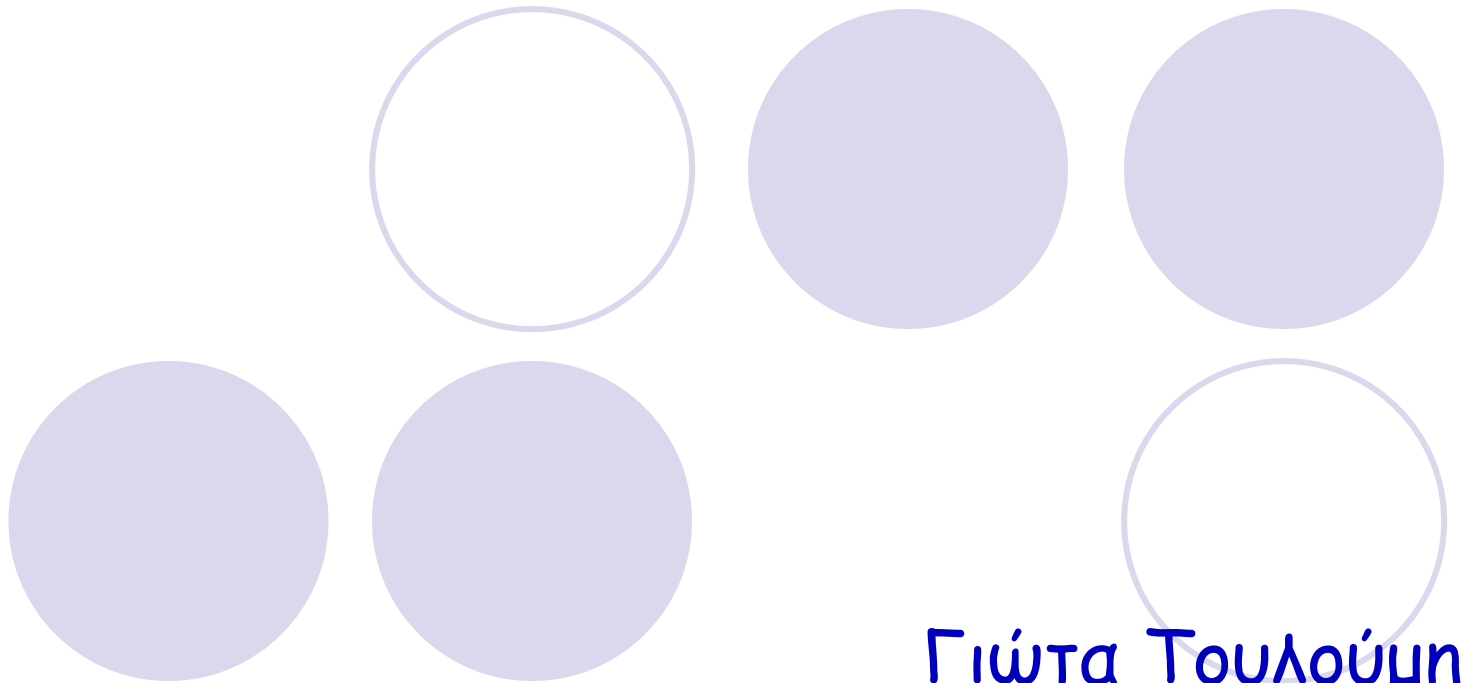


GENERALIZED LINEAR MODELS



Γιώτα Τουλούμη

Καθηγήτρια Βιοστατιστικής και Επιδημιολογίας
Εργ. Υγιεινής, Επιδημιολογίας και Ιατρικής Στατιστικής
Ιατρική Σχολή Πανεπιστημίου Αθήνας

gtouloum@med.uoa.gr

The components of a GLM

Ordinary linear regression

To introduce GLMs we turn to the familiar general linear model. There, the observations y_1, y_2, \dots, y_n are distributed according to normal distributions $N(\mu_i, \sigma^2)$

More importantly, we have that the expected value $\mu_i = E(y_i)$ is related to a p -dimensional vector $\mathbf{x}_i' = (x_1, \dots, x_p)$ as follows:

$$\mu_i = \mathbf{x}_i' \boldsymbol{\beta} = \sum_{j=1}^p x_{ij} \beta_j$$

with $\boldsymbol{\beta}$ being an unknown p -dimensional parameter vector.

