Exam 1 Review

Dr. Daniel B. Rowe
Professor of Computational Statistics
Department of Mathematical and Statistical Sciences
Marquette University

Observational Studies vs. Randomized Studies

Study Design is the methodology that is used to collect the information to address the research question.

Observational Studies: Studies based on observations collected on subjects.

Randomized Studies: Studies in which a sample of subjects are selected, and then an intervention is applied to a random subsample.

Bias – A systematic error that introduces uncertainty in estimates of effect or association.

Blind/Double Blind – A participant is unaware of his or her treatment status. Double blind when both the participant and outcome assessor are unaware of the treatment status.

2.1 Vocabulary

Clinical Trial – A specific type of study involving human participants and randomization to the comparison groups.

Cohort – A group of participants who usually share some common characteristics and who are monitored or followed over time.

Concurrent – At the same time; optimally comparison treatments are evaluated concurrently or in parallel.

Confounding – Complex relationship among variables that can distort relationships between the potential risk factors and the outcome.

Cross Sectional – At a single point in time.

2.1 Vocabulary

Incidence – The number of new cases (of disease) over a period of time.

Placebo – An inert substance designed to look, feel, and taste like the active or experimental treatment.

Prospective – A study in which information is collected looking forward in time.

Retrospective – A study in which information is collected looking backward in time.

Quasi-Experimental design – A design in which subjects are not randomly assigned to treatments.

2.1 Vocabulary

Randomization – A process by which participants are assigned to receive different treatments.

Per Protocol – An analytic strategy whereby only participants that adhered to the study protocol are analyzed.

Placebo – An inert substance designed to look, feel, and taste like the active or experimental treatment.

Prevalence – The proportion of individuals with the condition at a single point in time.

Stratification – A process whereby participants are partitioned or separated into mutually exclusive or non-overlapping groups.

Nothing on Section 2.2 Observational Study Designs.

3.1 Prevalence

Prevalence refers to the proportion of participants with disease at a particular point in time.

An estimate of the prevalence of disease at baseline is

$$\text{Point Prevalence} = \frac{\text{Number of persons with disease}}{\text{Number of persons examined at baseline}}$$

3.1 Prevalence

Example 3.1 Computing Prevalence of Cardiovascular Disease (CVD)

TABLE 3–1 Men and Women with Diagnosed CVD

	Free of CVD	History of CVD	Total
Men	1548	244	1792
Women	1872	135	2007
Total	3420	379	3799

$$\text{Prevalence} = \frac{\#\text{ with disease}}{\#\text{ examined at baseline}}$$

Prevalence of CVD = 379/3799 = 0.0998 → 9.98%

Prevalence of CVD in Men = 244/1792 = 0.1362 → 13.62%

Prevalence of CVD in Women = 135/2007 = 0.0673 → 6.73%

3.2 Incidence

Incidence reflects the likelihood of developing disease among a group of participants free of the disease who are at risk of developing the disease over a specified observation period.

$$\text{Cumulative Incidence} = \frac{\text{Number of persons who develop disease during a specified period}}{\text{Number of persons at risk at baseline}}$$

$$\text{Incidence Rate} = \frac{\text{Number of persons who develop disease during a specified period}}{\text{Sum of the lengths of time during which persons are disease-free}}$$

3.2 Incidence

Cardiovascular Disease

Incidence of CVD?

