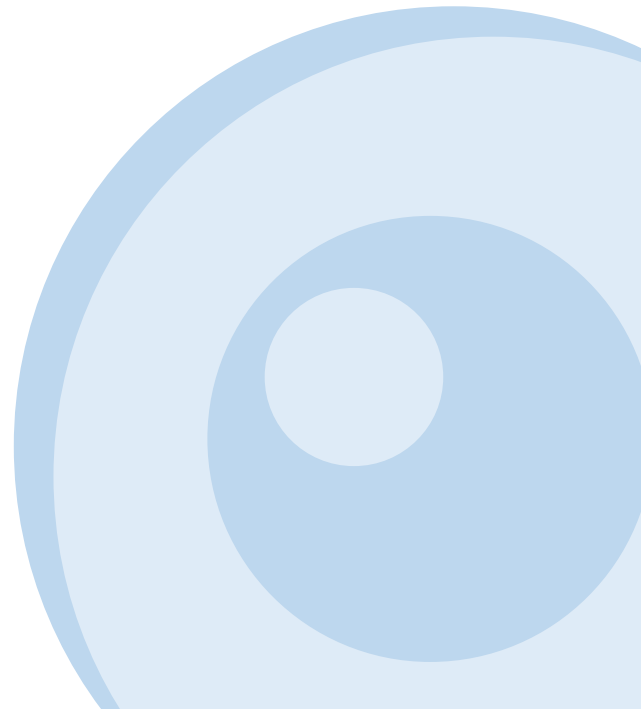


PSYC 193R:

