

Correlation, Simple & Multiple Regression

PSY3100 – A General 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 covariance to correlation
- Simple linear regression – continuous *and* dummy predictors
- Multiple linear regression – continuous, categorical, and mixed predictors
- Interaction (moderation) effects – continuous, dummy, and mixed moderators, simple slopes, centering
- Goodness of fit: R^2 , adjusted R^2 , RSE/RMSE, the F -test
- Relative importance of predictors: standardized β , partial/part correlations, LMG
- Assumptions (Kline's LINE) and regression diagnostics – checking them on a fitted model
- Common practical challenges
- Suppression effects – classical, net, cooperative
- Bridge to lavaan: regression as your first path model

Our running example

A simulated clinical sample ($N = 280$)

We will use one simulated dataset, `clinical_sample.csv`, so that every technique builds on the same numbers.

Variable	Type	Description	<i>M</i>	<i>SD</i>
<code>self_esteem</code>	continuous	Rosenberg Self-Esteem Scale, 10–40	28.68	5.18
<code>neg_affect</code>	continuous	PANAS Negative Affect, 10–50	22.17	7.00
<code>depression</code>	continuous	PHQ-9-like severity, 0–27 (outcome)	9.11	5.15
<code>treatment</code>	dummy (0/1)	0 = waitlist control, 1 = CBT	–	–
<code>gender</code>	dummy (0/1)	0 = man, 1 = woman	–	–

In R

```
clinical_sample <- read.csv("clinical_sample.csv")
```

Covariance: the starting point

What covariance tells us

Covariance quantifies the **direction** of the linear relationship between two variables, and how much they **vary together**.

$$s_{XY} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{N - 1} \quad (1.1)$$

where N = sample size, X_i, Y_i = the value of X/Y for person i , $\bar{X}, \bar{Y}$ = sample means.

- Positive s_{XY} : as X increases, Y tends to increase too.
- Negative s_{XY} : as X increases, Y tends to decrease.
- Problem: the **magnitude** of s_{XY} depends on the scales of X and Y – it cannot be compared across variable pairs.

Our covariance matrix

In R

```
cov(clinical_sample[, c("self_esteem", "neg_affect", "depression")])
```

	self_esteem	neg_affect	depression
self_esteem	26.83	-11.90	-8.37
neg_affect	-11.90	48.99	16.81
depression	-8.37	16.81	26.52

- Self-esteem and depression covary **negatively**: -8.37 .
- Negative affect and depression covary **positively**: 16.81 .
- The diagonal holds each variable's own **variance** (Equation [1.1](#) with $Y = X$).

Correlation: standardized covariance

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} \quad (1.2)$$

where s_{XY} = covariance of X and Y (Equation 1.1), s_X, s_Y = standard deviations of X and Y .

- Correlation rescales covariance by the standard deviations of both variables.
- Result: a **scale-free** index that always ranges from -1 to $+1$.
- $|r|$ close to 1: strong linear association. r near 0: weak/no linear association.

Our data

$$r_{\text{self-esteem, depression}} = \frac{-8.37}{5.18 \times 5.15} = -0.314$$

Our correlation matrix

In R

```
cor(clinical_sample[, c("self_esteem", "neg_affect", "depression")])
```

	self_esteem	neg_affect	depression
self_esteem	1.000	-0.328	-0.314
neg_affect	-0.328	1.000	0.466
depression	-0.314	0.466	1.000

Now directly comparable: negative affect ($r = .466$) has a **stronger** linear association with depression than self-esteem does ($r = -.314$).

Factors affecting correlation

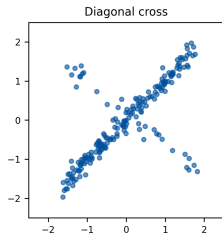
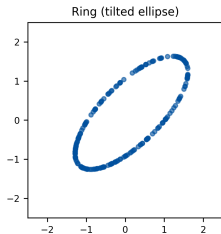
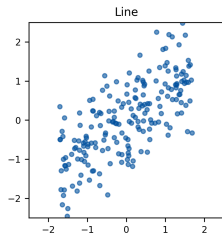
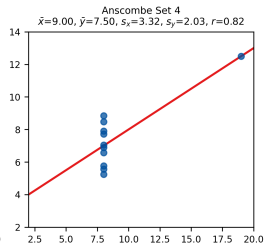
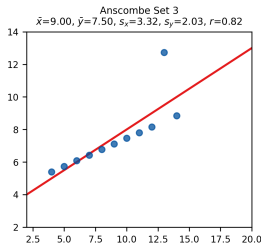
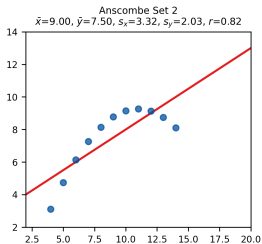
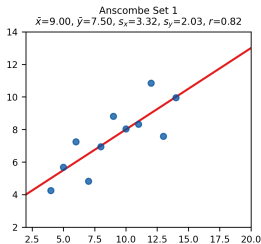
Factor	Effect
Nonlinearity	r only captures linear association – a strong curved relationship can show $r \approx 0$
Range restriction	Sampling from a narrow slice of X (e.g., only high scorers) shrinks $ r $
Outliers	A handful of extreme points can inflate <i>or</i> deflate r
Measurement error	Unreliable measures attenuate (weaken) the observed r relative to the true association

Modern practice

Always **plot** your data before trusting a correlation coefficient – Anscombe's quartet (1973) and the more recent "Datasaurus Dozen" (Matejka & Fitzmaurice, 2017) both show wildly different scatterplots sharing the same r .

Why we plot our data

Same summary statistics, very different data



All 3 panels:
mean=(0,0)
SD=(1,1)
 $r=0.70$

From correlation to regression

The key shift

Correlation is **symmetric** – X and Y play identical roles. Regression is **asymmetric** – we assign one variable the role of **predictor** (X) and the other the role of **criterion** (Y).

$$Y_i = B_0 + B_1X_i + \varepsilon_i \quad (2.1)$$

where Y_i = the *observed* value of Y for person i , B_0 = intercept, B_1 = slope, X_i = observed value of X for person i , ε_i = residual/error for person i (how far person i 's actual score falls from the line).

Predicted vs. observed: do we always need ε ?

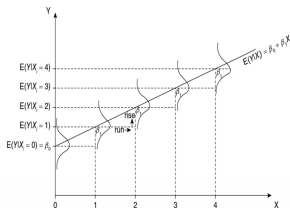
No – only when writing an equation for an actual data point

The **predicted (expected) value** equation has **no error term**, because ε has an expected value of zero:

$$E(Y | X) = \hat{Y} = B_0 + B_1 X \quad (2.2)$$

where $E(Y | X)$ = the expected (average) value of Y at a given value of X ; $\hat{Y}$ = the model's prediction (the "hat" always means "predicted/estimated," never an actual observation).

- Equation 2.1 (with ε_i) describes an **individual, observed** data point.
- Equation 2.2 (no ε) describes the **line itself** – what we predict *on average* at a given X .



Ordinary Least Squares (OLS)

How the line is chosen

OLS picks B_0 and B_1 to **minimize the sum of squared residuals (SSR)**:

$$SSR = \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \rightarrow \min_{B_0, B_1} \quad (2.3)$$

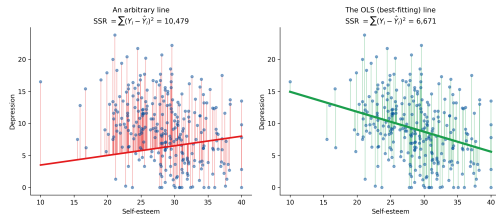
where $Y_i - \hat{Y}_i$ = the residual for person i (observed minus predicted); squaring and summing across all N people gives one number (SSR) that OLS minimizes by choice of B_0, B_1 . Squaring makes all deviations positive and penalizes large errors more heavily; the resulting line passes through $(\bar{X}, \bar{Y})$.

Where the slope formula comes from

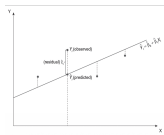
Take Equation 2.3, do the calculus (set the derivative with respect to B_1 to zero and solve), and this falls out: the unstandardized slope, rewritten in terms of the correlation and the two SDs.

$$B_1 = r_{XY} \frac{s_Y}{s_X} \quad (2.4)$$

OLS, visualized: why "least squares"



Left: an arbitrary line through the data (red), $SSR = 10,479$. Right: the actual OLS line (green), $SSR = 6,671$ – smaller than *every* other line you could draw, which is what "least squares" means.



For a single point: the vertical gap between the observed Y_i and the predicted $\hat{Y}_i$ on the line is the residual $\hat{\epsilon}_i$ – the quantity Equation 2.3 squares and sums across all N people.

