DATA ANALYSIS

Week 14: Repeated Measures (paired t-test)

logistics

- now available: how to report statistical tests
- PS policy revised, submissions are being graded as they come in
- Midterm 1 & 2 analysis now available
- last week to submit memes!
- next week:
  - memer contest opens
  - last extra credit survey opens

Reporting Statistical Test Results

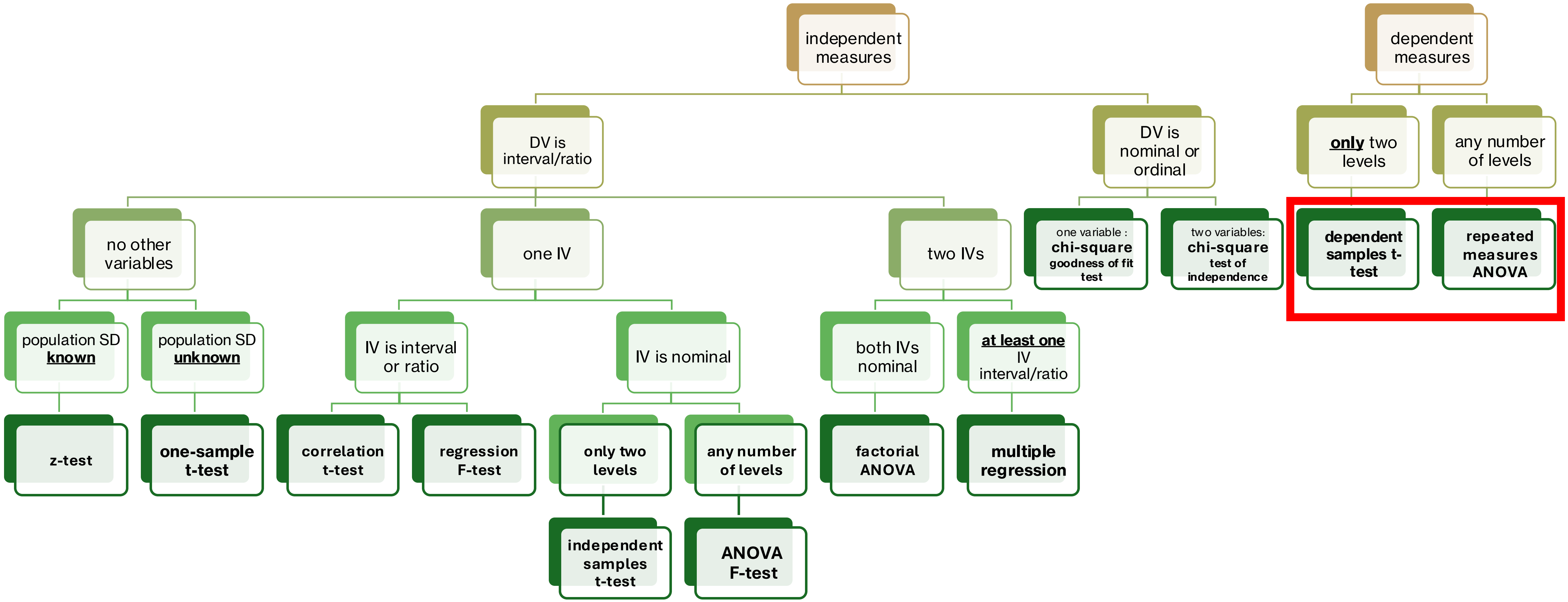
Test	APA Style Reporting
z-test	Adolescents who participated in these types of childhood group activities had significantly higher self-esteem, on average (M= 53.8), than found in the general population, z= 2.53, p = .011, d = 0.25.
one-sample t-test	Well-being scores for people with a history of frequent moves as children are significantly different (M = 37) from scores for the general population, t(11) = -3.17, p = .009, d = -0.92. Those who are frequent movers tend to have lower well-being scores than the general population.

