LECTURE 3

Regression – note that residuals are sample estimates of population errors

Linear models:

1. **Easy to fit**
2. **Commonly used**
3. **Practical application**
4. **Descriptive model**
5. **Assumptions reasonable**

Normal errors:

- **Ontological justification** – random fluctuations have good chance of resembling normal distributions
- **Epistemological** – normal distributions represent state of knowledge

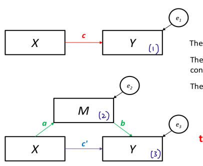
Assumptions:

1. **Validity**
   - Relevant IVs underlying DV (researcher)
   - Data will generalise to population
2. **Additivity & Linearity**
   - Data (excluding errors) should be linear (important math assumption)
   - Transform non-linear data into linear data if linear modelling
3. **Independence of errors**
4. **Equal variance of errors (homogeneity & homoscedasticity/sphericity)** impact p values & estimates
5. **Normality of errors (around mean 0)**
   - Least important = can overcome with bootstrapping

MEDIATION – inherently causal whereby IV acts on DV through MV

Mediation Regression equations:
(Baron & Kenny, 1986)

$$Y = \beta_1 + cX + e_1 \quad (1)$$
$$M = \beta_2 + aX + e_2 \quad (2)$$
$$Y = \beta_3 + c'X + bM + e_3 \quad (3)$$



The **total effect** of IV to DV (*c*)

The **direct effect** of IV to DV controlling for M (*c'*)

The **indirect effect** IV to DV via M (*a*b*)

c* = *a*b* + *c'

total = **indirect** + **direct**

Rearranging,

a*b* = *c* – *c'

Baron & Kenny approach

1. c is significant (IV predicts DV)
2. a is significant (IV predicts MV)
3. b is significant (MV predicts DV)

4a. c' is significantly smaller than c (partial mediation c' < c)

4b. c' is not significant (full mediation a*b and c' > c)

Issues with causal steps approach

1. **NHST**
   - ❖ Low power (each NHST conditioned on prior NHST – prone to false negatives = not rejecting false null hypothesis/Type II error)
   - ❖ Dichotomous – accept/reject (vs. CI)
2. **Indirect effect (a*b) not quantified**
   - ❖ Only testing if a and b are individually significant but does not test if a*b is significant
   - ❖ a and b both being significant does not mean that a*b is significant
3. **Preconditioning that IV predicts DV**
   - ❖ a*b can be significant although c is non-significant ie. there can be a significant mediation effect without significant total effect
   - ❖ Problematic because testing stops when c is non-significant

Critique of partial and complete mediation

1. Attempt at degree of mediation but no numerical importance = subjective
2. a*b can be significant although c is non-significant ie. there can be a significant mediation effect without significant total effect
3. Complete mediation implies that mediator completely accounts for IV's effects on DV – doesn't consider other possible mediators or models
4. Partial and complete is dependent on sample size (small-ish n with enough power to detect sig indirect effect but not enough power to detect sig direct effect) – also non-sig c' doesn't mean c' = 0 since NHST don't work like this.

Hayes approach: Testing a*b

- Instead of testing significance of c' < c this test looks at

1. Significance of a*b
2. Relative size of a*b, a and b

Sobel Test (assumes normality)

- H_0: a*b = 0 at population level
- H_1: a*b not equal to 0 indicating significant mediating/indirect effect

Issues:

1. **Low power**
   - ❖ Difficult to pick up significant effects when there is one at population level
2. **Non-normal distributions**
   - ❖ Difficult to pick up significant effects with non-normal data

Bootstrapping (does not assume normality)

- Builds CIs around estimates of a*b through repeated sampling from current sample
- Single observation in sample data for one particular case might either be used more than once in a sample or not at all in one particular bootstrap sample
- Robust to non-normal data
- Takes the means of bootstrap samples and plots a distribution
- Compare means of bootstrap distribution to original distribution
- Forms lower and upper bounds of CI

Output

************** TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y **************

Total effect of X on Y

Effect	se	t	p	LLCI	ULCI	c_ps	c_cs
.1970	.0185	10.6231	.0000	.1605	.2335	.0696	.5443

Direct effect of X on Y

Effect	se	t	p	LLCI	ULCI	c'_ps	c'_cs
.0984	.0194	5.0612	.0000	.0601	.1367	.0348	.2720

Indirect effect(s) of X on Y:

	Effect	BootSE	BootLLCI	BootULCI
na	.0985	.0186	.0629	.1358

Bootstrap CI does not contain 0!

Normal theory test for indirect effect(s):

	Effect	se	Z	p
na	.0985	.0141	6.9870	.0000

Regular Sobel test

Partially standardized indirect effect(s) of X on Y:

	Effect	BootSE	BootLLCI	BootULCI
na	.0348	.0056	.0240	.0458

So, na is a mediator between n and p6:
There is a significant path a*b through na

Completely standardized indirect effect(s) of X on Y:

	Effect	BootSE	BootLLCI	BootULCI
na	.2723	.0461	.1829	.3625

*if results for Sobel test and Bootstrap do not align, trust Bootstrap

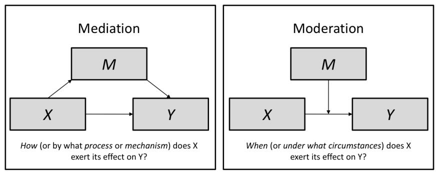
Issues for mediation

1. **Causal inference**
   - ❖ Regression modelling does not suffice causal effect – need more rigorous methods
2. **Confounding association**
   - ❖ Existence of other possible mediating variables
   - ❖ Mediating variable may not even account for causal effect
3. **Causal order**
   - ❖ Difficult to determine causal order
   - ❖ If IV isn't determined through random assignment or manipulation, any sequence of IV, MV and DV must be tested

