

Logistic Regression

PSY3100 – A Generalized Linear Model (GLM) Perspective

Frederick Anyan

Associate Professor of Quantitative Research Methods

[NTNU](#) | [Personal website](#)

NTNU, Department of Psychology

Fall 2026

Today's plan

- From a straight line to an S-curve: why OLS breaks on a binary outcome (stretching the x-axis)
- From probability to log-odds: stretching the y-axis until the S-curve becomes a straight line
- Maximum likelihood: how that straight line is fitted
- A worked example, built up predictor by predictor
- Interpreting coefficients: odds ratios
- Model selection with likelihood-ratio tests, and pseudo- R^2
- Logistic regression for classification
- Practical logistic regression in R

Examples of binary outcomes in psychology

- whether or not someone relapses after treatment: 0 = no relapse, 1 = relapse
- whether someone meets criteria for clinical caseness: 0 = below cutoff, 1 = above cutoff
- whether a participant drops out of a study: 0 = completed, 1 = dropped out

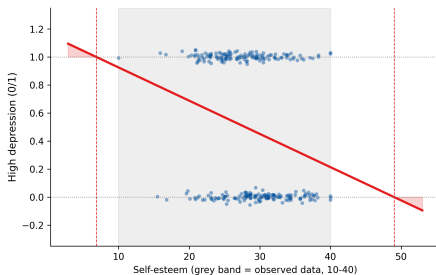
Our running example, extended

Today, we add exactly this kind of variable to the `clinical_sample` dataset:

$$\text{high_depression} = \begin{cases} 1 & \text{if depression} > \overline{\text{depression}} = 9.11 \\ 0 & \text{otherwise} \end{cases}$$

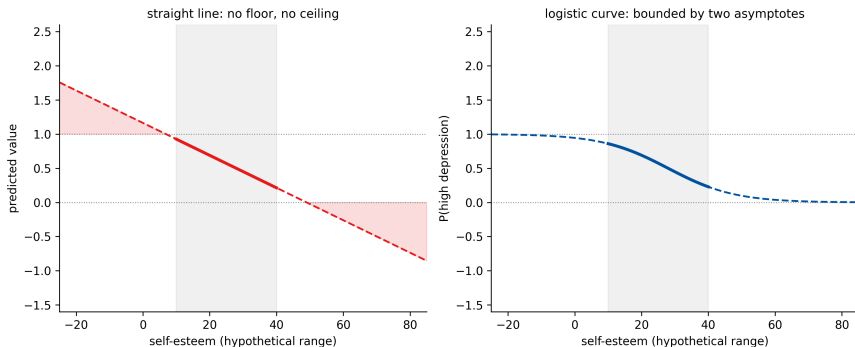
135 of 280 participants (48.2%) score above the mean.

What if we just fit a straight line?



- “linear probability modeling”: fit OLS straight to the 0/1 outcome, as if it were any other continuous DV – the fitted red line is $\widehat{\text{high_depression}} = 1.163 - 0.024(\text{self-esteem})$
- over our *observed* self-esteem values (grey band, 10-40) the line actually stays inside $[0,1]$ – but only just: it runs from .93 down to .21, hugging both edges
- extend self-esteem just a little past what we happened to sample (red dashed lines) and the line *does* cross: above 1 below self-esteem ≈ 6.9 , below 0 above self-esteem ≈ 49 (shaded red) – entirely plausible values for a real person, just not ones in this particular sample of 280
- a line has one slope, everywhere, forever – nothing in the equation tells it that p must stay in $[0,1]$, so sooner or later, for *some* value of the predictor, it always breaks

Stretching the x-axis



- push self-esteem hypothetically far outside our observed range (grey band)
- **left:** the line has no floor or ceiling – far enough out, it predicts impossible “probabilities” (shaded)
- **right:** the curve has two horizontal asymptotes, 0 and 1 – it only ever gets closer, never crosses them

What are the odds??

Two related but different questions

Odds: how much more likely is one event relative to another?

Probability: how likely am I to observe a particular event?

Reading odds notation, “a:b”

“a:b” means a ways for the event to happen against b ways for it not to. As a single *number*, odds = a/b – this is the number that matters below, not a by itself.

Odds – number between 0 and ∞

- Odds < 1 : first event less likely; e.g., rolling a six on a fair die: 1 way to succeed, 5 ways not to, written 1:5, i.e. $1/5 = 0.2$
- Odds $= 1$: equally likely; e.g., a fair coin landing heads, 1:1, i.e. $1/1 = 1$
- Odds > 1 : first event more likely; e.g., a strong free-throw shooter scoring, 4:1, i.e. $4/1 = 4$

Probability – number between 0 and 1

- $p = 0$: it can never happen
- $p = 0.5$: it happens 50% of the time
- $p = 1$: it happens always

Odds in logistic regression

Our event of interest: $Y = 1$ (a person scores `high_depression`). The complementary event: $Y = 0$ (they do not).

$$\text{Odds} = \frac{P(Y = 1)}{P(Y = 0)} \quad (1)$$

- read this as: the probability of the event happening, divided by the probability of it not happening
- because someone either is or is not above the depression cutoff, these two probabilities always add to 1:

$$P(Y = 0) = 1 - P(Y = 1)$$

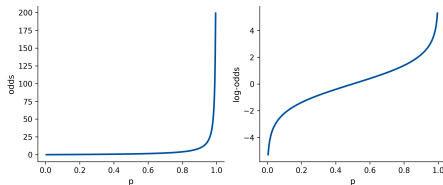
$$\text{Odds} = \frac{P(Y = 1)}{1 - P(Y = 1)} = \frac{1 - P(Y = 0)}{P(Y = 0)} = \frac{p}{1 - p} \quad (2)$$

- so once we write $p = P(Y = 1)$, the odds of high depression collapse to a single, familiar formula: $p/(1 - p)$

Log-odds: the logit

Odds still have a floor at 0, so a straight line could still undershoot them. Take the logarithm of the odds ($\log$ is the natural logarithm throughout – the $\log()$ function in R) and that last bound disappears too. This function – probability in, log-odds out – is called the **logit**:

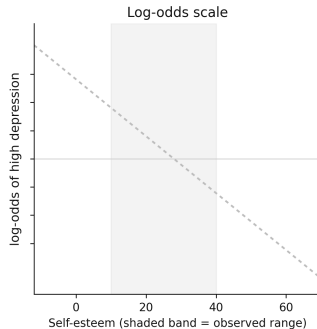
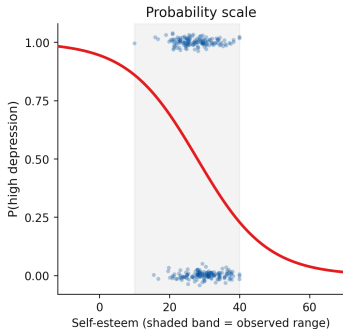
$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) \quad (3)$$