Example: self-esteem predicting depression

$$\hat{Y}_{\text{Depression}} = B_0 + B_1 X_{\text{Self-esteem}} \quad (2.5)$$

In R

```
m1 <- lm(depression ~ self_esteem, data = clinical_sample)
summary(m1); confint(m1)
```

Coefficients:

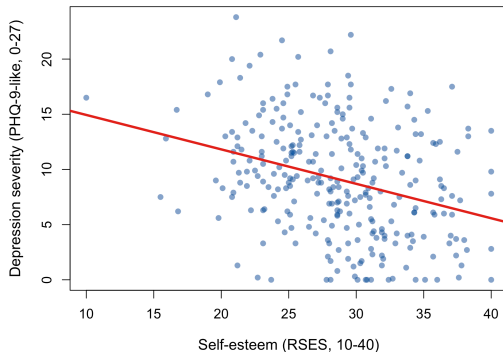
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	18.0614	1.6500	10.947	< 2e-16 ***
self_esteem	-0.3121	0.0566	-5.512	8.09e-08 ***

Residual standard error: 4.899 on 278 degrees of freedom
Multiple R-squared: 0.0985, Adjusted R-squared: 0.0953
F-statistic: 30.38 on 1 and 278 DF, p-value: 8.088e-08

Interpretation

Each 1-point increase in self-esteem is associated with a 0.312-point **decrease** in predicted depression severity (95% CI -0.42 to -0.20 , which excludes 0). The intercept ($B_0 = 18.06$) is the predicted depression score at self-esteem=0, outside the observed range here, so not directly meaningful on its own. The model overall is significant, $F(1, 278) = 30.38$, $p < .001$, and explains $R^2 = .099$ (9.9%) of the variance in depression ($R^2_{\text{adj}} = .095$); RSE = 4.90 on 278 df is the typical size of a prediction error, in the same units as depression.

Simple regression: visualized



$$\hat{Y}_{\text{Depression}} = 18.06 - 0.312 X_{\text{Self-esteem}}, \quad R^2 = .099$$

At self-esteem= 10: $\hat{Y} = 18.06 - 0.312(10) = 14.94$. At self-esteem= 40: $\hat{Y} = 18.06 - 0.312(40) = 5.58$ – the two ends of the red line above.

Unstandardized vs. standardized β

Unstandardized (B)

- In the original units of X and Y
- Directly interpretable: "1 point of $X \rightarrow B$ points of Y "
- Cannot compare across predictors on different scales

Standardized (β)

- Both X and Y are z-scored first
- In standard-deviation units
- In simple regression, $\beta = r_{XY}$ exactly

Check (Equation 2.4, rearranged)

$$\beta = B_1 \frac{s_X}{s_Y} = -0.312 \times \frac{5.18}{5.15} = -0.314 = r_{XY}$$

In R

```
lm.beta::lm.beta(m1)
```

A predictor does not have to be continuous

Dummy coding

A 2-level categorical predictor can enter a regression as a **dummy variable**: one group coded $X = 0$ (the **reference** category), the other $X = 1$.

$$\hat{Y}_{\text{Depression}} = B_0 + B_1 X_{\text{Treatment}}, \quad X_{\text{Treatment}} \in \{0, 1\} \quad (2.6)$$

- $X = 0$ (waitlist control): $\hat{Y} = B_0$ – the intercept **is** the control-group mean.
- $X = 1$ (CBT): $\hat{Y} = B_0 + B_1$ – and B_1 **is** the mean difference between groups.

This is equivalent to an independent-samples t -test

A simple regression with one 2-level dummy predictor is mathematically identical to an independent-samples t -test: B_1 equals the group mean difference, and its t -value equals the t -test's t -value exactly.

Example: treatment group predicting depression

$$\hat{Y}_{\text{Depression}} = B_0 + B_1 X_{\text{Treatment}} \quad (2.7)$$

In R

```
m2 <- lm(depression ~ treatment, data = clinical_sample)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.6754	0.4190	25.480	< 2e-16 ***
treatment	-3.0852	0.5883	-5.244	3.12e-07 ***

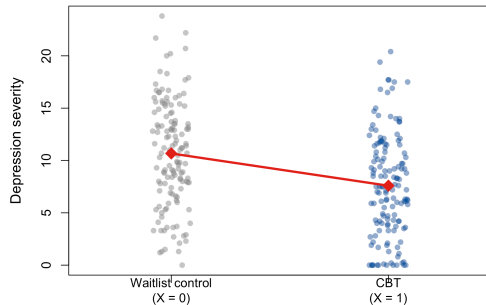
Residual standard error: 4.922 on 278 degrees of freedom
Multiple R-squared: 0.0900, Adjusted R-squared: 0.0867

```
> aggregate(depression ~ treatment, clinical_sample, mean)
  treatment depression
1         0      10.68
2         1       7.59
```

Interpretation

$B_0 = 10.68$: predicted depression for a waitlist-control participant. $B_1 = -3.09$: CBT participants score 3.09 points lower on average (95% CI -4.24 to -1.93) – and $B_0 + B_1 = 7.59$ exactly matches the CBT-group mean.

Dummy regression: visualized



$$\hat{Y}_{\text{Depression}} = 10.68 - 3.09 X_{\text{Treatment}}$$

At treatment= 0 (control): $\hat{Y} = 10.68 - 3.09(0) = 10.68$. At treatment= 1 (CBT): $\hat{Y} = 10.68 - 3.09(1) = 7.59$ – exactly the two group means.

Why stop at one predictor?

The GLM idea

Outcomes (like depression severity) are rarely explained by a single variable. **Multiple regression** extends the same linear model to several predictors at once.

$$\hat{Y} = B_0 + B_1X_1 + B_2X_2 + \cdots + B_kX_k \quad (3.1)$$

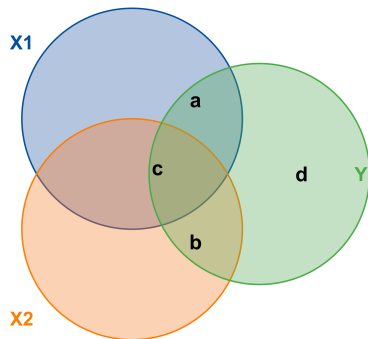
where k = number of predictors, B_j = the partial regression coefficient for predictor X_j ($j = 1, \dots, k$).

- Each B_j is now a **partial** effect: the expected change in Y per unit of X_j , **holding all other predictors constant**.
- Predictors can be any mix of continuous and categorical (dummy-coded).

Unique vs. shared variance

- When predictors **correlate with each other**, some of what each predictor explains about Y overlaps.
- A regression coefficient B_j reflects only X_j 's **unique** contribution, over and above the other predictors.
- This is exactly why B_1 in a multiple regression is usually *not* the same number as B_1 from the simple regression of Y on X_1 alone.

Partial vs. part correlation – the Venn picture



Partial r : the correlation between X_1 and Y once the variance shared with X_2 is removed from *both*. **Part (semipartial) r :** the same, but removed only from X_1 – Y is left untouched.

In our example: X_1 =self-esteem, X_2 =negative affect, Y =depression.

- a = variance self-esteem shares **uniquely** with depression (not via negative affect)
- b = variance negative affect shares **uniquely** with depression
- c = variance self-esteem and negative affect **both** explain in depression (their overlap)
- d = variance in depression **neither** predictor explains
- Partial r for self-esteem $\approx \sqrt{\frac{a}{a+d}}$: self-esteem's unique slice, relative to what is *left over* after removing the c shared with negative affect
- Part r for self-esteem $\approx \sqrt{\frac{a}{a+b+c+d}}$: self-esteem's unique slice, relative to *all* of depression's variance

Example 1: two continuous predictors

$$\hat{Y}_{\text{Depression}} = B_0 + B_1 X_{\text{Self-esteem}} + B_2 X_{\text{Neg. affect}} \quad (3.2)$$

In R

```
m3 <- lm(depression ~ self_esteem + neg_affect, data = clinical_sample)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	7.6087	2.0721	3.672	0.00029	***
self_esteem	-0.1792	0.0549	-3.264	0.00123	**
neg_affect	0.2996	0.0406	7.373	1.93e-12	***

Residual standard error: 4.487 on 277 degrees of freedom
Multiple R-squared: 0.2464, Adjusted R-squared: 0.2410
F-statistic: 45.29 on 2 and 277 DF, p-value: 9.598e-18

No special classical name

Unlike the examples on the next two slides, two continuous predictors together is just "plain" multiple regression – there is no separate classical-test name for this combination.

