Summary table

This summary table represents the output of a logistic regression model, comparing the unadjusted (bivariate) and adjusted (multivariate) effects of various predictors on an outcome variable, in this case "Adequate delivery care". The predictors include variables like partner's education, respondent's own education, wealth, insurance coverage, media exposure, and age.

In the table, each row represents a different predictor variable, and each column represents a different measure of the association between that predictor and the outcome:

* **p-value**: The statistical significance of each predictor. In general, a p-value of less than 0.05 is considered statistically significant, indicating strong evidence that the predictor is associated with the outcome.
* **OR (Odds Ratio)**: This represents the odds of the outcome occurring given the presence of the predictor, compared to the odds of the outcome occurring in the absence of that predictor. An OR greater than 1 indicates a positive association (i.e., the outcome is more likely with the predictor), while an OR less than 1 indicates a negative association (i.e., the outcome is less likely with the predictor).
* **LL (Lower Limit) and UL (Upper Limit)**: These represent the lower and upper bounds of the 95% confidence interval for the odds ratio. If this interval includes 1, then the association is not considered statistically significant at the 0.05 level.
* **AOR (Adjusted Odds Ratio)**: This is similar to the odds ratio, but it's adjusted for the effects of all other predictors in the model.

For instance, considering the variable 'Patner Education', women whose partners have 'No Education' have half the odds of having 'Adequate delivery care' compared to women whose partners have 'Higher' education in the adjusted model (AOR = 0.50, p-value = 0.0091).

Please note that the interpretation of the odds ratios and p-values depend on how each predictor is coded. For instance, for categorical variables, one level of the variable is usually chosen as the reference category (e.g., 'Higher' for 'Patner Education'), and the odds ratios for the other categories represent the odds of the outcome for that category relative to the reference category.

Also, note that these interpretations are specific to logistic regression, which is used when the outcome is binary (i.e., has two possible values). Different types of regression models may use different measures of association.

In the context of generalized linear models (GLMs), Maximum Likelihood Estimation (MLE) is a method used to estimate the parameters of the model. This method works by finding the parameter values that maximize the likelihood of the data given the model. Essentially, MLE finds the values that make the observed data most probable under the specified model.

GLMs consist of three components: a random component (the distribution of the response variable), a systematic component (the linear predictor), and a link function (which connects the random and systematic components). MLE is used to estimate the parameters in the systematic component, i.e., the coefficients of the predictor variables in the linear predictor.

However, there are several challenges and issues that can arise when using MLE to estimate parameters in GLMs:

1. **Convergence Issues:** The MLE process involves an iterative procedure to find the parameter values that maximize the likelihood function. However, this procedure may fail to converge, or it may converge to a local rather than global maximum, especially if the model is complex or the data are sparse.
2. **Overfitting:** MLE can lead to overfitting, especially if the model includes a large number of parameters relative to the number of observations. Overfitted models may have high variance and poor predictive performance on new data.
3. **Misspecification of the Model:** If the chosen GLM does not fit the data well, the MLEs may be biased or inconsistent. This could happen if, for example, the wrong link function or distribution is chosen for the GLM.
4. **Separation or Quasi-Separation:** In logistic regression, a type of GLM, separation occurs when the response variable can be perfectly predicted from the predictors. In this case, the MLEs do not exist because they would be driven to infinity in an attempt to achieve perfect prediction. Quasi-separation is a similar, but less extreme, situation where the response variable can be almost perfectly predicted.
5. **Multicollinearity:** When predictor variables are highly correlated, it can cause instability in the estimates of the coefficients, leading to inflated standard errors. This is a common problem in many types of regression models, including GLMs.
6. **Inference and Interpretation:** Although MLE can provide point estimates of the parameters, making statistical inferences (e.g., hypothesis tests, confidence intervals) based on these estimates can be challenging. Furthermore, the interpretation of the estimates can also be complex, especially for certain link functions.

These challenges underscore the importance of careful model selection, diagnostics, and validation when using MLE to estimate parameters in GLMs.

**show the mathematical derivation of the hessian and score function of bernoulli distribution using GLMs. step by step**

Sure, here's a step-by-step derivation of the score function and Hessian for a GLM with a Bernoulli response variable and a logit link function. This is a common setup for logistic regression.

Suppose we have a binary response variable Y, where Y\_i = 1 with probability p\_i and Y\_i = 0 with probability 1 - p\_i. We have a predictor vector X\_i for each observation i, and we want to model log(p\_i / (1 - p\_i)) = β^T X\_i. This is a GLM with a Bernoulli distribution and a logit link.

**Score Function:**

The score function U(β) is the derivative of the log-likelihood with respect to β.

For a single observation, the likelihood is p\_i^Y\_i \* (1 - p\_i)^(1 - Y\_i). Hence, the log-likelihood is l(β) = Y\_i log(p\_i) + (1 - Y\_i) log(1 - p\_i).

The derivative of l(β) with respect to β is given by:

scssCopy code

U(β) = ∂l(β) / ∂β = (Y\_i - p\_i) X\_i.

