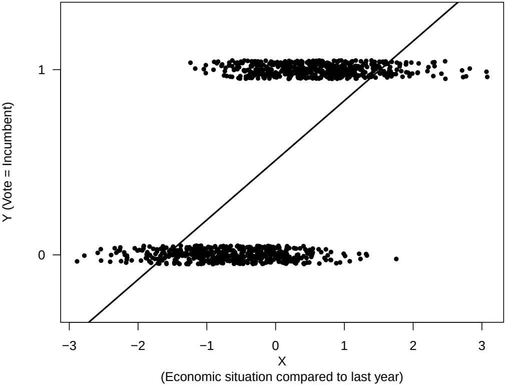
Intro to GLM – Day 3: Quantities of interest

Federico Vegetti
Central European University

ECPR Summer School in Methods and Techniques
10 August 2016

Reporting the model results

- Let's recall the LPM.



- Where $\beta_0 = 0.51$ and $\beta_1 = 0.32$.
- What do these numbers mean?

LPM vs Logit

LPM

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.51057    0.01223   41.73   <2e-16 ***
X            0.32185    0.01240   25.95   <2e-16 ***
```

Logit

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.07675    0.08449   0.908    0.364
X            2.25346    0.14165  15.908   <2e-16 ***
```

- Where:
  - `exp(0.07675) = 1.079772`
  - `exp(2.25346) = 9.52062`.
- What do these numbers mean?

Odds

- The odds are a ratio of the probability that $y_i = 1$ over the probability that $y_i = 0$.
  - When we have probability $p = 0.5$, then `0.5/0.5 = 1`. The odds are 1 to 1.
  - If we apply for a job where we have 80% probability of success, then `0.8/0.2 = 4`. The odds are 4 to 1: the chance of success is 4 time bigger than the chance of failure.
- Recall:

$$logit(\pi) = log\left(\frac{\pi}{1-\pi}\right) = X\beta$$

- Odds are what we obtain if we exponentiate the coefficients in a logistic regression model output.

Odds

- The odds are a ratio of the probability that $y_i = 1$ over the probability that $y_i = 0$.
  - When we have probability $p = 0.5$, then `0.5/0.5 = 1`. The odds are 1 to 1.
  - If we apply for a job where we have 80% probability of success, then `0.8/0.2 = 4`. The odds are 4 to 1: the chance of success is 4 time bigger than the chance of failure.
- Recall:

$$logit(\pi) = log\left(\frac{\pi}{1-\pi}\right) = X\beta$$

- Odds are what we obtain if we exponentiate the coefficients in a logistic regression model output.
- Odds of *what* against *what*?
- What do the odds expressed by the coefficient of `X` mean?

Odds ratios

- Let's consider a variable Y measuring on a population of 500 students whether they passed an English language test (1) or not (0).

Y=0	Y=1
147	353

- Here `353/147 = 2.40` means that the odds to pass the test are about 2.40 to 1.
- If we run a logit regression with intercept only, we get

```
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.87604    0.09816   8.924   <2e-16 ***
```

- This makes sense since `log(353/147) = 0.8760355`.

Odds ratios – dummy variables

- Now let's consider a dummy variable Z recording whether the students attended the English conversation sessions offered by the student union (1) or not (0).

	Y=0	Y=1	Total
Z=0	111	204	315
Z=1	36	149	185
Total	147	353	500

- Here, the odds of $Y = 1$ are:
  - `204/111 = 1.837838` when $Z = 0$.
  - `149/36 = 4.138889` when $Z = 1$.
- While the *odds ratio* of passing the test $(Y = 1)$ for those who went to the conversation sessions $(Z = 1)$ in respect to those who didn't $(Z = 0)$ is `(149/36)/(204/111) = 2.25`.
- In other words, having some English conversation makes it 2.25 times more likely to pass the language test.