The components of \mathbf{Y} are independent Normal variates with variance σ^2 and $E(\mathbf{Y}) = \boldsymbol{\mu}$ where $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ (1)

The generalization

To simplify the transition to *GLM*, we shall rearrange (1) slightly to produce the following three-part specification:

- **The random part:** the components of \mathbf{Y} have independent Normal distributions with $E[\mathbf{Y}] = \boldsymbol{\mu}$ and constant variance σ^2
- **The systematic component:** covariates $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ produce a linear predictor $\boldsymbol{\eta}$ given by
$$\boldsymbol{\eta} = \sum_{j=1}^p x_{ij} \boldsymbol{\beta}_j$$
- **The link between the random and the systematic components:** $\boldsymbol{\mu} = \boldsymbol{\eta}$

The link function

- This generalization introduces a new symbol η for the **linear predictor** and the third component that specifies that μ and η are in fact identical. If we write $\eta_i = g(\mu_i)$ then $g(\cdot)$ is called the **link function**.
- Therefore, the association of the mean $\mu_i = E[y_i]$ with X_i is modeled by introducing a *link* function g so that $\eta_i = g(\mu_i)$ where η_i is a linear combination and is called *the linear predictor*.

$$\eta_i = \mathbf{x}_i' \boldsymbol{\beta} = \sum_{j=1}^p x_{ij} \beta_j$$

The canonical link

A special link function is called the *canonical link*. This is when the canonical parameter is equal to the linear predictor:

$$\theta = \eta = \sum_{j=1}^p x_{ij} \beta_j$$

The canonical links have attractive statistical properties because in this case there exists a sufficient statistic (McCullagh and Nelder, 1989, p. 32)

$$\sum_i x_{ij} y_i, j = 1, \dots, p$$

The canonical links for Normal, Poisson and Binomial:

○ Normal	$\eta = \mu$	<i>Identity</i>
○ Poisson	$\eta = \log(\mu) = \log(\lambda)$	<i>Log of rate</i>
○ Binomial	$\eta = \log[\pi / (1 - \pi)]$	<i>Logit</i>

Notes on the linear predictor

- The linear predictor is linear in the parameters, not necessarily in the explanatory variables. For example, in polynomial regression, powers of the variables occur: $\eta_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_p x_i^p$
- The 'log likelihood of the model' is usually shorthand for 'the log likelihood function of the model for the given data evaluated at the MLE's of the parameters' i.e. the maximum of the log likelihood function.
- With minor exceptions the underlying theory (parameter estimation, inference, model assessment and comparison) proceeds in the same way for all GLM's; it is just the distribution and the link that differ.
- The link function is chosen to provide a suitable scale for the effects of explanatory variables to operate in a linear manner. Typically the range of μ will be transformed to the whole real line ($-\infty$ to ∞).

Model fit



- Fitting a *GLM* to data means defining the log likelihood of the parameters β given the data and maximizing this to give the MLE's $\hat{\beta}$
- The absolute value of maximized log likelihood will be **very** dependent on the particular data observed and so is not of interest itself.
- However, the difference in this quantity of two models, one nested to other, does provide a measure of the comparative fit of the two models.
- The more general model will necessarily have the greater log likelihood, so the question is whether the difference in log likelihood is large enough to indicate that the more general model provides a 'real' improvement in it.

Likelihood ratio test

- Two models are called *nested* if one is “contained” in some sense in the other. For example, in multiple regression, the larger model would contain all effects of the smaller model.
- The likelihood-ratio test compares the maximized likelihood of the two nested models, i.e.,

$$\lambda = \frac{L(\hat{\theta}_1; \mathbf{y})}{L(\hat{\theta}_2; \mathbf{y})} \quad \log(\lambda) = l(\hat{\theta}_1; \mathbf{y}) - l(\hat{\theta}_2; \mathbf{y})$$

