Contents

1 Logistic Regression

In its easiest form the logistic regression model permits the modeling of a binary response variable in dependency on a selection of explanatory variables.

- Random component: Binomial distribution of the response variable
- Systematic component: Linear predictor
- Link function: Logit function

1.1 Interpretation

For the interpretation again consider first the case of one numerical explanatory variable, x, and let $\pi(x)$ be the success probability at value x.
Then

$$Logit(\pi(x)) = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x \qquad (1)$$

Which implies that for every increase by one unit in x the log odds increase by β. But what does that mean?

Equation (1) is equivalent to

$$\frac{\pi(x)}{1-\pi(x)} = \exp(\alpha + \beta x) = e^{\alpha}(e^{\beta})^{x} \qquad (2)$$

Therefore the odds are multiplied by e^{β} for every one unit increase in x. For example $e^{\beta} = 1.5$ means that the odds increase by 50% for every increase by one unit in x.

Further, equations (1) and (2) are equivalent to

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \qquad (3)$$

It is obvious from all three equations that β determines the rate of increase in $\pi(x)$, but the increase is not the same for all x, since the relation ship between $\pi(x)$ and x is not linear (straight line), but S-shaped.
The steepest slope occurs at x for which $\pi(x) = 0.5$, this is when $x = -\alpha/\beta$. Therefore it is called the median effective level

$$EL_{50} = -\alpha/\beta.$$

The median effective level is the value of x, where the probability for success equals 0.5. If $\beta > 0$ then larger (smaller) values than EL_{50} relate to higher(lower) success probabilities than 0.5.
Of course one can also find p% effective level, EL_p.

$$EL_p = \frac{\ln(p/(100-p)) - \alpha}{\beta}$$

Example 1
High school and Beyond (from C.J. Anderson, University of Idaho, introduced in a course on categorical data analysis)

- Question: Is the proportion of students attending an academic program related to achievement (x).
- Data from high school seniors (N=600).
- Response: Did students attend an academic high school program type or a non-academic program type (Y).
- Explanatory variables: Total score on 5 standardized achievement tests (Reading, Writing, Math, Science, and Civics), x being the total of the scores.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval Lower	95% Wald Confidence Interval Upper	Hypothesis Test Wald Chi-Square	Hypothesis Test df	Hypothesis Test Sig.	Exp(B)	95% Wald Confidence Interval for Exp(B) Lower	95% Wald Confidence Interval for Exp(B) Upper
(Intercept)	-7.055	.6948	-8.417	-5.693	103.104	1	.000	.001	.000	.003
x	.027	.0027	.022	.033	106.411	1	.000	1.028	1.022	1.033
(Scale)	1[a]									

Dependent Variable: program
Model: (Intercept), x
a. Fixed at the displayed value.

The estimated model equation is

$$\log(\hat{odds}) = \ln\left(\frac{\hat{\pi}(x)}{1-\hat{\pi}(x)}\right) = -7.055 + .027x$$

A one point increase in the total test score multiplies the odds for a student to be in an academic program by an estimated factor of $e^{0.027} = 1.02$, saying that the odds for being in an academic program increases by 2% for a one unit increase in the total score.

Using equation (3) the estimated proportion of students with total score of $x = 260$ in an academic program equals

$$\hat{\pi}(260) = \frac{\exp(-7.055 + .027(260))}{1 + \exp(-7.055 + .027(260))} \approx 0.491$$

(This is what you get when you carry three decimal places, but it will change when you carry more.)
The $EL_{50} = 7.055/0.027 = 261.30$.
To find the 90% effective level, i.e. the value of x so that an estimated 90% of students with this score are in academic programs, is $EL_{90} = (ln(90/10) - (-7.055))/0.027 = 342.67$.

The estimated model equation permits a comparison of the odds for different values of x. The odds ratio for success when $x = x_1$ versus $x = x_2$ equals

$$\theta(x_1, x_2) = (e^{\beta})^{x_1 - x_2}$$

Example: The odds ratio for being in the program for students with scores of $x = 300$ versus $x = 200$ equals $\theta(300, 200) = 1.02^{100} = 7.24$ The odds for being on an academic program are 7.24 times higher for students with a score of 300 than for students with a score of 200.