- odds (left panel) have a hard floor at 0; log-odds (right panel) have neither a floor nor a ceiling – the whole real line, exactly what a straight line $B_0 + B_1 X$ can predict without ever breaking a bound
- a quick example: $p = 0.8$ gives odds $= 0.8/0.2 = 4$ and log-odds $= \log(4) = 1.386$ – three numbers, one and the same chance
- the next slides apply the logit one probability at a time and watch what it does to the y-axis

Stretching the y-axis, step by step

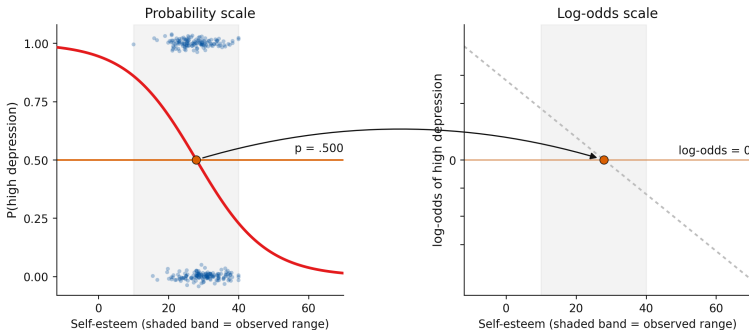
two rulers



- **left:** the fitted S-curve for $P(\text{high depression})$ against self-esteem (the model we fit as Model 1 in the worked example), with all 280 people drawn at exactly 0 or 1
- **right:** the same model on a new, still unlabelled, log-odds ruler – over the next slides we carry probabilities across one at a time with the logit (Equation 3) and watch the ruler fill in

Stretching the y-axis, step by step

$$p = 0.5$$

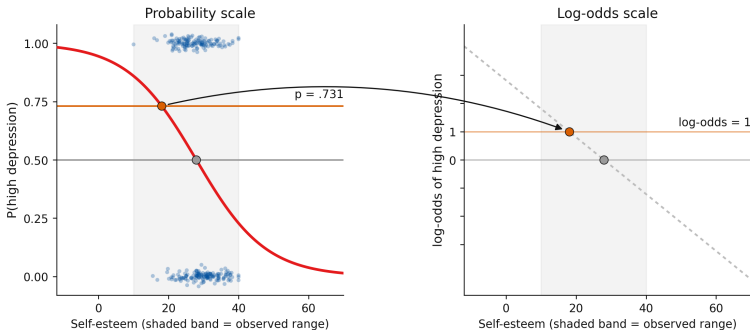


$$\text{logit}(0.5) = \log\left(\frac{0.5}{1 - 0.5}\right) = \log\left(\frac{0.5}{0.5}\right) = \log(1) = 0$$

- the middle of the probability ruler becomes the *centre* of the log-odds ruler: a 50/50 chance is odds of 1, and $\log(1) = 0$
- for Model 1, the curve crosses $p = 0.5$ at self-esteem ≈ 27.9 – so the straight line on the right crosses 0 at the same self-esteem value

Stretching the y-axis, step by step

$$p = 0.731$$

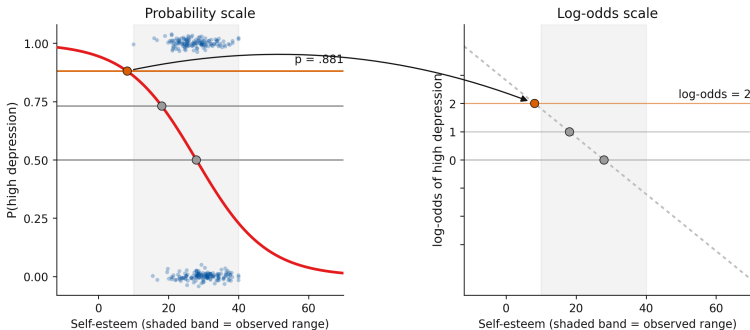


$$\text{logit}(0.7311) = \log\left(\frac{0.7311}{1 - 0.7311}\right) = \log\left(\frac{0.7311}{0.2689}\right) = \log(2.718) = 1$$

- a probability of 0.731 is odds of 2.718, and $\log(2.718) = 1$
- going from 0 to 1 on the new ruler took $0.731 - 0.500 = 0.231$ of the old ruler (Model 1 reaches this point at self-esteem ≈ 18.0)

Stretching the y-axis, step by step

$$p = 0.881$$

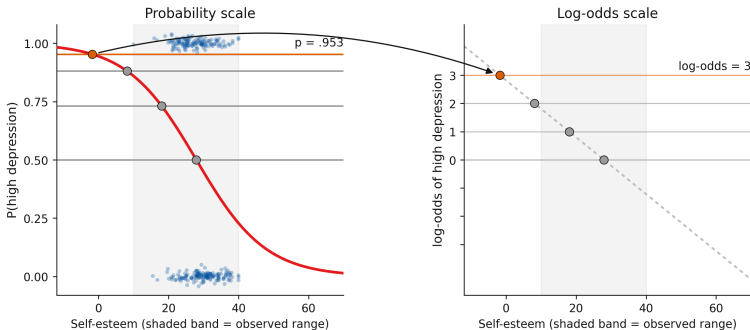


$$\text{logit}(0.8808) = \log\left(\frac{0.8808}{0.1192}\right) = \log(7.389) = 2$$

- one more unit on the new ruler ($1 \rightarrow 2$) now takes only $0.881 - 0.731 = 0.150$ of the old ruler
- on the probability scale the points are getting *closer together*; on the log-odds scale they stay exactly one unit apart – the ruler is being stretched

Stretching the y-axis, step by step

$$p = 0.953$$

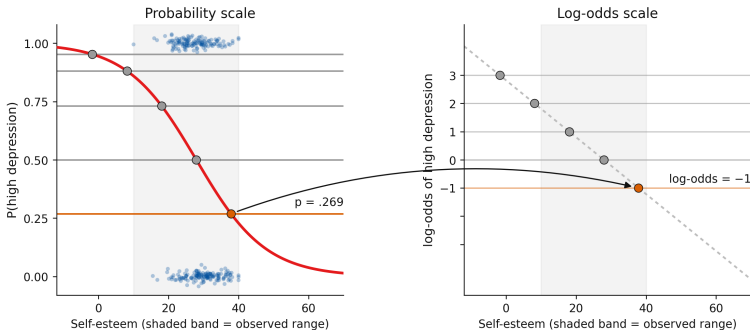


$$\text{logit}(0.9526) = \log\left(\frac{0.9526}{0.0474}\right) = \log(20.09) = 3$$