Odds ratios – dummy variables (2)

- If we run a logit of `Y` on `Z` we get

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)     0.6086     0.1179    5.16 2.47e-07 ***
Z               0.8118     0.2200    3.69 0.000224 ***
```

- Here the intercept
  - `exp(0.6086) = 1.84` refers to the odds that $Y = 1$ against $Y = 0$ <u>when $Z = 0$</u>.
  - This means that when $Z = 0$, the probability that $Y = 1$ is about 84% higher then the probability of $Y = 0$.
- And the slope coefficient
  - `exp(0.8118) = 2.25` refers to the odds ratio of $Y = 1$ (against $Y = 0$) <u>between $Z = 1$ and $Z = 0$</u>.
  - As we already saw, the odds that $Y = 1$ against $Y = 0$ is about 125% bigger when $Z = 1$ than when $Z = 0$.

Odds ratios – continuous variables

- Further, let's look at the effect of students' standardized score on an "extrovert personality" test, `X` ($\mu = 0.04$; $\sigma = 0.95$).

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.1768     0.1278   9.206   <2e-16 ***
X             1.5834     0.1639   9.662   <2e-16 ***
```

- Here the intercept refers to the odds that $Y = 1$ against $Y = 0$ when $X = 0$, so then `exp(1.1768) = 3.24`.
- The slope coefficient refers to the difference in log-odds for one unit increase of `X`.
  - `exp(1.5834) = 4.87` means that every unit increase of `X` multiplies the odds that $Y = 1$ by a factor of 4.9.
  - For instance, the odds when $X = 1$ are `exp(1.1768 + 1.5834*1) = 15.8`, meaning that students who are 1 SD more extroverted than the average are about 16 times more likely to pass the test than the average.
  - When $X = 2$ the odds are `exp(1.1768 + 1.5834*2) = 76.98`, meaning that students who are 2 SD more extroverted than the average are about 77 times more likely to pass the test.
  - Note that `log(76.98/15.8) = 1.58`.

Odds ratios – interactions

- Let's consider a full interaction model of `Y` on `X`, `Z` and `X*Z`.

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)     0.7787     0.1417   5.497 3.86e-08 ***
X               1.3745     0.1813   7.582 3.39e-14 ***
Z               1.6502     0.3894   4.238 2.26e-05 ***
X:Z             1.2022     0.4831   2.488   0.0128 *
```

- Here we basically have two equations, one for $Z = 0$ and one for $Z = 1$.
- The odds ratio of $Z = 1$ versus $Z = 0$ are `exp(1.6502) = 5.21`. This ratio applies only when $X = 0$.
- In other words, for an average-extroverted student, the mere fact of attending some conversation sessions makes it 5 times more likely to pass the English language test.

Odds ratios – interactions (2)

- What about the effect of extroversion for those who did and did not attend the conversation sessions?
- The odds ratios of 1 point increase of `X` are
  - `exp(1.3745) = 3.95` when $Z = 0$.
  - `exp(1.3745 + 1.2022) = 13.15` when $Z = 1$.
- Hence, students who are 1 SD more extroverted than the average are about 4 times more likely to pass the test even if they did *not* attend the conversation.
- However, the same type of students who went to the conversation sessions are 13 times more likely to pass the test *than the average-extroverted student*.
- Finally, the ratio between these two odds ratios `13.15366/3.9531 = 3.33` turns out to be the exponentiated coefficient of the interaction term: `exp(1.2022) = 3.33`
- This means that, *among the more extroverted students*, those who attended the sessions are 3.3 times more likely to pass the test.

Reporting quantities of interest

- To talk in terms of odds ratios can be frustrating, next to being difficult for the reader.
- This becomes more problematic the more our model gets complex.
  - When we include interaction effects in the model, interpreting the coefficients in terms of odds ratio becomes cumbersome.
- Moreover, even without interactions, coefficients in logit models can't be interpreted as unconditional marginal effects: they depend on the position of the predictors.
- Finally, the non-linearity of the logit transformation makes it tricky to present quantities that help the reader understand the magnitude of the phenomenon that we are observing.
  - To talk about "one point increase" may be inappropriate, as it depends on <u>where</u> that increase happens.
- Better to present quantities of interest.

Predicted probabilities

- Let's consider the same model we saw, just without interaction.

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)     0.8411     0.1471   5.719 1.07e-08 ***
X               1.6330     0.1683   9.702  < 2e-16 ***
Z               1.0592     0.2616   4.049 5.14e-05 ***
```