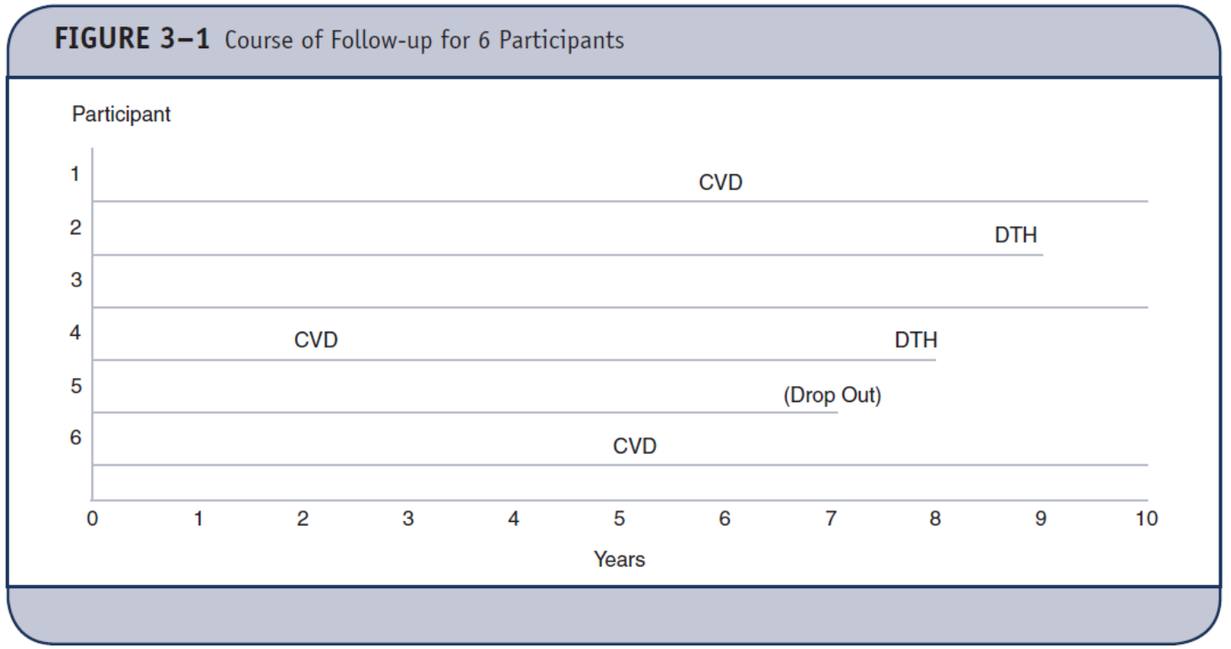
Incidence Rate of CVD

IR= 3/(6+9+10+2+7+5)

IR= 3/39

IR= 0.0769

7.7 per 100 person-years



3.4 Comparing Extent of Disease Between Groups

Cardiovascular Disease

Risk Difference of prevalent CVD in smokers versus nonsmokers

$$RD = \text{Prevalence}_{\text{smokers}} - \text{Prevalence}_{\text{nonsmokers}}$$

TABLE 3–2 Smoking and Diagnosed CVD

	Free of CVD	History of CVD	Total
Nonsmoker	2757	298	3055
Current smoker	663	81	744
Total	3420	379	3799

$$\text{Prevalence} = \frac{\#\text{ with disease}}{\#\text{ examined at baseline}}$$

RD= 81/744 – 298/3055 = 0.1089 – 0.0975 = 0.0114

3.4 Comparing Extent of Disease Between Groups

Relative Risk (RR) of CVD in smokers versus nonsmokers

$$RR = \frac{\text{Prevalence}_{\text{smokers}}}{\text{Prevalence}_{\text{nonsmokers}}} = \frac{81/744}{298/3055} = \frac{0.1089}{0.0975} = 1.12$$

TABLE 3–2 Smoking and Diagnosed CVD

	Free of CVD	History of CVD	Total
Nonsmoker	2757	298	3055
Current smoker	663	81	744
Total	3420	379	3799

$$\text{Prevalence} = \frac{\#\text{ with disease}}{\#\text{ examined at baseline}}$$

3.4 Comparing Extent of Disease Between Groups

Odds Ratio of CVD in hypertensives vs. non-hypertensives.

$$OR = \frac{181/840 \,/\, (1-181/840)}{188/2942 \,/\, (1-188/2942)} = \frac{0.275 \,/\, 0.725}{0.068 \,/\, 0.932} = 4.04$$

