***ODDS Ratio***

[***https://stats.stackexchange.com/questions/133623/help-me-understand-adjusted-odds-ratio-in-logistic-regression***](https://stats.stackexchange.com/questions/133623/help-me-understand-adjusted-odds-ratio-in-logistic-regression)

***Odds* are a way to express chances. *Odds ratios* are just that: one odds divided by another.**That means an odds ratio is what you multiply one odds by to produce another. Let's see how they work in this common situation.

Converting between odds and probability

The odds of a binary response YY are the ratio of the chance it happens (coded with 11), written Pr(Y=1)Pr(Y=1), to the chance it does not (coded with 00), written Pr(Y=0)Pr(Y=0):

Odds(Y)=Pr(Y=1)Pr(Y=0)=Pr(Y=1)1−Pr(Y=1).Odds(Y)=Pr(Y=1)Pr(Y=0)=Pr(Y=1)1−Pr(Y=1).

The equivalent expression on the right shows it suffices to model Pr(Y=1)Pr(Y=1) to find the odds. Conversely, note that we can solve

Pr(Y=1)=Odds(Y)1+Odds(Y)=1−11+Odds(Y).Pr(Y=1)=Odds(Y)1+Odds(Y)=1−11+Odds(Y).

Logistic regression

Logistic regression models the *logarithm* of the odds of YY as a linear function of explanatory variables. Most generally, writing these variables as x1,…,xpx1,…,xp, and including a possible constant term in the linear function, we may name the coefficients (which are to be estimated from the data) as β1,…,βpβ1,…,βp and β0β0. Formally this produces the model

log(Odds(Y))=β0+β1x1+⋯+βpxp.log⁡(Odds(Y))=β0+β1x1+⋯+βpxp.

The odds themselves can be recovered by undoing the logarithm:

Odds(Y)=exp(β0+β1x1+⋯+βpxp).Odds(Y)=exp⁡(β0+β1x1+⋯+βpxp).

Using categorical variables

Categorical variables, such as age group, gender, presence of Glaucoma, *etc.*, are incorporated by means of "dummy coding." To show that how the variable is coded does not matter, I will provide a simple example of one small group; its generalization to multiple groups should be obvious. In this study one variable is "pupil size," with three categories, "Large", "Medium", and "Small". (The study treats these as purely categorical, apparently paying no attention to their inherent order.) Intuitively, each category has its own odds, say αLαL for "Large", αMαM for "Medium", and αSαS for "Small". This means that, all other things equal,

Odds(Y)=exp(αL+β0+β1x1+⋯+βpxp)Odds(Y)=exp⁡(αL+β0+β1x1+⋯+βpxp)

for anybody in the "Large" category,

Odds(Y)=exp(αM+β0+β1x1+⋯+βpxp)Odds(Y)=exp⁡(αM+β0+β1x1+⋯+βpxp)

for anybody in the "Medium" category, and

Odds(Y)=exp(αS+β0+β1x1+⋯+βpxp)Odds(Y)=exp⁡(αS+β0+β1x1+⋯+βpxp)

for those in the "Small" category.

Creating identifiable coefficients

I have colored the first two coefficients to highlight them, because I want you to notice that they allow a simple change to occur: we could pick any number γγ and, by adding it to β0β0 and subtracting it from each of αLαL, αMαM, and αSαS, *we would not change any predicted odds.* This is because of the obvious equivalences of the form

αL+β0=(αL−γ)+(γ+β0),αL+β0=(αL−γ)+(γ+β0),

*etc.* Although this presents no problems for the model--it still predicts exactly the same things--it shows that the parameters are not in themselves interpretable. What stays the same when we do this addition-subtraction maneuver are the *differences* between the coefficients. Conventionally, to address this *lack of identifiability,* people (and by default, software) choose one of the categories in each variable as the "base" or "reference" and simply stipulate that its coefficient will be zero. This removes the ambiguity.

The paper lists reference categories first; "Large" in this case. Thus, αLαL is subtracted from each of αL,αM,αL,αM, and αSαS, and added to β0β0 to compensate.

The log odds for a hypothetical individual falling into all the base categories therefore equals β0β0plus a bunch of terms associated with all other "covariates"--the non-categorical variables:

Odds(Base category)=exp(β0+β1X1+⋯+βpXp).Odds(Base category)=exp⁡(β0+β1X1+⋯+βpXp).