Data Analysis & Visualization with R

Lecture 19: Review

Jonas S.-H. Lau [silau -et- ucsd.edu]
psyc193r.ucsd.edu



Two main families of tests

Comparing different
manipulations

Describing the linear
relationships
between variables



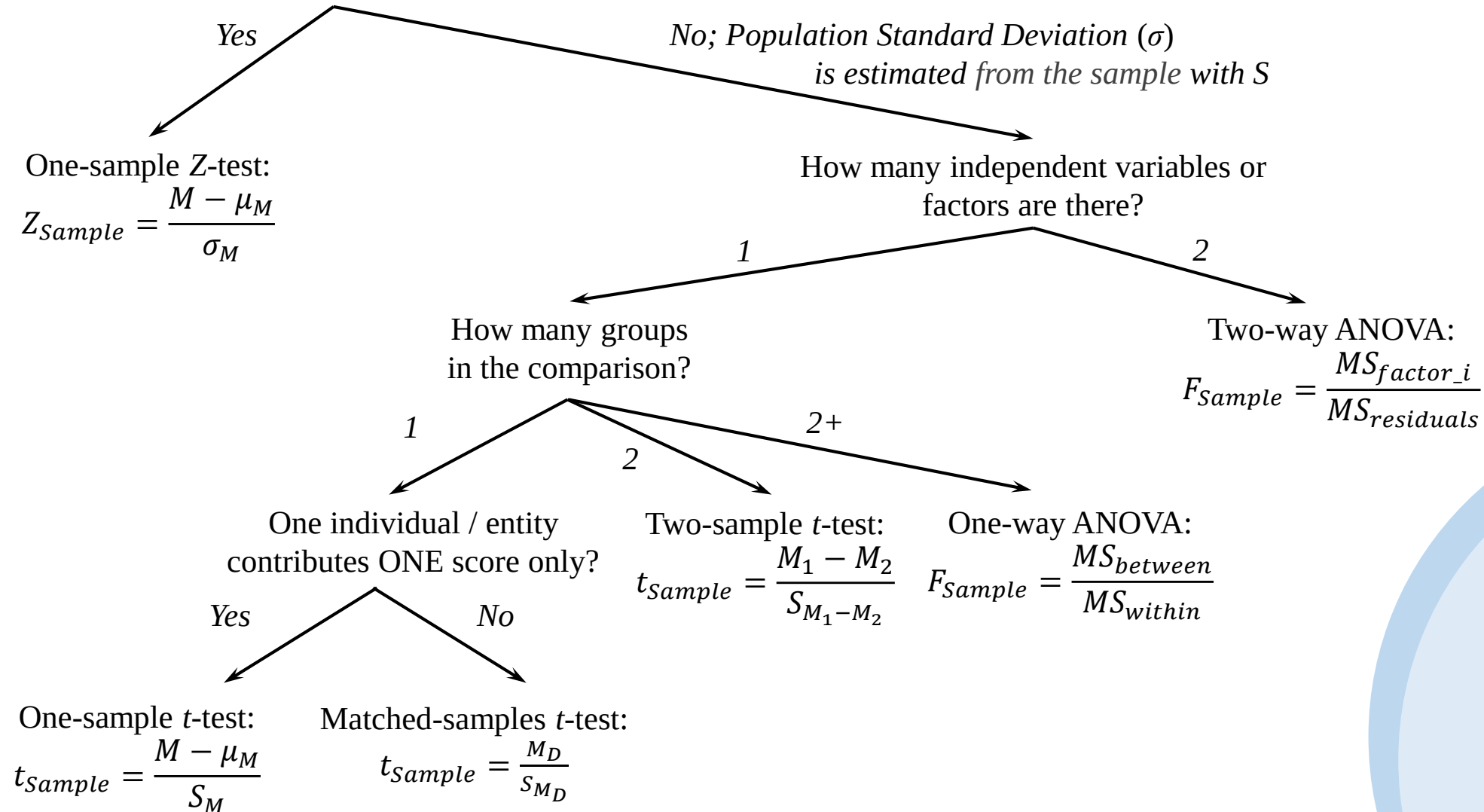
Picking the right test

Scale of Measurement

Independent variable:

Dependent variable:

Population Standard Deviation (σ) is *known*



Is drinking red wine
good for your health?

Is drinking red wine
good for your health?

Is drinking red wine
good for your health
... compared to not drinking any alcohol at all?

Is drinking red wine
good for your health
... compared to not drinking any alcohol / any liquid at all?

Is drinking red wine
good for your health
... compared to not drinking any alcohol at all?

Is drinking red wine
good for your **mental health**
... compared to not drinking any alcohol at all?

Is drinking red wine
good for your mental health?

(the *more* you drink, the *better* your mental health)

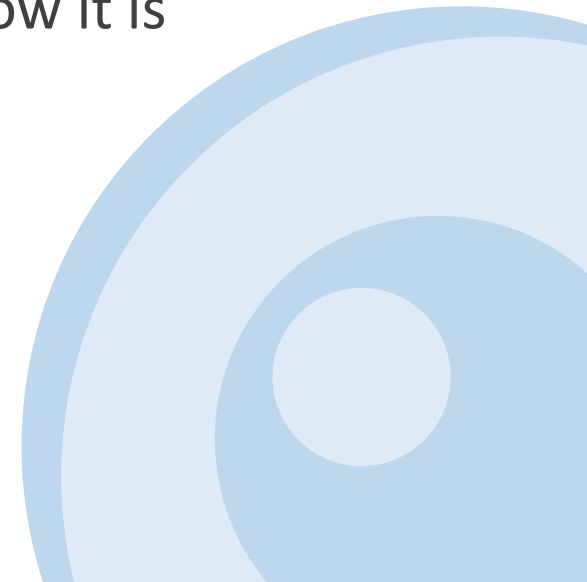
Design the study

→ Design the data analysis plan

→ Run the study

Your Tasks

- Design of the experiment / study
 - Estimate Power / Effect size
 - Check data structure
 - Check assumptions
 - Perform statistical test
 - Or interaction / simple effects
 - Post hoc tests
 - Effect Size
 - Confidence Interval
 - Interpreting the statistics, communicate the results
- The research questions
 - Best way to answer them
 - Target population vs sample
 - Level of Measurements
 - Operational Definitions
 - Theoretical IV vs how it is manipulated
 - Theoretical DV vs how it is measured



Your Tasks

- Design of the experiment / study
 - Estimate Power / Effect size
 - Check data structure
 - Check assumptions
 - Perform statistical test
 - Or interaction / simple effects
 - Post hoc tests
 - Effect Size
 - Confidence Interval
 - Interpreting the statistics, communicate the results
- Random sampling / Random Assignment
 - Experiment / Quasi-Experiment / Observational Study
 - Between-within subjects design
 - Law of large numbers
 - Central Limit Theorem
 - Sampling distributions
 - Standard Error