Comment:
Logistic Regression is appropriate for many different type of studies, since it models odds. Even data from retrospective studies (e.g. case-control studies), can be analyzed fitting this model.

1.2 Inference

1.2.1 Confidence Interval for $\boldsymbol{\beta}$

To estimate the effect of changes in x on $\pi(x)$ or the $odds(x)$ we should give confidence intervals for the slope parameter β.
Based on the point estimator, $\hat{\beta}$, and the estimated standard error, SE, from the SPSS output one can obtain a large sample Wald confidence interval for β

$$\hat{\beta} \pm z_{\alpha/2} SE$$

From here we can obtain a confidence interval for e^{β} by exponentiating the endpoints, giving the multiplicative effect on the odds of increasing x by one unit.
For smaller sample sizes or probabilities close to 0 or 1, the likelihood ratio confidence interval is the better choice. It includes all β_0 for which the likelihood ratio test for $H_0 : \beta = \beta_0$ is not significant, you can get these from SPSS, when using the GLM interface and choosing in the Statistics Tab the Profile Likelihood radio button.
Example in handout 3.

1.2.2 Statistical Test for $\boldsymbol{\beta}$

Already the Wald confidence interval is based on the fact that if the sample size is large then the z-score

$$Z = \frac{\hat{\beta} - \beta}{SE}$$

is approximately standard normally distributed, when β is the true population slope.

Wald test for the slope parameter $\boldsymbol{\beta}$ in a logistic regression model

1. Hypotheses:

Type of test	Hypotheses
2-tail	$H_0 : \beta = 0$ versus $H_a : \beta \neq 0$
lower tail	$H_0 : \beta \geq 0$ versus $H_a : \beta < 0$
upper tail	$H_0 : \beta \leq 0$ versus $H_a : \beta > 0$

   Choose α.

2. Assumptions: random samples, large sample size

3. Test Statistic:

$$z_0 = \frac{\hat{\beta}}{SE}$$

4. P-value:

Type of test	P-value
2-tail	$2P(Z > \lvert z_0 \rvert)$
lower tail	$P(Z < z_0)$
upper tail	$P(Z > z_0)$

In many situations the log-likelihood test is the better choice but does not permit for upper or lower tailed hypotheses. Remember, it is based on the log-likelihood, L_0, when $\beta = 0$, and the log-likelihood when β is not restricted.

Likelihood ratio test for the slope parameter $\boldsymbol{\beta}$ in a logistic regression model

1. Hypotheses:

   $H_0 : \beta = 0$ versus $H_a : \beta \neq 0$

   Choose α.

2. Assumptions: random samples, large sample size

3. Test Statistic:
$$\chi_0^2 = -2(L_0 - L_1), \quad df = 1$$

4. P-value $= P(\chi^2 > \chi_0^2)$

Example in handout 3.

1.2.3 Confidence Intervals for the Probabilities

Based on
$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$
the model can be used to estimate the probability of success for a given value of x.
In addition software like SPSS can provide confidence intervals for $\pi(x)$.
For $x = 260$ a 95% confidence interval for the true probability $\pi(260)$ is given as [0.471,0.561] (from SPSS).

The calculations are:
$$logit(\hat{\pi}(x)) \pm z_{\alpha/2} SE(logit(\hat{\pi}(x)))$$
with
$$SE(logit(\hat{\pi}(x))) = \sqrt{Var(\hat{\alpha}) + x^2\ Var\hat{\beta} + 2x\ Cov(\hat{\alpha}, \hat{\beta})}$$
The standard error for the confidence interval depends on the variance and covariance of $\hat{\alpha}$ and $\hat{\beta}$. Once you have this interval convert it to an interval for $\pi(x)$.

1.2.4 The Covariance Matrix for the Model estimates

One can request SPSS to print the covariance matrix for the parameter estimates. This matrix is the source for the standard errors for the confidence intervals.

Covariances of Parameter Estimates

	(Intercept)	x
(Intercept)	.48271	-.00183
x	-.00183	7.0E-006

Dependent Variable: program
Model: (Intercept), x

This matrix gives on the diagonal the variances of the parameter estimates, $\hat{\alpha}$ and $\hat{\beta}$. The off diagonal entries state the covariance for the two parameter estimates. (The covariance measures how strong the two variables are related, but is not standardized like the correlation, so the covariance can assume any value between $-\infty$ and ∞.)