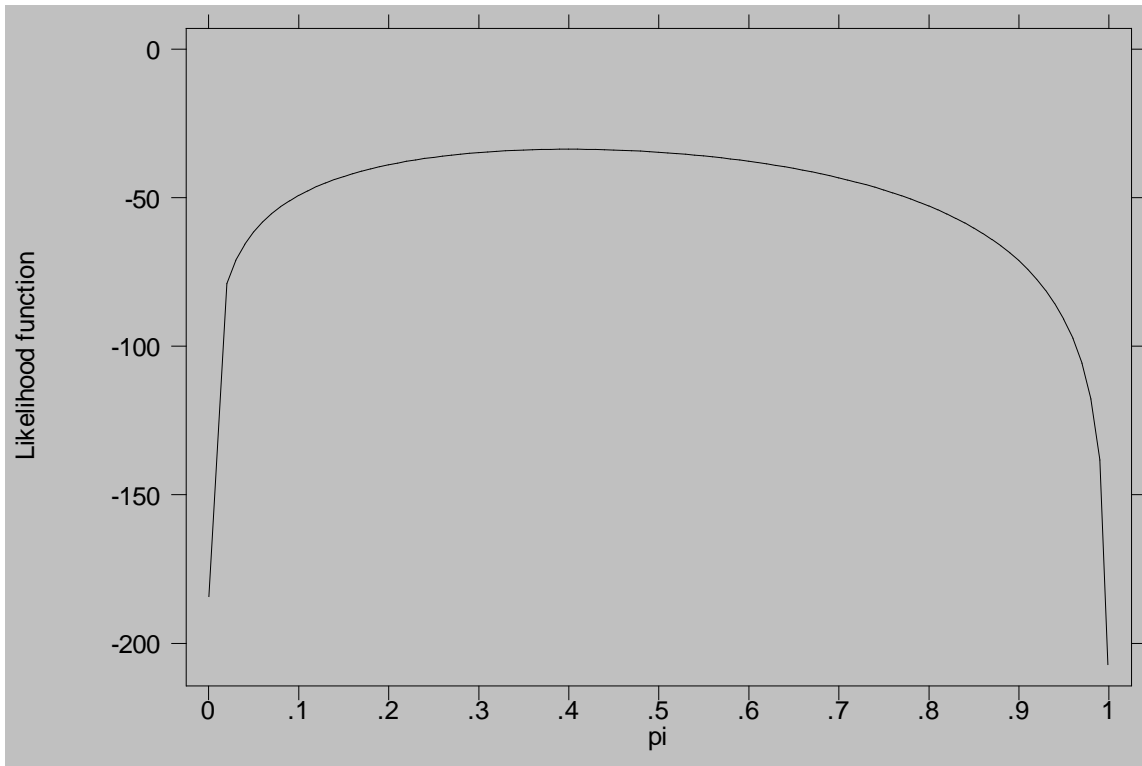
where $\hat{\theta}_1$ and $\hat{\theta}_2$ are the MLEs in the smaller and larger models respectively. λ takes values from 0 to 1 with lower values favoring the larger model.

In large samples $-2 \log \lambda \sim \chi_v^2$

where v is the difference in dimension in the two models.

Example

Consider the situation $n_1=n_2=n_3=n_4=n_5=10$
and $y_1=2, y_2=1, y_3=1, y_4=3, y_5=3$. A plot of the log-likelihood is as follows:



The log likelihood
is maximized at

$$\hat{\pi} = \bar{y} = 0.2$$

$$L(\theta; y) \propto P(y; \theta)$$

Log likelihood, usually negative

Example

- Consider the example: Binomial, $n_1=n_2=n_3=n_4=n_5=10$; $y_1=2, y_2=1, y_3=1, y_4=3, y_5=3$; $\hat{\pi} = 0.2$
- In the previous example, testing the hypothesis $H_0: \pi=0.3$ by the likelihood-ratio test, we compare the log-likelihood of the binomial model under the null hypothesis $l(0.3; \mathbf{y}) = -26.307$, while its maximum, evaluated at the MLE is $l(0.2; \mathbf{y}) = -25.020$.
- Then, the likelihood ratio under the two models is,
$$-2 \log \lambda = -2(-26.307 + 25.020) = 2.573$$
which is asymptotically distributed as a chi-square distribution with one degree of freedom. The p-value associated with the above value of λ is 0.109, which is non-significant.