Your Tasks

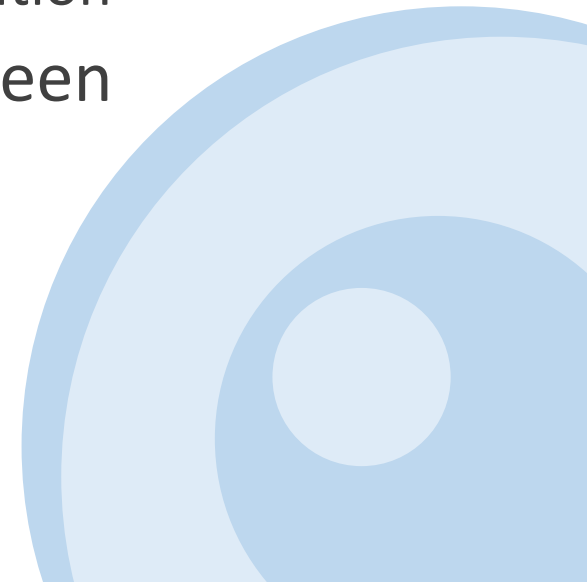
- Design of the experiment / study
 - Estimate Power / Effect size
 - Check data structure
 - Check assumptions
 - Perform statistical test
 - Or interaction / simple effects
 - Post hoc tests
 - Effect Size
 - Confidence Interval
 - Interpreting the statistics, communicate the results
- Depends on the test you plan to use
 - Power:
How many subjects do we need in a study
 - Effect Size:
Whether the study has practical differences between conditions

Your Tasks

- Design of the experiment / study
 - Estimate Power / Effect size
 - Check data structure
 - Check assumptions
 - Perform statistical test
 - Or interaction / simple effects
 - Post hoc tests
 - Effect Size
 - Confidence Interval
 - Interpreting the statistics, communicate the results
- Data format that support the analysis you want to do
 - One column for:
 - Participant identification
 - Each IV / factor
 - DV
 - (except for paired sample t-test)
 - Correct "class" for each variable
 - Esp important for ANOVA

Your Tasks

- Design of the experiment / study
 - Estimate Power / Effect size
 - Check data structure
 - Check assumptions
 - Perform statistical test
 - Or interaction / simple effects
 - Post hoc tests
 - Effect Size
 - Confidence Interval
 - Interpreting the statistics, communicate the results
- Sample Size
 - Shape of sampling distribution
 - Representativeness of the target population
 - Outliers
 - Effects on mean and sd
 - Normality
 - Of sampling distribution
 - Equal variance between conditions



Your Tasks

- Design of the experiment / study
 - Estimate Power / Effect size
 - Check data structure
 - Check assumptions
 - Perform statistical test
 - Or interaction / simple effects
 - Post hoc tests
 - Effect Size
 - Confidence Interval
 - Interpreting the statistics, communicate the results
- Use the correct test
 - Hypotheses
 - Always testing for population parameters (μ , ρ , σ , etc)
 - Decision / Conclusion
 - Reject / Not reject H_0
 - Meaningful conclusion: avoid “there is a statistical significant effect” without elaboration
 - Caution on *not* rejecting H_0
 - Report relevant statistics, e.g., means and SDs of conditions, test statistics and p-values

Your Tasks

- Design of the experiment / study
 - Estimate Power / Effect size
 - Check data structure
 - Check assumptions
 - Perform statistical test
 - Or interaction / simple effects
 - Post hoc tests
 - Effect Size
 - Confidence Interval
 - Interpreting the statistics, communicate the results
- Logic of post hoc tests
 - When / Why do we need it
 - Interpretation
 - Avoid “there is a statistical significant effect” without elaboration
 - Mindful of Type I error inflation



Your Tasks

- Design of the experiment / study
 - Estimate Power / Effect size
 - Check data structure
 - Check assumptions
 - Perform statistical test
 - Or interaction / simple effects
 - Post hoc tests
 - Effect Size
 - Confidence Interval
 - Interpreting the statistics, communicate the results
- Describe the relationship between criterion variable (DV) and the predictors
 - Parsimonious solution
 - Meaningful Interpretation
 - Avoid...