TABLE 3–5 Prevalent Hypertension and Prevalent CVD

	No CVD	CVD	Total
No hypertension	2754	188	2942
Hypertension	659	181	840
Total	3413	369	3782

$$\text{Prevalence} = \frac{\text{\# with disease}}{\text{\# examined at baseline}}$$

$$OR = \frac{\text{Prevalence}_{\text{exposed}} \,/\, (1-\text{Prevalence}_{\text{exposed}})}{\text{Prevalence}_{\text{unexposed}} \,/\, (1-\text{Prevalence}_{\text{unexposed}})}$$

Data

The **population** is the collection of all individuals about whom we wish to make generalizations.

The **sample** is a subset of individuals from the population.

Dichotomous variables have only two possible responses. Yes/No

Ordinal variables have more than two possible ordered responses.

Categorical variables sometimes called nominal variables are similar to ordinal variables except that the responses are unordered.

Data

Continuous variables take on an unlimited number of responses between defined minimum and maximum values.

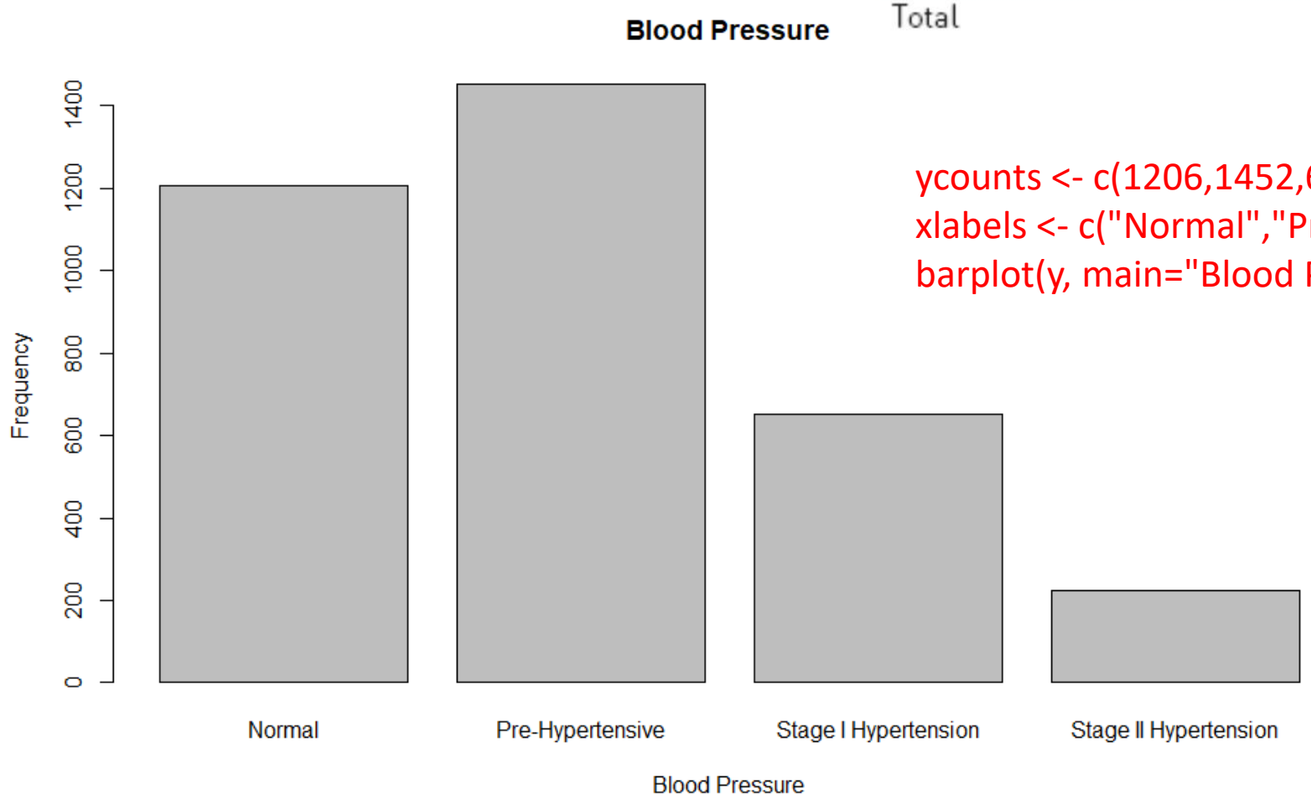
Statistics: Numerical summary measures computed on samples.

