

**Understanding the Hypothesis**

Course developed by  
Deborah H. Glueck and Keith E. Muller

Slides developed by Jessica R. Shaw, Keith E. Muller,  
Albert D. Ritzhaupt and Deborah H. Glueck

© Copyright by the Regents of the University of Colorado 1

---

---

---

---

---

---

---

---

**Learning Objectives**

Define statistical models.

Describe how statistical models facilitate  
outcome prediction.

Describe how statistical models facilitate  
hypothesis testing.

2

---

---

---

---

---

---

---

---

**Learning Objectives**

Map  
scientific goals to research hypotheses,  
research hypotheses to statistical models,  
and  
statistical models to study designs.

3

---

---

---

---

---

---

---

---

**Learning Objectives**

Define between-ISU hypothesis.

Define within-ISU hypothesis.

Define concept of interaction.

4

---

---

---

---

---

---

---

---

**Defining statistical models for prediction and testing**

5

---

---

---

---

---

---

---

---

**Statistical models help summarize associations between outcomes and predictors**

Accurate power and sample size calculation requires a basic understanding of statistical models.

6

---

---

---

---

---

---

---

---

**Models are used to predict outcomes and generate testable hypotheses**

Statistical models calculate a **predicted value** of a response variable for each value or level of a predictor.

Predicted values inform **hypotheses tests**, which ask whether observed values are significantly different than we would expect.

---

---

---

---

---

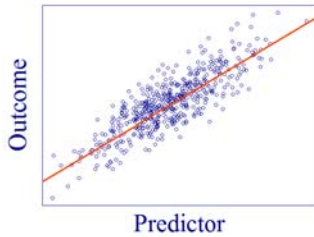
---

---

---

**Plotting the predicted values produces a line of prediction**

The **line of prediction** is the line that fits best when drawn through observed values of the response variable.



---

---

---

---

---

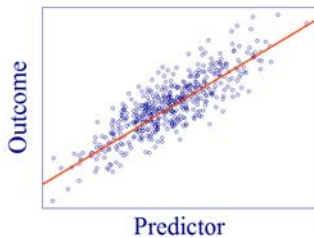
---

---

---

**Hypothesis tests often evaluate whether the slope of the predicted line is equal to zero**

The standardized slope of the predicted line reflects the magnitude of the association between the predictor and the response variables.



---

---

---

---

---

---

---

---

**Statisticians aim to select a good enough statistical model to represent reality**

Statistical models approximate reality using information available.

“Essentially, all models are wrong, but some are useful.” Box and Draper (1987)

Box et al., 1987  
Chang et al., 2010

10

---

---

---

---

---

---

---

---

**Mapping scientific goals to research hypotheses, statistical models, and study designs.**

11

---

---

---

---

---

---

---

---

**The last step requires accounting for different terminology for statistical models and study designs**

Understanding diverse terminology can ease interdisciplinary collaboration.

12

---

---

---

---

---

---

---

---

**Understanding statistical terminology will prepare you to use GLIMMPSE**

Recall, **GLIMMPSE** is a free, point-and-click, open-source software package for multilevel and longitudinal power and sample size calculations.

13

---

---

---

---

---

---

---

---

**Terminology changes but the role of each study element remains the same**

**Differential Terminology by Context**

Study design	Statistical model
Level	Cluster
Unit of observation	Single observed response
Between- ISU factors	Predictors or covariates
Within-ISU factors	Repeated measures

14

---

---

---

---

---

---

---

---

**GLIMMPSE uses the terminology for a statistical model**

**Statistical Model Terminology**

Term	Example
Cluster	Students in classrooms
Single observed response	Test scores measured 6 months after treatment
Predictor	Instructional method
Repeated measures	Test scores measured once per year for three years

15

---

---

---

---

---

---

---

---

**Predictors are ISU factors that affect the outcome(s) of interest**

**Predictors** may:

- Be interventional or observational
- Be fixed or chosen by design
- Be random and unknown before the experiment is observed, but known without appreciable error once observed

Note: **ISU** = Independent sampling unit

16

---

---

---

---

---

---

---

---

**Fixed predictors are chosen by design**

Most often, the scientific goal is to understand the association between fixed **predictors** and response variables.

17

---

---

---

---

---

---

---

---

**Fixed predictors are explanatory variables of interest defined by sample selection or randomization**

Example	Predictor	Design Decision
Sample selection	Sex	Inclusion of 100 male participants and 100 female participants
Randomization	Treatment	Random assignment of 100 participants to the treatment condition and 100 to the control condition

18

---

---

---

---

---

---

---

---

**Random predictors are quantifiable factors that vary by chance but are of experimental interest**

Examples include:

- Blood pressure
- Alcohol intake

Note: ISU = Independent sampling unit

19

---

---

---

---

---

---

---

---

**Random covariates are factors which impact the outcome but are not the primary explanatory variable**

The value of a random covariate is not known before an experiment is observed.