14	T: April 22, 2025	W14: Repeated Measures
14	Th: April 24, 2025	W14 continued…
14	Su: April 27, 2025	**Week 14 Quiz due**
15	M: April 28, 2025	**PS5+ PS6 revision due**
15	T: April 29, 2025	W15: Miscellaneous Data
15	Th: May 1, 2025	W15 continued…
16	M: May 5, 2025	**PS7 due**
16	T: May 6, 2025	W16: Last Class / Final Exam review
17	Th: May 15, 2025	**PS7 revision + Computational Exam Due by 1.30 pm**
17	Th: May 15, 2025	**Conceptual Exam (1.30-3 pm, VAC South)**

special cases

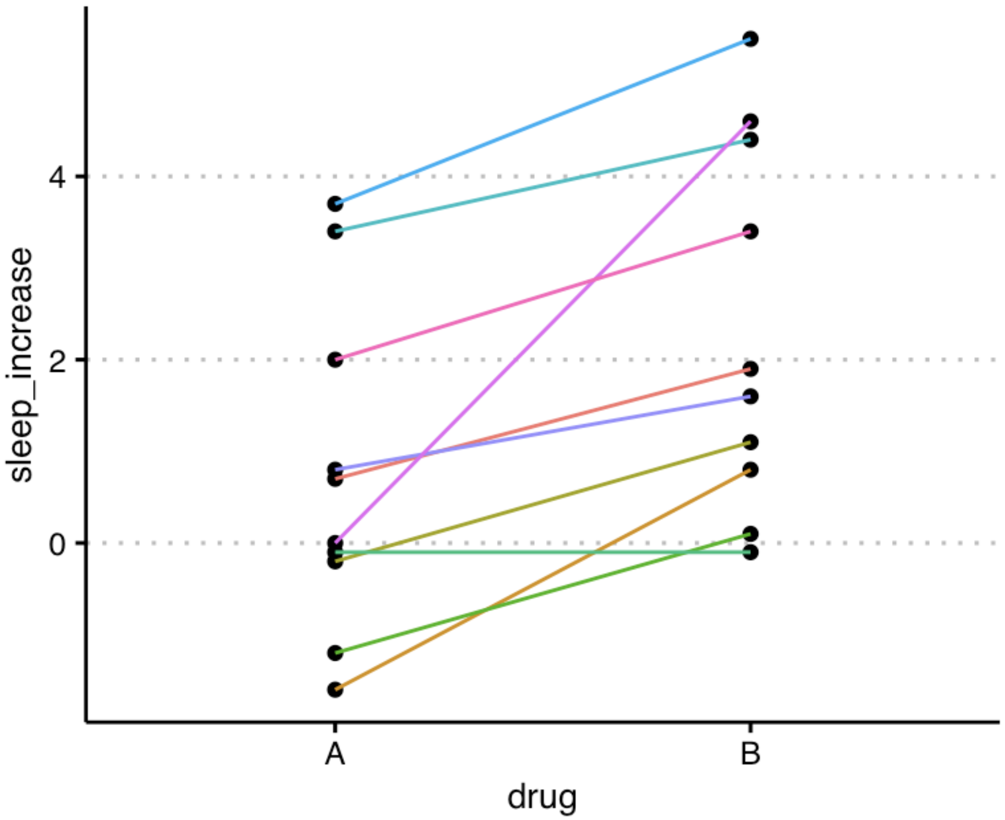
single means	• one-sample z and t-tests
two groups, independent samples	• independent samples / unpaired t-test
two groups, dependent samples	• dependent samples / paired t-test

hypothesis testing flowchart



sleep dataset

- data about “the effect of two soporific drugs (increase in hours of sleep compared to control) on 10 patients”
- last time, we did a repeated measures ANOVA for these data
- We found a significant effect of the type of drug ($M_{drug:A} = 0.75, M_{drug:B} = 2.33$) on sleep increase, *F* (1,9) = 16.5, *p* = .003. Drug B seems to be more effective than Drug A.
- since our IV only has two levels, this is a special case of the repeated measures ANOVA and one can also conduct a paired or dependent measures t-test for these data



assuming independence

- how would we have proceeded if the data were independent?
- we have 2 groups of scores, so we could have conducted an independent groups t-test