The SE for the estimates is the square root of their variances (the diagonal entries. For example, the SE in the estimation of β is according to this matrix $SE = \sqrt{0.000007} = 0.0026$

The SE^2 for $\hat{\alpha} + \hat{\beta}x$ (needed for the confidence interval for $logit(\pi(x))$) is

$$\hat{Var}(\hat{\alpha}) + x^2\hat{Var}(\hat{\beta}) + 2x\hat{Cov}(\hat{\alpha}, \beta)$$

.

Therefore a $(1-\alpha) \times 100\%$ confidence interval for $logit(\pi(x))$ is given by

$$logit(\hat{\pi}(x)) \pm z_{\alpha/2}\sqrt{\hat{Var}(\hat{\alpha}) + x^2\hat{Var}(\hat{\beta}) + 2x\hat{Cov}(\hat{\alpha}, \hat{\beta})}$$

Example 2
A 95% confidence interval for $logit(\pi(260))$

$$\begin{array}{l} logit(\hat{\pi}(x)) \pm z_{\alpha/2}\sqrt{\hat{Var}(\hat{\alpha}) + x^2\hat{Var}(\hat{\beta}) + 2x\hat{Cov}(\hat{\alpha}, \beta)} \\ 0.065 \pm 1.96 \times \sqrt{0.00849} \end{array}$$

(This result requires carrying 8 significant digits). Thus the the 95% confidence interval for $logit(\pi(x))$ is [-0.116 , 0.246].
Using

$$\pi(x) = \frac{exp(logit(\pi(x)))}{1 + exp(logit(\pi(x)))}$$

will return [.471, .561] for a 95% confidence interval for $\pi(260)$.

1.3 Logistic Regression with Categorical Predictors

In this section we will present the interpretation of a logistic regression in the case that the predictor x is categorical.

One categorical predictor
Introduce indicator (dummy) variables. If x has I levels, $I-1$ dummy variables are needed:

I	Variable
1	$d_1 = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$
2	$d_2 = \begin{cases} 1 & \text{if } x = 2 \\ 0 & \text{otherwise} \end{cases}$
$\vdots$	$\vdots$
$I-1$	$d_{I-1} = \begin{cases} 1 & \text{if } x = I-1 \\ 0 & \text{otherwise} \end{cases}$

When choosing these dummies the category I, the category not represented by a dummy, becomes the reference category.
The model is

$$logit(\pi(x)) = \alpha + \beta_1 d_1 + \ldots \beta_{I-1} d_{I-1}$$

So, one categorical predictor with I categories creates a model with $I-1$ slope parameters!

The logits for the different categories are

Category, x	$logit(\pi(x))$
1	$\alpha + \beta_1$
2	$\alpha + \beta_2$
$\vdots$	$\vdots$
$I-1$	$\alpha + \beta_{I-1}$
I	α

The β_i is the difference in the logit for categories i and I of variable x.

$$\begin{aligned} \beta_i &= logit(\pi(i)) - logit(\pi(I)) \\ &= ln(odds(i)) - ln(odds(I)) \\ &= ln(odds(i)/odds(I)) \end{aligned}$$

Thus

$$e^{\beta_i} = \frac{odds(i)}{odds(I)}$$

the odds ratio for success comparing categories i and I of x.
The odds for a success are e^{β_i} times higher in category i than in category I.

Also, from

$$logit(\pi(1)) - logit(\pi(2)) = \beta_1 - \beta_2$$

we can conclude that

$$e^{\beta_i - \beta_j} = \frac{odds(i)}{odds(j)}$$

is the odds ratio for success comparing categories i and j of x.
The odds for a success are $e^{\beta_i - \beta_j}$ times higher in category i than in category j.

