Lecture 11
Segment 2

Details: Centering and dummy coding

Centering predictors

- To center means to put in deviation form
  - $X_C = X - M$
- Why center?
  - Two reasons
    - Conceptual
    - Statistical

Centering predictors

- Conceptual reason
  - Suppose
    - Y = child's language development
    - X1 = mother's vocabulary
    - X2 = child's age

Centering predictors

- Conceptual reason
  - The intercept, B_0, is the predicted score on Y when all X are zero
  - If X = zero is meaningless, or impossible, then B_0 will be difficult to interpret
  - If X = zero is the average then B_0 is easy to interpret

Centering predictors

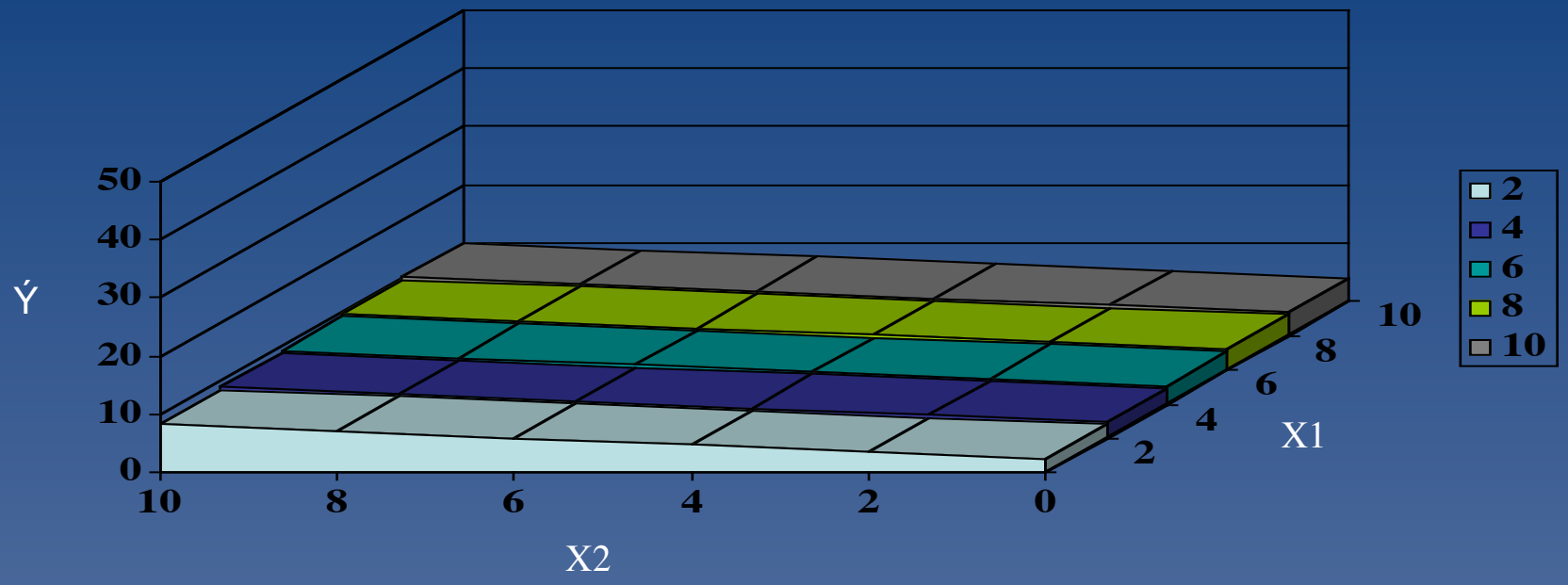
- Conceptual reason
  - The regression coefficient B_1 is the slope for X1 assuming an average score on X2
  - No moderation implies that B_1 is consistent across the entire distribution of X2

Centering predictors

- Conceptual reason
  - However, moderation implies that B_1 is NOT consistent across the entire distribution of X2
  - Where in the distribution of X2 is B_1 most representative?
  - Let's look at this graphically