- $2 \rightarrow 3$ on the new ruler: only $0.953 - 0.881 = 0.072$ of the old ruler
- the closer p gets to 1, the more log-odds it takes to move the same amount of probability – the top of the probability ruler is being pulled upward without limit

Stretching the y-axis, step by step

$$p = 0.269$$

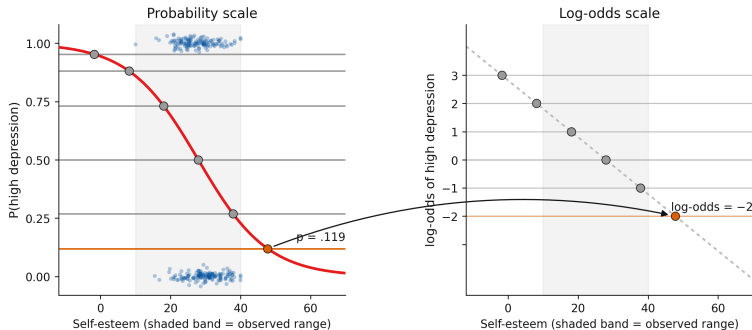


$$\text{logit}(0.2689) = \log\left(\frac{0.2689}{1 - 0.2689}\right) = \log\left(\frac{0.2689}{0.7311}\right) = \log(0.368) = -1$$

- below $p = 0.5$ the odds are below 1 (here 0.368), and the log of a number below 1 is negative – so the lower half of the probability ruler becomes the *negative* half of the log-odds ruler
- $0.269 = 1 - 0.731$ is the mirror image of $p = 0.731$, and it lands at -1 , exactly opposite $+1$ (Model 1 reaches this point at self-esteem ≈ 37.8)

Stretching the y-axis, step by step

$$p = 0.119$$

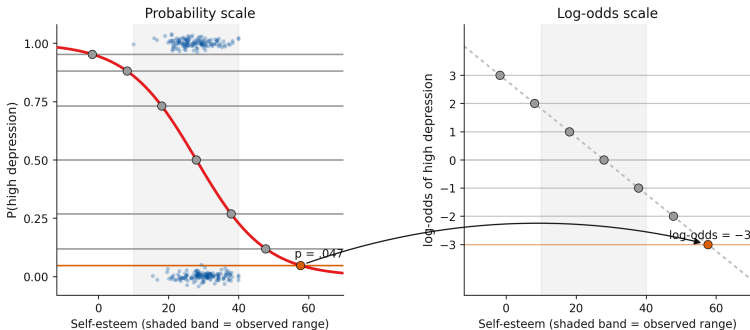


$$\text{logit}(0.1192) = \log\left(\frac{0.1192}{1 - 0.1192}\right) = \log\left(\frac{0.1192}{0.8808}\right) = \log(0.135) = -2$$

- one more unit down on the new ruler ($-1 \rightarrow -2$) takes only $0.269 - 0.119 = 0.150$ of the old ruler – the same gap as between 0.731 and 0.881 above
- $0.119 = 1 - 0.881$, the mirror image of $p = 0.881$ (Model 1 reaches this point at self-esteem ≈ 47.8)

Stretching the y-axis, step by step

$$p = 0.047$$



$$\text{logit}(0.0474) = \log\left(\frac{0.0474}{1 - 0.0474}\right) = \log\left(\frac{0.0474}{0.9526}\right) = \log(0.0498) = -3$$

- $-2 \rightarrow -3$ on the new ruler: only $0.119 - 0.047 = 0.072$ of the old ruler
- the closer p gets to 0, the more log-odds it takes to move the same amount of probability – the bottom of the probability ruler is pulled downward without limit ($0.047 = 1 - 0.953$, the mirror image of $p = 0.953$)

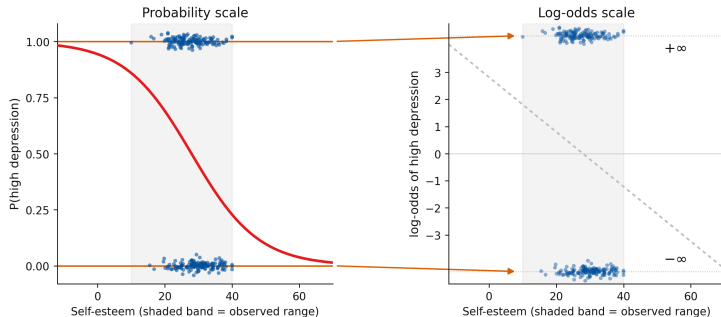
The new ruler, tick by tick

Probability p	Odds $p/(1-p)$	Log-odds	Model 1 reaches it at self-esteem
0.047	0.050	-3	57.7
0.119	0.135	-2	47.8
0.269	0.368	-1	37.8
0.500	1.000	0	27.9
0.731	2.718	1	18.0
0.881	7.389	2	8.1
0.953	20.09	3	-1.8

- **log-odds column:** equally spaced – every row is one unit apart
- **self-esteem column:** also equally spaced, about 9.9 points per row – because on the log-odds ruler, Model 1 is a straight line
- **probability column:** *not* equally spaced – the gaps shrink towards 0 and towards 1, which is exactly the S-shape

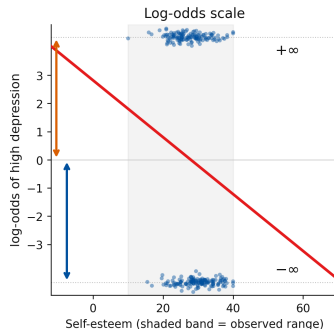
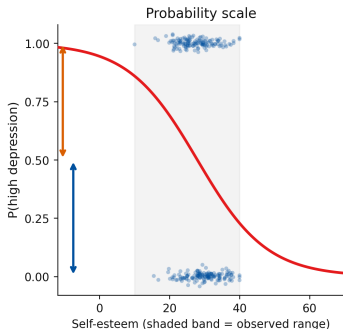
Stretching the y-axis, step by step

$p = 1$ and $p = 0$



$$\text{logit}(1) = \log\left(\frac{1}{0}\right) = \log(1) - \log(0) = 0 - (-\infty) = +\infty \quad \text{logit}(0) = \log\left(\frac{0}{1}\right) = \log(0) = -\infty$$

- strictly, we cannot divide by 0 – but writing the log of a fraction as a difference of logs shows where it is heading: $\log(x) \rightarrow -\infty$ as $x \rightarrow 0$
- every observed person is exactly 0 or 1, so on the log-odds ruler all 280 of them sit at $+\infty$ or $-\infty$: there is no finite distance between a data point and the line, so least squares cannot be done on this scale – which is why the line is estimated by maximum likelihood instead (next section)