Example 3
HSB,
x = SES (socio economic status, categories: 1=low,2=middle,3=high)
Y = academic program (no=0, yes=1)
the dummies:

I	Variable
1	$d_1 = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$
2	$d_2 = \begin{cases} 1 & \text{if } x = 2 \\ 0 & \text{otherwise} \end{cases}$

The model

$$logit(\pi(x)) = \alpha + \beta_1 d_1 + \beta_2 d_2$$

The estimates:

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	.956	.1754	.612	1.299	29.673	1	.000
[ses=1.00]	-1.725	.2530	-2.221	-1.229	46.488	1	.000
[ses=2.00]	-.989	.2101	-1.401	-.577	22.152	1	.000
[ses=3.00]	0[a]	.	.	.	.	.	.
(Scale)	1[b]						

Dependent Variable: y
Model: (Intercept), ses

Therefore the estimated model equation is

$$logit(\hat{\pi}(x)) = 0.956 - 1.725 d_1 - .989 d_2$$

The odds for being in an academic program for students with low SES (=1) is $e^{-1.725} = .178$ times the odds for students with high SES (=3).

(odds(1)/odds(3)=.178)

According to the output we can reject $H_0 : \beta_1 = 0$ at significance level of 5% ($\chi^2_0 = 46.49, df = 1$) and conclude that the odds for being in an academic program are significantly different for students with SES of 1 and 3.

More than one categorical predictor

Assume we have two binary predictors, x and z, including a binary response Y, the data could be presented in 3-way contingency table. Assume that all variables are coded with 0 and 1.

Let $\pi(x, z) = P(Y = 1|x = 1, z = 1)$, the logistic regression model becomes

$$logit(\pi(x, z)) = \alpha + \beta_1 x + \beta_2 z$$

including main effects for x and z.

x and z are called dummy or indicator variables (coded with 0 and 1). The consequence for the logit are summarized in the following table

x	z	$logit(\pi(x, z))$
0	0	α
1	0	$\alpha + \beta_1$
0	1	$\alpha + \beta_2$
1	1	$\alpha + \beta_1 + \beta_2$

This model does not include interaction, meaning the relationship between x and Y is the same for all levels of z, implying a homogeneous association between x, y and Z.

The β_1 gives the difference in the logit for categories 1 and 0 of variable x, and β_2 gives the difference in the logit for categories 1 and 0 of variable z.

$$\begin{aligned}\beta_1 &= logit(\pi(1,z)) - logit(\pi(0,z)) \\ &= ln(odds(1,z)) - ln(odds(0,z)) \\ &= ln(odds(1,z)/odds(0,z))\end{aligned}$$

Thus

$$e^{\beta_1} = \frac{odds(1,z)}{odds(0,z)}, \quad z = 0, 1$$

gives the odds ratio for success comparing categories 0 and 1 of x while controlling for z.
The odds for "success" $(Y = 1)$ at $x = 1$ is e^{β_1} times the odds for "success" at $x = 0$, controlling for z.

A test for $\beta_1 = 0$ is equivalent to a test for odds ratio =1, or a test whether x and Y are independent when controlling for z, which we have called conditional independence before.
If $\beta_1 = 0$ the model would reduce to

$$logit(\pi(z)) = \alpha + \beta_2 z$$

Example 4
HSB data
Again the response will be the variable identifying the program (1=academic, 0 = general or vocational),
the factors considered will be gender (0=female, 1=male) and school type (0=public, 1=private).
The model:

$$logit(\pi(gender, school)) = \alpha + \beta_1 gender + \beta_2 school$$

The SPSS output:

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
38.596	2	.000

Dependent Variable: program
Model: (Intercept), gender, school

a. Compares the fitted model against the intercept-only model.

Goodness of Fit[b]

	Value	df	Value/df
Deviance	.321	1	.321
Scaled Deviance	.321	1	
Pearson Chi-Square	.318	1	.318
Scaled Pearson Chi-Square	.318	1	
Log Likelihood[a]	-396.377		
Akaike's Information Criterion (AIC)	798.754		
Finite Sample Corrected AIC (AICC)	.		
Bayesian Information Criterion (BIC)	796.913		
Consistent AIC (CAIC)	799.913		

Dependent Variable: program
Model: (Intercept), gender, school

a. The full log likelihood function is displayed and used in computing information criteria.

b. Information criteria are in small-is-better form.

The omnibus test indicates that at least one slope is different from 0. At significance level of 5% the data provide sufficient evidence ($\chi^2_0 = 38.569, df = 2, P-value < 0.001$) that either gender or school type or both have an effect on the chance of a student being in an academic program.

The Deviance is reported being 0.321 with df=1 indicating a very good model fit.

