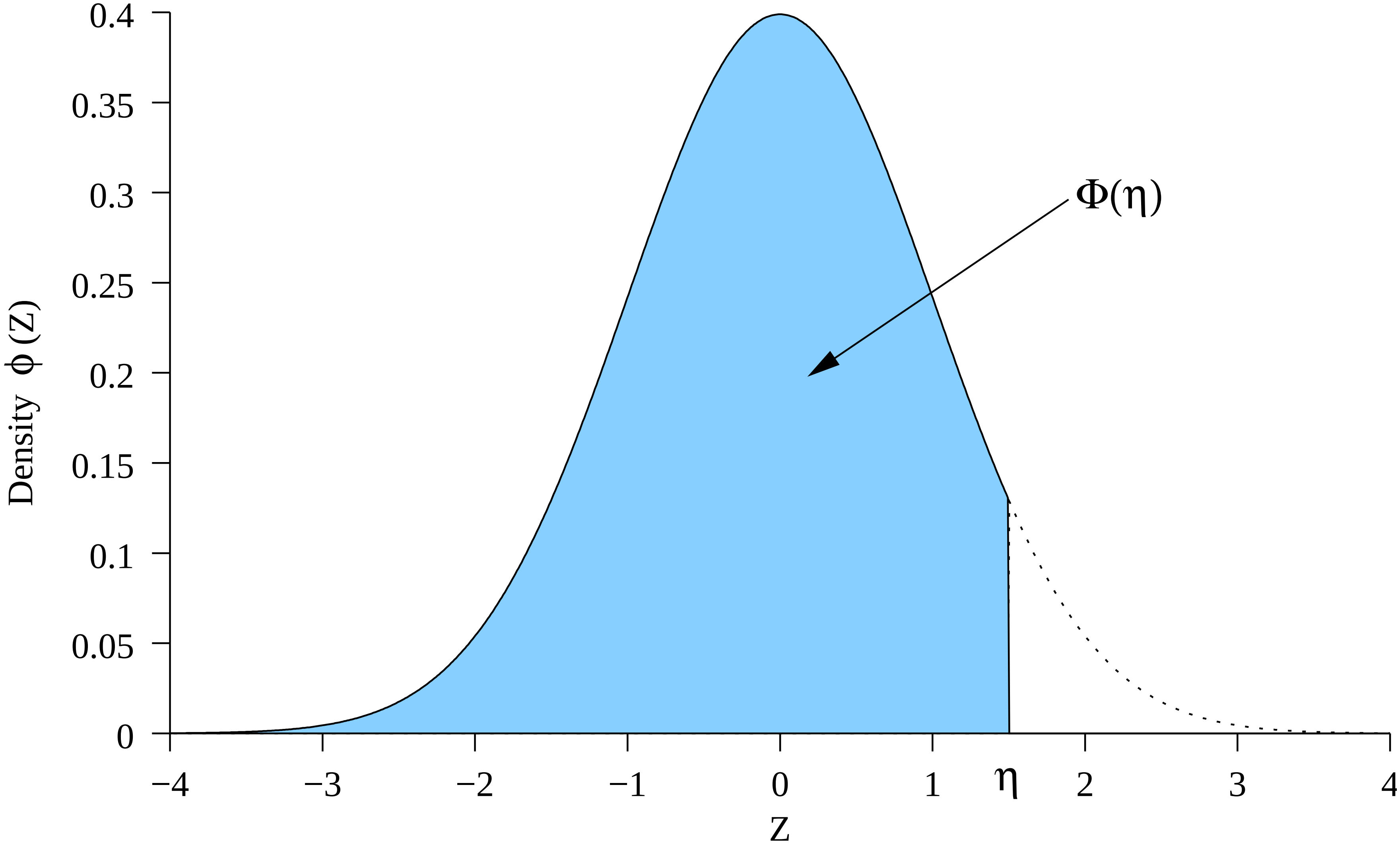
Probit Models

- Another very common link function used with binomial data is the probit.

- The mean, or the chance of survival, is the probability that a standard normal random variable will be less than the linear predictor

$$\mu_{ij} = \Phi(\eta_{ij}) = \int_{-\infty}^{\eta_{ij}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz.$$

Which is illustrated in the following figure.



Φ(η)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Density ϕ(Z)
−4
−3
−2
−1
0
1
η
2
3
4
Z

- As the linear predictor, η, increases the corresponding chance of survival, the shaded area, also increases.

- Intuitively this has appeal as we can think of some underlying continuous variate, $y \sim \mathrm{N}(\eta, 1)$ growth, which will result in a survival if it falls above the threshold value of zero.

