 dose*i_diagII -1;
run;
```

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	2.7345	6	0.4557	0.8414
Pearson	2.7046	6	0.4508	0.8449

Parameter Estimates

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
i_diagI	1	-2.2735	0.4573	24.7197	<.0001
i_diagII	1	-1.4341	0.3742	14.6887	0.0001
i_diagI*dose	1	0.1654	0.0414	15.9478	<.0001
i_diagII*dose	1	0.3064	0.0486	39.8186	<.0001

Hypothesis Test

Contrast Test Results

Label	DF	Wald Chi-Square	Pr > ChiSq
Equal slope	1	4.8787	0.0272

With p=0.0272, The hypothesis of a common slope is rejected.

Try fitting model based on log doses:

```
proc logistic data=adverse;
   freq count;
   model status = i_diagI i_diagII
                  ldose*i_diagI ldose*i_diagII /
                  noint scale=none aggregate;
   contrast 'equal slope' ldose*i_diagI 1 ldose*i_diagII -1;
run;
```

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	4.8774	6	0.8129	0.5596
Pearson	4.4884	6	0.7481	0.6109

The goodness-of-fit tests are not as supportive of this model as they are for the actual dose model, but they are still entirely satisfactory.

Hypothesis Test

Contrast Test Results			
Label	DF	Wald Chi-Square	Pr > ChiSq
Equal slope	1	2.4034	0.1211

- With p=0.1211, we do not reject the hypothesis of a common slope.

- Thus, both models fit the data, and one model offers the possibility of a parallel lines model.

11.5 Estimating Tolerance Distributions

• Quantal response data analysis techniques can be applied in other areas as well.

• Example: ED50 can be applied to describe the median ages at which certain types of physical development occur

Age	Onset Number	Total Girls
5	5	209
6	8	126
7	21	136
8	54	143
9	72	115
10	90	112
11	121	126
12	90	91

```
data development;
  input age onset total @@;
datalines;
5.5    5  209
6.5    8  126
7.5   21  136
8.5   54  143
9.5   72  115
10.5  90  112
11.5 121  126
12.5  90   91
;

data development2;
   set development;
   notonset=total-onset;
do;
   Response='yes'; count=onset; output;
   Response='no';  count=notonset;  output;
end;
run;
```