This can be derived by first differentiating l(β) with respect to p\_i, then using the chain rule to differentiate p\_i with respect to β.

The score function for the entire dataset is then the sum of the individual score functions: U(β) = Σ (Y\_i - p\_i) X\_i.

**Hessian:**

The Hessian is the matrix of second derivatives of the log-likelihood. For our model, the Hessian for a single observation is given by:

scssCopy code

H(β) = ∂²l(β) / ∂β ∂β^T = - p\_i (1 - p\_i) X\_i X\_i^T.

Again, this can be derived by first differentiating U(β) with respect to p\_i, then using the chain rule to differentiate p\_i with respect to β.

The Hessian for the entire dataset is the sum of the individual Hessians: H(β) = - Σ p\_i (1 - p\_i) X\_i X\_i^T.

In practice, these derivatives would be computed and summed over all observations in the dataset to find the MLE of β. These computations are often done using software, as they can be complex and computationally intensive.

**Derive step by step the analytical solution for the poisson distribution**

The Poisson distribution is commonly used to model the number of events (such as the number of phone calls at a call center) in a fixed period of time. Here, we will derive the maximum likelihood estimate for the parameter λ of a Poisson distribution.

Let's say you have a random sample X\_1, X\_2, ..., X\_n from a Poisson distribution with parameter λ (the average rate of events). The probability mass function of a Poisson distribution is given by:

P(X = k) = (λ^k \* e^-λ) / k!, for k = 0, 1, 2, ...

The likelihood function of the sample is given by the product of the probabilities of the individual observations:

L(λ) = Π (λ^X\_i \* e^-λ) / X\_i!, for i = 1 to n.

It's easier to work with the natural logarithm of the likelihood function, which is the log-likelihood function:

l(λ) = log(L(λ)) = Σ [X\_i \* log(λ) - λ - log(X\_i!)], for i = 1 to n.

The maximum likelihood estimate (MLE) for λ is found by taking the derivative of the log-likelihood function with respect to λ, setting it to zero, and solving for λ. The derivative is:

dl/dλ = Σ [X\_i/λ - 1], for i = 1 to n.

Setting this equal to zero gives:

Σ X\_i/λ - n = 0.

Solving for λ gives the maximum likelihood estimate:

λ̂ = (Σ X\_i) / n.

This is the sample mean of the observations, which makes intuitive sense: the best estimate for the average rate of events is simply the observed average rate in the sample.

**Show step by step why we can't find the analytical solution of the poisson distribution but can only use the numerical solution.**

The Poisson distribution is used to model the number of events (such as the number of phone calls to a call center) in a fixed period of time. The Poisson distribution has one parameter, λ, which represents the average rate of events. As we have seen in the previous steps, the maximum likelihood estimate for λ is the sample mean of the observations, which is an analytical solution.

However, in practice, especially in the context of Generalized Linear Models (GLMs) where the Poisson distribution is often used with a log link function, the likelihood equations often cannot be solved analytically for the parameters. This is due to the fact that the likelihood equations involve both the parameters of the linear predictor (the β coefficients in GLMs) and the exponential function (due to the log link function), leading to a set of nonlinear equations.

Let's consider a simple Poisson regression model:

y ~ Poisson(λ) log(λ) = Xβ

The likelihood function for this model is:

L(β) = Π [ (e^(X\_i β))^(y\_i) \* exp(-e^(X\_i β)) ] / y\_i! , for i = 1 to n.

Taking the logarithm, the log-likelihood function is:

l(β) = Σ [ y\_i X\_i β - e^(X\_i β) - log(y\_i!) ], for i = 1 to n.

To find the MLE, we take the derivative of l with respect to β and set it equal to zero:

dl/dβ = Σ X\_i (y\_i - e^(X\_i β)) = 0.

This equation involves both the β's and the exponential function of the β's, and generally cannot be solved analytically for β. Therefore, a numerical method such as Newton-Raphson or Iteratively Reweighted Least Squares (IRLS) is used to find the MLE of β.

The need for a numerical solution arises from the fact that the link function (the log function in this case) introduces a nonlinearity into the likelihood equations, making it impossible to solve them directly for the parameters. However, these numerical methods are well-established and implemented in statistical software, so it is straightforward to fit a Poisson regression model in practice.

**Fisher scoring methods**

Fisher scoring is an iterative method used to estimate the parameters of a statistical model when the maximum likelihood estimates cannot be found directly. It's named after Sir Ronald Fisher, a pioneering statistician of the 20th century. Fisher scoring is very similar to the Newton-Raphson method, but instead of using the Hessian matrix (second derivatives of the log-likelihood function), it uses the expected or average information matrix.