Including a random covariate can help isolate the true effect of the predictor on the outcome(s) of interest by reducing the error variance.

20

---

---

---

---

---

---

---

---

**A repeated measures model describes a study in which a response variable is measured multiple times**

Repeated measures may take many forms:

- Longitudinal repeated measures
- Multivariate repeated measures
- Spatial repeated measures

21

---

---

---

---

---

---

---

---

**Longitudinal** repeated measures are within-independent sampling unit factors that vary with time

Example:

Distance walked by each participant, on each day, over one month.

Bray and Maxwell, 1985  
Doncaster and Davey, 2007

22

---

---

---

---

---

---

---

---

**Spatial** repeated measures are within-independent sampling unit factors that vary over space

Example:

Pollution levels measured once per kilometer along a river.

Bray and Maxwell, 1985  
Doncaster and Davey, 2007

23

---

---

---

---

---

---

---

---

**Multivariate** repeated measures are observations measured multiple times for multiple outcomes

Important note:

**MultiVARIATE** ≠ **MultiVARIABLE**

24

---

---

---

---

---

---

---

---



**MultiVARIATE ≠ MultiVARIABLE**

**Multivariate** → Multiple outcomes

**Multivariable** → Multiple predictors or covariates

25

---

---

---

---

---

---

---

---

**COMMON STATISTICAL MODELS**

26

---

---

---

---

---

---

---

---

**Three statistical models are often used for multilevel and longitudinal studies**

Three models:

1. The univariate linear model
2. The multivariate linear model
3. The mixed linear model

We will **not** discuss how to fit the models for data analysis.

27

---

---

---

---

---

---

---

---

**Univariate, multivariate and mixed linear models have a common form**

A model is a statement about a **population**:

$$\text{response} = \text{prediction} + \text{error}.$$

The adjective **linear** reflects the additive error and says how the parameters act on predictors.

Data analysis uses a **sample** of data to find **estimates** of the parts of the model.

We observe the response values.

28

---

---

---

---

---

---

---

---

**The univariate linear model is used to analyze experiments with a single outcome measured at only one point in time**

Univariate models can only be applied to **uncorrelated** data, as found in studies with no clustering and a single outcome measure.

29

---

---

---

---

---

---

---

---

**The multivariate linear model is used to analyze experiments with multiple outcomes or hypotheses**

Multivariate outcomes may be multiple time points, multiple outcome variables measured at one time point, or multiple outcome variables measured at multiple time points.

30

---

---

---

---

---

---

---

---

**The mixed linear model is a more flexible model for analyzing experiments with multiple outcomes**

The mixed model is more flexible because it makes fewer assumptions than the multivariate model.

We will discuss model assumptions in the next lecture.

31

---

---

---

---

---

---

---

---

**Both multivariate and mixed linear models are used often for correlated data**

Multivariate and mixed models are employed in studies with:

A single level and multiple outcomes;

Multiple levels and a single outcome;

Multiple levels and multiple outcomes.

32

---

---

---

---

---

---

---

---

**DEFINING STUDY HYPOTHESES**

33

---

---

---

---

---

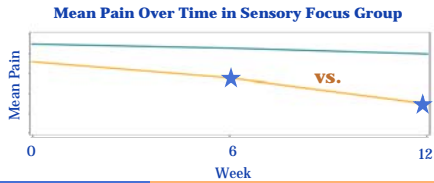
---

---

---

**Within-ISU hypotheses are useful for tracking changes in an ISU over time**

Within-independent sampling unit hypotheses compare measurements over time or space.



34

---

---

---

---

---

---

---

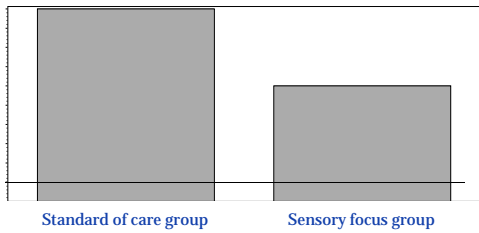
---

---

---

**Between-ISU hypotheses help identify differences between study groups**

**Mean Pain Reported**



35

---

---

---

---

---

---

---

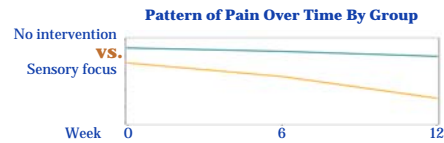
---

---

---

**Within by between-ISU hypotheses evaluate outcomes over multiple dimensions**

Within by between-independent sampling unit hypotheses compare the patterns of an outcome over repeated measures across groups.



36

---

---

---

---

---

---

---

---

---

---

**Review Summary**

- Statistical models calculate a predicted value of a response variable for each value or level of a predictor
- Predicted values then inform hypotheses tests, which ask whether observed values are significantly different than we would expect
- You can test between-ISU hypotheses, within-ISU hypotheses or an interaction

37

---

---

---

---

---

---

---

---



**End of Video**

38

---

---

---

---

---

---

---

---