Logistic Regression: Used for binary or multinomial outcomes (nominal or ordinal).

Linear Regression Review:

* Example Interpretation of Y intercept: average birthweight of a mother who is 0 years old is -1163 grams
* Least squares estimate:
* Example interpretation of slope: for every 1 year increase in age, birthweight increases by (slope) grams on average

Interested in testing association between predictor and outcome:

Assumptions of Linear Regression:

* Specification of the correct mean function. Make sure your data actually exhibit a linear relationship
* Homoscedasticity: equal variance
* For each value of the predictor X, the distribution of the response (or errors) is normally distributed (important for small samples)
* Independent observations

Logistic Regression Model:

* For every unit increase in our “outcome” increases by

Interpretation of Y-int.: the log odds of disease for an unexposed person

* **Can only interpret for cohort or cross-sectional studies**

Interpretation of Beta: the log of the odds ratio of disease comparing the exp group to the unexp group

Logistic Regression with One k-level Categorical Predictor:

* **For a k-level categorical predictor, we need to create k – 1 dummy variables**

1. Decide which level you wish to be the reference level. This is usually the lowest level of the level that corresponds to being unexposed
2. Create k – 1 dummy variables such that each of the remaining k-1 categories are compared to the reference group.

How can we compare low dose to high dose in example model?

Logistic Regression with One Continuous Predictor:

* Interpret alpha example: log odds of disease for those aged 0
* Interpret beta example: for every 1 unit increase in “X” out outcome (log odds of disease) increases by
* Increasing the units of continuous predictor, for example by 10 gives you:
  + For every 10 unit increase in your odds if getting disease increase by \_\_%

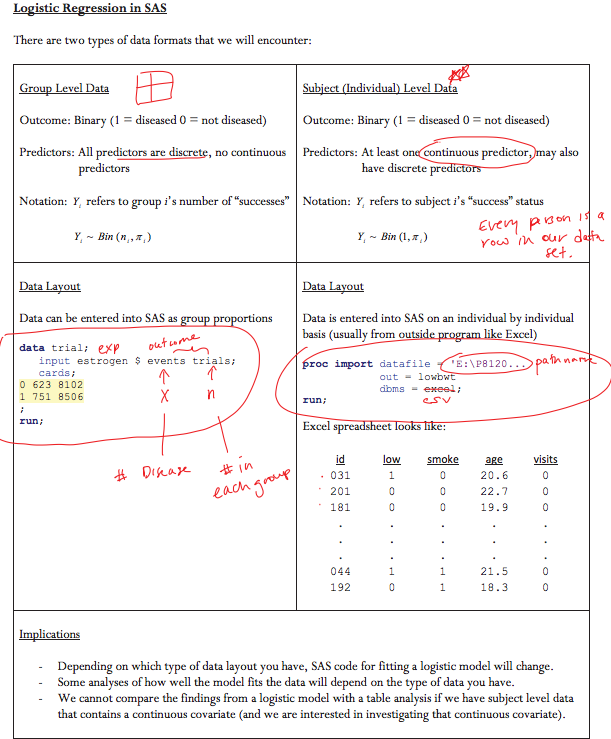
Logistic Regression w/ Multiple Predictors:

Ex: model probability of disease with a binary exposure variable and a 3-level categorical predictor age.

Need to make dummy variables for age:

Let

Regression model:



\*\* IN SAS remember to make sure the convergence status is satisfied

Confidence Intervals for Parameters in Logistic Regression:

* z= 1.96 and the beta & s.e. are pulled from SAS
* 🡪 C.I. for OR

Assumptions for C.I. and Hypothesis Tests in LR:

1. independent observations (no repeated measurements on same people, no one is related etc.)
2. Correctly specified the model
   1. All important variables are in the model
   2. You are using the correct functional form for each variable
3. Sufficiently large sample size (at least 10 events per covariate in the model)

Wald Tests: used to test a single parameter or several at once

* **To test a single parameter**:
* There are 2 test statistics we can use:
* **To test multiple parameters:**
* \*\* USE TYPE 3 ANALYSIS IN SAS

1. : the true log odds ratio describing the association between mother’s smoking status and baby’s low birthweight status.
2. Set =0.05
3. Assumptions: independent observations, correctly specified the model, sufficiently large sample size
4. Use a Wald Test:

Test Statistic:

p-value=

1. Decision:
2. Conclusion: We have sufficient evidence in these data to suggest that smoking and birth weight status are associated at the 5% significance level.

Likelihood Ratio Tests:

* The numerator is always less than or equal to the denominator
* When the numerator is much less than the denominator, there is a larger probability of observing the data that we actually observed when the parameters are not restricted to H0. Therefore, this gives evidence against H0.

LRT Statistic:

Degrees of freedom: )

1. State the true log odds ratios you are testing
2. Set =0.05
3. Assumptions: independent observations, correctly specified the model, sufficiently large sample size
4. Use a LRT: under H0

Test statistic :

Critical value = (look up on X^2 table)

p-value:

1. Decision:
2. Conclusion: these data do not provide sufficient evidence to suggest that the number of visits is associated with birth weight status at the 5% significance level, after accounting for age and smoking status

Wald Test vs. Likelihood Ratio Test:

* Likelihood ratio test is more reliable and more powerful than the Wald test
* When testing one single variable it is more common to use a Wald test
* You must use the LRT when you wish to test the significance of groups of variables

1. Interaction is present when stratum-specific estimates of association **differ from one another.** We want to describe how the effect changes across strata. We can think of interaction as “a finding to be reported”
   1. When interaction present, focus on stratum specific ORs.
   2. The parameters for the main effects take on different meanings and should not be interpreted in the same way as originally
2. Confounding is present when stratum specific estimates of association differ from the crude estimate (but not from each other). This is **bias**. We want to remove the effect of the confounder by adjusting for it in our model.

Modeling Interaction: effect modification or interaction occurs when an effect is thought to vary across strata. The relationship of interest is different at different levels

* Marginal, pooled, or crude tables include the unstratified relationship
* Stratified tables are conditional or partial tables

Ex:

Formal way to test interaction term:

1. N/A
2. vs

\*Where is the interaction

1. Set
2. Assumptions: Same as Wald, and at least 10 cases and 10 controls (or w/ and w/out exposure) per variable
3. Test Statistic:

= 4.093

1. P-value =

Reject if P<.05

1. At the 5% level of significance, we have sufficient evidence to conclude that the association between smoking and cancer depends on (or varies by) number of partners.

*\*If the interaction is significant, report the stratum specific ORs and 95% CIs.*

*\*When there is an interaction term in the model, it is not meaningful to interpret/test the coefficients for the main effects*

*\*Testing for an interaction requires more power than testing for an association*

*\*If the interaction term is not significant, we should remove it from the model*

When comparing crude estimate to the adjusted estimate, there is confounding when the difference is greater than 10%