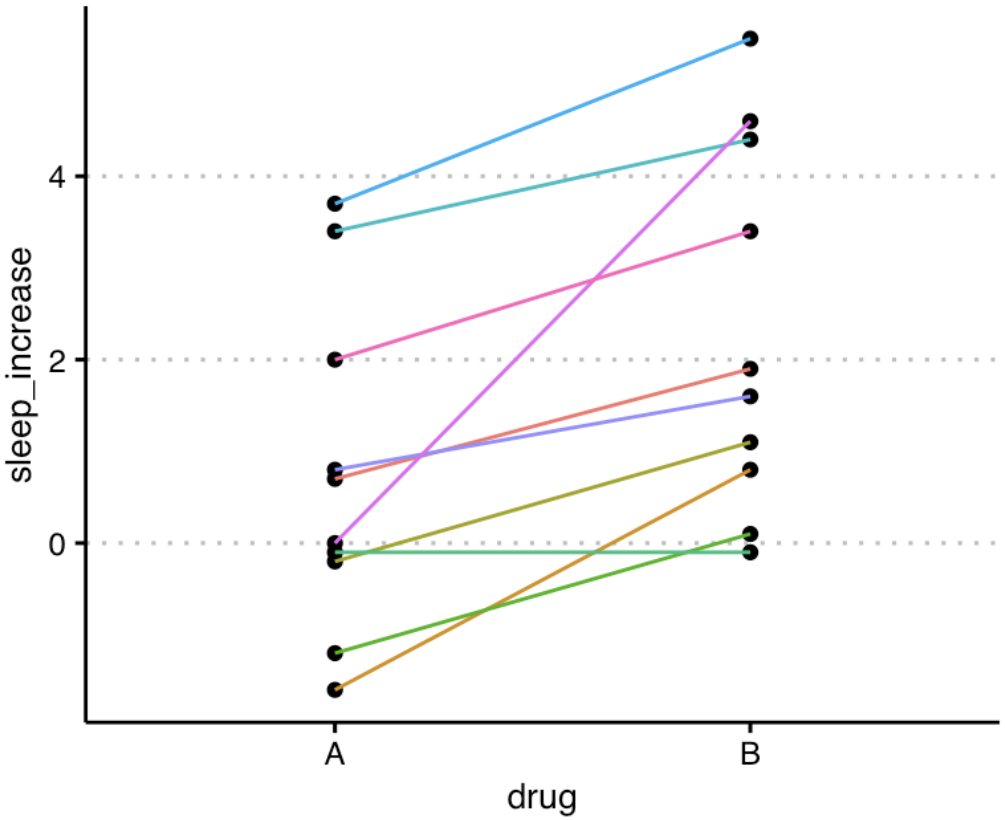
$$t_{df} = \frac{sample\ statistic\ (b)\ - population\ parameter\ (\beta)}{standard\ error} = \frac{(M_2 - M_1)\ - 0}{s_{M_2 - M_1}}$$

$$s_{M_2 - M_1} = \sqrt{\frac{s_p{}^2}{n_1} + \frac{s_p{}^2}{n_2}}$$

- **note**: we could also conduct the F-test (independent groups one-way ANOVA) where we substitute the group means and compare to the grand mean model

paired / dependent groups t-test

- similar idea but now we compute difference scores for each participant
  - D = $X_A - X_B$
  - if the drug had no effect on a participant, what should the value of D be?
- compute the mean of these differences
  - $M_D = \frac{\sum D}{n}$
  - if the drug had no effect overall, what should M_D be?



hypothesis testing (paired t-test)

- **step 1: state the hypotheses**
  - H_0: $\mu_D = 0$
  - H_1: $\mu_D \neq 0$
- **step 2: set criteria for decision**

$$t_{df} = t_{critical}$$

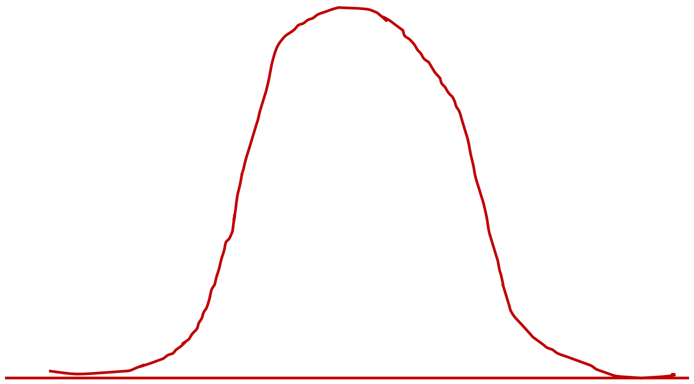
- **step 3: collect data**

$$t_{observed} = \frac{sample\ statistic\ (M_D)\ - population\ parameter\ (\mu_D)}{standard\ error}$$