- **orange**: the stretch from 0.5 to 1 on the old ruler becomes 0 to $+\infty$ on the new one
- **blue**: the stretch from 0 to 0.5 becomes $-\infty$ to 0
- with both halves stretched, the bounded S-curve becomes an unbounded **straight line** – and the coefficients that `glm()` reports are the intercept and slope of *this* line, not of the S-curve

The logistic regression model

The finished model is the straight-line equation you know from OLS, now predicting log-odds instead of raw Y :

$$\log\left(\frac{p}{1-p}\right) = B_0 + B_1X_1 + \cdots + B_kX_k \quad (4)$$

and the **inverse logit** carries any log-odds value z back to a probability:

$$p = \text{logit}^{-1}(z) = \frac{\exp(z)}{1 + \exp(z)} \quad (5)$$

- the logit (probability $\rightarrow$ log-odds) and the inverse logit (log-odds $\rightarrow$ probability) are one function, seen from both directions
- the coefficients $B_0, \dots, B_k$ describe the straight line; to talk about probabilities we pass its height through Equation 5
- one question is left: how are $B_0, \dots, B_k$ chosen? That is the next section

Fitting the straight line: maximum likelihood

- least squares needs a distance between each data point and the line – but on the log-odds scale every observed person sits at $+\infty$ or $-\infty$, so there is no distance to square
- instead, logistic regression is fit by **maximum likelihood**, in four steps:
 - 1 propose a candidate straight line (a value for B_0 and B_1)
 - 2 turn it into a predicted probability p_i for each of the 280 people, with the inverse logit
 - 3 ask: how probable are the outcomes we *actually* observed, if these p_i were true? That number is the **likelihood**
 - 4 try other lines, and keep the one with the highest likelihood
- R does this search for us; what we need to understand is *what* it is maximizing

The likelihood, person by person

- someone who *has* high depression ($Y = 1$) contributes p_i – a good line gives them a p_i close to 1
- someone who does *not* ($Y = 0$) contributes $1 - p_i$ – a good line gives them a p_i close to 0

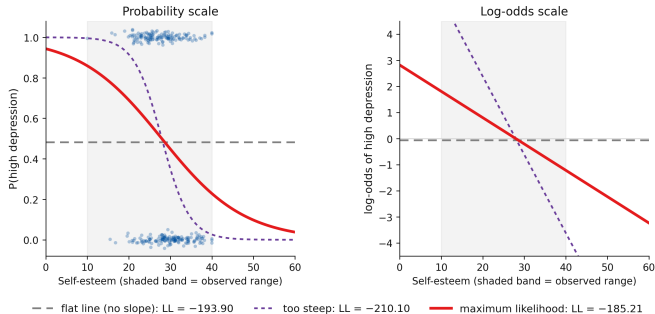
Two real participants, under the line $\log\left(\frac{p}{1-p}\right) = 2.818 - 0.101(\text{self-esteem})$

	Self-esteem	Observed Y	Predicted p_i	Contribution
Participant 1	19.0	1	0.711	0.711
Participant 3	37.6	0	0.274	$1 - 0.274 = 0.726$

- the likelihood multiplies all 280 contributions together ($0.711 \times 0.726 \times \dots$); a product of 280 numbers below 1 is vanishingly small, so we add up their logarithms instead – the **log-likelihood**:

$$LL = \sum_{i=1}^{280} \log(\text{contribution}_i) \quad (6)$$

Maximum likelihood: trying candidate lines



Candidate line	B_0	B_1	Log-likelihood LL
flat line (ignores self-esteem)	-0.071	0.000	-193.90
too steep	8.380	-0.300	-210.10
maximum likelihood	2.818	-0.101	-185.21

- the flat line gives everyone the same $p = 0.48$; the steep line makes confident predictions that are often wrong, and every confident miss is heavily penalized
- no other pair of B_0 , B_1 gets closer to 0 than -185.21 – these are exactly the coefficients R reports for Model 1 in the worked example

From likelihood to deviance

R reports *deviance* alongside the log-likelihood

$$D = -2 LL \quad (\text{smaller deviance} = \text{better fit})$$

- same information, flipped: log-likelihood is “higher = better”, deviance is “lower = better” – similar in spirit to the residual sum of squares in OLS
- our maximum-likelihood line: $D = -2 \times (-185.21) = 370.41$ – the “Residual deviance” R prints for Model 1
- the flat line: $D = -2 \times (-193.90) = 387.81$ – the “Null deviance” R prints under every model
- the drop from 387.81 to 370.41 is the improvement in fit that comes from adding self-esteem; whether a drop like this is more than chance is the job of the likelihood-ratio test (model selection)

Our worked example: three research questions

- we build up the same model three times, each time adding one predictor, each addition answering its own question

Model 1

RQ: Does self-esteem predict whether someone scores above the sample's average depression level (`high_depression`)?

Model 2

RQ: Does gender predict `high_depression`, over and above self-esteem?

Model 3

RQ: Does the self-esteem $\rightarrow$ `high_depression` relationship itself differ by gender (a self-esteem $\times$ gender interaction)?

Model 1: self-esteem predicting high depression

Research question: Does self-esteem predict high_depression?

In R

```
mod1 <- glm(high_depression ~ self_esteem, data = clinical_sample,  
             family = binomial(link = "logit"))
```

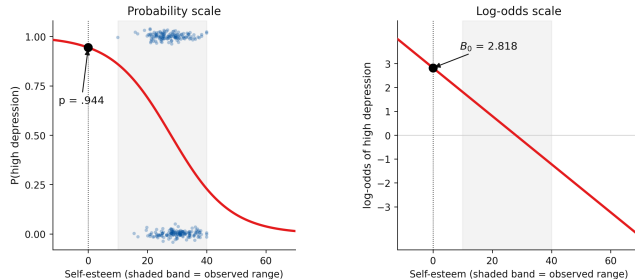
Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.81832	0.73310	3.844	0.000121	***
self_esteem	-0.10090	0.02527	-3.994	6.51e-05	***

Null deviance: 387.81 on 279 degrees of freedom
Residual deviance: 370.41 on 278 degrees of freedom
AIC: 374.41

- **Estimate:** coefficient in log-odds – self-esteem's is -0.101 , negative, so higher self-esteem predicts *lower* log-odds of high depression
- **z, $\Pr(>|z|)$:** tests $H_0: B = 0$ on the z-distribution (ML estimates are large-sample-normal, not exactly t) – here $z = -3.994$, $p < .001$
- **deviance:** 370.41 here vs. 387.81 for the flat, intercept-only line – the two deviances from the maximum-likelihood section