- However, we need to distinguish between intuitive appeal and the actual existence of an underlying continuous variate.

- To fit the probit model we will need:
  - the mean (μ)
  - the variance ($\boldsymbol{R}$)
  - the partial derivatives of the mean ($\boldsymbol{H}$).

- The formula for the mean is given above.

- The variance is the same as for the logit.

- The partial derivative matrix is

$$\boldsymbol{H} = \mathrm{Diag}\left(\frac{1}{\sqrt{2\pi}} e^{-\eta_{ij}^2/2}\right).$$

ASReml

```
Probit Analysis
  Type * !A
  Partners !I
  Day50
  Num
ff.dat !SKIP=1
Day50 !BINOMIAL !PROBIT !TOTAL Num ~ mu Type*Partners
0 0 0
predict Type Partners !TDIFF
```

```
  1 LogL=-57.5483     S2=  1.0000          5 df    1.000
  2 LogL=-60.0108     S2=  1.0000          5 df    1.000
  3 LogL=-60.0108     S2=  1.0000          5 df    1.000
  4 LogL=-60.0108     S2=  1.0000          5 df    1.000
  5 LogL=-60.0108     S2=  1.0000          5 df    1.000
  6 LogL=-60.0108     S2=  1.0000          5 df    1.000
Final parameter values                         1.0000
Deviance from GLM fit                5      142.45
Variance heterogenity factor [Deviance/DF]     28.49

Source                Model  terms     Gamma     Component    Comp/SE   % C
Variance                 10      5   1.00000       1.00000       0.00   0 F

Analysis of Variance             NumDF     DenDF    F_inc             Prob
  5 mu                               1       5.0     7.14            0.044
  1 Type                             2       5.0     5.85            0.049
  2 Partners                         1       5.0     4.94            0.077
  6 Type.Partners                    1       5.0     4.82            0.079
Notice: The DenDF values are calculated ignoring fixed/boundary/singular
            variance parameters using numerical derivatives.

Warning: This Analysis of Variance based on the working variable is not
         equivalent to the Analysis of Deviance.  Standard errors are scaled
         by the variance of the working variable, not the residual deviance.
```

Predict File:

```
Ecode is E for Estimable, * for Not Estimable


---- ---- ---- ---- ---- ---- ---- ----   1 ---- ---- ---- ---- ---- ---- ---- ----
Predicted values of Day50
The cells of the hypertable are calculated from all model terms constructed
               solely from factors in the averaging and classify sets.
Warning:     4 non-estimable [aliased] cell(s) may be omitted from the table.
               The Overall SED statistic includes non-estimable predictions.

Type            Partners            Probit_value  Stand_Error Ecode Retransformed_value  approx_SE
P                          1           0.7063        0.2748 E           0.7600         0.0849
P                          8           0.7063        0.2748 E           0.7600         0.0849
V                          1           0.3585        0.2566 E           0.6400         0.0951
V                          8          -0.8416        0.2858 E           0.2000         0.0797
N                          0           0.5828        0.2668 E           0.7200         0.0891
SED: Standard Error of Difference: Min     0.3701   Mean     0.3844   Max     0.3964

Predicted values with t statistics
 0.7063
 0.7063      0.00
 0.3585     -0.93  -0.93
-0.8416     -3.90  -3.90  -3.12
 0.5828     -0.32  -0.32   0.61   3.64
```

- Logit
  - Effects are in terms of log odds ratios.
  - For example the estimated difference between P1 and V1 on the logit scale was $1.1527 - 0.5754 = 0.5773$

$$\ln(\widehat{p}_{P1}/1-\widehat{p}_{P1}) - \ln(\widehat{p}_{V1}/1-\widehat{p}_{V1}) = \ln\left[\frac{\widehat{p}_{P1}}{1-\widehat{p}_{P1}} \times \frac{1-\widehat{p}_{V1}}{\widehat{p}_{V1}}\right]$$

  - converted into an odds ratio

$$\frac{\widehat{p}_{P1}}{1-\widehat{p}_{P1}} \times \frac{1-\widehat{p}_{V1}}{\widehat{p}_{V1}} = \frac{.76}{.24} \times \frac{.36}{.64}$$
$$\exp(0.5773) = 1.7813$$

- Odds-ratios are very easy to interpret when $p_{ij} \simeq 0$

- Probit

- Effects are in terms of random effects on the underlying scale

- For example the estimated difference between P1 and V1 on the probit scale was $0.7063 - 0.3585 = .3978$ standard deviations.

- This works well if we have other normal random effects.