$$t_{observed} = \frac{M_D - \mu_D}{SE} = \frac{M_D - \mu_D}{s_{M_D}}$$

- **step 4: make a decision!**

null hypothesis
sampling distribution
of ALL <u>mean differences</u>



NHST for two dependent groups (paired t-test)

step 1: state the hypotheses

$H_0: \mu_D = 0$
$H_1: \mu_D \neq 0$

step 2: set criteria for decision

$\alpha = .05$
find $t_{critical}$ based on one vs. two tailed test and degrees of freedom
$df = n - 1$

step 3: collect data

(1) compute $s_{M_D} = \frac{s_D}{\sqrt{n}}$
(2) compute $t_{observed} = \frac{M_D - \mu_D}{s_{M_D}}$
(3) find p-value for t-score

step 4: make a decision!

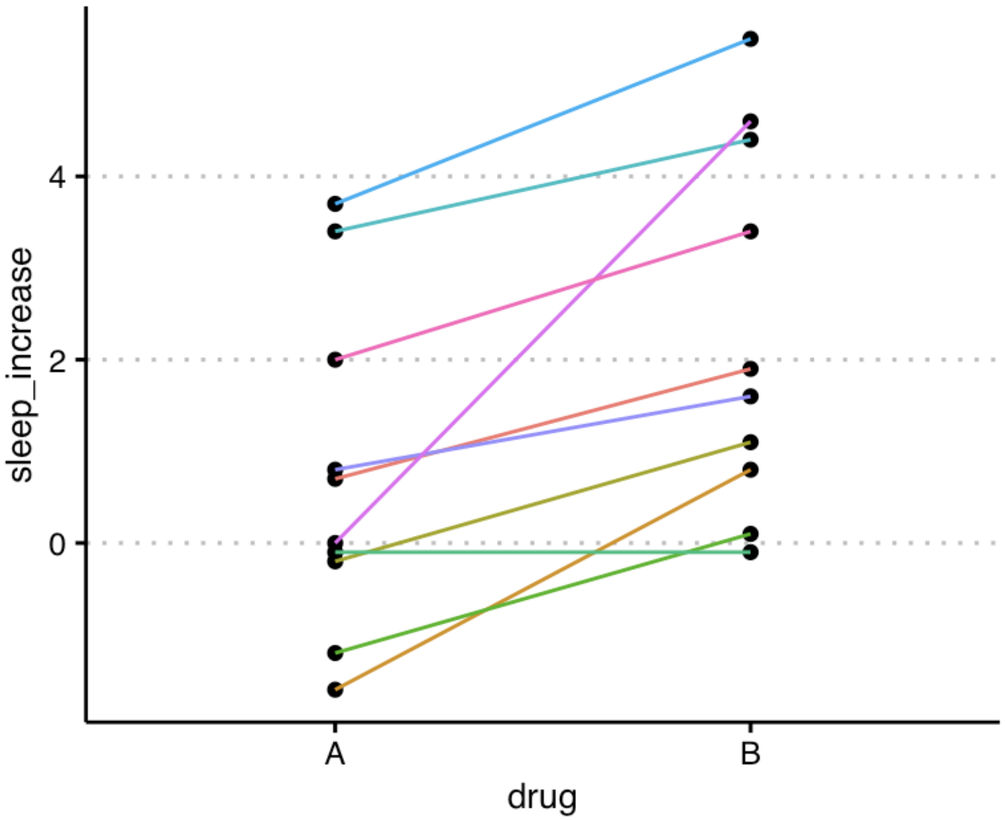
check whether $t_{observed}$ is beyond $t_{critical}$ and p-value < .05. if so, reject null hypothesis!

effect size for paired t-test

- effect sizes represent how extreme is the mean difference relative to the “standard” difference that is to be expected under the null hypothesis
- $d = \frac{\mu_D}{\sigma_D}$
- when the original population standard deviation is unknown, we estimate it using the sample standard deviation of mean differences
- estimated $d = \frac{M_D}{s_D}$

