Statistics One

Lecture 12
The General Linear Model (GLM)

1

Two segments

- The General Linear Model (GLM)
- Dummy coding

2

Lecture 12 ~ Segment 1

The General Linear Model (GLM)

3

General Linear Model (GLM)

- GLM is the mathematical framework used in many common statistical analyses, including multiple regression and ANOVA
  - ANOVA is typically presented as distinct from multiple regression but it IS a multiple regression

4

Characteristics of GLM

- *Linear*: pairs of variables are assumed to have linear relations
- *Additive*: if one set of variables predict another variable, the effects are thought to be additive

5

Characteristics of GLM

- BUT! This does not preclude testing non-linear or non-additive effects

6

Characteristics of GLM

- GLM can accommodate such tests, for example, by
  - Transformation of variables
    - Transform so non-linear becomes linear
  - Moderation analysis
    - Fake the GLM into testing non-additive effects

7

GLM example

- Simple regression
  - $Y = B_0 + B_1X_1 + e$
  - Y = faculty salary
  - X1 = years since PhD

8

GLM example

- Multiple regression
  - $Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + e$
  - Y = faculty salary
  - X1 = years since PhD
  - X2 = number of publications
  - X3 = (years x pubs)

9

GLM example

- One-way ANOVA
  - $Y = B_0 + B_1X_1 + e$
  - Y = faculty salary
  - X1 = gender

10

GLM example

- Factorial ANOVA
  - $Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + e$
  - Y = faculty salary
  - X1 = gender
  - X2 = race
  - X3 = interaction (gender x race)

11

Analysis of Variance (ANOVA)

- Appropriate when the predictors (IVs) are all categorical and the outcome (DV) is continuous
  - Most common application is to analyze data from randomized experiments

12

Analysis of Variance (ANOVA)

- More specifically, randomized experiments that generate more than 2 means
  - If only 2 means then use:
    - Independent t-test
    - Dependent t-test

13

General Linear Model (GLM)

- GLM is the mathematical framework used in many common statistical analyses, including multiple regression and ANOVA
  - ANOVA is typically presented as distinct from multiple regression but it IS a multiple regression

14

Characteristics of GLM

- *Linear*: pairs of variables are assumed to have linear relations
- *Additive*: if one set of variables predict another variable, the effects are thought to be additive

15

END SEGMENT

16

Lecture 12 ~ Segment 2

Dummy coding

17

Dummy coding

- A system to code categorical predictors in a regression analysis

Dummy coding

- Example
  - IV: Area of research in a Psychology department
    - Cognitive
    - Clinical
    - Developmental
    - Social
  - DV: Number of publications

Dataframe

ProfID	Group	Pubs
NU	Cognitive	83
ZH	Clinical	74
MK	Developmental	80
RH	Social	68

Dummy coding

	D1	D2	D3
Cognitive	0	0	0
Clinical	1	0	0
Developmental	0	1	0
Social	0	0	1

Dataframe

ProfID	Group	Pubs	D1	D2	D3
NU	Cognitive	83	0	0	0
ZH	Clinical	74	1	0	0
MK	Developmental	80	0	1	0
RH	Social	68	0	0	1

Summary statistics

Group	M	SD	N
Cognitive	93.31	29.48	13
Clinical	60.67	11.12	8
Developmental	103.50	23.64	6
Social	70.13	21.82	9
Total	81.69	27.88	36

23

Regression model

- $\hat{Y} = B_0 + B_1(D1) + B_2(D2) + B_3(D3)$

Coefficients

	B	SE	B	t	p
	93.31	6.50	0	14.37	<.001
D1 (Clinical)	-32.64	10.16	-.51	-3.21	.003
D2 (Devel)	10.19	11.56	.14	0.88	.384
D3 (Social)	-23.18	10.52	-.35	-2.20	.035

25

Unweighted effects coding

	C1	C2	C3
Cognitive	-1	-1	-1
Clinical	1	0	0
Developmental	0	1	0
Social	0	0	1

Coefficients

	B	SE	B	t	p
	81.90	4.06	0	14.37	<.001
D1 (Clinical)	-21.23	6.85	-.51	-3.21	.003
D2 (Devel)	21.60	7.88	.14	0.88	.384
D3 (Social)	-11.78	7.12	-.35	-2.20	.035

27

Weighted effects coding

	C1	C2	C3
Cognitive	$-N_{Clin}/N_{Cog}$	$-N_{Dev}/N_{Cog}$	$-N_{Soc}/N_{Cog}$
Clinical	N_{Clin}/N_{Cog}	0	0
Developmental	0	N_{Dev}/N_{Cog}	0
Social	0	0	N_{Soc}/N_{Cog}

Segment summary

- Dummy coding
  - A system to code categorical predictors in a regression analysis

END SEGMENT

30

END LECTURE 12

31