```
ods graphics on;
proc probit order=data plots=predpplot;
   model onset/total = age / lackfit;
run;
ods graphics off;
```

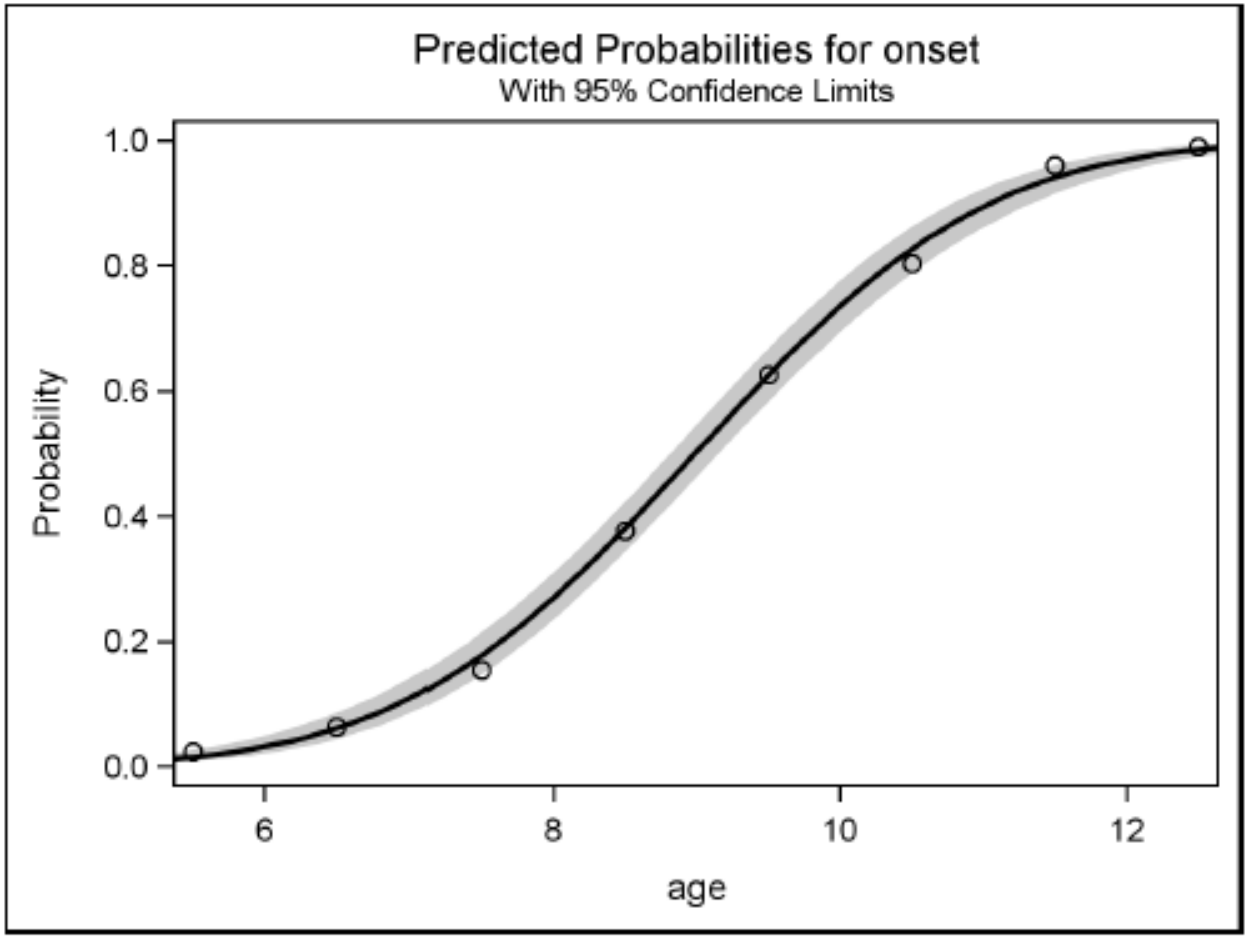
Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-5.5836	0.2776	-6.1277	-5.0395	404.55	<.0001
age	1	0.6215	0.0310	0.5607	0.6823	401.44	<.0001

Probit Model in Terms of Tolerance Distribution

MU	SIGMA
8.98464317	1.6091091

Predicted probabilities of onset by age



Odds Ratio As Criterion For Relative Risk

Let $\pi_0(\theta) = \frac{\theta}{(\theta+1)}$ be event proportion for non-exposed relative to role θ of background factors.

Let $\pi_1(\theta) = \frac{\psi\theta}{(\psi\theta+1)}$ be event proportion for exposed.

$$\frac{\pi_1(\theta)\{1-\pi_0(\theta)\}}{\pi_0(\theta)\{1-\pi_1(\theta)\}} = \psi \text{ for all } \theta$$

$$\pi_0(\theta) = \pi_1\left(\frac{\theta}{\psi}\right), \text{ or } \pi_0(\psi\theta) = \pi_1(\theta)$$

$$\theta(x_1, x_2, ..., x_t) = \theta_0 \left\{ \prod_{k=1}^{t} \psi_k^{x_k} \right\} = \frac{\pi(x_1, x_2, ..., x_t)}{1-\pi(x_1, x_2, ..., x_t)}$$

$$= \exp(\beta_0) \prod_{k=1}^{t} \exp(\beta_k x_k)$$

$$= \exp\left(\beta_0 + \sum_{k=1}^{t} \beta_k x_k \right)$$

where $\theta_0 = \exp(\beta_0)$ and $\psi_k = \exp(\beta_k)$.