Example 2: two categorical predictors

$$\hat{Y}_{\text{Depression}} = B_0 + B_1 X_{\text{Treatment}} + B_2 X_{\text{Gender}} \quad (3.3)$$

In R

```
m4 <- lm(depression ~ treatment + gender, data = clinical_sample)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.8103	0.4895	20.040	< 2e-16 ***
treatment	-3.1276	0.5785	-5.407	1.39e-07 ***
gender	1.8948	0.5796	3.269	0.00122 **

Residual SE: 4.838 on 277 df Mult. R2: 0.1238 Adj. R2: 0.1175
F-statistic: 19.57 on 2 and 277 DF, p-value: 1.120e-08

Reading the coefficients

equivalent to a 2×2 factorial ANOVA

$B_0 = 9.81$: reference combination, waitlist control + man ($X_1 = X_2 = 0$). $B_1 = -3.13$: holding gender constant, CBT predicts 3.13 points lower depression. $B_2 = 1.89$: holding treatment constant, women score 1.89 points higher than men. Two categorical predictors together is exactly a 2×2 factorial ANOVA (without the interaction term) – same linear model, different name.

Example 3: mixed predictors

$$\hat{Y}_{\text{Depression}} = B_0 + B_1 X_{\text{Self-esteem}} + B_2 X_{\text{Treatment}} \quad (3.4)$$

In R

```
m5 <- lm(depression ~ self_esteem + treatment, data = clinical_sample)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	19.5410	1.5927	12.269	< 2e-16 ***
self_esteem	-0.3096	0.0539	-5.747	2.39e-08 ***
treatment	-3.0583	0.5571	-5.489	9.12e-08 ***

Residual SE: 4.661 on 277 df Mult. R2: 0.1870 Adj. R2: 0.1811
F-statistic: 31.85 on 2 and 277 DF, p-value: 3.544e-13

Reading the coefficients

this is an ANCOVA

$B_0 = 19.54$: waitlist-control participant with self-esteem = 0 (an anchor, outside observed range). $B_1 = -0.310$: holding treatment constant, each self-esteem point predicts 0.310 points lower depression. $B_2 = -3.06$: holding self-esteem constant, CBT predicts 3.06 points lower depression. A continuous covariate plus a categorical group predictor is exactly what an **ANCOVA** is – same linear model, different name.

Same model, many names

- **Linear regression model** – the general statistical term
- **Ordinary Least Squares (OLS)** – the estimation method used to fit it
- **General Linear Model (GLM)** – the broader family; t -test, ANOVA, ANCOVA, and regression are all special cases

Note: two different "GLMs"

Do not confuse the *General* Linear Model (this lecture) with the *Generalized* Linear Model (logistic, Poisson regression, etc.) – confusingly similar names, genuinely different things. We cover the generalized version (logistic regression) next.

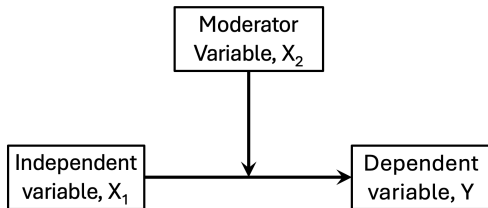
Interaction and Moderation Effects

What this section covers

Does the effect of one predictor on the outcome depend on the level of another predictor?

- What an interaction (moderation) effect is, and the product-term regression model
- Main effects vs. **conditional** main effects, once an interaction is in the model
- Simple slopes analysis – the effect of X_1 on Y at chosen values of the moderator
- Continuous $\times$ continuous, continuous $\times$ dummy, and dummy $\times$ dummy moderators
- When to keep vs. drop a non-significant interaction, and how to center your predictors

What is an interaction effect?



- So far, every model in this lecture has assumed each predictor's effect on Y is **constant**, regardless of the other predictors.
- An **interaction** (moderation) effect happens when the relationship between X_1 and Y depends on the value of X_2 .
- We test for this by adding a third predictor to the model: the **product term** $X_3 = (X_1 \times X_2)$.
- If X_3 's coefficient is significant, X_1 's effect on Y genuinely depends on X_2 – an interaction is present.

Interaction using regression

Main-effects (additive) model

Additive: X_1 and X_2 each contribute independently to Y – their effects simply add together, with no term capturing whether one predictor's effect changes depending on the other.

$$\hat{Y} = B_0 + B_1X_1 + B_2X_2$$

Interaction (product-term) model

$$\hat{Y} = B_0 + B_1X_1 + B_2X_2 + B_3X_1X_2 \quad (4.1)$$

- X_1X_2 is the **product term** – literally X_1 multiplied by X_2 , added as a third predictor.
- B_3 is the **interaction coefficient**: how much the effect of X_1 on Y changes for every 1-unit increase in X_2 .
- An interaction is **over-additive** (enhancing, B_3 same sign as B_1) or **under-additive** (dampening, B_3 opposite sign), relative to just adding the two main effects.

Main effects become *conditional* main effects

Without the interaction term (Equation 3.1)

B_1 is X_1 's effect on Y , assumed to be the **same at every value** of X_2 .

With the interaction term (Equation 4.1)

B_1 is now only X_1 's effect on Y **when $X_2 = 0$** – a *conditional* main effect, not a universal one.

Set $X_2 = 0$ in Equation 4.1: $\hat{Y} = B_0 + B_1X_1 + B_2(0) + B_3X_1(0) = B_0 + B_1X_1$

- Symmetric statement for B_2 : it is X_2 's effect on Y when $X_1 = 0$.
- If $X_2 = 0$ is outside the observed range (e.g., a continuous variable that never actually hits 0), B_1 is no longer a meaningful number on its own – exactly why **centering** (later in this section) matters once an interaction is in the model.

Simple slopes analysis

The question

What is the effect of X_1 on Y , **specifically** when the moderator X_2 takes some chosen value w ?

$$\text{Simple slope of } X_1 \text{ at } X_2 = w: \quad B_1 + B_3 w \quad (4.2)$$

- Substitute $X_2 = w$ into Equation 4.1 and collect the terms multiplying X_1 : $\hat{Y} = (B_0 + B_2 w) + (B_1 + B_3 w)X_1$.
- The conditional **intercept** is $B_0 + B_2 w$; the conditional **slope** is $B_1 + B_3 w$ – both change as w changes.
- Conventionally evaluated at a few meaningful values of w : the moderator's mean, ± 1 SD, or (for a dummy moderator) its two possible values, 0 and 1.
- Each simple slope has its own standard error (accounting for the covariance between B_1 and B_3) and can be tested for significance in its own right.

A second dataset for interaction effect

New dataset

If we test an interaction on `clinical_sample.csv` (our running dataset all lecture), it comes back **non-significant** – self-esteem $\times$ negative affect: $B_3 = -0.012$, $p = .120$. That is not a mistake: this dataset was simulated to be purely **additive**, so every interaction we could test on it is null by construction.

`interaction_demo.csv`

A second, purpose-built companion dataset – same variables, same scales, same story ($N = 280$), but with three true interaction effects built in.

The next three worked examples use `interaction_demo.csv`. We return to the real `clinical_sample.csv` result above at the end of this section, once we have built the tools to interpret it properly.

Interaction between continuous predictor and continuous moderator

Does the (protective) effect of self-esteem on depression depend on how much negative affect someone reports? A **buffering** hypothesis: self-esteem should matter more when negative affect is high.

$$\hat{Y}_{\text{Depression}} = B_0 + B_1 X_{\text{Self-esteem}} + B_2 X_{\text{Neg. affect}} + B_3 (X_{\text{Self-esteem}} \times X_{\text{Neg. affect}}) \quad (4.3)$$