Your Tasks

- Design of the experiment / study
 - Estimate Power / Effect size
 - Check data structure
 - Check assumptions
 - Perform statistical test
 - Or interaction / simple effects
 - Post hoc tests
 - **Effect Size**
 - Confidence Interval
 - Interpreting the statistics, communicate the results
- Practical importance of the research
 - Effect size from data:
 - Cohen's d
 - η^2
 - r^2



Your Tasks

- Design of the experiment / study
 - Estimate Power / Effect size
 - Check data structure
 - Check assumptions
 - Perform statistical test
 - Or interaction / simple effects
 - Post hoc tests
 - Effect Size
 - Confidence Interval
 - Interpreting the statistics, communicate the results
- Confidence Interval:
 - As precision of the estimated population parameter
“the true mean is likely to be within the interval ___% of the time”
 - As a reliability measure
“if we repeated the experiment, ___% of the time we will get a mean between that interval”

Your Tasks

- Design of the experiment / study
 - Estimate Power / Effect size
 - Check data structure
 - Check assumptions
 - Perform statistical test
 - Or interaction / simple effects
 - Post hoc tests
 - Effect Size
 - Confidence Interval
 - Interpreting the statistics, communicate the results
- Highlight the decisions
 - Research questions
 - Sample size estimation
 - Theoretical / Practical significance of the research
 - Avoid empty statements, e.g., “there is a significant effect...”
 - Highlight the results
 - Describe differences between conditions
 - Relationship between variables

Programming

- General Programming / Clarity / Commenting
 - General probability
 - Power
 - Data processing
 - Plotting
 - Assumption checking
 - Statistical tests
- `pnorm()` / `qnorm()` / `pt()` / `qt()`
 - `for( i in ____ ){ ... }`
 - `dat[dat$gender = "...",]`
 - `aggregate()`
 - `t.test()` / `aov()` / `summary()`
 - `pwr` package
 - `var.test()` / `TukeyHSD()`
 - `cor()` / `lm()` / `anova()` / `predict()` / `resid()`
 - `vif()` / `step()`
 - `hist()` / `plot()` / `ggplot()`

Programming

- General Programming / Clarity / Commenting
 - General probability
 - Power
 - Data processing
 - Plotting
 - Assumption checking
 - Statistical tests
- `pnorm()` / `qnorm()` / `pt()` / `qt()`
 - `for( i in ____ ){ ... }`
 - `dat[dat$gender = "...",]`
 - `aggregate()`
 - `t.test()` / `aov()` / `summary()`
 - `pwr` package
 - `var.test()` / `TukeyHSD()`
 - `cor()` / `lm()` / `anova()` / `predict()` / `resid()`
 - `vif()` / `step()`
 - `hist()` / `plot()` / `ggplot()`

Programming

- General Programming / Clarity / Commenting
 - General probability
 - **Power**
 - Data processing
 - Plotting
 - Assumption checking
 - Statistical tests
- `pnorm()` / `qnorm()` / `pt()` / `qt()`
 - `for( i in ____ ){ ... }`
 - `dat[dat$gender = "...",]`
 - `aggregate()`
 - `t.test()` / `aov()` / `summary()`
 - **pwr package**
 - `var.test()` / `TukeyHSD()`
 - `cor()` / `lm()` / `anova()` / `predict()` / `resid()`
 - `vif()` / `step()`
 - `hist()` / `plot()` / `ggplot()`

Programming

- General Programming / Clarity / Commenting
 - General probability
 - Power
 - **Data processing**
 - Plotting
 - Assumption checking
 - Statistical tests
- `pnorm()` / `qnorm()` / `pt()` / `qt()`
 - `for( i in ____ ){ ... }`
 - `dat[dat$gender = "...",]`
 - `aggregate()`
 - `t.test()` / `aov()` / `summary()`
 - `pwr` package
 - `var.test()` / `TukeyHSD()`
 - `cor()` / `lm()` / `anova()` / `predict()` / `resid()`
 - `vif()` / `step()`
 - `hist()` / `plot()` / `ggplot()`

Programming

- General Programming / Clarity / Commenting
 - General probability
 - Power
 - Data processing
 - **Plotting**
 - Assumption checking
 - Statistical tests
- `pnorm()` / `qnorm()` / `pt()` / `qt()`
 - `for( i in ____ ){ ... }`
 - `dat[dat$gender = "...",]`
 - `aggregate()`
 - `t.test()` / `aov()` / `summary()`
 - `pwr` package
 - `var.test()` / `TukeyHSD()`
 - `cor()` / `lm()` / `anova()` / `predict()` / `resid()`
 - `vif()` / `step()`
 - `hist()` / `plot()` / `ggplot()`

