

Choosing the Test

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

Review models used for multilevel and longitudinal studies.

Describe which hypothesis test statistics are used for each model.

2

Learning Objectives

Evaluate criteria for comparing tests

Learn how to choose an appropriate test for data analysis, and aligned power and sample size analysis

3

Recall, two statistical models are used most often for multilevel and longitudinal studies

Common models:

1. The multivariate model
2. The mixed model

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

4

Reversible mixed models are mixed models that can also be expressed as multivariate models

The mixed model is more general.

All multivariate models are special cases of mixed models.

A mixed model meeting the conditions on the next slide is **reversible**.

5

GLIMMMPSE calculates power and sample size for a mixed model under the following assumptions

1. Each ISU contains the same number of units of observation.
2. Units of observation are measured at the same times, the same locations, or are measured for the same variables.
3. Predictor variables are only measured once.

6

**SELECT THE HYPOTHESIS
TEST STATISTIC USED FOR
EACH MODEL**

7

**Tests for multivariate models can be
divided into two groups**

1. Univariate approach to repeated measures
2. Multivariate approach to repeated measures

8

**Univariate approach to repeated
measures tests assume commensurate
data**

Commensurate data have values observed on the same measurement scale.

Example:

Alcohol intake is measured 3, 6, and 9 months following treatment.

9

Multivariate approach to repeated measures tests allow the analysis of multivariate data

Examples:

Researchers measure intake of fats, proteins, and carbohydrates at weeks 0, 2, and 4.

Researchers measure height, weight and head circumference.

10

UNDERSTAND CRITERIA FOR COMPARING TESTS

11

Recall the definition of the null hypothesis

A **null hypothesis** claims that the two events are **not related**.

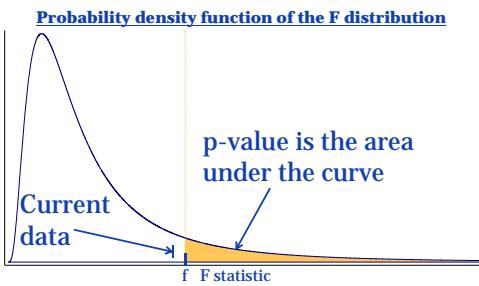
12

Small p-values lead researchers to achieve their goal of rejecting the null hypothesis

A Type I error occurs when a statistical test mistakenly rejects the null hypothesis when in fact the null hypothesis is true.

13

Recall, a p-value is the probability of seeing data more extreme than the current data if the null is true



14

The claimed or nominal Type I error rate is chosen by investigators

The rate is a number between 0 and 1.

It represents the hoped for probability that a test will reject the null hypothesis if the null hypothesis is true.

15

The actual Type I error rate can be estimated by simulation

Consider running the experiment many times and recording the results.

$$\text{Empirical Type I error} = \frac{\text{\# rejections}}{\text{\# experimental replications}}$$

16

Some testing paradigms have accurate Type I error rates

A Type I error rate is **accurate** if the actual rate is equal to the **claimed** rate.

Example:

The claimed rate is 0.05 and the actual rate is 0.05.

17

Some testing paradigms do not achieve the claimed Type I error rate

The actual Type I error rate may be higher or lower than the claimed rate.

Example:

The claimed rate may be 0.05, but the actual rate may be 0.25.

18

Inflated Type I error rates decrease the chance of replication of results

An **inflated** Type I error rate occurs when a test rejects the null hypothesis (which is true) at a higher rate than the desired alpha level chosen by the investigator.

19

Why is it important whether a test has an accurate Type I error rate?

When Type I errors occur, scientists draw wrong conclusions more often than desired.

20

Type I errors are not just statistical problems

Mistaken conclusions can directly affect patient care.

Example:

A Type I error may result in the approval of a drug that is ineffective.

21

Inflated Type I error rates decrease the chance of replication of results

Failure to replicate results means that two studies draw different conclusions.

22

Conflicting recommendations make patients lose trust in science

Example:

A study tests the null hypothesis that eating cranberries does not increase risk for cancer. Researchers reject the null and conclude that eating cranberries causes cancer. A second study fails to reject the null hypothesis and concludes that eating cranberries does not cause cancer. The public is unsure of the safety of eating cranberries.