Deviance



- A **saturated** model is one with the maximum possible number of parameters, with no redundancies. Usually this implies the same number of parameters as observations, and a model whose fitted values exactly reproduce the observations.
- Consider $l(\tilde{\theta}; \mathbf{y})$ the maximized likelihood associated with a model that fits as many parameters as observations (saturated model). This is the maximum achievable log-likelihood. The discrepancy of any other model (with maximized likelihood) from this saturated model is measured by:

$$S = \frac{D(\hat{\theta}; \tilde{\theta})}{\phi} = -2 \left\{ l(\hat{\theta}; \mathbf{y}) - l(\tilde{\theta}; \mathbf{y}) \right\}$$

Deviance (continued)

$$S = \frac{D(\tilde{\theta}; \hat{\theta})}{\phi} = -2 \left\{ l(\hat{\theta}; \mathbf{y}) - l(\tilde{\theta}; \mathbf{y}) \right\}$$

- The numerator of the fraction on the left is called the **deviance** for the model under consideration. The scaled version of this $S = \frac{D(\tilde{\theta}; \hat{\theta})}{\phi}$ for some appropriate scale parameter ϕ , is called the **scaled deviance**.
- If the current model for n observations has p parameters, and is appropriate for the data, then the sampling distribution of the scaled deviance is asymptotically X^2 with $n-p$ degrees of freedom.
- This can be used as the basis for the test of the null hypothesis that the current model is an adequate one.

Example: The normal distribution

$$l(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu_i)^2$$

In the saturated model, the MLE of μ_i is just y_i , so

$$l(\tilde{\theta}, y) = -\frac{n}{2} \ln(2\pi\sigma^2)$$

In the current model we have some linear predictor:

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$$

and the fitted values are then $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip}$

$$\text{and } l(\hat{\theta}, y) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Hence the scaled deviance (and thus the likelihood-ratio criterion) is

$$S = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sigma^2}$$

Example: The normal distribution (continued)

$$S = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sigma^2}$$

- S has a $X^2_{n-(p+1)}$ distribution if the current model is appropriate. In practice σ^2 is unknown. (This is why in the normal distribution case the F-test is more appropriate than the X^2 test).
- For other GLM's (remember the scale parameter is equal to 1 for Poisson and for binomial for ungrouped data) φ is known to be equal to 1, so we can use the scale deviance to assess fit.
- Note that for the normal model, the deviance $D = \sigma^2 S$ is the residual sum of squares.

Nested models

- Let's see an example for how the scaled deviance can be used to compare two nested models.
- **Example:** If model 1 has linear predictor for i -th unit $\mu_i = \beta_0 + \beta_1 x_{i1}$ and model 2, with the same distribution, link function and scale parameter, has linear predictor $\mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2}$ then model 1 is nested in model 2 because it can be derived from it by setting $\beta_2 = \beta_3 = 0$.
- The relative fit of two nested models can be compared by comparing their respective scaled deviances. The change in scaled deviance comparing models 1 and 2 is:
$$S_1 - S_2 = -2(l_1 - l_s) + 2(l_2 - l_s) = -2(l_1 - l_2)$$
or minus twice the log likelihood ratio (l_1 , l_2 and l_s denote the log likelihood of models 1, 2 and the saturated model)

Nested models (continued)

$$S_1 - S_2 = -2(l_1 - l_s) + 2(l_2 - l_s) = -2(l_1 - l_2)$$

- If model 1 has p_1 parameters and model 2 has $p_2 > p_1$ parameters, then under the null hypothesis that both models fit equally well, and given that each is an adequate fit, then asymptotically

$$S_1 - S_2 \sim X_{p_2 - p_1}^2$$

Example: Normal distribution

- We have seen that in this case the scaled deviance is RSS / σ^2 , so

$$S_1 - S_2 = \frac{RSS_1 - RSS_2}{\sigma^2}$$

- Under the null hypothesis this has a $X^2_{p_2 - p_1}$ distribution.
- In practice, σ^2 has to be estimated, and the residual mean square from the more complex model is used for this: $\hat{\sigma}^2 = \frac{RSS_2}{n - p_2}$
- and the ratio $\frac{(RSS_1 - RSS_2) / (p_2 - p_1)}{\hat{\sigma}^2}$

is referred to an $F_{p_2 - p_1, n - p_2}$ distribution.