In R

```
m7 <- lm(depression ~ self_esteem * neg_affect, data = interaction_demo)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.1604	5.1780	-0.417	0.67683
self_esteem	0.2485	0.1718	1.447	0.14917
neg_affect	0.9199	0.2312	3.979	0.00009 ***
self_esteem:neg_affect	-0.0221	0.0079	-2.803	0.00543 **

Residual standard error: 4.131 on 276 degrees of freedom

Multiple R-squared: 0.2576, Adjusted R-squared: 0.2495

F-statistic: 31.92 on 3 and 276 DF, p-value: 9.584e-18

Interpretation

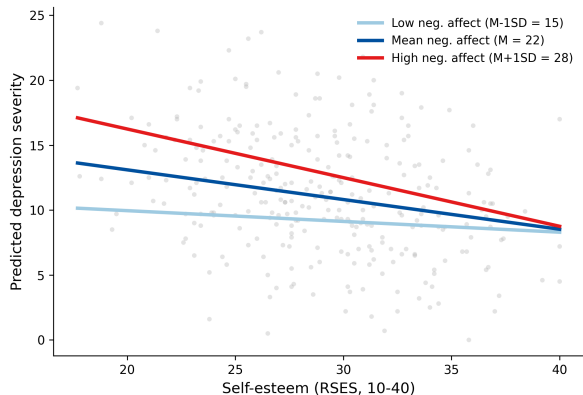
The interaction is significant, $B_3 = -0.0221$, $p = .005$: the already-negative (protective) slope of self-esteem gets **more negative** as negative affect increases, consistent with the buffering hypothesis. The model explains $R^2 = .258$ of the variance in depression, $F(3, 276) = 31.92$, $p < .001$, $RSE = 4.13$.

Interaction between continuous predictor and continuous moderator

Negative affect: $M = 21.56$, $SD = 6.59$. Simple slope of self-esteem at chosen levels (Equation 4.2, $B_1 + B_3w$):

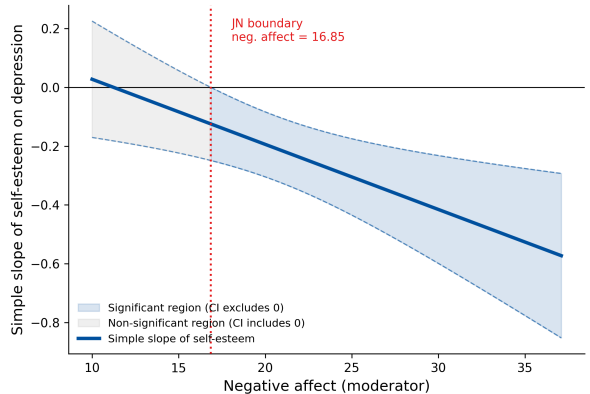
Neg. affect	slope	t	p
Low ($M-1SD=15$)	-0.083	-1.16	.246
Mean ($M=22$)	-0.229	-4.05	< .001
High ($M+1SD=28$)	-0.375	-4.57	< .001

As negative affect rises, self-esteem's protective slope gets steeper (more negative) – exactly the buffering pattern the interaction coefficient implied.



The Johnson-Neyman Technique

- Below 16.85: slope not significant. Above it: significantly negative (protective)
- Explains the previous slide directly: Low ($w = 15$) sat just below the boundary; Mean/High ($w = 22, 28$) sat above it



Interaction between continuous predictor and dummy moderator

Does the effect of self-esteem on depression depend on gender?

$$\hat{Y}_{\text{Depression}} = B_0 + B_1 X_{\text{Self-esteem}} + B_2 X_{\text{Gender}} + B_3 (X_{\text{Self-esteem}} \times X_{\text{Gender}}) \quad (4.4)$$

In R

```
m8 <- lm(depression ~ self_esteem * gender, data = interaction_demo)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	24.5674	2.2210	11.062	0.00000 ***
self_esteem	-0.5078	0.0765	-6.634	0.00000 ***
gender	-7.1588	3.2670	-2.191	0.02927 *
self_esteem:gender	0.3407	0.1112	3.064	0.00240 **

Residual standard error: 4.271 on 276 degrees of freedom
Multiple R-squared: 0.2066, Adjusted R-squared: 0.1980
F-statistic: 23.96 on 3 and 276 DF, p-value: 8.272e-14

Interpretation

The interaction is significant, $B_3 = 0.341$, $p = .002$: self-esteem's protective slope differs by gender – $B_1 = -0.508$ for men (gender= 0), $B_1 + B_3 = -0.167$ for women (gender= 1). The model explains $R^2 = .207$ of the variance in depression, $F(3, 276) = 23.96$, $p < .001$, $RSE = 4.27$.

Interaction between continuous predictor and dummy moderator

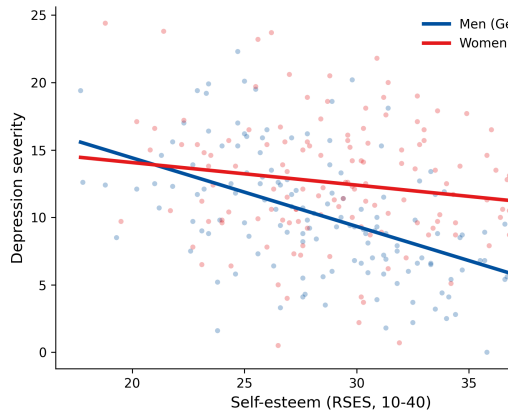
Equation 4.4, with $X_{\text{Gender}} = 0$ (men) and $= 1$ (women):

$$\hat{Y} = B_0 + B_1 X_{\text{SE}} + B_2 X_{\text{Gender}} + B_3 (X_{\text{SE}} \times X_{\text{Gender}})$$

$$\text{Men: } \hat{Y} = B_0 + B_1 X_{\text{SE}} = 24.57 - 0.508 X_{\text{SE}}$$

$$\begin{aligned} \text{Women: } \hat{Y} &= (B_0 + B_2) + (B_1 + B_3) X_{\text{SE}} \\ &= (24.57 - 7.16) + (-0.508 + 0.341) X_{\text{SE}} = 17.41 - 0.167 X_{\text{SE}} \end{aligned}$$

Men's slope $B_1 = -0.508$; women's slope $B_1 + B_3 = -0.167$. Self-esteem is a much stronger protective factor for men than for women in this (simulated) sample – $B_3 = 0.341$ is exactly the gap between the two slopes.



Interaction between dummy predictor and dummy moderator

Does CBT's effect on depression depend on gender?

$$\hat{Y}_{\text{Depression}} = B_0 + B_1 X_{\text{Treatment}} + B_2 X_{\text{Gender}} + B_3 (X_{\text{Treatment}} \times X_{\text{Gender}}) \quad (4.5)$$