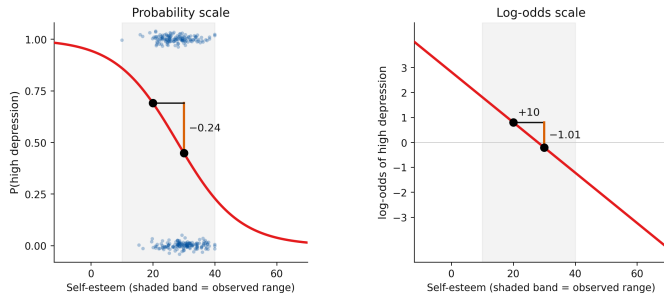
$$\log\left(\frac{p}{1-p}\right) = 2.818 - 0.101 (\text{self-esteem}) \quad (7)$$



- **log-odds:** $B_0 = 2.818$ is where the straight line crosses self-esteem = 0 – the predicted log-odds of high depression for someone with a self-esteem score of 0
- **odds:** exponentiate the log-odds, $\exp(2.818) = 16.75$ – $\exp()$ undoes the log, taking us from log-odds back to odds
- **probability:** $p = \text{odds} / (1 + \text{odds}) = 16.75 / 17.75 = 0.944$ – the black point on the S-curve

Read with care

Self-esteem was observed only from 10 to 40 (shaded band), so self-esteem = 0 is an extrapolation: B_0 anchors the line, it does not describe anyone in our sample. Centering self-esteem would move this anchor to a meaningful person, exactly as in multiple regression.



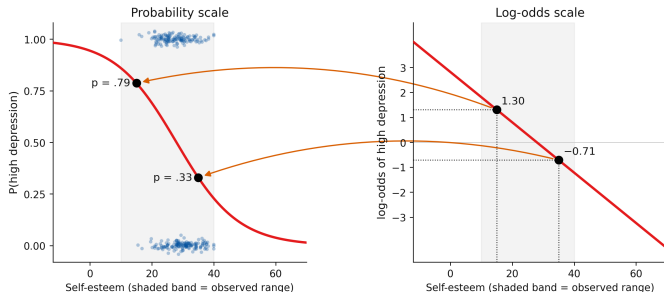
- $B_1 = -0.101$: each one-point increase in self-esteem changes the log-odds of high depression by -0.101 – rise over run on the straight line, exactly as in OLS
- self-esteem 20: $2.818 - 0.101 \times 20 = 0.80$ self-esteem 30: $2.818 - 0.101 \times 30 = -0.21$
- change in log-odds: $-0.21 - 0.80 = -1.01 = 10 \times (-0.101)$ – and the *same* -1.01 for any 10-point step, anywhere on the line

Self-esteem	Log-odds = $2.818 - 0.101 \times \text{SE}$	Odds = $\exp(\text{log-odds})$	Probability = $\frac{\text{odds}}{1+\text{odds}}$
10	1.81	6.11	0.86
20	0.80	2.23	0.69
30	-0.21	0.81	0.45

- **log-odds:** drop by the same 1.01 at each 10-point step ($1.81 \rightarrow 0.80 \rightarrow -0.21$)
- **probability:** drops by 0.17 from 10 to 20 ($0.86 \rightarrow 0.69$), but by 0.24 from 20 to 30 ($0.69 \rightarrow 0.45$) – the change in p depends on where you start, so B_1 cannot be read as “a change in probability”
- **odds:** each 10-point step *multiplies* the odds by the same number, 0.365 ($6.11 \times 0.365 = 2.23$, and $2.23 \times 0.365 = 0.81$) – this constant multiplier is the idea behind the **odds ratio** (next section)

Reading Model 1 off the log-odds graph

using the equation



1. Plug in (Equation 7): log-odds
 $= 2.818 - 0.101(\text{self-esteem})$ – read the height of the straight line

2. Carry it back to probability with the inverse logit (Equation 5): $\text{odds} = \exp(\text{log-odds})$, then
 $p = \text{odds} / (1 + \text{odds})$

Self-esteem	Log-odds	Odds	Probability
15 (low)	$2.818 - 0.101(15) = 1.30$	3.69	0.79
35 (high)	$2.818 - 0.101(35) = -0.71$	0.49	0.33

Model 2: adding gender

Research question: Does gender predict high_depression, over and above self-esteem?

In R

```
mod2 <- glm(high_depression ~ self_esteem + gender, data = clinical_sample,  
            family = binomial(link = "logit"))
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.62169	0.74861	3.502	0.000462	***
self_esteem	-0.10397	0.02582	-4.026	5.66e-05	***
gender	0.60978	0.25086	2.431	0.015048	*

Null deviance: 387.81 on 279 degrees of freedom
Residual deviance: 364.42 on 277 degrees of freedom
AIC: 370.42

Model 2: the fitted equation

$$\log \left(\frac{p}{1-p} \right) = 2.622 - 0.104 (\text{self-esteem}) + 0.610 (\text{gender}_{\text{woman}}) \quad (8)$$

Equation 8 is Model 2's output written as an equation: each Estimate becomes the number in front of its predictor (gender: 0 = man, 1 = woman).

- **2.622** (intercept): the log-odds for a man (gender = 0) with self-esteem 0 – outside the data, like Model 1's intercept
- **-0.104**: each extra point of self-esteem lowers the log-odds by 0.104, for men and women alike
- **0.610**: at any given self-esteem, a woman's log-odds are 0.610 higher than a man's
- on the log-odds graph this is two *parallel* straight lines, one per gender, 0.610 apart

Example: self-esteem 35

man: $2.622 - 0.104 \times 35 = -1.02$
on the next slide

woman: $-1.02 + 0.610 = -0.41$ – turned into probabilities

Model 2: worked predictions at self-esteem=35

$$\log\left(\frac{p}{1-p}\right) = 2.622 - 0.104(\text{self-esteem}) + 0.610(\text{gender}_{\text{woman}})$$

Woman (gender=1)

Man (gender=0)

$$\log\left(\frac{p}{1-p}\right) = 2.622 - 0.104(35) + 0.610 = -0.41 \qquad \log\left(\frac{p}{1-p}\right) = 2.622 - 0.104(35) = -1.02$$

$$\frac{p}{1-p} = \exp(-0.41) = 0.67$$

$$p = \frac{0.67}{1.67} = 0.40$$

$$\frac{p}{1-p} = \exp(-1.02) = 0.36$$

$$p = \frac{0.36}{1.36} = 0.27$$

At the same (high) self-esteem level...

...a woman's predicted probability of high depression (.40) is still noticeably higher than a man's (.27). Gender adds real information beyond self-esteem alone.

Model 3: does the self-esteem effect differ by gender?

Research question: Does the self-esteem $\rightarrow$ high_depression relationship itself differ by gender (a self-esteem $\times$ gender interaction)?

In R

centering first: `c_self_esteem = self_esteem - 28.68`