sleep data: paired t-test

- **step 1: state the hypotheses**
  - H_0: $\mu_D = 0$, no difference in drug effectiveness
  - H_1: $\mu_D \neq 0$, there is a difference in drug effectiveness
- **step 2: set criteria for decision**
  - find $t_{critical}\ (n-1)$ = ± 2.2621
- **step 3: collect data**
  - compute M_D = −1.58 $and\ s_{M_D} = \frac{s_D}{\sqrt{n}}$ = 0.388
  - $t_{observed} = \frac{M_D - \mu_D}{SE} = \frac{M_D - \mu_D}{s_{M_D}} = -4.062$
  - compute p-value = 0.0028
- **step 4: decide!**
  - There is a significant difference in drug effectiveness ($M_{drug:A} = 0.75$, $M_{drug:B} = 2.33$), $t\ (9) = -4.062$, p = .003. Drug B seems to be more effective than Drug A.
- Note, again, t^2 from paired t-test = F from repeated measures ANOVA!



Sheets solution
Video tutorial

W14 activity 2: your scores!

- did conceptual exam scores change across the two midterms?
- summarize the data
- conduct the t-test

complex designs and datasets

- often, we want more than a single measurement within each level/manipulation of our independent variable
- 25 trials per prime condition, 4 prime conditions: how many rows of data per participant?
- what should the data look like for a repeated measures ANOVA?
- what if the software glitches and we lose data for some trials or some conditions?

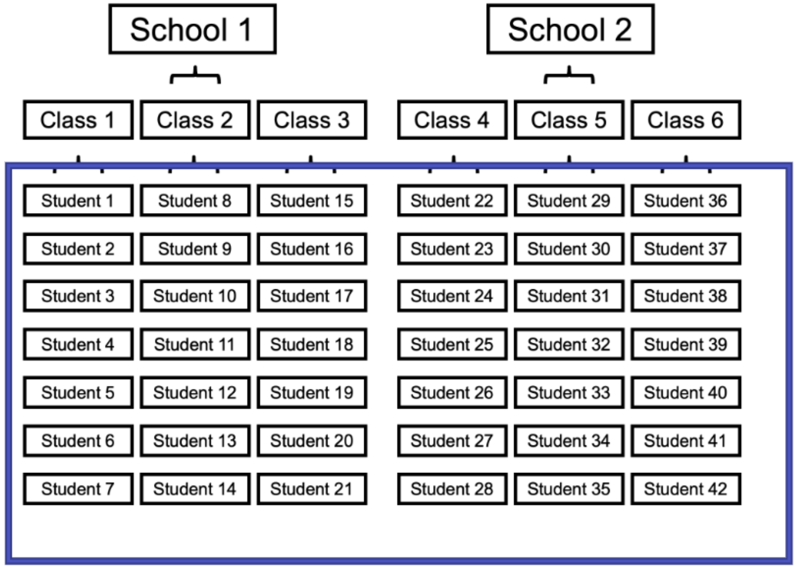
Younger and older adults attempted to retrieve words (e.g., abdicate) from low-frequency word definitions (e.g., *to formally renounce a throne*). Retrieval was preceded by primes that were "both" semantically and phonologically related(e.g., *abandon*), phonologically related (e.g., *abdomen*), semantically related (e.g., *resign*), or unrelated (e.g., *pink*). Each participant received definitions for all 100 target words, presented in a random order, with each type of prime word shown 25 times (i.e., they received the phonological prime on 25 of the 100 trials, the semantic prime on 25 of the 100 trials, etc.). Younger and older adults both benefited from phonological primes in retrieval, and also showed reduced, but reliable, facilitation from "both" primes compared to semantic and unrelated primes. There was no difference in retrieval for semantic and unrelated primes.