In R

```
m9 <- lm(depression ~ treatment * gender, data = interaction_demo)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	10.8053	0.4940	21.873	0.00000	***
treatment	-1.6406	0.7164	-2.290	0.02276	*
gender	3.6663	0.6856	5.348	0.00000	***
treatment:gender	-3.1738	1.0325	-3.074	0.00232	**

Residual standard error: 4.278 on 276 degrees of freedom
Multiple R-squared: 0.2038, Adjusted R-squared: 0.1952
F-statistic: 23.55 on 3 and 276 DF, p-value: 1.327e-13

Interpretation

The interaction is significant, $B_3 = -3.174$, $p = .002$: CBT's benefit is larger for women than for men. Men: treatment effect = $B_1 = -1.64$; women: $B_1 + B_3 = -4.81$. The model explains $R^2 = .204$ of the variance in depression, $F(3, 276) = 23.55$, $p < .001$, $RSE = 4.28$.

Interaction between dummy predictor and dummy moderator

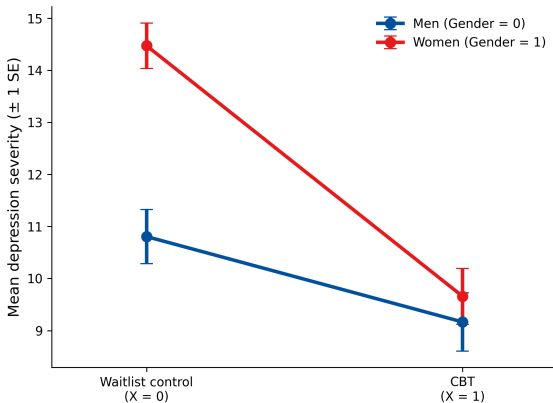
	Control	CBT
Men	10.81	9.16
Women	14.47	9.66

Simple effect of treatment (Control $\rightarrow$ CBT):

- Men: $B_1 = -1.64$
- Women: $B_1 + B_3 = -4.81$

CBT reduces depression for everyone, but the drop is roughly three times larger for women – a “difference of differences”:

$$(14.47 - 9.66) - (10.81 - 9.16) \approx 3.16 \approx |B_3|.$$



Sig. vs. non-sig. interaction: what to report

If you hypothesized the interaction *a priori*

Keep it in the model and report it, even if it turns out non-significant – that is itself a meaningful (null) result for a pre-registered hypothesis.

If you did not

Dropping a non-significant interaction is recommended – it gives a more parsimonious, less complex model, and the main effects become easier to interpret again.

- Significant interaction → interpret the interaction (simple slopes, as above).
- Non-significant interaction → drop the product term and interpret the (now unconditional again) main effects.

Back to our real data

On `clinical_sample.csv`, self-esteem $\times$ negative affect was non-significant ($B_3 = -0.012$, $p = .120$) – so we drop it. That leaves exactly the additive model from **Example 1** earlier in this lecture: $B_{\text{self-esteem}} = -0.179$, $B_{\text{neg. affect}} = 0.300$. The interaction test confirms the additive model was the right model.

What and why

Centering = subtracting a constant (usually the sample mean) from a continuous predictor before fitting the model: $X_c = X - \bar{X}$. You can technically center at any value, but the mean is the standard choice.

- Recommended for **continuous** predictors in an interaction model – the raw-score intercept and main effects are anchored at $X = 0$, often outside the observed range and not meaningful (self-esteem = 0 never occurs on a 10–40 scale).
- **Not** recommended for categorical/dummy predictors – $X = 0$ is already a real, observed group (e.g., men, or waitlist control), so there is no ambiguity to resolve by centering.
- Centering changes *what the intercept and main effects mean* – it does **not** change the interaction coefficient, R^2 , RSE, or the model's predicted values.

Centering, worked: raw vs. mean-centered

Self-esteem $\times$ negative affect model (`interaction_demo.csv`), mean-centered:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	11.0427	0.2565	43.049	0.00000	***
se_c	-0.2287	0.0565	-4.047	0.00007	***
na_c	0.2782	0.0392	7.089	0.00000	***
se_c:na_c	-0.0221	0.0079	-2.803	0.00543	**

Residual standard error: 4.131 on 276 degrees of freedom

Multiple R-squared: 0.2576, Adjusted R-squared: 0.2495

- **Intercept** (11.04): predicted depression for someone *average* on both self-esteem and negative affect – a real, meaningful point, unlike the raw model's $B_0 = -2.16$ (self-esteem = 0 and negative affect = 0, both outside the data).
- **Main effects** become **conditional-at-the-mean**: $B_{se_c} = -0.229$ is self-esteem's slope specifically when negative affect is at its mean – and matches the “Mean neg. affect” row of the simple-slopes table on slide 34 exactly.
- **Interaction** (-0.0221), R^2 (.2576), and RSE (4.131) are **identical** to the raw-score model – centering only reparameterizes, it does not refit a different model.

What this section covers

How well does the model, as a whole, fit the data?

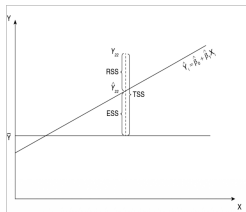
- R^2 – proportion of variance explained
- Adjusted R^2 – penalized for the number of predictors
- Residual standard error (RSE) / RMSE – typical prediction error, in the outcome's original units
- The F -test – does the whole model beat an intercept-only (mean-only) baseline?

R^2 : proportion of variance explained

$$R^2 = \frac{SS_{\text{regression}}}{SS_{\text{total}}} = 1 - \frac{SS_{\text{residual}}}{SS_{\text{total}}} \quad (5.1)$$

where $SS_{\text{total}} = \sum (Y_i - \bar{Y})^2$ (total variation in Y), $SS_{\text{regression}} = \sum (\hat{Y}_i - \bar{Y})^2$ (variation explained by the model), $SS_{\text{residual}} = \sum (Y_i - \hat{Y}_i)^2$ (unexplained variation, = SSR from Equation 2.3).

- Ranges from 0 (model explains nothing) to 1 (model explains everything).
- Equivalently: the squared correlation between observed Y and predicted $\hat{Y}$.



The problem with R^2 alone

R^2 never decreases

Adding *any* variable to a model – even pure noise – cannot decrease R^2 , and will typically increase it slightly by chance alone.

Adjusted R^2

$$R^2_{\text{adj}} = 1 - (1 - R^2) \frac{n - 1}{n - k - 1} \quad (5.2)$$

where n = sample size, k = number of predictors. Penalizes R^2 for the number of predictors relative to sample size.

Check, using our two-predictor model ($n = 280, k = 2$)

$$R^2_{\text{adj}} = 1 - (1 - .246) \frac{279}{277} = .241$$

In R

```
summary(m3)$r.squared; summary(m3)$adj.r.squared
```

Residual standard error (RSE) and RMSE

A goodness-of-fit measure in the outcome's own units

R^2 is unitless (0 to 1); RSE/RMSE instead answer "on average, how far off is a prediction, in the original scale of Y ?"

$$\text{RSE} = \sqrt{\frac{SS_{\text{residual}}}{n - k - 1}}, \quad \text{RMSE} = \sqrt{\frac{SS_{\text{residual}}}{n}} \quad (5.3)$$

where SS_{residual} = residual sum of squares (Equation 5.1), $n - k - 1$ = residual df (RSE, what R's `summary.lm` reports), n = sample size (RMSE, common in machine-learning contexts). Nearly identical whenever $n \gg k$.

Our two-predictor model

RSE = 4.49 on 277 df – compare to simply guessing the mean every time, which leaves a typical error of $SD_{\text{depression}} = 5.15$: the model's predictions are, on average, about 0.66 points closer to the truth.

In R

```
summary(m3)$sigma # RSE sqrt(mean(resid(m3)^2)) # RMSE
```

The F -test for overall model fit

$$H_0 : B_1 = B_2 = \dots = B_k = 0 \quad H_1 : \text{at least one } B_j \neq 0$$

$$F = \frac{R^2/k}{(1 - R^2)/(n - k - 1)} \sim F_{k, n-k-1} \quad (5.4)$$

where k = number of predictors, $n - k - 1$ = residual degrees of freedom.

Our model

$$F(2, 277) = \frac{.246/2}{.754/277} = 45.29, \quad p < .001$$

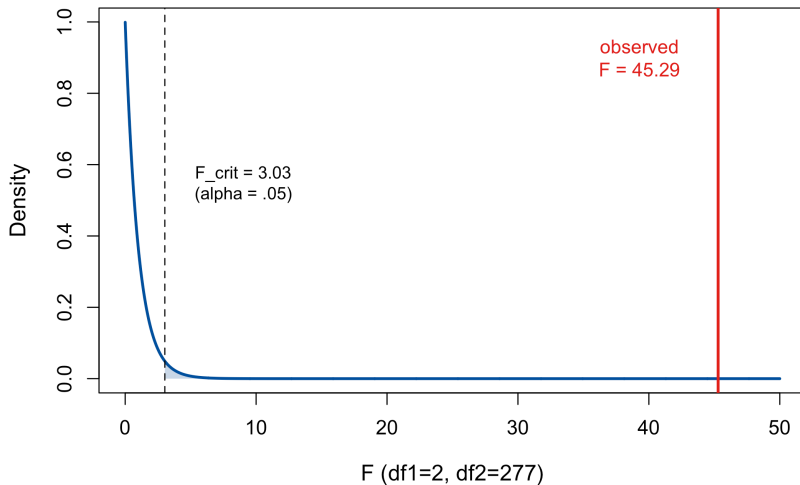
In R

```
summary(m3)$fstatistic # or: anova(m3)
```

Conclusion

We reject H_0 : self-esteem and negative affect jointly predict depression significantly better than the intercept-only (mean) model.

The F -distribution, visualized



Which predictor matters more?

The challenge

Once predictors correlate with each other, "how important is X_1 ?" no longer has one single correct answer – it depends on **how** you define importance.

Three common answers, this section works through each in turn – each answers a **different research question**:

- 1 Standardized β weights – e.g., "controlling for negative affect, how strong is self-esteem's effect, in comparable units?"
- 2 Partial and part (semipartial) correlations – e.g., "how much of depression's variance does self-esteem explain that nothing else does?"
- 3 Variance-decomposition methods (e.g., LMG / relative importance analysis) – e.g., "of all the variance my model explains, what share comes from each predictor?"

Standardized β in multiple regression

$$\beta_{\text{Neg. affect}} = \frac{r_{\text{NA,Dep}} - r_{\text{SE,Dep}} r_{\text{NA,SE}}}{1 - r_{\text{NA,SE}}^2} \quad (6.1)$$

where NA = negative affect, SE = self-esteem, Dep = depression; all r 's are simple bivariate correlations.

- Adjusts the bivariate correlation for the correlation between predictors.
- If predictors are **uncorrelated**, β reduces to the simple bivariate r .

In R

```
lm.beta::lm.beta(m3)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0000	0.0521	0.000	1.00000
self_esteem_z	-0.1803	0.0552	-3.264	0.00123 **
neg_affect_z	0.4071	0.0552	7.373	1.93e-12 ***

Partial and part (semipartial) correlations

Partial correlation

Remove the influence of X_2 from **both** X_1 and Y before correlating them.

$$r_{Y,X_1 \cdot X_2} = \frac{r_{Y,X_1} - r_{Y,X_2}r_{X_1,X_2}}{\sqrt{(1 - r_{X_1,X_2}^2)(1 - r_{Y,X_2}^2)}} \quad (6.2)$$

Part (semipartial) correlation

Remove X_2 's influence from X_1 **only**; Y stays raw.