The estimates:

Parameter Estimates

Parameter	B	Std. Error	95% Profile Likelihood Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-.191	.1184	-.424	.041	2.596	1	.107
[gender=1.00]	.070	.1692	-.261	.402	.173	1	.678
[gender=.00]	0[a]	.	.	.	.	.	.
[school=1.00]	1.533	.2720	1.020	2.092	31.791	1	.000
[school=.00]	0[a]	.	.	.	.	.	.
(Scale)	1[b]						

Dependent Variable: program
Model: (Intercept), gender, school

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

The estimated model equation:

$$logit(\hat{\pi}(gender, school)) = -0.191 + 0.070gender + 1.533school$$

The gender has at significance level of 5% no effect on the chance of a student being in an academic program when controlling for school type ($\chi^2_0 = 0.173, df = 1, P = 0.678$). (At significance level of 5% gender and being in an academic program are conditionally independent (conditioned on the school type)).
The odds for being in an academic program for students in private schools is $e^{1.533} = 4.63$ times the odds for students in public schools.

odds(0)/odds(1)=odds(private)/odds(public) $= e^{1.533} = 4.63$.

The effect of school type on the chance of being in an academic program is significant ($\chi^2_0 = 31.791, df = 1, P < 0.001$) when controlling for gender. (At significance level of 5% school type and being in an academic program are not conditionally independent (conditioned on gender)).

Comment: The model introduced above does not include interaction terms. I.e. the effect of x on Y is the same for all levels of z and the effect of z on Y is the same for all levels of x.
Thus this model implies homogeneous associations between x and Y given z and between z and Y given x.

1.4 Multiple Logistic Regression

Consider now the general logistic regression model with k predictors, $x_1, x_2, \ldots, x_k$.
The model equation is then

$$logit(\pi(x_1, x_2, \ldots, x_k)) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

β_i gives the effect of a one unit increase in x_i on the log odds for $Y = 1$ controlling for the remaining predictors, thus e^{β_i} is the multiplicative effect of a one unit increase in x_i on the odds for $Y = 1$ at fixed levels of the remaining variables. And α are the log odds for $Y = 1$, when all predictors are zero.

1.4.1 Test for an association with a factor

To check if a certain factor is useful in the model we can use the difference in the log likelihood for the model including the factor and the otherwise same model excluding the factor as a test statistic for conducting a test if the parameter(s) for the factor is(are) different from 0.

1. H_0 : "success" is independent from factor versus $H_a : H_0$ is not true. Choose α.

2. Random sample, large sample

3. Test Statistic: Let L_0 be the log likelihood for the model excluding the factor, and L_1 the log likelihood for the same model adding the factor

$$\chi^2_0 = -2(L_0 - L_1), df = \text{difference in the number of parameters in the two models}$$

4. P-value=$P(\chi^2 > \chi^2_0)$

Example 5
HSB data
Consider the model including predictors:

- x_1 academic score (total on the exams)
- x_2 school type (0=public, 1=private)
- d_1, d_2 SES (1=low, 2=middle, 3=high)

Question: When correcting for academic score and school type does the SES have an effect on the chance of a student being in an academic program?
Model M_0:

$$logit(\pi) = \alpha + \beta_1 x_1 + \beta_2 x_2$$

SPSS gives $L_0 = -332.058$ for this model in "Goodness of fit table".

Model M_1:

$$logit(\pi) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 d_1 + \beta_4 d_2$$

SPSS gives $L_1 = -325.980$ for this model.

1. H_0 : being in an academic program is independent from SES when controlling for academic score and school type $H_a : H_0$ is not true. ($H_0 : \beta_3 = \beta_4 = 0$), $\alpha = 0.05$.
2. Assumptions are met
3. Test Statistic:
$$\chi_0^2 = -2(L_0 - L_1) = 12.156, df = 5 - 3 = 2$$
4. P-value=$P(\chi^2 > 12.156)$, which is according to the table less than 0.005.
5. Reject H_0 since the P-value is smaller than α.
6. At significance level of 5% the data suggests that the SES is associated with the chance of a student being in an academic program even when controlling for academic success and school type.

Estimates: $\hat{\beta}_3 = -0.891$ and $\hat{\beta}_4 = -.706$ indicate that the higher the SES the higher the chance for a student to be in an academic program when keeping the other factors unchanged.
$e^{\hat{\beta}_3} = 0.41, e^{\hat{\beta}_4} = 0.49, e^{\hat{\beta}_3 - \hat{\beta}_4} = 0.83$: Interpretation???