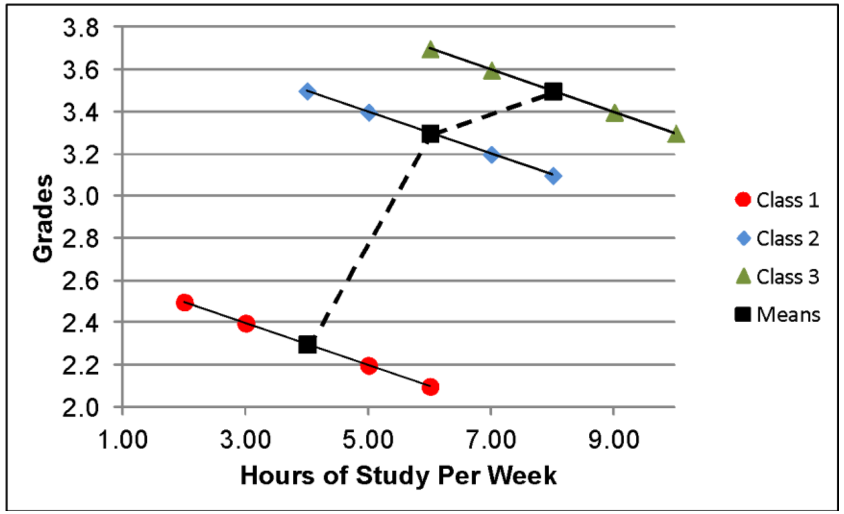
t-test & ANOVA limitations

- require simple designs, complete data, and normal residuals
- not equipped to handle missing data
- difficult to accommodate differing number of repeats/trials
- cannot capture nested/clustered designs

an example

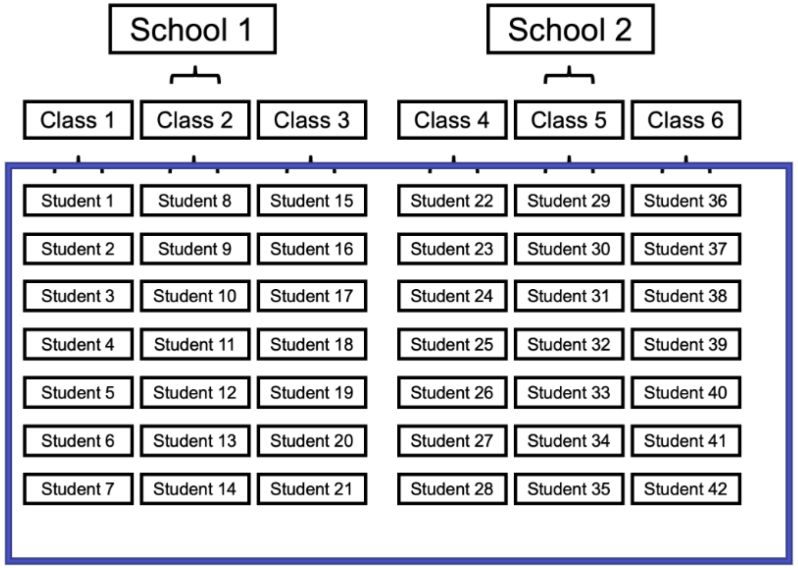
- when data exist at multiple levels there are potentially several levels of analysis
- "Is student math achievement related to hours of study per week, teacher experience, and availability of school resources?"
- ignoring the different levels of a research design or aggregation of data across levels can lead to serious flaws in analysis (e.g., Simpson's paradox)