$$r_{Y(X_1 \cdot X_2)} = \frac{r_{Y,X_1} - r_{Y,X_2}r_{X_1,X_2}}{\sqrt{1 - r_{X_1,X_2}^2}} \quad (6.3)$$

Working: self-esteem, controlling for negative affect

Required: $r_{\text{Dep,SE}} = -.314$, $r_{\text{Dep,NA}} = .466$, $r_{\text{SE,NA}} = -.328$

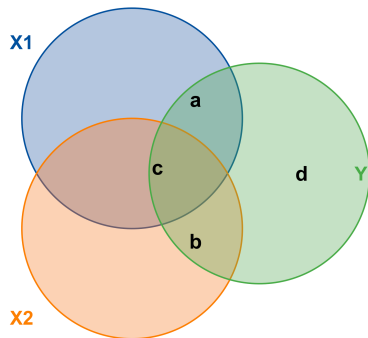
$$r_{\text{partial}} = \frac{-.314 - (.466)(-.328)}{\sqrt{(1 - (-.328)^2)(1 - .466^2)}} = \frac{-.314 + .153}{\sqrt{.892 \times .783}} = \frac{-.161}{.836} = -.192$$

$$r_{\text{part}} = \frac{-.314 - (.466)(-.328)}{\sqrt{1 - (-.328)^2}} = \frac{-.161}{\sqrt{.892}} = \frac{-.161}{.944} = -.170$$

In R

```
ppcor::pcor(clinical_sample[,c("depression","self_esteem","neg_affect")])  
ppcor::spcor(clinical_sample[,c("depression","self_esteem","neg_affect")])
```

Partial vs. part correlation – the Venn picture



In our example: X_1 =self-esteem, X_2 =negative affect, Y =depression.

- a = variance self-esteem shares **uniquely** with depression (not via negative affect)
- b = variance negative affect shares **uniquely** with depression
- c = variance self-esteem and negative affect **both** explain in depression (their overlap)
- d = variance in depression **neither** predictor explains
- Partial r for self-esteem $\approx \sqrt{\frac{a}{a+d}}$ (in sign, matches $-.192$): self-esteem's unique slice, relative only to what is *left over* after removing the c shared with negative affect
- Part r for self-esteem $\approx \sqrt{\frac{a}{a+b+c+d}}$ (matches $-.170$): self-esteem's unique slice, relative to *all* of depression's variance

Incremental R^2

Another way to ask "does X_2 add anything?"

Compare R^2 from a model with X_2 added against a model without it.

$$\Delta R^2 = R_{\text{full}}^2 - R_{\text{reduced}}^2 \quad (6.4)$$

Adding negative affect after self-esteem

$$R_{\text{full}}^2(.246) - R_{\text{self-esteem only}}^2(.099) = .147$$

In R

```
m_se <- lm(depression ~ self_esteem, clinical_sample)
anova(m_se, m3) # F-change test comparing nested models
```

- ΔR^2 is the squared part (semipartial) correlation for the added predictor: $(-.385)^2 \approx .148$, matching (rounding) the .147 above.
- Commonly tested with an F -change test when comparing nested models.

Relative importance analysis: the relaimpo package

Why we need something beyond β , partial, or part r

Standardized β , partial r , and part r all answer "unique contribution, holding the others fixed" – so none of them credit a predictor for variance it **shares** with others, even though that shared variance is real explanatory power too.

LMG method (Lindeman, Merenda & Gold, 1980)

Averages each predictor's R^2 contribution across **every possible entry order** into the model.

In R

```
library(relaimpo)
m3 <- lm(depression ~ self_esteem + neg_affect,
         data = clinical_sample)
calc.relimp(m3, type = c("lmg", "first", "last",
                        "betasq", "pratt"))
```

relaimpo output

Response surface: depression ~ self_esteem + neg_affect

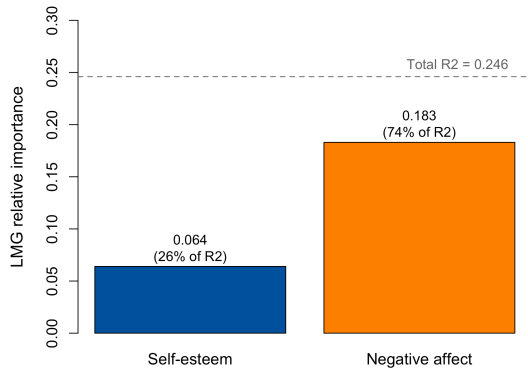
	lmg	last	first	betasq	pratt
self_esteem	0.0638	0.0290	0.0985	0.0325	0.0566
neg_affect	0.1827	0.1479	0.2174	0.1658	0.1898

- **first** = R^2 when entered *alone* (self-esteem's .0985 matches the simple-regression R^2 from earlier)
- **last** = R^2 gain when entered *last* (after the other predictor) – equals ΔR^2 /squared part correlation
- **lmg** = average of **first** and **last** across all orderings (with 2 predictors, just their mean)
- **betasq** = squared standardized β (does *not* sum to R^2 on its own)
- **pratt** = standardized $\beta \times$ bivariate r (the Hoffman product measure, next slide)

In R

```
calc.relimp(m3, type=c("lmg","first","last","betasq","pratt"))
```

LMG results for our model



The downside of LMG

- **Computationally expensive:** with p predictors, LMG averages over all $p!$ possible entry orders (or, more efficiently, 2^p subset models) – this grows fast and can be slow with many predictors.
- **No simple significance test:** there is no standard p -value for “is this predictor’s LMG share significantly different from zero/from another predictor’s” – bootstrapped confidence intervals are used instead (`boot.relimp()`), which adds complexity.
- **Loses direction:** LMG is always non-negative – it tells you *how much* variance a predictor accounts for, but not the **sign** of the relationship. You still need the regular coefficient table for that.
- **Less familiar to reviewers/readers:** standardized β is the default in most published psychology papers; LMG requires more explanation for an unfamiliar audience.

A cautionary note: Hoffman's product measure

$$r_{Y, X_1 \cdot (X_2 \dots X_k)}^2 \approx \beta_{X_1} r_{Y, X_1} \quad (6.5)$$

- Simple to compute: standardized β times the bivariate correlation (this is relaimpo's "pratt" column).
- Sums to R^2 exactly, **always** – by algebraic construction, not because it is a good measure.
- **Flaw:** can produce *negative* "importance" values when predictors are correlated in complex ways (e.g., suppression, coming up next) – LMG cannot.

Verify with our data

$$\beta_{SE} \times r_{SE} = -0.180 \times -0.314 = 0.0566, \quad \beta_{NA} \times r_{NA} = 0.407 \times 0.466 = 0.1898$$

$$\text{sum} = 0.2464 \approx R^2(.246)$$

Putting it all together

Measure	Answers	Self-esteem value
Standardized β	unique effect, SD units	−0.180
Partial r	unique assoc., both vars. adjusted	−0.192
Part (semipartial) r	unique assoc., only X ad- justed	−0.170
LMG relative impor- tance	share of total R^2	.064 (26%)

The takeaway

These are **different questions**, not competing estimates of one "true" number. Report the one that matches your research question – and say which one you used.

Matching your research question to a strategy

Research question	Strategy	Why
"Controlling for X2, how strong is X1's effect, in units I can compare across predictors?"	Standardized β	Puts every predictor on the same (SD) scale; the default in most published psychology papers
"How much of Y's variance does X1 explain that no other predictor also explains?"	Part (semipartial) r , squared	Directly interpretable as a share of <i>total</i> Y variance (ΔR^2)
"Of the variance not already claimed by X2, how much does X1 uniquely explain?"	Partial r	Smaller, more conservative denominator – useful when you want to know what is left to explain <i>after</i> a known predictor
"Of all the variance my model explains, what share comes from each predictor (including what they share)?"	LMG (relative importance)	Only measure that sums exactly to total R^2 ; captures shared variance too
"X1 and X2 are uncorrelated (e.g., one is randomized) – which one matters?"	Any of the above – they all agree	With $r_{X_1, X_2} = 0$, β , partial r , and part r collapse to the same bivariate r (Equation 7.1, next slide)

Orthogonal predictors: the special case

When $r_{X_1, X_2} = 0$

All of the measures on the previous slide – β , partial r , part r – collapse to the **same** value: the simple bivariate correlation.

$$r_{X_1, X_2} = 0 \Rightarrow \beta_{X_1} = r_{X_1 \cdot X_2} = r_{X_1(X_2 \cdot)} = r_{Y, X_1} \quad (7.1)$$

A concrete demonstration, using our own data

Our randomized treatment variable is (by design) almost perfectly uncorrelated with self_esteem: $r = .008$. So we can recover the joint-model treatment coefficient in two completely different ways:

- 1 **Jointly:** `coef(lm(depression ~ self_esteem + treatment))["treatment"] = -3.058`
- 2 **Sequentially:** residualize depression on self-esteem *first*, *then* regress those residuals on treatment alone:
`r <- resid(lm(depression ~ self_esteem))`
`coef(lm(r ~ treatment))["treatment"] = -3.058 – identical`

Standardization: when and why to standardize

$$z_X = \frac{X - \bar{X}}{s_X} \quad (8.1)$$

where $\bar{X}$ = sample mean, s_X = sample SD; a z-scored variable always has mean 0, SD 1.