- We want to know how the probability that $Y = 1$ changes as `X` goes from -2 to $+2$.
- To transform our coefficients into probabilities we need to use the inverse logit function:

$$\pi = \frac{exp(X\beta)}{1 + exp(X\beta)}$$

- Which sometimes is written as:

$$\pi = \frac{1}{1 + exp(-X\beta)}$$

Predicted probabilities – bivariate

- Given our output, when `X = -2` we have

$$P(Y|X=-2) = \frac{exp(0.8411 - 2 * 1.6330)}{1 + exp(0.8411 - 2 * 1.6330)} = 0.08$$

- When `X = 0` we have

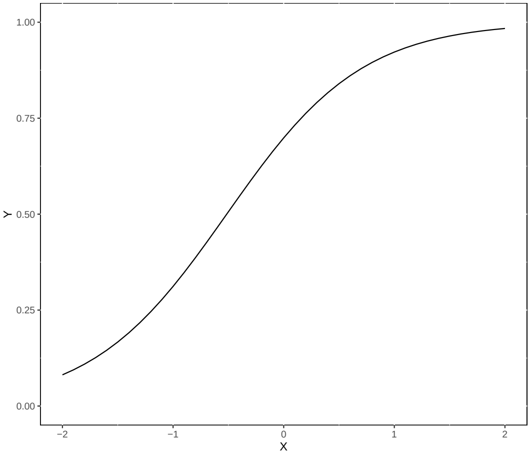
$$P(Y|X=0) = \frac{exp(0.8411)}{1 + exp(0.8411)} = 0.699$$

- And when `X = +2` we have

$$P(Y|X=2) = \frac{exp(0.8411 + 2 * 1.6330)}{1 + exp(0.8411 + 2 * 1.6330)} = 0.98$$

- Notice the non-linearity: one increase of two points from `-2` to `0` produced a change in probability of `0.62`, while an increase of the same magnitude from `0` to `+2` produced a change in probability of `0.28`.

Predicted probabilities – bivariate (2)



Predicted probabilities – multivariate

- What if we take `Z` into account? For `X = -2` we have

$$P(Y|X=-2, Z=0) = \frac{exp(0.8411 - 2*1.6330)}{1 + exp(0.8411 - 2*1.6330)} = 0.08$$

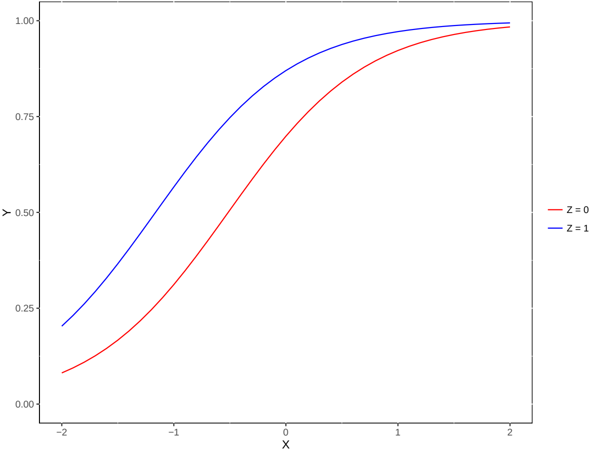
$$P(Y|X=-2, Z=1) = \frac{exp(0.8411 - 2*1.6330 + 1.0592)}{1 + exp(0.8411 - 2*1.6330 + 1.0592)} = 0.20$$

- For `X = +2` we have

$$P(Y|X=2, Z=0) = \frac{exp(0.8411 + 2*1.6330)}{1 + exp(0.8411 + 2*1.6330)} = 0.98$$

$$P(Y|X=2, Z=1) = \frac{exp(0.8411 + 2*1.6330 + 1.0592)}{1 + exp(0.8411 + 2*1.6330 + 1.0592)} = 0.99$$

Predicted probabilities – multivariate (2)



Predicted probabilities – interactions

- Let's consider now the model with the interaction `X*Z` that we already saw.