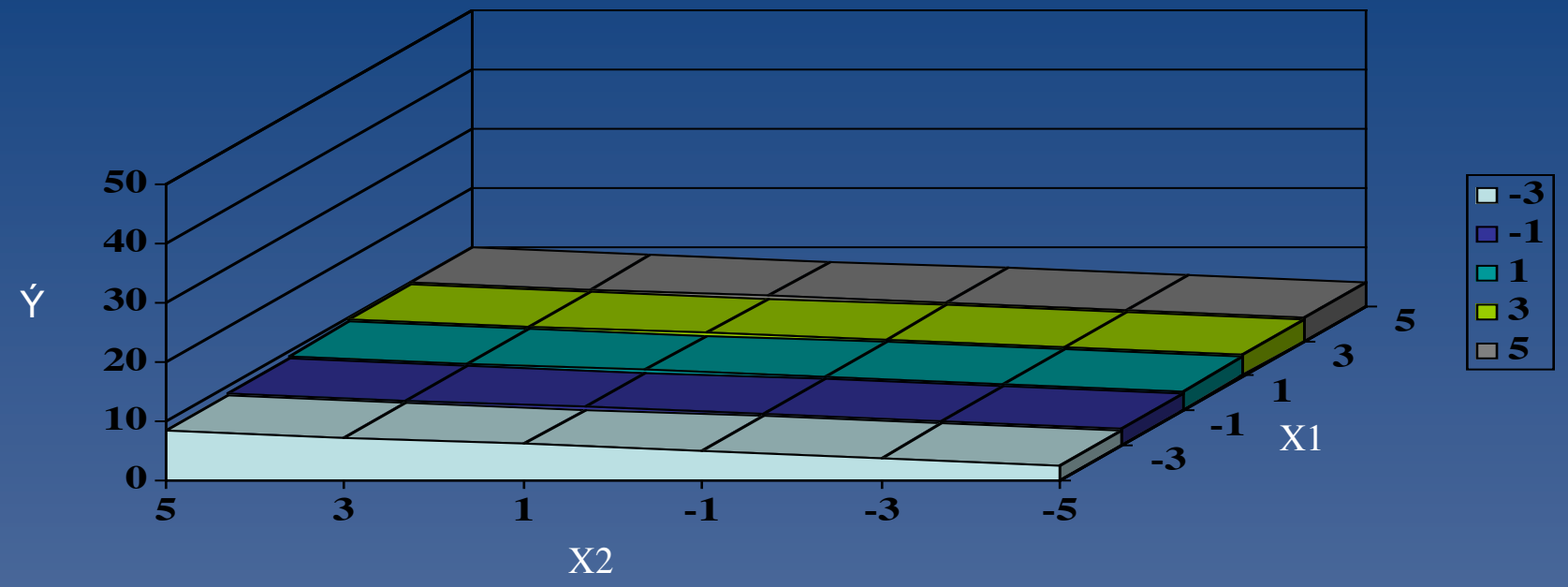
Uncentered, Additive

$\acute{Y} = 2 + .2X1 + .6X2$



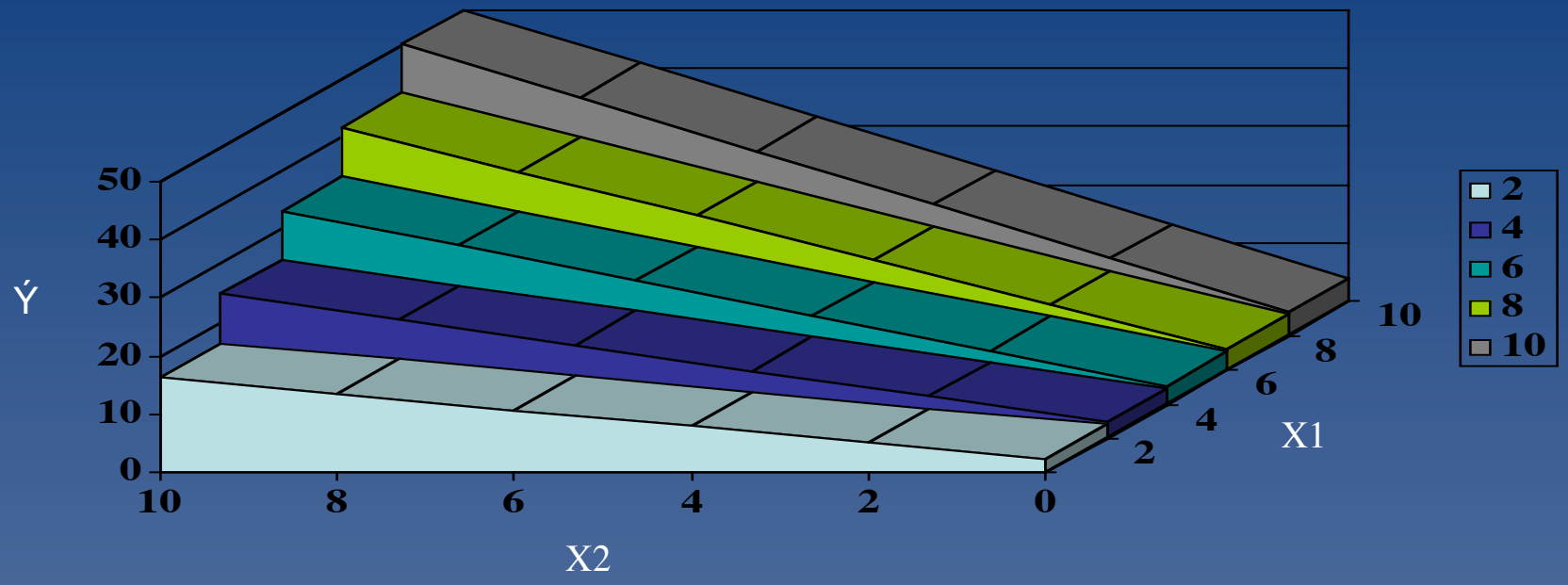
Centered, Additive

$\acute{Y} = 6 + .2X1 + .6X2$



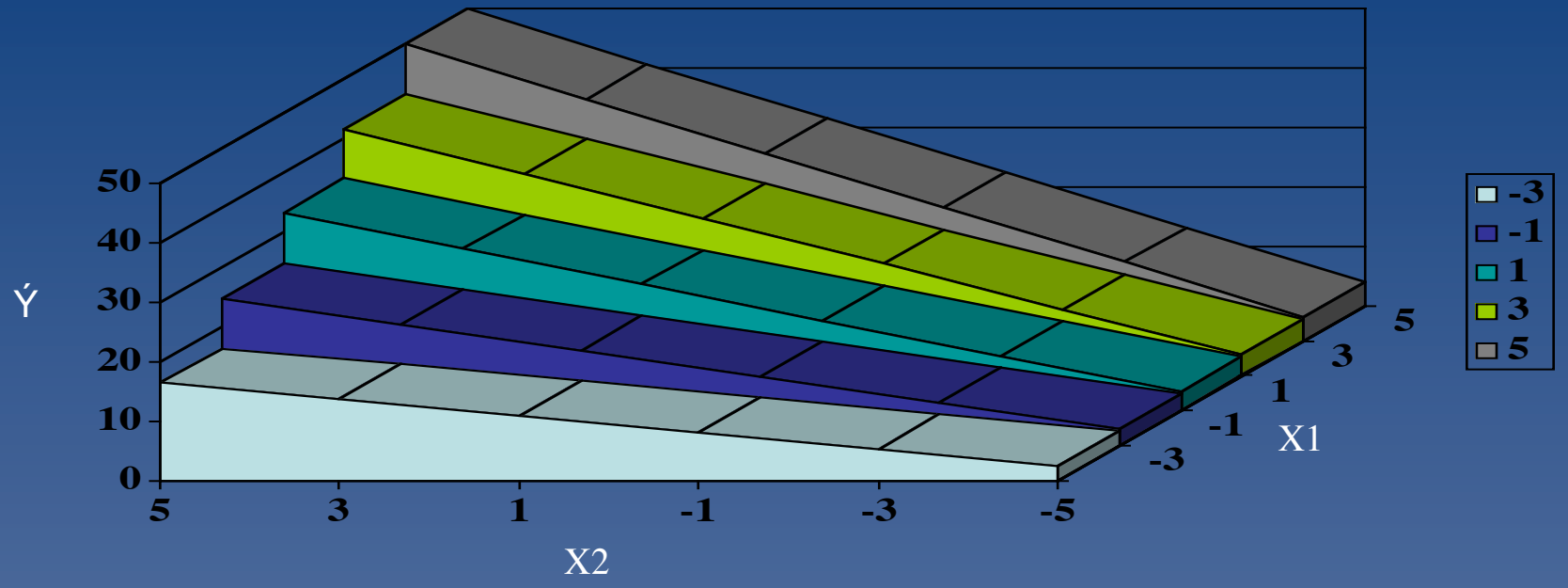
Uncentered, Moderation

$\acute{Y} = 2 + .2X1 + .6X2 + .4X1*X2$



Centered, Moderation

$\acute{Y} = 16 + 2.2X1 + 2.6X2 + .4X1*X2$



Centering predictors

- Statistical reason
  - The predictors, X1 and X2, can become highly correlated with the product, X1*X2
    - Can result in multi-colinearity

Centering for moderation: Summary

- Center predictors
- Run sequential regression (2 steps)
  - Step 1: Main effects
  - Step 2: Moderation effect
    - Evaluate B for PRODUCT or ΔR^2 from Model 1 to Model 2