Worked example: standardizing self-esteem for one person

A participant with self-esteem = 35 (sample $\bar{X} = 28.68$, $s_X = 5.18$): $z = \frac{35 - 28.68}{5.18} = 1.22$ (1.22 SDs above the mean)

In R

```
clinical_sample$self_esteem_z <- scale(clinical_sample$self_esteem)
m3_std <- lm(scale(depression) ~ scale(self_esteem) + scale(neg_affect))
```

Standardize when...

Predictors are on different scales and you want relative magnitude; you want effect sizes comparable across studies.

Keep unstandardized when...

Raw units are meaningful to your audience; predictors are dummy-coded (a 0/1 "SD" is often not meaningful).

Interpreting standardized coefficients here

- $\beta_{\text{self-esteem}} = -0.180$: a 1-*SD* increase in self-esteem predicts a 0.180-*SD* decrease in depression, **holding negative affect constant**.
- $\beta_{\text{neg. affect}} = 0.407$: a 1-*SD* increase in negative affect predicts a 0.407-*SD* increase in depression, **holding self-esteem constant**.

Common misinterpretation to avoid

Standardized β is **not** automatically comparable across different *samples* or *populations* – it depends on the variance of X in *this* sample. A restricted-range sample will shrink β even if the underlying relationship is identical.

Assumptions of regression: LINE

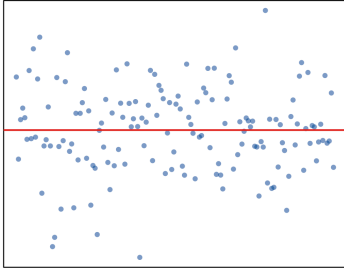
	Assumption	What it means
L	Linearity	The relationship between each X and Y is (approximately) linear
I	Independence	Residuals are uncorrelated with each other (no clustering/autocorrelation)
N	Normality	Residuals are approximately normally distributed
E	Equal variance	Residual spread does not depend on the predicted value (homoscedasticity)

Kline's fuller list adds

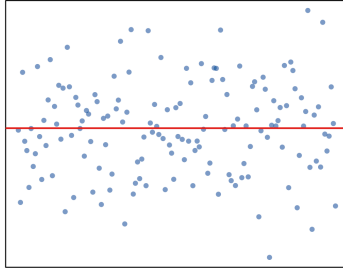
Predictors measured without (much) error; no unmodeled causal effects among predictors; no specification error (right functional form, no omitted/irrelevant variables).

LINE: what violations look like

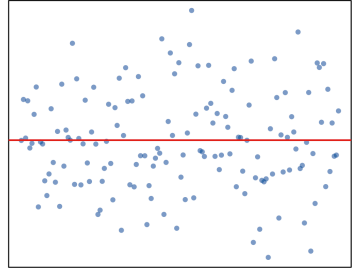
Linear (OK)



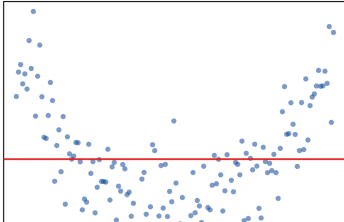
Equal variance (OK)



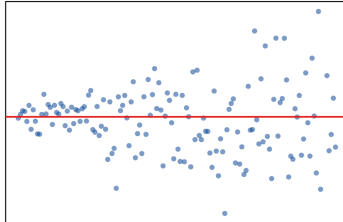
Independent (OK)



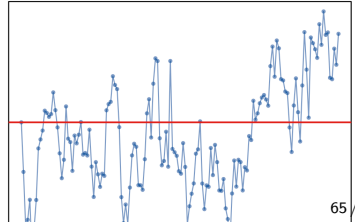
Not linear



Heteroscedastic



Not independent
(e.g. clustered/time data)

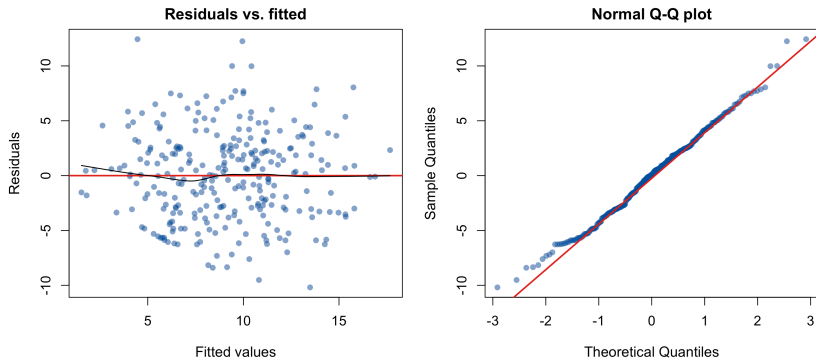


What this section covers

Are the LINE assumptions actually met – on *our* model?

- Residual-vs-fitted and Normal Q-Q plots
- The Shapiro-Wilk test for normality
- Consequences of each type of violation
- Outliers, leverage, and influence (Cook's distance, Mahalanobis distance)
- Fixing a violation – a worked example

Checking assumptions on our own model



Reading these two panels

Left (residuals vs. fitted): points scattered evenly above/below zero with no funnel or curve → linearity and equal variance both look fine. **Right** (Normal Q-Q): points tracking the diagonal closely → residuals are close to normally distributed. Shapiro-Wilk confirms: $W = .994$, $p = .295$ (fails to reject normality).

Consequences of violated assumptions

Violation	Consequence
Nonlinearity	Biased coefficients; model systematically over/under-predicts in parts of the range
Heteroscedasticity	SE s (and thus p -values, CIs) are wrong, even if coefficients themselves stay unbiased
Non-independence	Standard errors badly underestimated $\rightarrow$ inflated Type-I error rate
Non-normal residuals	Mostly a problem in <i>small</i> samples; large- n inference is fairly robust (CLT)

Practical fixes

Transform variables (e.g., log), add polynomial/interaction terms, use robust or cluster-robust SE s, or move to a model built for the data structure (e.g., multilevel models for clustered/longitudinal data – covered later in the course).

Outliers, leverage, and influence

Three different questions about one data point

Outlier: unusual on Y given its X (big residual). **High leverage:** unusual on X (far from the predictor means).

Influential: removing it would actually change the fitted model – needs *both* some leverage *and* some residual.

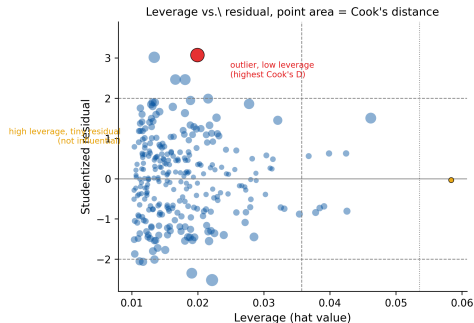
$$h_{ii} = \left[X(X^T X)^{-1} X^T \right]_{ii}, \quad D_i = \frac{e_i^2}{p s^2} \cdot \frac{h_{ii}}{(1 - h_{ii})^2} \quad (10.1)$$

where h_{ii} = leverage (hat value) of case i (0 to 1); D_i = Cook's distance, folding leverage and the residual e_i into one influence score; $p = k + 1$ = number of estimated parameters (predictors + intercept); s^2 = residual variance.

Rules of thumb (n = sample size, k = number of predictors)

Diagnostic	Flags	Typical cutoff
Leverage h_{ii}	Unusual X	$> 2(k+1)/n$ (or $3(k+1)/n$, stricter)
Cook's D_i	Influence	$> 4/n$ worth checking; > 1 rare, serious
Mahalanobis D^2	Multivariate X outlier	$> \chi^2_{k, \alpha=.001}$

Checking influence on our own model



Two contrasting real cases (full model, $n = 280$, $k = 4$)

A participant with an unremarkable predictor profile but a much higher depression score than predicted ($z_{\text{resid}} = 3.07$) has the highest Cook's $D = 0.038$ – an *outlier* driving real influence, though still well under the $D = 1$ danger zone. Separately, the single highest-leverage case (low self-esteem *and* high negative affect together, Mahalanobis $D^2 = 13.2$) has Cook's $D \approx 0$: its actual depression score lands almost exactly where the model already predicted, so despite its unusual X it barely moves the fit. *Leverage is not the same thing as influence.*

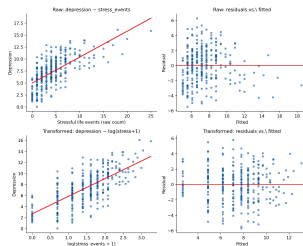
What to do about an influential point

Not: just delete it

Removing points because they are inconvenient inflates R^2 and biases inference – decide *before* looking at how it changes the results.

