**Logistic Regression**

Estimate and maximum likelihood

**Generalized Linear Models**

- *Linear component*:

**- Link function:** let univariate function g be monotonic (strictly decreases or increases) and differentiable

- *Random component*: are independent and from an exponential family, which impose the variance of depends on through a variance function where is called the *dispersion parameter.*

Classic linear regression assumes normal dist.

Logistic regression assumes and is a Bernoulli trial. Other possible links for binary responses include

- Probit:

, where is the cumulative standard normal distribution

- Complementary log-log:

**Maximum Likelihood Estimation of Proportions**

Suppose we draw a random sample of size from some population, send them an offer, and respond. What is our best guess of the response probability?

- Let, *log-likelihood­­*

**Maximum Likelihood Estimation of Means**

Suppose we draw a random sample of size from a normal population with unknown mean and known variance. The observed values are. What is the best guess of?

Pick so that the probability of observing the three values given is maximized:

Note: thus maximum

**Likelihood Function for Logistic Regression**

Probability distribution:

Log-Likelihood

Maximize this with respect to and

**Residual versus null deviance**

**Residual deviance:** deviance for full model

**Null deviance:** deviance for the intercept-only model (think of as SST)

Difference between them measures how much variation is explained by the model and plays the role of extra sums of squares. Can test overall significance of the model b/c it has a **Chi-squared distribution**.

https://i.gyazo.com/c44aaac42dc68ce4705ac986e772923a.png

**Likelihood-Ratio Test**

We can test with the test statistic, which has a chi square distribution with degrees of freedom (the D’s represent deviance)

**Predictive Accuracy of Classifiers**

- GLMs minimize deviance

- common penalized measure

- **Classification rate**:

- **Recall:**

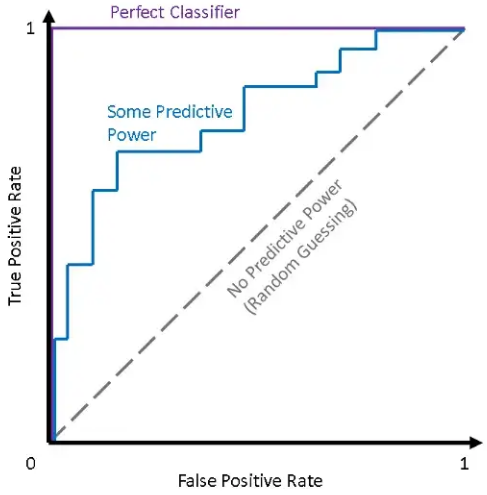
- **TPR (True Positive Rate) (Sensitivity):**

- **FPR (False Positive Rate) (1-Specificity):**

- **Precision:**

- **F1:**

- **AUC** (Area under curve)**:** ROC curve is produced by calculating and plotting the true positive rate against the false positive rate for a single classifier at a variety of thresholds



**Scoring Model:**

- ***Proxy behavior***: behavior that has been observed in the past that is similar to future behavior you would like to predict, e.g., response to similar offer sent yesterday

Warning: your model may not work because this is only a proxy behavior. Seasonality, the state of the economy, what competition is doing, etc. usually all affect response

- ***Target period***: time period when proxy offer was active

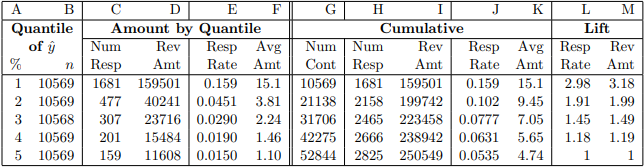
- ***Base period***: a period of time prior to the target period. Information from the base period will be used to predict proxy behavior

Performance usually assessed with a *gains table*

1. Find quantiles of predicted values

2. Compute number of responders and revenue by quantile, also averages

3. Compute cumulative counts and revenues by quantile, also averages and lifts



- **Columns A and B**: Quantiles of

- **Columns C and D**: number of responders and total revenue by quantile

- **Columns E and F**: response rate and average revenue in quantile, e.g., and per contact

- **Column G**: *Depth* of contacts or cumulative counts, e.g.,

- **Columns H and I**: cumulative responders and revenue by quantile:

Row 2: responders and revenue at

Last row: total responders and total revenue

- **Columns J and K**: cumulative response rate and average revenue in quantile:

Row 2: and per contact at

Last row: **contacting at random** gives respond rate, per contact

- **Columns L and M**: lift of model over random guessing, e.g. indicates the response rate from using model to pick best of the names is improved by over picking names at random. Revenue more than tripled!