```
mod3 <- glm(high_depression ~ c_self_esteem * gender, data = clinical_sample,  
            family = binomial(link = "logit"))
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.36927	0.17394	-2.123	0.03373	*
c_self_esteem	-0.12363	0.03812	-3.243	0.00118	**
gender	0.61259	0.25102	2.441	0.01466	*
c_self_esteem:gender	0.03715	0.05165	0.719	0.47200	

Null deviance: 387.81 on 279 degrees of freedom
Residual deviance: 363.90 on 276 degrees of freedom
AIC: 371.90

Model 3: the fitted equation

$$\log\left(\frac{p}{1-p}\right) = -0.369 - 0.124(\text{cSE}) + 0.613(\text{gender}) + 0.037(\text{cSE} \times \text{gender}) \quad (9)$$

Equation 9 is Model 3's output written as an equation. cSE = self-esteem – 28.68 (centered), so cSE = 0 is a person with average self-esteem.

- **–0.369** (intercept): the log-odds for a man with average self-esteem
- **–0.124**: the self-esteem slope for men – each extra point lowers the log-odds by 0.124
- **0.613**: at average self-esteem, a woman's log-odds are 0.613 higher than a man's
- **0.037** (interaction): how much the self-esteem slope differs for women – their slope is $-0.124 + 0.037 = -0.087$

Is the interaction real?

0.037 is small and not significant ($p = .472$): no evidence that the self-esteem effect differs between men and women. On the log-odds graph, the two lines are almost parallel.

Odds ratios: one number for the slope

A one-point change in self-esteem moves p by different amounts depending on where you start, so B cannot be read as a change in probability. Instead, interpret $\exp(B)$ – the **odds ratio (OR)**.

Definition: the odds ratio

The **odds ratio** is the number the odds are multiplied by when the predictor increases by one unit:

$$\text{OR} = \frac{\text{odds}(X+1)}{\text{odds}(X)} = \exp(B).$$
 Unlike the change in probability, it is the same number wherever you start.

How to read an OR

OR < 1: odds *decrease* as the predictor increases. OR = 1: no association. OR > 1: odds *increase*. % change = $(\text{OR} - 1) \times 100$.

Example: $B_{\text{self-esteem}} = -0.101 \Rightarrow \exp(B_{\text{self-esteem}}) = \text{OR} = 0.904$

- OR = 0.904 < 1: odds *decrease* as self-esteem rises; $(0.904 - 1) \times 100 = -9.6\%$, “odds reduced 9.6% per point”
- for 5 points: $0.904^5 = 0.604$ (reduced 40%)
- for 10 points: $0.904^{10} = 0.365$ (reduced 64%)

Odds ratios for regression coefficients

In R

`exp(coef(mod3))` then `exp(confint(mod3))`

(Intercept)	c_self_esteem	gender	c_self_esteem:gender
0.6912402	0.8837109	1.8452026	1.0378529

	2.5 %	97.5 %
(Intercept)	0.4887982	0.9687986
c_self_esteem	0.8174421	0.9499398
gender	1.1317452	3.0308171
c_self_esteem:gender	0.9377623	1.1493832

- an OR of 1 means no change – a 95% CI is “significant” exactly when it *excludes* 1 (the OR-scale equivalent of a log-odds CI excluding 0)
- **self-esteem**: OR=0.884, CI [0.82, 0.95] – entirely *below* 1: a significant association, odds of high depression fall as self-esteem rises, each point multiplies the odds by somewhere between 0.82 and 0.95
- **gender**: OR=1.85, CI [1.13, 3.03] – entirely *above* 1: a significant association, women’s odds of high depression are higher than men’s, 13-203% higher (point estimate: 85% higher)
- **interaction**: OR=1.04, CI [0.94, 1.15] – straddles 1: *not* significant, we cannot rule out that the self-esteem effect is the same for women and men

Model 3, on the log-odds scale and (exponentiated) on the odds scale

$$\log\left(\frac{p}{1-p}\right) = -0.369 - 0.124(\text{cSE}) + 0.613(\text{gender}) + 0.037(\text{cSE} \times \text{gender})$$

$$\frac{p}{1-p} = 0.691 \times 0.884^{\text{cSE}} \times 1.845^{\text{gender}} \times 1.038^{\text{cSE} \times \text{gender}}$$

Start with the reference person: a man (gender = 0) with average self-esteem (cSE = 0). Every term with cSE or gender drops out, leaving only the intercept.

Scale	Calculation	Value
log-odds	B_0	-0.369
odds	$\exp(-0.369)$	0.691
probability	$0.691/(1 + 0.691)$	0.409

In words

Odds of high depression of 0.69 to 1 (about 7 to 10): a 41% chance. Note that $\exp(B_0) = 0.691$ is the *baseline odds*, not an odds ratio – it is the starting point that every odds ratio multiplies.

Now switch gender to 1 (woman), keeping self-esteem at average ($cSE = 0$): the gender term is added.

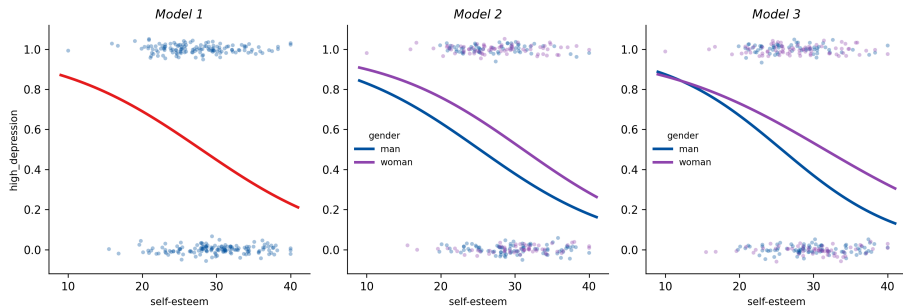
Scale	Calculation	Value
log-odds	$-0.369 + 0.613$	0.243
odds	$\exp(0.243) = 0.691 \times 1.845$	1.275
probability	$1.275 / (1 + 1.275)$	0.561

Odds ratio for gender: $\exp(0.613) = 1.845$

A woman's odds are 1.85 times a man's (85% higher): adding 0.613 to the log-odds is the same as multiplying the odds by 1.845. In probabilities: 0.41 for the average man vs. 0.56 for the average woman.

Model selection

which model is best?



