HS 267D "Checklist"

The Labs encompass the complete curriculum. Use this checklist to help organize your studies.

Tutorial: measurement scales (quantitative, ordinal, and categorical); data table (variable, value, observation); SPSS basics (variable creation, data entry, basic menus); descriptive stats (stemplot; frequencies, mean, standard deviation, median, boxplot); Normal distribution; binomial distribution; confidence interval basics (z and t); hypothesis testing basics (z and t).

Lab 1: Inference about a proportion (Ch 16)

a. Data conditions: one sample, binary response; information quality; sample quality
b. Notation: p ≡ proportion parameter; $\hat{p}$ = sample proportion
c. CI for p (large sample and small sample)
d. Hypothesis test of H_0: p = some value (large sample and small sample)
e. Sample size for estimating p with given margin of error

Lab 2: Comparing two proportions (Ch 17)

a. Data conditions: two (independent) groups; binary response
b. Notation: RR ≡ relative risk parameter; $\hat{RR}$ ≡ relative risk estimator
c. Interpretation of $\hat{RR}$
d. CI for RR
f. Hypothesis test of H_0: $RR = 1$
e. Systematic error (info bias, selection bias, confounding) in analytic research

Lab 3: Naturalistic and cohort samples (Ch 18)

a. Data conditions: naturalistic sample or purposive cohorts, categorical response; no systematic error
b. Chi-square distribution (characteristics, table, *df*, relation to *z*)
c. Cross-tabulation (incidences/prevalences, relative risks, chi-square test of association)
d. Fisher's test
e. Sample size requirements

Lab 5: Stratified tables: confounding and interaction (Ch 19)

a. Data conditions: binary explanatory variable, binary response, categorical confounders
b. Confounding
c. Crude (overall) vs. strata-specific results
d. Mantel-Haenszel summary RR
e. Test for interaction

Lab 6: Correlation and regression (Ch 14)

a. Data conditions: quantitative explanatory variable, quantitative response variable
b. Scatterplot
c. Notation: ρ ≡ correlation parameter; r ≡ correlation estimator (strength and direction)
d. Test of H_0: $\rho = 0$
e. Notation: β ≡ slope parameter; b ≡ slope estimator (effect of X on Y)
f. CI for β
g. Coefficient of determination r^2
h. Dealing with nonlinear and non-normal relationships

Notions worth repeating before Final (Google Doc)

- Interpretations necessary. Integrate descriptive and inferential results.
- Inferential results (confidence intervals and *P*-values) should be cognizant of parameter being inferred. Examples of parameters:
    - population correlation coefficient (rho); estimator is *r*
    - slope coefficient parameter(s) ($beta_i$); estimator are b_i
    - binomial parameter (*p*); estimator is "p-hat"
    - relative risk parameter (*RR*); estimator is "RR-hat"
    - odds ratio parameter (*OR*); estimator is "OR-hat"
- Interpretation of confidence intervals
    - Intends to capture location of *parameter* (consider entire interval)
    - CI length quantifies precision (half confidence interval length = margin of error)
    - With due caution, and not over-simplifying, can be used to judge significance at various levels, e.g., a 95% CIs can be used to judge statistical significance at alpha = .05 level, a 90% CIs can be used to judge statistical significance at the alpha = .10 level, and so on.
- Interpretation of *P* values - Quantifies evidence *against* the null hypothesis, and nothing else (therefore, you must know and be aware of H_0)
    - "Significance" language without context will surely be misinterpreted; see: Cohen,1994 (link active)
    - *Not* a measure of effect size
    - No sharp boundaries (surely God loves *P* = .05 as nearly much as *P* = .06)
- When using inferential methods, i.e., *P*-values and CIs, be aware of assumptions (e.g., L.I.N.E.; expected values more than 5; etc. etc.)
- Systematic errors in public health research (e.g., information bias, selection bias, confounding) more important than random errors (CIs and *P*-values do *not* address systematic error)
- Be aware of sample types
    - Single (e.g., Ch 16), independent (e.g., Ch 17), paired (e.g., §18.6), case-control (18.5)
    - Experimental vs. observational designs
    - Naturalistic vs. cohort vs. case-control observational samples
- Computer doesn't replace knowledge:
    - Know what you are looking for
    - Computer output can be misleading: Illustration: SPSS output: Despite what output says, these are not risk statistics (RR NOT risk) + *RR* statistics do not apply because this is a case-control study.
- Illustrations
    - Sample size for estimating *p* (exercises 16.19 and 16.20, p. 371)
    - Sample size for comparing proportions: Lab 3B
    - Naturalistic/cohort sample (prison.sav), SPSS output -- Lab 3B
    - Stratified analysis: confounding and interaction -- Lab 5
    - Labs 6: simple regression

Inferential Methods: HS 267D **Computational Public Health Statistics**

Categorical outcomes

Sampling technique (Chapter)	**Explanatory Variable**	**Estimator ⇒ parameter**	**Confident Interval (method)**	**Null hypothesis (test procedure)**	**Additional points**
Single sample (Ch 16)	None	$\hat{p} \Rightarrow p$	CI for *p* (Plus-four or exact)	H_0: *p* = some value p_0 (z or binomial)	Selection bias Info. bias
Two samples (Ch 17)	Binary	$\hat{RR} \Rightarrow RR$	CI for *RR* (traditional method)	H_0: *RR* = 1 (z, chi-sq, or Fisher's)	Confounding Misclassification
Naturalistic & cohort (Ch 18)	Categorical	$\hat{p}_i \Rightarrow p_i$		H_0: *no association* (chi-sq or Fisher's)	Types of samples
Stratified tables – confounding (Ch 19)	Categorical	$\hat{RR}_{MH} \Rightarrow RR$	CI for RR (Mantel-Haenszel)	H_0: *RR* = 1 M-H chi-square	Simpson's paradox M-H Summary RR
Stratified tables - interaction (Ch 19)	Categorical	$\hat{RR}_{stratumk}$	N/A	H_0: $RR_1 = RR_2$ Heterogeneity chi-sq	Heterogeneity

Quantitative outcomes

Sampling technique (Chapter)	**Explanatory variable**	**Estimator ⇒ Parameter**	**Confidence interval (method*)**	**Null hypothesis (test procedures*)**	**Additional points**
Single sample (§11.1 – §11.4)	None	$\bar{x} \Rightarrow \mu$	CI for μ *t* procedure	H_0: μ = some value μ_0 (*t* proc.)	Valid sample Valid information
Paired samples (§11.5)	Binary	$\bar{x}_d \Rightarrow \mu_d$	CI for μ_d Paired *t*	H_0: $\mu_d = 0$ Paired *t* proc	Matched pairs
Two samples (Ch 12)	Binary	$\bar{x}_1 - \bar{x}_2 \Rightarrow \mu_1 - \mu_2$	CI for $\mu_1 - \mu_2$ Unequal variance *t*	H_0: $\mu_1 - \mu_2 = 0$ Unequal variance *t*	Independent samples
Independent (Ch 14)	Quantitative	$r \Rightarrow \rho$ (correlation)	Not covered	H_0: $\rho = 0$	Linearity Bivariate Normality
Independent (Ch 14)	Quantitative	$b \Rightarrow \beta$ (regression)	CIs for β	not covered	L.I.N.E. Math transforms

Turquoise shading ≡ techniques covered in tutorial.

* Assumes Normality or central limit theorem (moderate to large samples). Transform or non-parametric methods with small non-symmetrical distribution.