**Generalized Logistic Regression Model**

Let be a multinomial r.v. for the outcome with values. The probability that observation comes from class is, where

**Approaches:**

- *One vs all*: fit separate logistic regression models

- *Generalized logit*: fit models. Pick class 1 as the base category, but, as with the binary logit, this choice is arbitrary (models equivalent with a different base category)

**Generalized logit model math**:

- Model:

**-** Solving for:

- The probabilities must sum to 1:

- Solving for:

- Substituting back into formula for:

**Generalized logit estimation**:

- Likelihood function:

Where when and otherwise

- The log-likelihood is the cost function

- Denote optimal values of by. Compute estimated probabilities that each observation is in class k as:

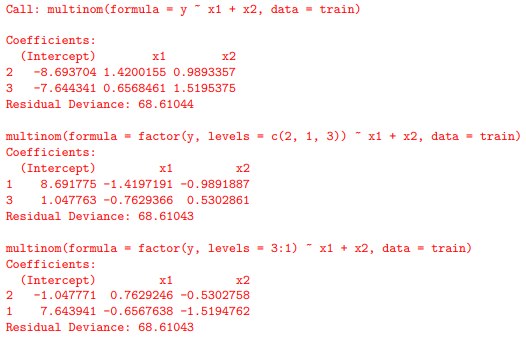
and

- The *maximum likelihood classifier* assigns to the class with the largest

Base category is

All four models give similar results

- There are three possible base categories and all give equivalent models



- All three models have the same deviance (cost function value)

- Let be slopes for. Note that

- Clearly,**,** e.g.,

- **,** e.g.,

**Making predictions**

- **Binary logit**: individual chooses between two options and selects the one that provides greater utility

- **Multinomial logit**: individual chooses among more than two alternatives and selects the one that provides the greatest utility

- **Ordered logit**: individual reveals the strength of his or her preferences with respect to a single outcome

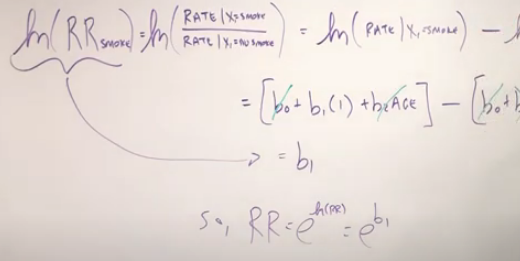
- **Conditional logit**: allows variables that vary across alternatives and possibly across the individuals as well, e.g., choice of mode of transportation (e.g., train, bus, car). Characteristics or attributes of these include time waiting, how long it takes to get to work, and cost.

- **Log-linear analysis**: class of models that subsumes (includes or absorbs) the logit models and more.

**Poisson**

- The PMF of a Poisson distribution:

- Suppose we observe a random sample of ordered pairs from where has a Poisson distribution with mean. More generally, could be a vector of predictors. Assume (log link function) (log(mu) -> rate ratio)



- The likelihood is:

- The log-likelihood of a Poisson model is:

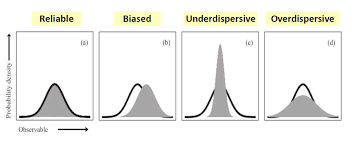
- The log-likelihood of a saturated Poisson model is:

- Let be the MLEs. The deviance is:

Where for

**Beyond Poisson**

- Recall that for Poisson,, but in practice we may find, called the problem of **overdispersion** (or **underdispersion** )



- **Negative binomial distribution** (NBD) **regression** models allow for overdispersion

- Beyond overdispersion, we often observe too many zero values for either Poisson or NBD. What to do?

*Zero-inflated Poisson* (ZIP) assumes a mixture distribution

,

Where is the probability of extra (structural) zeros

*Zero-inflated negative binomial* (ZINP)

*Hurdle* models

- GLMs can accommodate other distributions including exponential and gamma

**Gamma regression**

- The PMF of a *gamma distribution* is:

- Where is the *shape* parameter and is the *scale* parameter. The mean of a gamma distribution is, so. The pdf can be written in terms of:

- The log pdf is given by:

- The log-likelihood is:

- The log-likelihood of the saturate model is:

- The deviance is, but notice how the last three terms of and are identical and thus cancel out. Thus the deviance is:

**Regression Terms and Symbols:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Term** | **ACT** | **JWHT** | **Other** |
| Sum of squared errors | **SSE** | **RSS** |  |
| Total sum of squares | **SST** | **TSS** |  |
| Mean squared error | **MSE** |  |  |
| Residual Standard Error |  | **RSE** |  |

**SSE =**

**MSE** =

**Penalized Estimates**

or AIC = deviance + 2p

**Multicollinearity**

**Pipe:** e.g., . If you are studying then do not control for

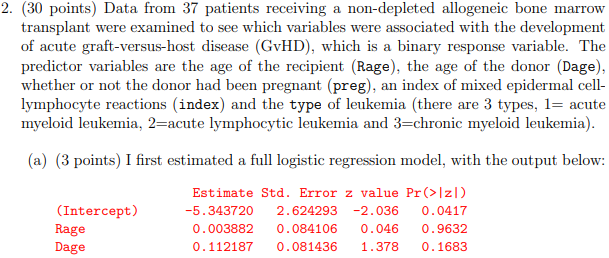
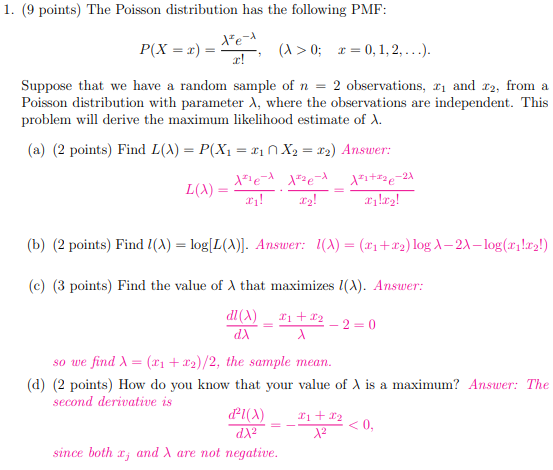
**Latent construct:** predictors manifestations of common, underlying **latent construct**, e.g., . Often, estimate *w* and use it instead of and

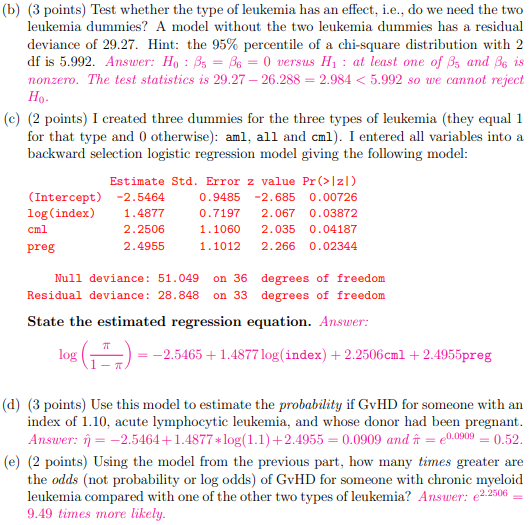
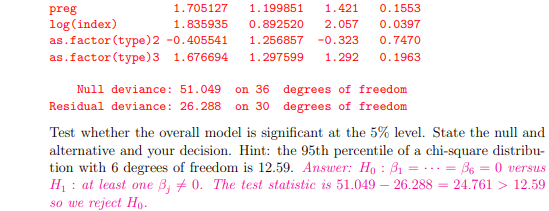
**Back door confound** (**fork**): include control to block back-door path, e.g., if and then control for *w* to study

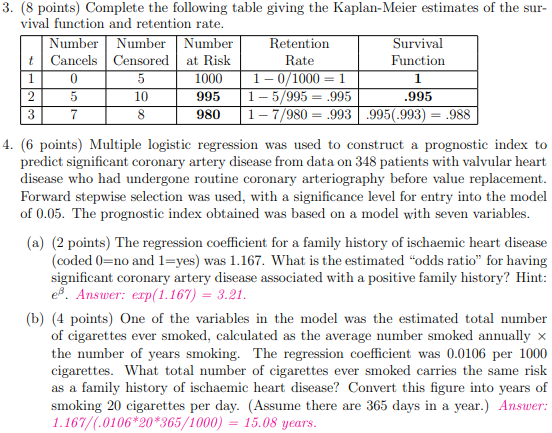
**Collider:** usually do not control for colliders, e.g., if and , then do not control for collider *w* when studying

**Extras**

F =, bottom from full model





https://i.gyazo.com/85d4f8c0e66ad9fffdeeccd5978e1a01.png