Model 1: log-Lik = -185.21 (2 estimated parameters) Model 2: log-Lik = -182.21 (3 estimated parameters) Model 3: log-Lik = -181.95 (4 estimated parameters)

- read top to bottom, log-Lik gets closer to 0 (better) each time: Model 3 looks “best” by this number alone
- but Model 3 also has the most estimated parameters (4), and adding *any* parameter, useful or not, can only raise (never lower) the likelihood – exactly as R^2 only ever rose as predictors were added in OLS
- so “highest likelihood” alone is not a fair basis for selection; the next few slides build the actual test: is Model 3’s small extra gain more than we would expect just from having one more parameter to play with?

Likelihood-ratio test: the logic

Problem

If we always pick the model with the higher likelihood, we always pick the more complicated model – taken to the extreme, a “saturated” model with one parameter per data point has $LL = 0$ (perfect fit) but is useless for inference.

Solution

Adding parameters always increases LL somewhat by chance alone. Ask: how much increase would we expect if the simpler model (H_0) were actually true?

$$\chi^2 = -2(LL_1 - LL_2) \sim \chi^2_{(df_2 - df_1)} \quad (10)$$

- LL_1, LL_2 : the log-likelihoods of the simpler and more complex model (in that order); $df_2 - df_1$: how many *extra* parameters the complex model has
- χ^2 : the test statistic itself (just a name – it is the same quantity as $-2(LL_1 - LL_2)$); it is compared to a χ^2 (chi-squared) distribution with $df_2 - df_1$ degrees of freedom to get a p -value

The likelihood-ratio test asks whether the observed increase in LL is significantly larger than that.

Model selection: starting from nothing

Where we always start: the null model

Before comparing our own models, we need a baseline to improve on: the **null model**, `high_depression ~ 1`, predicting the same log-odds for everyone (no predictors – just the sample’s overall log-odds). Every model we fit should, at minimum, improve on this.

Careful: R’s own “Model 1”/“Model 2” labels reset every call

`lrtest()` calls whichever model you list *first* “Model 1” and the next “Model 2”, regardless of our own Model 1/2/3 numbering below – in the left block “Model 2” means the null model; on the right it means self-esteem+gender. Read the formulas, not the labels.

```
Model 1: ~ self_esteem
Model 2: ~ 1
#Df LogLik Df  Chisq  Pr
1  2 -185.21
2  1 -193.90 -1  17.394 ***
```

```
Model 1: ~ self_esteem
Model 2: ~ self_est.+gender
#Df LogLik Df  Chisq  Pr
1  2 -185.21
2  3 -182.21  1  5.991  *
```

- our Model 1 (self-esteem) improves on the null model: $\chi^2(1) = 17.39$, $p < .001$ – self-esteem alone is worth adding
- our Model 2 (self-esteem + gender) improves on our Model 1: $\chi^2(1) = 5.99$, $p = .014$ – gender adds real information

Model selection: Model 2 vs. Model 3

Likelihood ratio test

Model 1: `high_depression ~ self_esteem + gender`

Model 2: `high_depression ~ c_self_esteem * gender`

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	3	-182.21			
2	4	-181.95	1	0.5173	0.4720

Model 3 is *not* significantly better than Model 2

$\chi^2(1) = 0.52$, $p = .472$ – the self-esteem $\times$ gender interaction does not improve fit. **Prefer the simpler Model 2.**

Model selection: but what about goodness-of-fit?

- in OLS, we have $R^2 = 1 - \text{RSS}/\text{TSS}$ (proportion of variance explained) – not possible in the same form for maximum-likelihood-based models
- define a “Pseudo”- R^2 with similar behaviour instead: $R^2_{\text{pseudo}} \in [0, 1]$, close to 0 = bad fit, close to 1 = excellent fit (perhaps overfit); many versions exist, McFadden’s (1977) is common:

$$R^2_{\text{McFadden}} = 1 - \frac{LL_{\text{model}}}{LL_{\text{null}}} \quad (11)$$

Our three models ($LL_{\text{null}} = -193.90$)

Model 1: $1 - (-185.21)/(-193.90) = .045$ Model 2: $1 - (-182.21)/(-193.90) = .060$ Model 3:
 $1 - (-181.95)/(-193.90) = .062$

Then why did we prefer Model 2 on the previous slide?

Pseudo- R^2 can only *rise* (never fall) as parameters are added – like OLS R^2 , it does not penalize complexity. Model 3’s rise (.060 → .062) is tiny, and the LR test already told us it is no more than chance alone would produce ($\chi^2(1) = 0.52$, $p = .472$). The two are not in conflict: pseudo- R^2 says “fits at least as well,” the LR test says “not significantly better” – the LR test is built to answer whether an improvement is real, so it decides.

Logistic regression as a classifier

- so far we used logistic regression to *explain*: which predictors matter, and by how much
- it can also *predict*: guess, for a new person, whether they will be high-depression or not
- the model already gives a rule: if the predicted $P(Y = 1) > 0.5$, classify as 1 (“high depression”), otherwise as 0
- this is “supervised” learning – the model learns from 280 people whose true outcome we already know – and logistic regression is a standard baseline that fancier classifiers must improve on
- the real test is not how well the rule fits those 280 people, but how well it works on people it has never seen

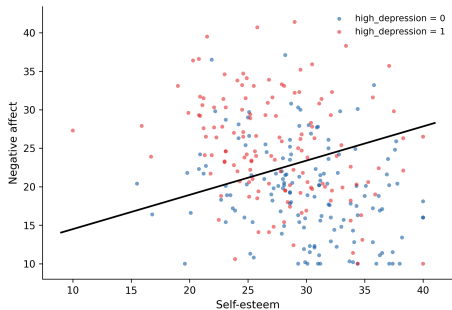
What is it?

A cross-tabulation: every case sorted into one of four boxes by *what the model predicted* vs. *what actually happened* – “confusion” because it shows exactly which kinds of cases get mislabeled, not just how often.

		Predicted	
		No	Yes
Observed	No	True Negative (TN)	False Positive (FP)
	Yes	False Negative (FN)	True Positive (TP)

- maximize TN and TP (correct) – the diagonal
- minimize FN and FP (errors) – off the diagonal

Decision boundary for our own data



`high_depression ~ self_esteem + neg_affect` – a new model, used for the rest of this section, chosen for two *continuous* predictors so both fit on the axes (not one of Models 1-3).