Some notes on nested models

- The F test for comparing normal linear models is 'exact', i.e. not just asymptotic.
- The X^2 test for change in deviance is typically a much better asymptotic approximation than the X^2 test for comparison with the saturated model.
- There is disagreement over the terminology for the deviance. The terms for D and S above are used as the standard reference, McGullagh and Nelder, and in Stata. Clayton and Hills call S the deviance (not scaled deviance).
- In a Poisson or binomial model, the scale parameter is 1, so S and D coincide and the changes in deviance can be referred directly to a X^2 distribution: these tests are log likelihood ratio tests.

Wald test

- **Wald** tests are based on the asymptotic distribution of the maximum-likelihood estimator $\hat{\theta}$ of θ . Thus, in large samples (under some regularity conditions), the distribution of θ is

$$\hat{\theta} \sim N\left(\theta, I^{-1}(\theta)\right)$$

- a multivariate normal distribution with mean equal to the unknown parameter, and variance-covariance matrix equal to the inverse of the *information matrix*. The latter is the $n \times n$ matrix of minus the expectation of the second derivatives,

i.e

$$I(\theta) = -E \left\{ \frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \right\} \quad \text{with } i = 1, \dots, n, j = 1, \dots, n$$

Wald test (continued)

- Using the asymptotic distribution of $\hat{\theta}$, under the null hypothesis $H_0 : \theta = \theta_0$, the quadratic form

$$W = (\hat{\theta} - \theta_0)' \hat{I}(\theta) (\hat{\theta} - \theta_0) \sim \chi_n^2$$

approximately, where $\hat{I}(\theta)$ is the *observed information matrix*, i.e., the matrix comprised of minus the second derivatives of the log likelihood without having taken their expectation.

- In the univariate case, the square root of the above test

$$z = \frac{\hat{\theta} - \theta_0}{\sqrt{\text{var}(\hat{\theta})}}$$

is asymptotically distributed according to a standard normal distribution.

Wald test (continued)

$$z = \frac{\hat{\theta} - \theta_0}{\sqrt{\text{var}(\hat{\theta})}}$$

- Notice that we do not assume normally distributed θ . In that case, the ratio above would be *exactly* distributed according to a t distribution.
- Unless this assumption holds, the small-sample properties of the above ratio are unknown. The long-term properties (whether one assumes a t or a standard normal distribution) are asymptotically identical.

Example of the Binomial distribution (continued)

The observed information in the case of the binomial distribution is:

$$I(\hat{\pi}) = -\frac{d^2 l(\pi; \mathbf{y})}{d\pi^2} = \frac{1}{\pi^2} \sum_{i=1}^n y_i - \frac{1}{(1-\pi)^2} \left(\sum_{i=1}^n y_i - \sum_{i=1}^n n_i \right) \Bigg|_{\pi = \hat{\pi}}$$
$$= \frac{1}{(0.2)^2} (10) - \frac{1}{(1-0.2)^2} (10-50) = 312.5$$

So given the data presented above, the Wald test for the null hypothesis $H_0: \pi=0.3$ would be:

$$\left(\frac{0.2 - 0.3}{\sqrt{1/I(0.2)}} \right) = \frac{-0.10}{\sqrt{1/312.5}} \approx -1.767$$

which asymptotically is a z test that in the case of a two-sided alternative hypothesis would produce a p-value . Note that the result of the Wald test is consistent, but not identical with the likelihood ratio results shown previously.

Example: Plasma levels of retinol

- Low plasma concentrations of retinol, beta-carotene, or other carotenoids might be associated with increased risk of developing certain types of cancer. The following data (from Therese Stuckel, Dartmouth University), were taken from $n=315$ subjects that had had a surgical procedure during a three-year period that involved biopsy or extraction of non-cancerous lesions. This cross-sectional study explored factors determining the levels of these substances.

Plasma levels of retinol - variables in dataset

AGE: Age in years

SEX: Sex (1=Male, 2=Female).

SMOKSTAT: Smoking status (1=Never, 2=Former, 3=Current Smoker)

QUETELET: Quetelet ($\text{weight}/(\text{height}^2)$)

VITUSE: Vitamin Use (1=Yes, fairly often, 2=Yes, not often, 3=No)

CALORIES: Number of calories consumed per day.