Comment: Ordinal variables can sometimes be included as numerical predictors with the model. The problem with this approach is that the result will depend on the scores being chosen for the different categories.
Unless there is a justification for the choice of scores the approach demonstrated in the example should be chosen.

1.4.2 Interaction

Two predictors, x and z, are interacting in their effect on the response variable Y, if the conditional relationships (conditioned on z) between x and Y differ for the different levels of z.
To accommodate interaction between factors with the model cross product terms between the predictors are added to the model.
When not including interaction terms automatically a homogeneous associations between the three variables is assumed (i.e. the value of z has NO effect on the relationship between x and Y).
To illustrate

Example 6
To permit for interaction between academic success and school type include the cross product term $x_1 \times x_2$ with the model

$$logit(\pi) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \times x_2$$

Testing $H_0 : \beta_3 = 0$ is testing for H_0 : association between academic program and academic success is the same for both school types.
To include interaction with SPSS either include this term in the model section, when specifying the GLM, or use transform>Compute Variable... and create a new variable including the products of x_1 and x_2. The new variable is then added to the predictor variables and added as main effect in the model section.

In both cases when testing $H_0 : \beta_3$ the Wald χ^2 statistic reported is $0.001, df = 1$ and $P = .971$. At significance level of 5% the data do not provide sufficient evidence that academic success and school type interact in their effect on the proportion of students in academic high school programs.

Question: How many interaction terms would be needed when permitting for interaction between SES and academic success?

1.4.3 Interpretation of Slopes in Interaction Models

Consider a logistic regression model for a binary response (success/failure) with one numerical predictor x and one categorical predictor z at two levels (treatment A and B). Introduce the dummy variable for z

$$d_z = \begin{cases} 1 & \text{if} \quad z = \text{treatment} \quad \text{A} \\ 0 & \text{if} \quad z = \text{treatment} \quad \text{B} \end{cases}$$

This makes treatment B the reference category.

Then the model equation for the interaction model becomes:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1 x + \beta_2 d_z + \beta_3 x d_z$$

Let's consider the model equations for the different treatments to better understand the interpretation of the parameters:
Treatment = B ($d_z = 0$): $\ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1 x$
Treatment = A ($d_z = 1$): $\ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1 x + \beta_2 + \beta_3 x = (\alpha + \beta_2) + (\beta_1 + \beta_3)x$

This implies the following interpretations:

- e^{α}: for treatment B these are the odds for success when $x = 0$
- e_1^{β}: for treatment B the odds for success are e_1^{β} times higher/lower for every one unit increase in x.
- $e^{\alpha+\beta_1}$: for treatment A these are the odds for success when $x = 0$
- $e^{\beta_1+\beta_3}$: for treatment A the odds for success are $e^{\beta_1+\beta_3}$ times higher/lower for every one unit increase in x.

Example 7
Illustrate the interpretation of the slopes in interaction model
Let π = probability a randomly chosen person votes for Trump.
x_1 = age of a person
x_2 = state the person lives in ($x_2 = 1 \Leftrightarrow$ state = NY, $x_2 = 0 \Leftrightarrow$ state = Montana

The interaction model:

$$\text{logit}(\pi) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \times x_2$$

Then the model equations for the different states looks like this:

- Montana ($x_2 = 0$):

$$\text{logit}(\pi) = \alpha + \beta_1 x_1$$

- New York ($x_2 = 1$):

$$\text{logit}(\pi) = \alpha + \beta_1 x_1 + \beta_2 + \beta_3 x_1 = (\alpha + \beta_2) + (\beta_1 + \beta_3) x_1$$

Assume $\beta_1 = 0.09$, then $e^{\beta_1} = 1.1$, and $\beta_3 = -0.08$, then $e^{\beta_1 + \beta_3} = 1.01$

- Montana: in Montana for every extra year in age the odds for a person to vote for Trump increases by 10%

- New York: in New York for every extra year in age the odds for a person to vote for Trump increases by 1%

The effect of age on the odds to vote for Trump is not the same in the two states, age and state interact in their effect on the odds of voting for Trump.