MODERATION (GLM: regression, two-way ANOVA with 2 categorical predictors and continuous outcome variable)

- Standardise IV and MV or centre them with mean of 0

Differences between mediation & moderation



- Mediation (how) and Moderation (when)
- Moderation involves an interaction between M and X (looks like multiple regression) $Y = b_0 + b_1X + b_2X + e$

Regression Coefficient
(Simple Slope)

Conditional effects of the focal predictor at values of the moderator(s):

Percentile	pa	Effect	se	t	p	LLCI	ULCI
16th	-8.1968	-.8397	.1429	-5.8761	.0000	-1.1211	-.5583
50th	1.2366	-.6166	.1414	-4.3599	.0000	-.8951	-.3381
84th	8.2366	-.4511	.2019	-2.2342	.0263	-.8486	-.0535

$$Y = b_0 + b_1X + b_2M + b_3XM + e$$

$$Y = (b_0 + b_2M) + (b_1 + b_3M)X + e$$

simple intercept simple slope

- Note that both slope and intercept depend on value of moderator

Output – Test for Significance of Moderator

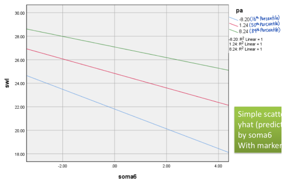
Model

	coeff	se	t	p	LLCI	ULCI
constant	24.4378	.3816	64.0406	.0000	23.6864	25.1892
soma6	-.6458	.1349	-4.7874	.0000	-.9115	-.3802
pa	.3211	.0479	6.6986	.0000	.2267	.4155
Int_1	.0236	.0135	1.7473	.0818	-.0030	.0503

*interaction of moderator is non-significant

Relationship between IV and DV depends on value of MV

- When positive affect is very **low** at its **16th percentile** (-8.20), X is a **strong**, negative predictor of Y (effect size = -.84, p < .001).
- When positive affect is **moderate** at its **50th percentile** (1.24), X is a **moderate,** negative predictor of Y (effect size -.62, p < .001)
- When positive affect is very **high** at its **84th percentile** (8.24), X is a **weak**, negative predictor of Y (effect size = -.45, p = .0283)



- At low levels of moderator, there is a stronger negative relationship between IV and DV
- At higher levels of moderator, there is a weaker negative relationship between IV and DV

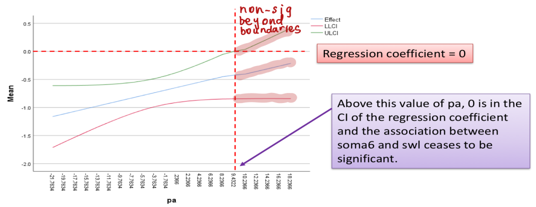
JOHNSON & NEWMAN APPROACH

Conditional effect of focal predictor at values of the moderator:

pa	Effect	se	t	p	LLCI	ULCI
-21.7634	-1.1605	.2796	-4.1505	.0000	-1.7111	-.6099
-19.7634	-1.1132	.2558	-4.3521	.0000	-1.6169	-.6095
-17.7634	-1.0659	.2327	-4.5810	.0000	-1.5241	-.6078
-15.7634	-1.0186	.2105	-4.8386	.0000	-1.4332	-.6041
-13.7634	-.9713	.1896	-5.1220	.0000	-1.3447	-.5979
-11.7634	-.9240	.1705	-5.4194	.0000	-1.2598	-.5883
-9.7634	-.8767	.1538	-5.7014	.0000	-1.1795	-.5739
-7.7634	-.8294	.1403	-5.9115	.0000	-1.1057	-.5532
-5.7634	-.7821	.1311	-5.9652	.0000	-1.0403	-.5240
-3.7634	-.7348	.1271	-5.7801	.0000	-.9852	-.4845
-1.7634	-.6876	.1288	-5.3366	.0000	-.9412	-.4339
.2366	-.6403	.1360	-4.7071	.0000	-.9081	-.3724
2.2366	-.5930	.1479	-4.0098	.0001	-.8841	-.3018
4.2366	-.5457	.1634	-3.3394	.0010	-.8674	-.2239
6.2366	-.4984	.1817	-2.7435	.0065	-.8561	-.1407
8.2366	-.4511	.2019	-2.2342	.0263	-.8486	-.0535
9.4322	-.4228	.2147	-1.9691	.0500	-.8456	.0000
10.2366	-.4038	.2236	-1.8059	.0721	-.8440	.0365
12.2366	-.3565	.2463	-1.4471	.1491	-.8415	.1286
14.2366	-.3092	.2699	-1.1455	.2530	-.8406	.2223
16.2366	-.2619	.2941	-.8906	.3740	-.8409	.3171
18.2366	-.2146	.3187	-.6733	.5013	-.8421	.4129

- Identify the value of the moderator variable where the relationship between IV and DV changes from significant to non-significant (p = .05) however this is somewhat arbitrary – we can assume the relationship just outside of this is still in the same direction (although not significant)

Regression line with upper & lower CI bounds



- The relationship between IV and DV are no longer significant at values of moderator outside of the 95% CI bounds
- Y = values of regression coefficients, X = values of moderator
- At **low** levels of positive affect, physical wellbeing is **strongly** negatively predicting satisfaction with life (R^2 = -1.0)
- At **higher** levels of positive affect, physical wellbeing is **weakly** negatively predicting satisfaction with life (R^2 = -0.2)
- If you are low in positive affect and high in physical wellbeing, the model predicts you will have low satisfaction with life
- In the upper right corner of the graph, 0 is captured in the confidence interval, which maps with at high levels of positive affect, physical wellbeing is no longer a significant predictor of satisfaction with life