- dots = the 280 people, colored by their *true* outcome; the model itself never sees this color, only the (x, y) position
- one predictor $\rightarrow$ the $p = 0.5$ cutoff is a point; two predictors $\rightarrow$ it is a whole **line**: solve $B_0 + B_{se}(se) + B_{na}(na) = 0$ for `neg_affect`
- above/left of the line: classified “high depression”; red dots below/right and blue dots above/left are mistakes

Confusion matrix for our model

In R

```
predicted <- as.integer(  
  predict(modC, type="response") > 0.5)  
confusionMatrix(factor(predicted),  
  factor(clinical_sample$high_depression),  
  positive="1")
```

	Reference	
Prediction	0	1
0	108	50
1	37	85

```
Accuracy : 0.6893  
No Information Rate : 0.5179  
Sensitivity : 0.6296  
Specificity : 0.7448  
Pos Pred Value : 0.6967  
Neg Pred Value : 0.6835  
Kappa : 0.3757  
Prevalence : 0.4821
```

What the matrix shows

Prediction (rows) = the model's guess; **Reference** (columns) = the truth. Correct: 108 correctly cleared (true negatives) and 85 correctly flagged (true positives), so $(108 + 85)/280 = 0.689$ of all cases. Errors: 37 false alarms (flagged, but not high-depression) and 50 missed cases (high-depression, but not flagged) – the model misses more cases than it falsely flags.

Interpreting the confusion matrix output

Metric	Plain-English meaning	Value
Accuracy	overall, proportion of all 280 cases classified correctly	.689
No Info. Rate	accuracy from just guessing the <i>majority</i> class every time (a baseline to improve on)	.518
Sensitivity	of those <i>actually</i> high-depression, proportion correctly flagged	.630
Specificity	of those <i>actually not</i> high-depression, proportion correctly cleared	.745
Pos Pred Value	of those <i>flagged</i> as high-depression, proportion who truly are (a.k.a. precision)	.697
Neg Pred Value	of those <i>cleared</i> , proportion who truly are not	.684
Kappa	agreement between prediction and reality, corrected for chance agreement (0 = chance, 1 = perfect)	.376
Prevalence	base rate: proportion <i>actually</i> high-depression in this sample	.482

Bold rows are the ones to report: accuracy (compared with the no-information rate), sensitivity and specificity. The others are useful for specific purposes, such as judging a screening tool.

Cross-validation: will the model work on new people?

- so far, the model was tested on the same 280 people it learned from – a bit like grading students on the exact questions they practised
- **cross-validation**: fit the model on one part of the data (the “training set”), then check its predictions on the part it never saw (the “test set”); repeat with different splits and average the results
- why it matters: a model can fit the quirks of its own sample (**overfitting**); cross-validation shows how well it works on new people

What this means for our results

Our 69% accuracy was measured on the same 280 people the model was fitted to, so it is probably a little too optimistic. Cross-validation would give a more honest estimate of how well the model classifies people it has not seen.

Practical logistic regression in R

```
library(tidyverse)
library(lmtest)      # lrtest()

mod1 <- glm(high_depression ~ self_esteem, data = clinical_sample,
            family = binomial(link = "logit"))
summary(mod1)        # coefficients (log-odds), deviances
exp(coef(mod1))       # odds ratios
exp(confint(mod1))    # confidence intervals for the odds ratios

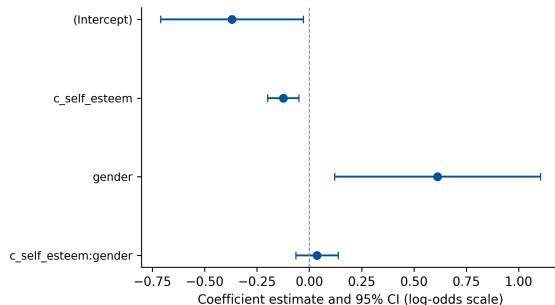
mod2 <- glm(high_depression ~ self_esteem + gender, data = clinical_sample,
            family = binomial(link = "logit"))
clinical_sample <- clinical_sample |>
  mutate(c_self_esteem = self_esteem - mean(self_esteem))
mod3 <- glm(high_depression ~ c_self_esteem * gender, data = clinical_sample,
            family = binomial(link = "logit"))

lrtest(mod1, mod2, mod3)  # likelihood-ratio tests, one step at a time
```

Coefficient plots

In R

```
library(modelsummary)
modelplot(mod3)
```

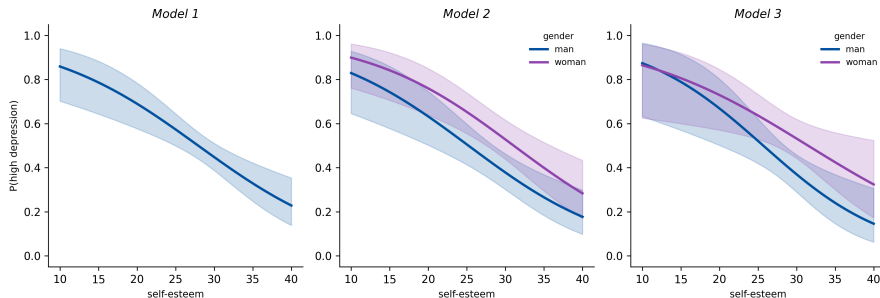


- each dot is a Model 3 coefficient (log-odds scale); each bar is its 95% confidence interval; the dotted line at 0 means “no effect”
- **c_self_esteem**: -0.12 , CI $[-0.20, -0.05]$, entirely below 0 – higher self-esteem, lower log-odds of high depression
- **gender**: 0.61 , CI $[0.12, 1.11]$, entirely above 0 – women have higher log-odds than men
- **interaction**: 0.04 , CI $[-0.06, 0.14]$, crosses 0 – no evidence that the self-esteem effect differs by gender
- **(Intercept)**: -0.37 , the log-odds for a man with average self-esteem

Why marginal predictions? Earlier we hand-computed p at two or three chosen values (e.g., self-esteem 15 vs. 35). This does the same computation at every value and draws the curve with a confidence band – one picture instead of many hand calculations.

In R

```
library(marginaleffects)
plot_predictions(mod1, condition=list("self_esteem"))
plot_predictions(mod2, condition=list("self_esteem","gender"))
plot_predictions(mod3, condition=list("c_self_esteem","gender"))
```



Summary

- a straight line cannot respect the 0/1 bounds of a binary outcome; the S-curve can (stretching the x -axis)
- the logit stretches the probability axis into log-odds, where the S-curve becomes a straight line (stretching the y -axis)
- that line is fitted by **maximum likelihood**, not least squares; deviance = $-2 \times \log\text{-likelihood}$
- coefficients are on the log-odds scale; **odds ratios** ($\exp(B)$) make them readable
- nested models are compared with **likelihood-ratio tests**; **pseudo- R^2** gives a rough sense of fit
- logistic regression doubles as a simple **classifier** – confusion matrices and cross-validation evaluate that use

Thank you!