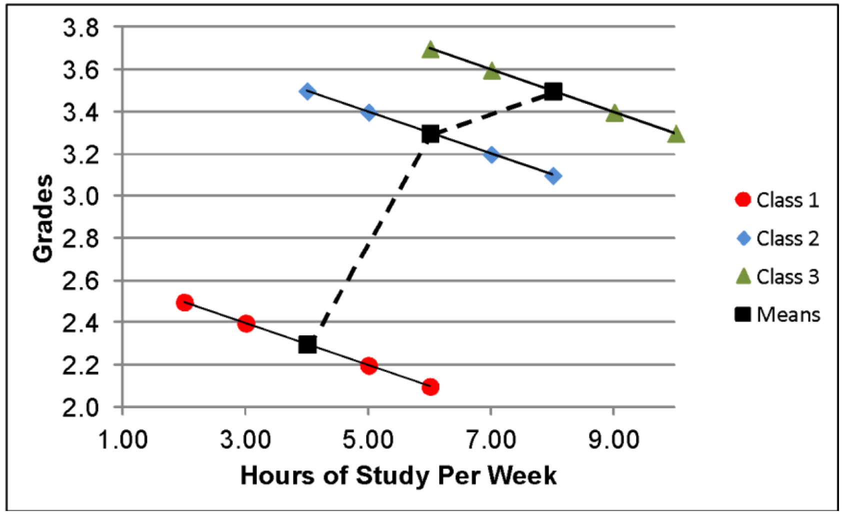




mixed effects models

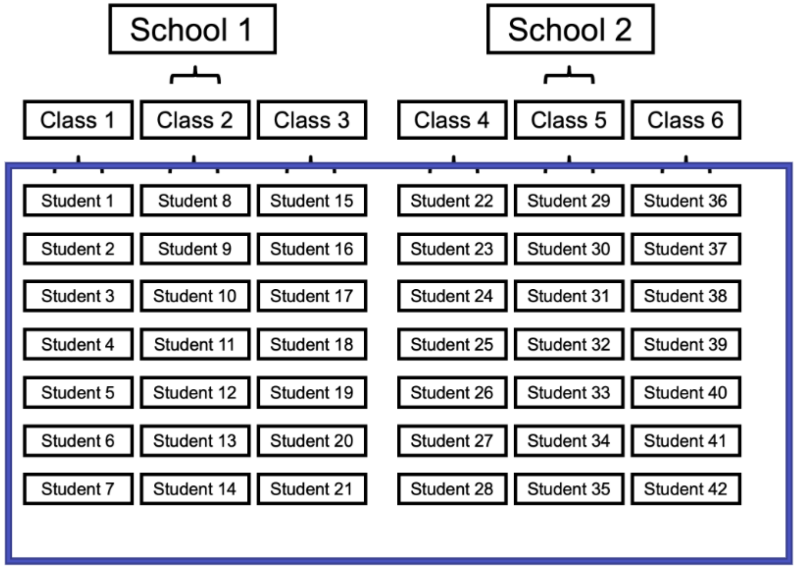
- linear/generalized mixed effects models consider the variability due to:
  - missing data
  - categorical/continuous IVs and DVs
  - unbalanced designs
  - clustered designs (no collapsing into means)
- think of them as the parent models from which special cases such as t-tests and ANOVAs are derived
- different ‘lines/curves’ are fit for each individual and for each item, with their own slope and intercept, instead of “averaging” across everyone

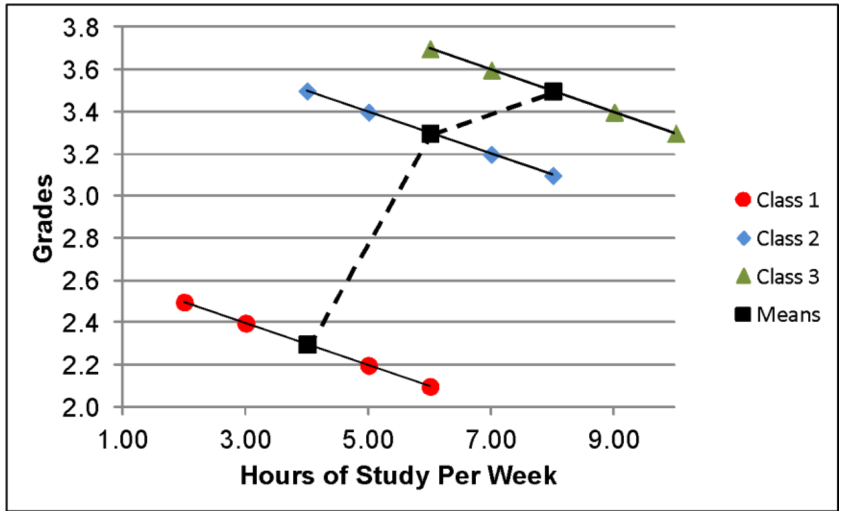




mixed effects models

- are appropriate when data are nested – when several units or levels of analysis are possible and their separate and joint influences need to be considered
- common alternative terms are multilevel models, random coefficient models, and hierarchical linear models
- data structures suitable for mixed models arise in a wide variety of common research problems
  - classes within schools within states
  - trials within subjects within age groups





W14 Activity 3

- reviewing all tests so far (plan for Week 15 & 16)
- review sheet

Reviewing Statistical Tests

data = model + error

Test	data	model	error
z-test			
one-sample t-test			

next time

- non-parametric tests

Here are the to-do's for this week:

- Submit Week 14 Quiz
- Submit revision for Problem Set 5
- Submit revision for Problem Set 6
- Submit any lingering questions here!
- Extra credit opportunities:
    - Submit Exra Credit Questions
    - Submit Optional Meme Submission

Before Tuesday

- Watch: Chi-square goodness of fit test.
    - Practice Data
    - Solution Sheet
- Watch: Chi-square test of independence.
    - Practice Data
    - Solution Sheet

Before Thursday

- Watch Bootstrapping Statistics
- Work on: Reviewing all tests.