FAT: Grams of fat consumed per day.

FIBER: Grams of fiber consumed per day.

ALCOHOL: Number of alcoholic drinks consumed per week.

CHOLESTEROL: Cholesterol consumed (mg per day).

BETADIET: Dietary beta-carotene consumed (mcg per day).

RETDIET: Dietary retinol consumed (mcg per day)

BETAPLASMA: Plasma beta-carotene (ng/ml)

RETPLASMA: Plasma Retinol (ng/ml)

Example: Plasma levels of retinol

In our example, consider the association of plasma levels of retinol (`retplasm`, the outcome) with alcohol consumption (`alcohol`, the predictor). An outlier has been removed from the data.

```
. reg retplasm alcohol
```

Source	SS	df	MS			
Model	671843.17	1	671843.17	Number of obs =	314	
Residual	12948338.7	312	41501.0855	F(1, 312) =	16.19	
Total	13620181.9	313	43514.958	Prob > F =	0.0001	
				R-squared =	0.0493	
				Adj R-squared =	0.0463	
				Root MSE =	203.72	

retplasm	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
alcohol	9.365251	2.327637	4.02	0.000	4.785401	13.9451
_cons	578.8857	13.04634	44.37	0.000	553.2158	604.5556

Higher alcohol consumption is associated with higher plasma levels of retinol, a possibly carcinogenic substance. The output lists the F test of the overall significance of regression, the R-square and a t test assessing the significance of the predictor (equal here to the square root of the F statistic).

Example (continued)

Contrast this with the output from the STATA command `glm`.

```
. glm  retplasm alcohol
Iteration 0:    log likelihood = -2113.9991
Generalized linear models                               No. of obs       =        314
Optimization    : ML: Newton-Raphson                   Residual df     =        312
Deviance        = 12948338.69                          Scale param     = 41501.09
Pearson         = 12948338.69                          (1/df) Deviance = 41501.09
Variance function: V(u) = 1                            (1/df) Pearson  = 41501.09
Link function   : g(u) = u                             [Gaussian]
Standard errors : OIM                                  [Identity]
Log likelihood  = -2113.999055                          AIC             = 13.4777
BIC            = 12948327.19
```

```
-----
      retplasm |          Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      alcohol |    9.365251    2.327637     4.02   0.000     4.803167    13.92734
       _cons  |   578.8857   13.04634    44.37   0.000    553.3153    604.4561
-----
```