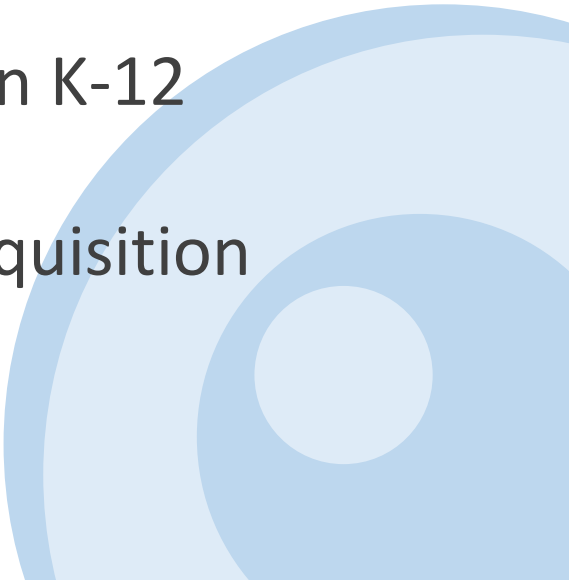
Programming

- General Programming / Clarity / Commenting
 - General probability
 - Power
 - Data processing
 - Plotting
 - Assumption checking
 - Statistical tests
- `pnorm()` / `qnorm()` / `pt()` / `qt()`
 - `for( i in ____ ){ ... }`
 - `dat[dat$gender = "...",]`
 - `aggregate()`
 - `t.test()` / `aov()` / `summary()`
 - `pwr` package
 - `var.test()` / `TukeyHSD()`
 - `cor()` / `lm()` / `anova()` / `predict()` / `resid()`
 - `vif()` / `step()`
 - `hist()` / `plot()` / `ggplot()`

Programming

- General Programming / Clarity / Commenting
 - General probability
 - Power
 - Data processing
 - Plotting
 - Assumption checking
 - Statistical tests
- `pnorm()` / `qnorm()` / `pt()` / `qt()`
 - `for( i in ____ ){ ... }`
 - `dat[dat$gender = "...",]`
 - `aggregate()`
 - `t.test()` / `aov()` / `summary()`
 - `pwr` package
 - `var.test()` / `TukeyHSD()`
 - `cor()` / `lm()` / `anova()` / `predict()` / `resid()`
 - `vif()` / `step()`
 - `hist()` / `plot()` / `ggplot()`

Scenarios

- Coffee drinking and study efficiency
 - Social Media Addiction
 - Preterm Babies' Health
 - M&M's in a bag
 - Mozart Effect
 - Financial Literacy
 - The Official SAT Practice
 - Face Reading
 - Lie Detection
 - Blind Tasting
 - Sleeping Pills and Sleep Quality
 - Productive Noise
 - Fitting Room Tricks
 - Video games for memory training
 - Brain Size and Intelligence
 - Wine Quality
 - Family involvement in K-12 Education
 - Second Language Acquisition
- 

A wine study

In 1990s, a winery would like to promote the benefit of responsible drinking on mental health. In a long-term study, they recruited a group of healthy adults in their 50s. The winery gave around half of the group a coupon each day for a glass of wine, and the rest of the group a coupon for a bottle of grape juice each day.

After ~20 years, the participants are now in their 70s. The winery employed an independent psychologist to test their memory capacity. The test has a full scale of 50 points (the higher the better). Each participants were tested twice (you might want to take the average).

Data at:

https://psyc193r.ucsd.edu/data/psyc193r_wineHealth_data.csv

(disclaimer: the data are, of course, fake)

A wine study

In 1990s, a winery would like to promote the benefit of responsible drinking on mental health. In a long-term study, they recruited a group of healthy adults in their 50s. The winery gave around half of the group a coupon each day for a glass of wine, and the rest of the group a coupon for a bottle of grape juice each day.

After ~20 years, the participants are now in their 70s. The winery employed an independent psychologist to test their memory capacity. The test has a full scale of 50 points (the higher the better). Each participants were tested twice (you might want to take the average).

wine:

0 -- grape_juice; 1 -- wine

gender:

0 -- female, 1 -- male

1. The estimated effect size is $d = 0.2$ or $\eta^2 = 0.1$ (not both), estimated the sample size needed for a 80% power
2. Check the assumptions
3. Perform a statistical test
4. Report the effect size

There are multiple ways to approach the problem... really!