Parameters: Summary measures computed on populations.

4.2 Ordinal and Categorical Variables

Example:

	Frequency	Relative Frequency (%)	Cumulative Frequency	Cumulative Relative Frequency (%)
Normal	1206	34.1	1206	34.1
Prehypertension	1452	41.1	2658	75.2
Stage I hypertension	653	18.5	3311	93.7
Stage II hypertension	222	6.3	3533	100.0
Total	3533	100.0		



Description

```
ycounts <- c(1206,1452,653,222)
xlabels <- c("Normal","Pre-Hypertensive","Stage I Hypertension", "Stage II Hypertension")
barplot(y, main="Blood Pressure",xlab="Blood Pressure", ylab="Frequency",names.arg=xlabels)
```

4.3 Continuous Variables

Example 1: Small Numbers

Data values: 1,2,2,3,4

Sample Mean

Notation for sum x's

$$\bar{X} = \frac{\sum X}{n} = \frac{12}{5} = 2.4$$

$$\sum X = 1+2+2+3+4 = 12$$

Notation for sum x's

```
x <- c(1,2,2,3,4)
sum(x)
mean(x)
```

4.3 Continuous Variables

Example 1: Small Numbers

Data values: 1,2,2,3,4

Sample Median

$$median = middle \quad value$$

$$median = 2$$

Order data from smallest to largest.
If the number of data values is odd, take the middle value as the median.
If the number of data values is even, take the average of the middle two.

Sample Mode

$$mode = most \quad frequent \quad value$$

$$mode = 2$$

Order data from smallest to largest.
Count how many time each value occurs. Take the one with the highest count.

4.3 Continuous Variables

Example 1: Small Numbers

Data values: 1,2,2,3,4

Sample Variance & Standard Deviation

	X	$\bar{X}$	$X-\bar{X}$	$(X-\bar{X})^2$
	1	2.4	-1.4	1.96
	2	2.4	-0.4	0.16
	2	2.4	-0.4	0.16
	3	2.4	0.6	0.36
	4	2.4	1.6	2.56
$\sum$	12			5.20

$$s^2 = \frac{1}{n-1}\sum (X-\bar{X})^2$$

$$s^2 = \frac{1}{5-1}\left[(1-2.4)^2+(2-2.4)^2+(2-2.4)^2+(3-2.4)^2+(4-2.4)^2\right]$$