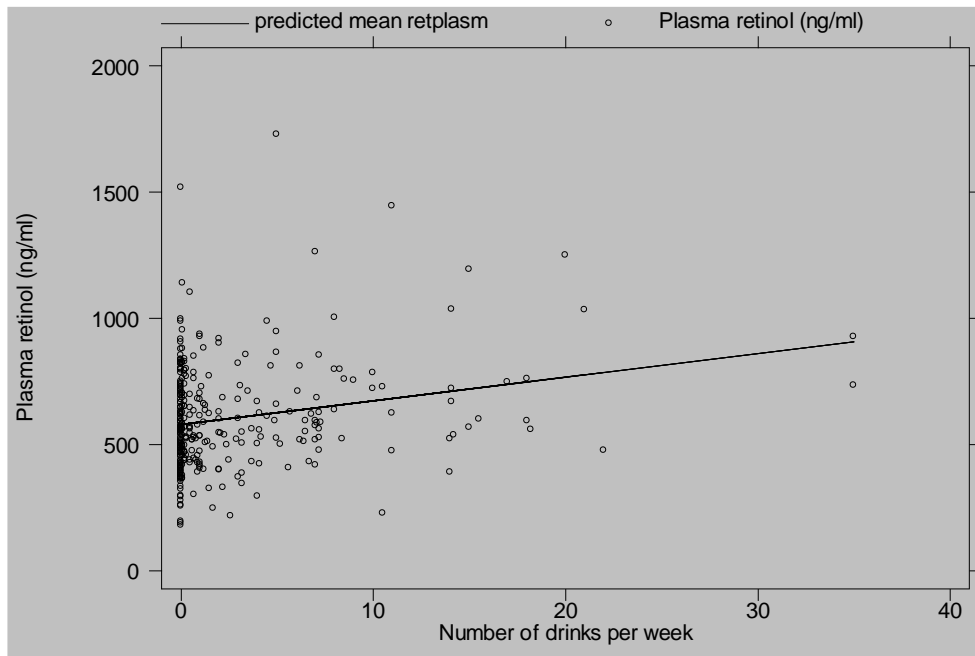
The STATA glm command

- The previous output includes the following:
 - Maximization history (number of iterations in the iterative maximization procedure)
 - The value of the maximized log-likelihood (-2113.9991 here)
 - The variance function and link (here 1 and identity respectively)
 - The AIC and BIC numbers (useful for comparing different models)
 - The MLE estimate of β , which is identical to the OLS estimate presented above
 - A z test of the significance of the predictor (`alcohol`). This is identical to the t test listed in the regression output above (i.e. $\hat{\beta}/\text{var}(\hat{\beta})$), but the distributional assumptions are slightly different in the generalized linear model and the general linear model as we mentioned earlier.
 - The deviance (12948338.69) and the (1/df) Deviance (41501.09), equal respectively to the residual sum of squares and mean squares in the linear model.

Example (continued)

The effect of alcohol consumption on plasma retinol levels is given graphically as follows:

```
. quietly glm retplasm alcohol
. predict yhat
(option mu assumed; predicted mean retplasm)
(1 missing value generated)
. graph yhat retplasm alcohol, xlab ylab c(1.) s(io) border
```



Example: Assessing the significance of fat as a predictor of plasma retinol levels

Consider the following output (note the use of *sequential* sums of squares - option *sequential*):

```
. anova retplasm alcohol fat, continuous(alcohol fat) seq
      Number of obs =      314      R-squared      = 0.0617
      Root MSE      = 202.716      Adj R-squared = 0.0556
```

Source	Seq. SS	df	MS	F	Prob > F
Model	839986.495	2	419993.248	10.22	0.0001
alcohol	671843.17	1	671843.17	16.35	0.0001
fat	168143.325	1	168143.325	4.09	0.0439
Residual	12780195.4	311	41093.8758		
Total	13620181.9	313	43514.958		

The sequential option produces $SSE(\text{fat}|\text{alcohol})=168143.325$. The F statistic associated with it is $F=4.09$. Comparing this to an F distribution with 1 and 311 degrees of freedom, we obtain a p-value of 0.0439, which is just significant at the 5% alpha level, arguing for addition of fat in the model.

Example (continued): The test command

There are several ways to obtain $SSE(\text{fat}|\text{alcohol}) = 168143.325$ besides using the `anova` command.

We could for example subtract $SSE(\text{fat}, \text{alcohol}) = 12780195.4$ from $SSE(\text{alcohol}) = 12948338.69$. The result would be $SSE(\text{fat}|\text{alcohol}) = 168143.325$.

Alternatively, we can use the `test` command after having run the `reg` command as follows:

```
. test fat
( 1)  fat = 0.0
      F( 1, 311) = 4.09
      Prob > F = 0.0439
```

or after the `anova` command as follows:

```
. test fat
```

Source	Partial SS	df	MS	F	Prob > F
fat	168143.325	1	168143.325	4.09	0.0439
Residual	12780195.4	311	41093.8758		

Model-building in GLM

Example: Consider the output from the STATA command `glm`:

```
. glm  retplasm alcohol fat
Iteration 0:    log likelihood = -2111.9469
Generalized linear models                               No. of obs      =          314
Optimization    : ML: Newton-Raphson                   Residual df    =          311
                                                        Scale param    = 41093.88
Deviance        = 12780195.36                          (1/df) Deviance = 41093.88
Pearson         = 12780195.36                          (1/df) Pearson  = 41093.88
Variance function: V(u) = 1                            [Gaussian]
Link function   : g(u) = u                             [Identity]
Standard errors : OIM
Log likelihood  = -2111.946946                          AIC             =          13.471
BIC             = 12780178.12
```

```
-----
retplasm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
alcohol  |   9.990265   2.336708    4.28   0.000    5.410401   14.57013
fat      |  -.6975524   .3448463   -2.02   0.043   -1.373439  -.0216661
_cons   |  630.7705   28.7483    21.94   0.000   574.4249   687.1162
```