*No* terms associated with any categorical variables appear here. (I have slightly changed the notation at this point: the betas βiβi now are the coefficients only of the *covariates*, while the full model includes the alphas αjαj for the various categories.)

Comparing odds

Let us compare odds. Suppose a hypothetical individual is a

male patient aged 80–89 with a white cataract, no fundal view, and a small pupil being operated on by a specialist registrar, ...

Associated with this patient (let's call him Charlie) are estimated coefficients for each category: α80-89α80-89 for his age group, αmaleαmale for being male, and so on. Wherever his attribute is the base for its category, the coefficient is zero *by convention*, as we have seen. Because this is a linear model, *the coefficients add.* Thus, to the base log odds given above, the log odds for this patient are obtained by adding in

α80-89+αmale+αno Glaucoma+⋯+αspecialist registrar.α80-89+αmale+αno Glaucoma+⋯+αspecialist registrar.

This is precisely the amount by which the log odds of this patient vary from the base. To convert from log odds, undo the logarithm and recall that this turns addition into multiplication. Therefore, the base odds must be multiplied by

exp(α80-89)exp(αmale)exp(αno Glaucoma)⋯exp(αspecialist registrar).exp⁡(α80-89)exp⁡(αmale)exp⁡(αno Glaucoma)⋯exp⁡(αspecialist registrar).

These are the numbers given in the table under "Adjusted OR" (adjusted odds ratio). (It is called "adjusted" because covariates x1,…,xpx1,…,xp were included in the model. They play no role in any of our calculations, as you will see. It is called a "ratio" because it is precisely the amount by which the base odds must be multiplied to produce the patient's predicted odds: see the first paragraph of this post.) In order in the table, they are exp(α80-89)=1.58exp⁡(α80-89)=1.58, exp(αmale)=1.28exp⁡(αmale)=1.28, exp(αno Glaucoma)=1.00exp⁡(αno Glaucoma)=1.00, and so on. According to the article, their product works out to 34.534.5. Therefore

Odds(Charlie)=34.5×Odds(Base).Odds(Charlie)=34.5×Odds(Base).

(Notice that the base categories all have odds ratios of 1.00=exp(0)1.00=exp⁡(0), because including 11 in the product leaves it unchanged. That's how you can spot the base categories in the table.)

Restating the results as probabilities

Finally, let us convert this result to probabilities. We were told the baseline predicted probability is 0.736%=0.007360.736%=0.00736. Therefore, using the formulas relating odds and probabilities derived at the outset, we may compute

Odds(Base)=0.007361−0.00736=0.00741.Odds(Base)=0.007361−0.00736=0.00741.

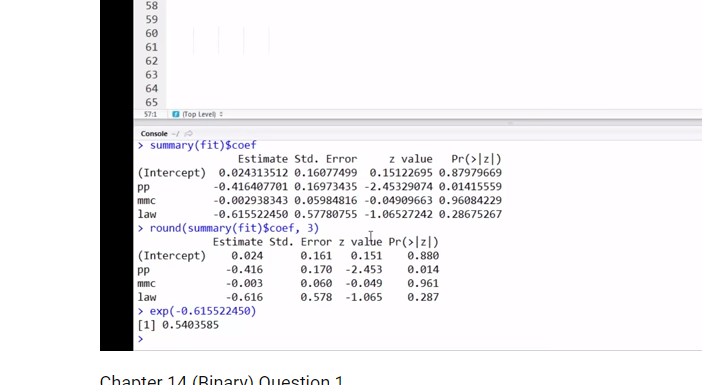
Consequently Charlie's odds are

Odds(Charlie)=34.5×0.00741=0.256.Odds(Charlie)=34.5×0.00741=0.256.

Finally, converting this back to probabilities gives

Pr(Y(Charlie)=1)=1−11+0.256=0.204.

<https://www.youtube.com/watch?v=CXWZqzKdkp4&list=PLpl-gQkQivXji7JK1OP1qS7zalwUBPrX0&index=52>



Glm(y ~ x, family = binmomial)

Odds ratio comparing 1 numerator (law enacted) , 0 demonm(no law) = 0.54

Or

1 – 0.54 – 46 % decrease in odds of drivers being kill after law enacted, holding others constant