Here's a broad overview of the Fisher scoring method:

1. **Initialization**: Start with an initial estimate of the parameters. This might be based on a simpler model, a method of moments estimator, or it could be arbitrary.
2. **Scoring Step**: Compute the score vector, which is the gradient (first derivative) of the log-likelihood function. This measures how steeply the log-likelihood is changing at the current estimate.
3. **Information Step**: Compute the Fisher information matrix, which is the expected value of the negative second derivative of the log-likelihood. This measures how quickly the log-likelihood function is curving at the current estimate.
4. **Update Step**: Update the parameter estimates by adding the inverse of the information matrix times the score vector to the current parameter estimates. This step is designed to move the parameters in the direction that most increases the likelihood, scaled by the curvature of the log-likelihood.
5. **Convergence Check**: Check if the estimates have converged. This can be done by checking if the score vector is close enough to zero, or if the parameter estimates have stopped changing significantly. If the estimates have not converged, return to step 2 with the updated estimates.

The Fisher scoring method has some nice theoretical properties. In particular, under certain regularity conditions, the Fisher scoring method is guaranteed to converge to the maximum likelihood estimates, and these estimates are asymptotically normal and efficient.

However, in practice, Fisher scoring can sometimes fail to converge, especially for poorly chosen initial estimates or for complex models with many parameters. When Fisher scoring fails to converge, other methods such as the Newton-Raphson method or the Expectation-Maximization (EM) algorithm may be used.

It's worth noting that for Generalized Linear Models (GLMs), the Fisher scoring method is equivalent to the Iteratively Reweighted Least Squares (IRLS) method, which is the standard method used to fit GLMs.

**Generalized Estimating Equations**

Generalized Estimating Equations (GEEs) are a technique used to estimate the parameters of a generalized linear model with a possible correlation among outcomes. GEEs, introduced by Liang and Zeger (1986), are an extension of Generalized Linear Models (GLMs) and have broad applicability, particularly in cases where data are clustered or there are repeated measures.

In GLMs, we assume that observations are independent. However, this assumption does not hold in many cases. For example, when we have repeated measures on a subject over time, or in clustered sampling, observations from the same cluster are likely to be more similar to each other than observations from different clusters.

GEEs come into play in these situations. GEEs use the concept of quasi-likelihood and a robust "sandwich" estimator of the variance to estimate the parameters of the model, allowing for the within-cluster or within-subject correlation.

Here's a broad overview of the GEE method:

1. **Specification of the Marginal Mean Model**: Just like in GLMs, we specify a systematic component (linear predictor) and a link function which connects the expected value of the response variable to the systematic component.
2. **Specification of the Working Correlation Matrix**: This is a matrix that describes the structure of the correlation between observations within the same cluster. It's called a "working" matrix because it does not need to be correctly specified in order to get consistent parameter estimates.
3. **Estimation of the Parameters**: Using an iterative process, the parameters of the mean model and the correlation structure are estimated. The parameters are typically estimated using the method of quasi-likelihood, and the correlation structure is estimated using the method of moments.
4. **Robust Variance Estimation**: The standard errors of the parameter estimates are adjusted for the within-cluster correlation using a robust "sandwich" estimator. This ensures that even if the working correlation matrix is misspecified, we still get valid standard errors.

One of the key advantages of GEEs is that they provide 'population-averaged' effects, as opposed to 'subject-specific' effects estimated by other approaches such as mixed models. This makes GEEs particularly useful when we are interested in estimating the effect of a treatment or exposure on a population, rather than on specific subjects.

However, GEEs have some limitations. They require large sample sizes to perform well, and the choice of the working correlation structure can have a big impact on the efficiency of the estimates. Also, while GEEs provide robust standard errors, they do not provide a full model of the correlation structure, which can limit their usefulness in some situations.

**How do you compare models in the case of GLMs with links**

Comparing models in the case of Generalized Linear Models (GLMs) with link functions can be done in several ways. Some of the most commonly used methods include:

1. **Akaike's Information Criterion (AIC)**: This is a measure of the relative quality of a statistical model for a given set of data. It provides a means for model selection by comparing models while taking into account the complexity of the model. A model with a lower AIC is typically preferred over a model with a higher AIC.
2. **Bayesian Information Criterion (BIC)**: Similar to AIC, BIC also penalizes model complexity but does so more heavily than AIC. When comparing models, the model with the lowest BIC is typically preferred.
3. **Deviance**: This is a measure of the goodness of fit of a statistical model. In the context of GLMs, the deviance is used to compare the fit of different models. A smaller deviance indicates a better fit to the data.
4. **Likelihood Ratio Tests**: These tests compare the likelihood of the data under two models - a simpler (null) model and a more complex (alternative) model. The test statistic follows a chi-squared distribution under the null hypothesis. If the p-value associated with this test statistic is small, it provides evidence against the null model in favor of the alternative model.
5. **Cross-Validation**: This involves partitioning the data into a training set and a validation set. The model is fit on the training data and its predictive accuracy is assessed on the validation data. This can be particularly useful when the goal is prediction.

It's important to note that these measures can only guide the model selection process and do not provide definitive answers. The best model depends not only on these statistical measures but also on the purpose of the model, the understanding of the underlying process, and the quality and quantity of the data.