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)     0.7787     0.1417   5.497 3.86e-08 ***
X               1.3745     0.1813   7.582 3.39e-14 ***
Z               1.6502     0.3894   4.238 2.26e-05 ***
X:Z             1.2022     0.4831   2.488   0.0128 *
```

Predicted probabilities – interactions (2)

- For `X = -2` we have

$$P(Y|X=-2, Z=0) = \frac{exp(0.7787 - 2 * 1.3745)}{1 + exp(0.7787 - 2 * 1.3745)} = 0.12$$

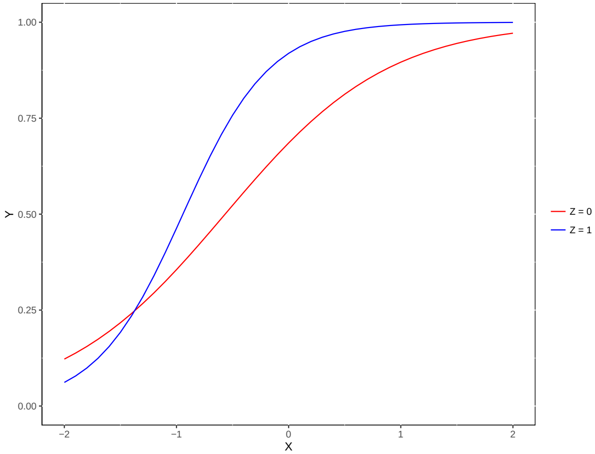
$$P(Y|X=-2, Z=1) = \frac{exp(0.7787 - 2 * 1.3745 + 1.6502 - 2 * 1.2022)}{1 + exp(0.7787 - 2 * 1.3745 + 1.6502 - 2 * 1.2022)} = 0.06$$

- For `X = +2` we have

$$P(Y|X=2, Z=0) = \frac{exp(0.7787 + 2 * 1.3745)}{1 + exp(0.7787 + 2 * 1.3745)} = 0.97$$

$$P(Y|X=2, Z=1) = \frac{exp(0.7787 + 2 * 1.3745 + 1.6502 + 2 * 1.2022)}{1 + exp(0.7787 + 2 * 1.3745 + 1.6502 + 2 * 1.2022)} = 0.999$$

Predicted probabilities – interactions (3)



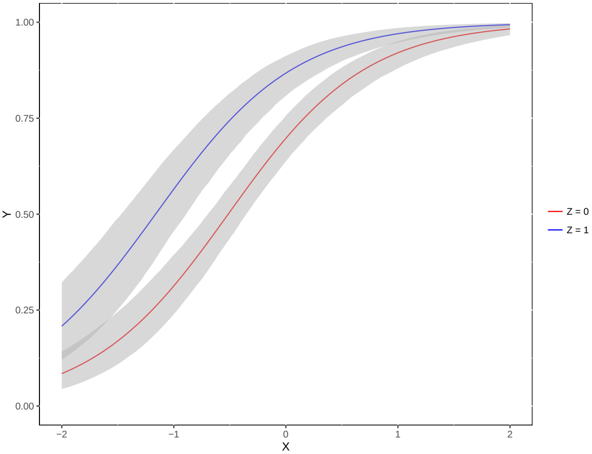
Confidence Intervals

- To report our results in a compelling way, we need to report also the uncertainty of our estimates.
- Recall how standard errors are found in the ML framework:
  - We have the matrix of second partial derivatives, called "Hessian".
  - The inverse is the variance/covariance matrix of our estimates.
  - Fortunately, `R` extracts this information for us via the `vcov()` function.
- Because the standard errors are on the same scale of the predictors, we can use them to add confidence intervals (CIs) to our odds ratios.
  - For instance, the coefficient of `Z` in our first model was `0.8118` with standard error `0.22`.
  - Thus, as we saw, the odds ratio of $Y = 1$ between $Z = 1$ and $Z = 0$ is `exp(0.8118) = 2.25`.
  - Moreover, its confidence interval goes from `exp(0.8118-1.96*0.22) = 1.46`, to `exp(0.8118+1.96*0.22) = 3.47`.

Confidence Intervals (2)

- One way to get CIs for our predicted probabilities is to simulate a distribution of values based on the means of our coefficients (the point estimates) and the variance/covariance matrix.
  - This method is often employed when conditional effects are involved, as it was invoked by Brambor et al. (2006).
- An alternative is to bootstrap.
  - Bootstrap means, you sample from our data (with replacement), run the model, calculate predicted probabilities, store them, do the same again and again and again.
  - As a result, you'll have a distribution of quantities of interest, and you can choose the interval to display.
  - Bootstrap is somewhat more conservative than the simulation-based approach. It is more accurate in some cases, for instance when you have outliers that might drive your results.

Predicted probabilities with CIs



Predicted probabilities with CIs (2)