Example (continued)

In the plasma retinol levels example, the likelihood-ratio criterion can be calculated by subtracting the deviance $D(X1, X2) = 12780195.36$ from $D(X1) = 12948338.69$ and dividing by 41093.88 ((1/df) Deviance). The result is 4.09, which compared to a chi-square distribution with one degree of freedom produces a p value of 0.0431 as seen from the following output:

```
. display chi2tail(1,4.09)
.04313765
```

Alternatively, we can use the `test` command as follows:

```
. test fat
( 1)  [retplasm]fat = 0.0
      chi2( 1) =      4.09
      Prob > chi2 =      0.0431
```


Example: Joint significance of fat and fiber after accounting for alcohol

To account for the joint significance of daily fat and fiber intake after alcohol consumption has been accounted for we proceed as follows:

```
. reg retplasm alcohol fat fiber
```

Source	SS	df	MS	Number of obs = 314		
Model	841663.312	3	280554.437	F(3, 310)	=	6.81
Residual	12778518.5	310	41221.0276	Prob > F	=	0.0002
-----				R-squared	=	0.0618
-----				Adj R-squared	=	0.0527
Total	13620181.9	313	43514.958	Root MSE	=	203.03

retplasm	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
alcohol	9.964244	2.343874	4.25	0.000	5.35233	14.57616
fat	-.6767318	.3604768	-1.88	0.061	-1.386023	.0325589
fiber	-.4526052	2.244069	-0.20	0.840	-4.868137	3.962927
_cons	635.0317	35.71259	17.78	0.000	564.762	705.3014

Example (continued)

We could calculate the F criterion

$$\frac{[SSE(\mathbf{X}_1) - SSE(\mathbf{X}_1, \mathbf{X}_2)]/p_2}{SSE(\mathbf{X}_1, \mathbf{X}_2)/(n - p_1 - p_2 - 1)}$$

manually but the test command simplifies things significantly:

```
. test fat fiber
( 1) fat = 0.0
( 2) fiber = 0.0
      F( 2, 310) = 2.06
      Prob > F = 0.1292
```

The result indicates that the two predictors are jointly non-significant at the 95% significance level. Compare the previous output to that of the test command after fitting a GLM with alcohol, fiber and fat as the predictors of retinol plasma levels (retplasm) (next slide)

Example (continued)

```
. glm  retplasm alcohol fat fiber
Iteration 0:  log likelihood = -2111.9263
Generalized linear models                No. of obs      =          314
Optimization      : ML: Newton-Raphson   Residual df     =          310
                                                Scale param     =  41221.03
Deviance          =  12778518.55         (1/df) Deviance =  41221.03
Pearson           =  12778518.55         (1/df) Pearson  =  41221.03
Variance function: V(u) = 1              [Gaussian]
Link function     : g(u) = u              [Identity]
Standard errors   : OIM
Log likelihood    = -2111.926345         AIC              =  13.47724
BIC               =  12778495.55
```

```
-----
retplasm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
alcohol  |    9.964244   2.343874     4.25   0.000     5.370336    14.55815
fat      |   -.6767318   .3604768    -1.88   0.060    -1.383253     .0297898
fiber    |   -.4526052   2.244069    -0.20   0.840    -4.850899     3.945688
_cons    |   635.0317    35.71259    17.78   0.000     565.0363    705.0271
-----
```

```
. test fat fiber
( 1)  [retplasm]fat = 0.0
( 2)  [retplasm]fiber = 0.0
      chi2( 2) =      4.12
      Prob > chi2 =      0.1275
```

Example (continued)

$$\frac{[SSE(\mathbf{X}_1) - SSE(\mathbf{X}_1, \mathbf{X}_2)]}{SSE(\mathbf{X}_1, \mathbf{X}_2) / (n - p_1 - p_2 - 1)} = \frac{169820.142}{12778518.5/310} = \frac{169820.142}{41221.0276} \approx 4.12 = 2 \times (2.06)$$

Thus, the multiple-regression model and the GLM are producing almost identical results.

Addition of the predictors fat and fiber jointly is not significant in terms of the overall reduction of unexplained variability (error) in the model, after the effect of alcohol has been accounted for.