$$s^2 = \frac{5.2}{4} = 1.3$$

Standard Deviation

$$s = \sqrt{s^2} = \sqrt{1.3} = 1.14$$

4.3 Continuous Variables

Example 1: Small Numbers

Data values: 1,2,2,3,4

Sample Variance & Standard Deviation

	X	X^2
	1	1
	2	4
	3	9
	3	9
	4	16
$\sum$	12	34

$$s^2 = \frac{1}{n-1}\left[\sum X^2 - \frac{1}{n}\left(\sum X\right)^2\right]$$

$$s^2 = \frac{1}{5-1}\left[34 - \frac{12^2}{5}\right]$$

$$s^2 = \frac{5.2}{4} = 1.3$$

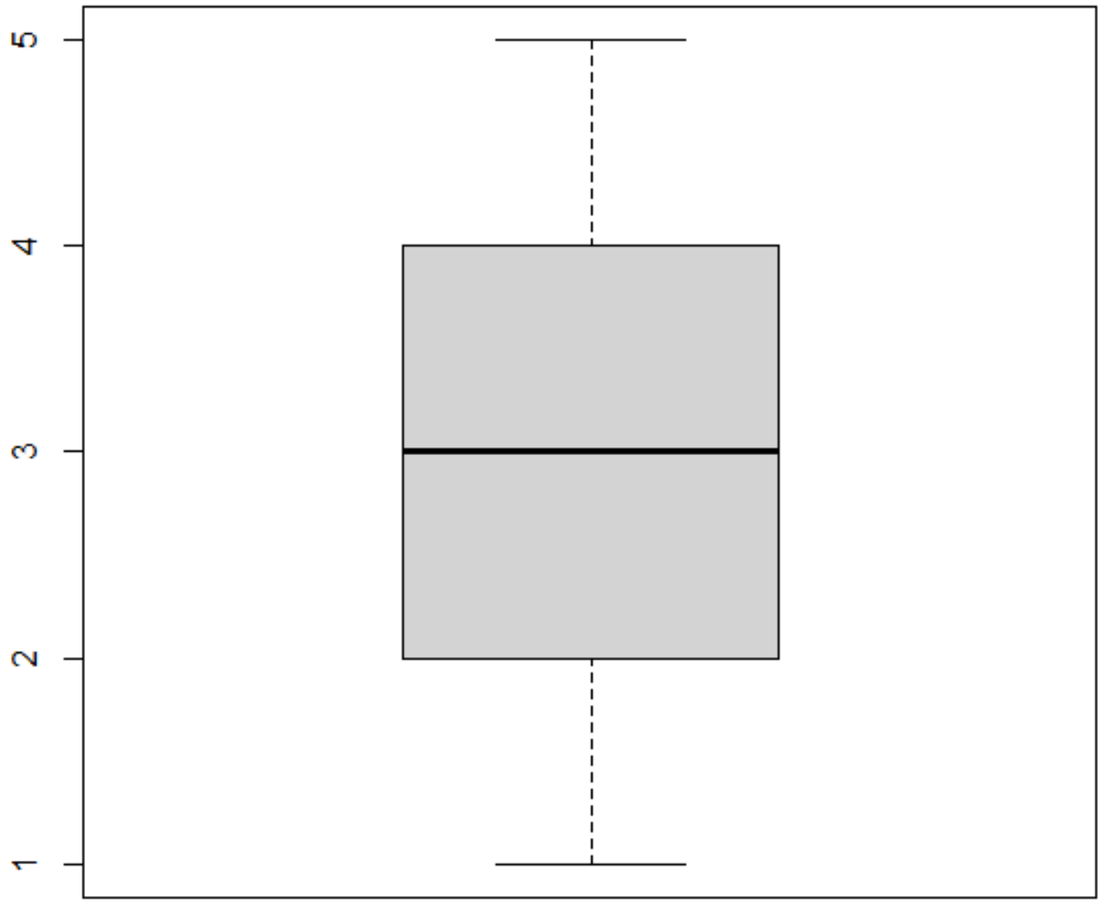
$$s = \sqrt{s^2} = \sqrt{1.3} = 1.14$$

4.3 Continuous Variables

Example 1: Small Numbers

Data values: 1,2,3,4,5

Box-Whisker Plot



5-number summary

1. L = minimum value
2. Q_1 = data value where 25% are smaller
3. Q_2 = median (where 50% are smaller)
4. Q_3 = data value where 75% are smaller
5. H = maximum value

$$IQR = Q_3 - Q_1$$

Q_1 = median of lower half.
Q_3 = median of upper half

0%	25%	50%	75%	100%
1	2	3	4	5

Questions?

Bring pencil, calculator, caffeinated beverage.

Will hand out exam and formula sheet.