- Check first for a **data or coding error** (impossible value, mis-scored item) – fix or exclude only these
- Run a **sensitivity analysis**: refit with and without the case, report both if conclusions differ
- Consider **robust regression** (e.g., M-estimation) if several points are influential, not just one
- If genuine and retained, at minimum **report it** – transparency lets readers judge for themselves

A fix, worked: log-transforming a skewed predictor



Before vs. after (new illustrative variable: stressful life events)

	B1	R2	residual pattern
Raw: depression ~ stress_events	0.540	0.444	funnel/curve visible
Log: depression ~ log(stress_events+1)	3.197	0.491	clean

The other two fixes, in R (same idea, not demoed in full here)

```
lm(depression ~ poly(stress_events, 2)) # polynomial  
lm(depression ~ x1 * x2) # interaction term  
coefTest(m, vcov=vcovHC(m, type="HC3")) # robust SEs
```

Common challenges in applied regression

- Not knowing the assumptions exist, let alone how to check them
- Running a model **without** subject-matter knowledge of the variables
- Not knowing the alternative when an assumption *is* violated
- **Extrapolating** predictions beyond the observed range of X

Extrapolation, concretely

Our self-esteem range is 10–40 (full RSES range). Predicting depression at self-esteem = 2 (impossible on this scale) is meaningless – the model was never trained on data anywhere near there.

Strategies to avoid these pitfalls

- 1 Plot before you model – scatterplots, histograms, boxplots by group
- 2 Pre-register your model specification when possible, rather than searching for significance
- 3 Check diagnostics every time, not just when results look "off"
- 4 Report effect sizes and confidence intervals, not just p -values – notice we have done this for every model in today's lecture
- 5 When in doubt, consult someone who knows the subject matter (not just the statistics)

What is suppression?

Definition (Cohen, Cohen, West & Aiken, 2003)

A predictor whose presence in the model **changes** another predictor's coefficient in a way that increases (rather than reduces) its apparent effect on Y .

- Counter-intuitive: we usually expect adding predictors to *shrink* each other's coefficients (shared variance).
- Suppression is when a predictor instead makes another predictor's effect **look bigger**.

Three patterns, each shown with a small illustrative simulated example:

- Classical suppression
- Net suppression
- Cooperative suppression

How the three patterns differ

	Classical	Net	Cooperative
Suppressor's own link to Y	near zero	positive, but weaker than the other predictor	positive (same direction as the other predictor)
Predictors correlated with each other?	yes	yes (strongly)	yes – but negatively
What changes when both are entered	focal predictor's β grows , suppressor gets a nonzero β despite $r \approx 0$	focal predictor's β flips sign	both predictors' β s grow beyond their bivariate r
Telltale sign	suppressor has $\beta \neq 0$ but bivariate $r \approx 0$	sign of β disagrees with sign of bivariate r	R^2 jumps dramatically above either predictor alone

Classical suppression – example

Negative affect, social desirability, depression

A predictor Z (social desirability) that is **uncorrelated with Y** but **correlated with X** increases X 's standardized effect once added.

In R

```
lm.beta(lm(depression ~ neg_affect + social_desirability))
```

	r w/ Dep.	beta(simple)	beta(full)
neg_affect	.502	.502	.597
social_desirability	-.001	--	-.238

R2_simple = .252 R2_full = .300

Net suppression – example

Rumination, negative affect, depression

A predictor's coefficient **flips sign** relative to its own bivariate correlation with Y , once a second, more strongly related predictor is added.

In R

```
lm.beta(lm(depression ~ rumination + neg_affect))
```

	r w/ Dep.	beta(simple)	beta(full)
rumination	.304	.304	-.097
neg_affect	.507	--	.575

```
R2_simple = .092      R2_full = .262
```

Cooperative suppression – example

Self-esteem, optimism, well-being

Two predictors that both correlate **positively** with Y, but **negatively** with each other – their combined standardized effects both **exceed** their bivariate correlations.

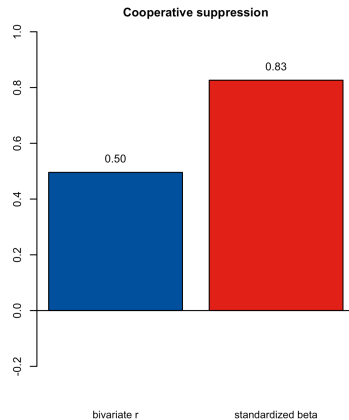
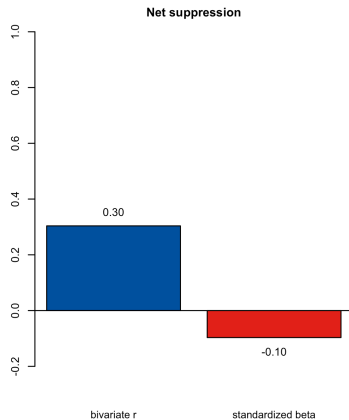
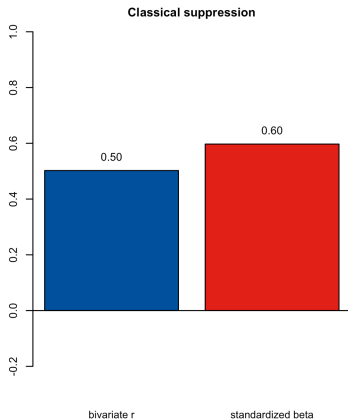
In R

```
lm.beta(lm(well_being ~ self_esteem + optimism))
```

	r w/ WB	beta(simple)	beta(full)
self_esteem	.496	.496	.826
optimism	.511	.511	.837

R2_either_alone ~= .25 R2_together = .837

The three suppression patterns, side by side



Regression is already a (simple) structural model

The key realization

A regression equation *is* a structural equation: it states a hypothesis about how variables relate. **Path analysis** and **SEM** simply chain more of these equations together.

- Regression: usually one equation, one outcome
- Path analysis: multiple equations, variables can be both predictor *and* outcome
- SEM: adds **latent variables** (measured indirectly via multiple indicators)

Fitting our regression in lavaan

Same model, new syntax

```
library(lavaan)

m_lm      <- lm(depression ~ self_esteem, data = clinical_sample)
m_lavaan <- sem('depression ~ self_esteem',
               data = clinical_sample, meanstructure = TRUE)

summary(m_lm)
summary(m_lavaan)
```

They agree exactly

Same slope (-0.312), same *SE* (0.057) – lavaan is fitting the identical linear model, just through the lens of covariance-structure estimation.

Why this matters going forward

What regression *cannot* do

A single regression equation cannot represent **indirect** effects: if X affects M , and M affects Y , that requires **two** linked equations.

$$X \rightarrow M \rightarrow Y$$

- This exact limitation is the motivating example for **path analysis** and **mediation** (next in the course sequence).
- Everything from today – coefficients, R^2 , standardization, assumptions – carries over largely unchanged.

Where this fits in the course

- 1 Correlation, simple & multiple regression (GLM) – today
- 2 Logistic regression (Generalized Linear Model)
- 3 Moderation/mediation and path analysis
- 4 The common factor model: EFA and CFA
- 5 Introduction to SEM
- 6 Model fit, evaluation, revision, comparison, equivalence testing
- 7 Multiple-group analysis and best practices in SEM

One toolkit, reused throughout

A small, consistent set of R packages (`lm`, `lm.beta`, `ppcor`, `relaimpo`, `lavaan`) carries you from today all the way to full SEM – less to relearn, more time to think about the substance.

Summary

- Correlation standardizes covariance; regression adds direction (predictor $\rightarrow$ criterion) and lets predictors be continuous **or** categorical (dummy coding)
- An equation with $\hat{Y}$ (hat) never needs $+\varepsilon$; an equation for an observed Y_i always does
- Multiple regression coefficients are **unique, partial** effects – not the same as simple-regression coefficients
- R^2 , adjusted R^2 , and the F -test evaluate the model as a whole; standardized β , partial/part r , and LMG evaluate individual predictors – pick the one matching your question
- Always check LINE assumptions and report CIs; watch for suppression when coefficients behave unexpectedly
- Regression is the foundation everything else in this course builds on

Thank you!