Dummy coding

- A system to code categorical predictors in a regression analysis

Dummy coding

- Example
  - IV: Area of research
    - Cognitive
    - Social
    - Neuroscience
    - Cognitive neuroscience
  - DV: # of publications

Dummy coding

	C1	C2	C3
Cognitive	0	0	0
Social	1	0	0
Neuro	0	1	0
Cog neuro	0	0	1

Data file

Case	Group	DV	C1	C2	C3
1	Cog	61	0	0	0
2	Soc	78	1	0	0
3	Neuro	47	0	1	0
4	CN	65	0	0	1
…					

Regression model

- $\hat{Y} = B_0 + B_1(C1) + B_2(C2) + B_3(C3)$

Coefficients

		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
Model		B	Std. Error	Beta		
1	(Constant)	93.308	6.495		14.366	.000
	Social (C1)	-32.641	10.155	-.514	-3.214	.003
	Neuro (C2)	10.192	11.558	.138	.882	.384
	Cog Neuro (C3)	-23.183	10.523	-.351	-2.203	.035

Descriptive Statistics
Dependent Variable: Publications

Area	Mean	Std. Deviation	N
Cognitive	93.3077	29.48272	13
Cog Neuro	70.1250	21.82029	8
Neuro	103.5000	23.64530	6
Social	60.6667	11.12430	9
Total	81.6944	27.88017	36

Unweighted effects coding

	C1	C2	C3
Cognitive	-1	-1	-1
Social	1	0	0
Neuro	0	1	0
Cog neuro	0	0	1

		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
Model		B	Std. Error	Beta		
1	(Constant)	81.900	4.055		20.198	.000
	Social (C1)	-21.233	6.849	-.598	-3.100	.004
	Neuro (C2)	21.600	7.883	.550	2.740	.010
	Cog Neuro (C3)	-11.775	7.122	-.322	-1.653	.108

Weighted effects coding

	C1	C2	C3
Cognitive	$-n_s/n_c$	$-n_n/n_c$	$-n_{cn}/n_c$
Social	1	0	0
Neuro	0	1	0
Cog neuro	0	0	1