23

Researchers must accept that some experiments will incorrectly reject the null hypothesis

When the actual Type I error rate is bigger than the claimed one, the study designer is increasing the chance that the study cannot be replicated.

24

We only trust findings if several studies come to the same conclusion

Replicated findings are findings from subsequent studies that match the original findings.

Scientific knowledge grows from the steady accretion of replicated studies.

25

A uniformly most powerful test achieves the greatest power among all reasonable tests of the same size

Uniformly most powerful tests only exist for certain special cases.

Example:

A uniformly most powerful test exists for the general linear hypothesis in the general linear univariate model (e.g., testing means equal).

26

Under certain conditions, different tests coincide

Coincidence occurs when two tests provide exactly same p-values, lead to the same inferences, and produce the same power and sample size when designing a study.

Coincidence can simplify the process of selecting a test.

27

We suggest two criteria for evaluating tests

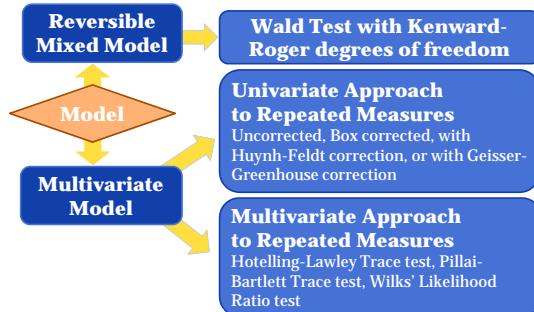
1. Check to see if the actual Type I error rate is close to claimed Type I error rate.
2. Choose the test with accurate Type I error rate that has the highest power.

28

CHOOSE AN APPROPRIATE TEST

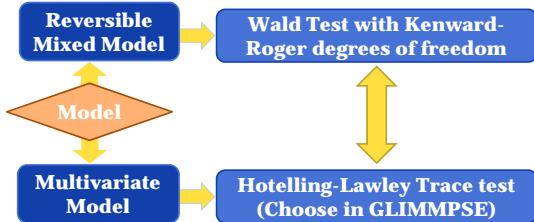
29

GLIMMMPSE incorporates a variety of tests, outlined in the map below



30

We will discuss why most testing decisions can be made using the following reduced map



31

Some tests may provide better power for certain designs/hypotheses

This lecture focuses on heuristics that apply to most, but not all, situations.

32

If you are uncertain of which test to use, we recommend using the Hotelling-Lawley Trace test

The Hotelling-Lawley Trace test often coincides with a popular mixed model test.

# of repeated measures	# of groups of ISUs	Recommended test
≤ 2	≤ 2	
≤ 2	≥ 3	
≥ 3	≤ 2	Hotelling-Lawley Trace test
≥ 3	≥ 3	

33

Recall, GLIMMPS only calculates power for reversible mixed models

The Wald test with Kenward-Roger degrees of freedom for reversible mixed models approximately coincides with the Hotelling-Lawley test.

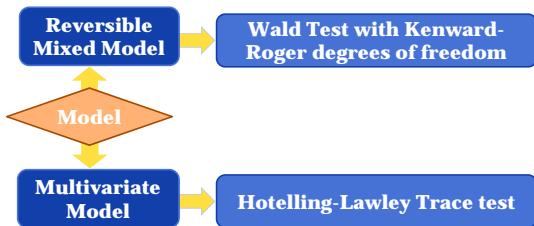
34

The Hotelling-Lawley trace test coincides with the Wald test under the following three conditions

1. No missing observations
2. Each ISU has same number of observations
3. No repeated predictors

35

Accurate power and sample size can be calculated for most study designs with the Hotelling-Lawley Trace test



36

Review Summary

- We often employ a multivariate model or a mixed model for multilevel and longitudinal studies
- Main criteria for comparing tests:
 - Actual and claimed Type I error rates should be close
 - Choose the test with accurate Type I error rate that has the highest power
- Because it can work in many different scenarios, we often recommend using the Hotelling-Lawley Trace test

37



38
