University of Saskatchewan

Biostatistics Review
T-test, ANOVA, Chi-square

Public Health and Preventative Medicine Residents
Bonnie Janzen (slides donated by D. Fuller and P. Pahwa)

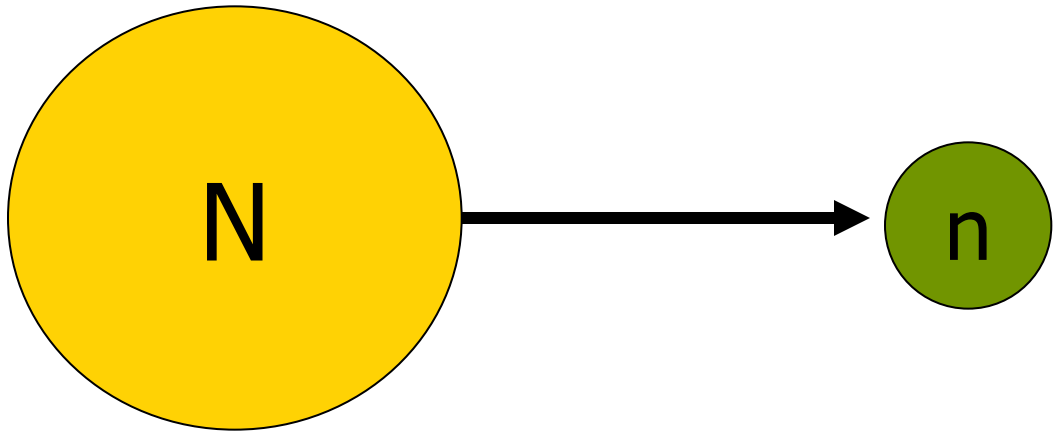
University of Saskatchewan

Objectives

- Describe how ANOVA, Chi-square, T-test are used to explain differences in health and health behaviors.
- Explore a health issue effectively by appropriately differentiating when to select each of these
- Explore a health issue effectively by critically appraising the use of each of these

Populations and Samples

- **Population**
  a) Entire groups of objects about which information is wanted
- **Sample:**
  a) A part of the population used to gain information about the whole



POPULATION

SAMPLE

UNIVERSITY OF SASKATCHEWAN

Populations and Samples

- Population Parameter (N)
  a) Characteristic of the population
     - Age = 45years
     - BP =115/75
- Sample Statistic (n)
  a) Estimates the population parameter
     - Age = 44years
     - BP =120/80
  b) May or may not be a good estimate

Populations and Samples

- Very difficult to measure everyone in an entire population
  a) **Census** = Measuring the entire populations
- Need to sample people from the population
  a) Two goals with samples
     - Big enough to represent the population
     - Not biased toward a certain group or variable
- **This is main reason we calculate confidence intervals, have *p* values, and use statistical tests. We want to understand uncertainty between our samples and the true population.**

Types of Data

- Data can be broadly categorized as **categorical** or **continuous**
  - a) **Categorical data**
    - Limited number of unique (“discrete”) possible values
    - Differences between categories can’t be described with consistent measures
  - b) **Continuous data**
    - Unlimited possible values on an unbroken scale
    - Measurable/quantifiable differences between values

Types of Data

- **Categorical**
  a) Limited number of possible values
  b) Difference between categories not quantifiable
- **Types of Categorial data**
  a) **Nominal:** Categories with *no ranking or order*
    - Men/Women, City A vs City B, blood types, marital status

  b) **Ordinal:** *Ranked/ordered* categories
    - A/B/C academic grades, cancer stages, self-rated health (excellent, very good, good…)

Types of Data

- **Continuous**
  a) Unbroken scale, Unlimited possible values
- **Types of continuous data**
  a) **Interval**
     - There *is* a consistent difference between values on a scale, but the ratio between values is not meaningful (not clear definition of “zero”)
       - Temperature in degrees C: 20℃ is not ‘twice as much heat’ as 10℃, and “0℃” does not mean “no heat”
  b) **Ratio**
     - Ratio of values has specific meaning; true zero point on scale
       - 20kg is twice as much mass as 10kg; “0kg” really does mean “no mass”

In general…

Qualitative/

Categorical

Nominal	Classification Only
Ordinal	Classification + Logical Order

Quantitative/

Continuous

Interval	Classification + Logical Order + Equal Intervals
Ratio	Classification + Order + Equal Intervals + Absolute Zero

UNIVERSITY OF SASKATCHEWAN

Statistical Tests

Exposure measure (independent variable)	Outcome measure (dependent variable)	Example	Test
Categorical, two groups only	Continuous	Comparing blood pressure between men and women	T-test
Categorical, two or more groups	Continuous	Comparing blood pressure between four med school classes	Analysis of Variance (ANOVA)
Categorical	Categorical	Comparing likelihood of dementia (Y/N) by sex/gender	Chi-square test (χ_2)
Any (multiple)	Categorical	Comparing likelihood of dementia (Y/N) by sex/gender *and* age	Logistic regression
Any (multiple)	Continuous	Comparing blood pressure by sex/gender and age	Linear regression

UNIVERSITY OF SASKATCHEWAN

T-test, ANOVA and Chi-Squared Test

Punam Pahwa, PhD
Dept. of Community Health and Epidemiology
Canadian Centre for Health and Safety in Agriculture

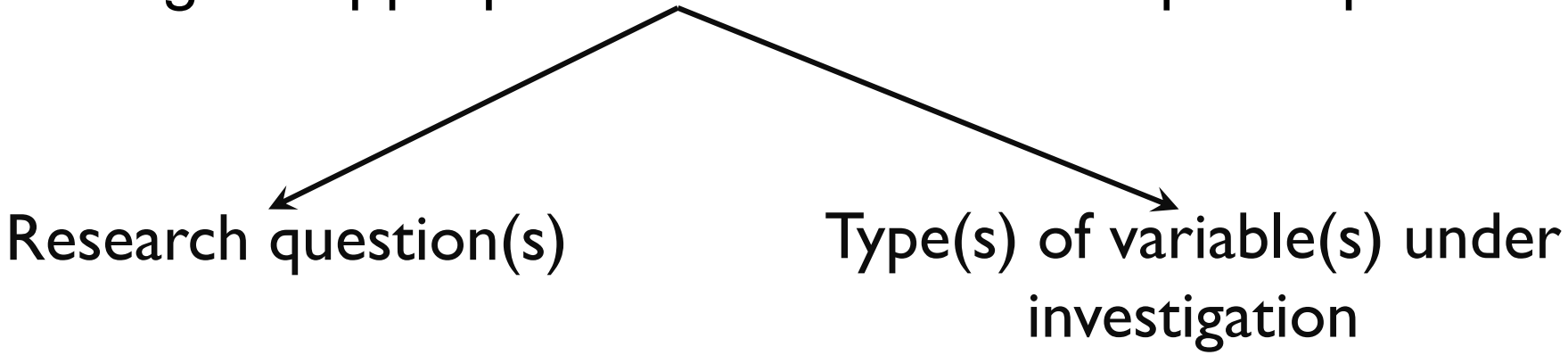
Introduction

- The purpose of most research is to :
  a) Compare two or more groups
     - male vs female (blood pressure), young vs old (mental health)
     - effectiveness of 3 treatments on acne lesions reduction
  b) Assess relationship/association among a set of variables.
     - Is there an association between smoking and lung cancer?
     - Is there an association between exposure to pesticides and non-Hodgkin's lymphoma?

- Diseases are multi-factorial – Statistical methods – not focus of this lecture

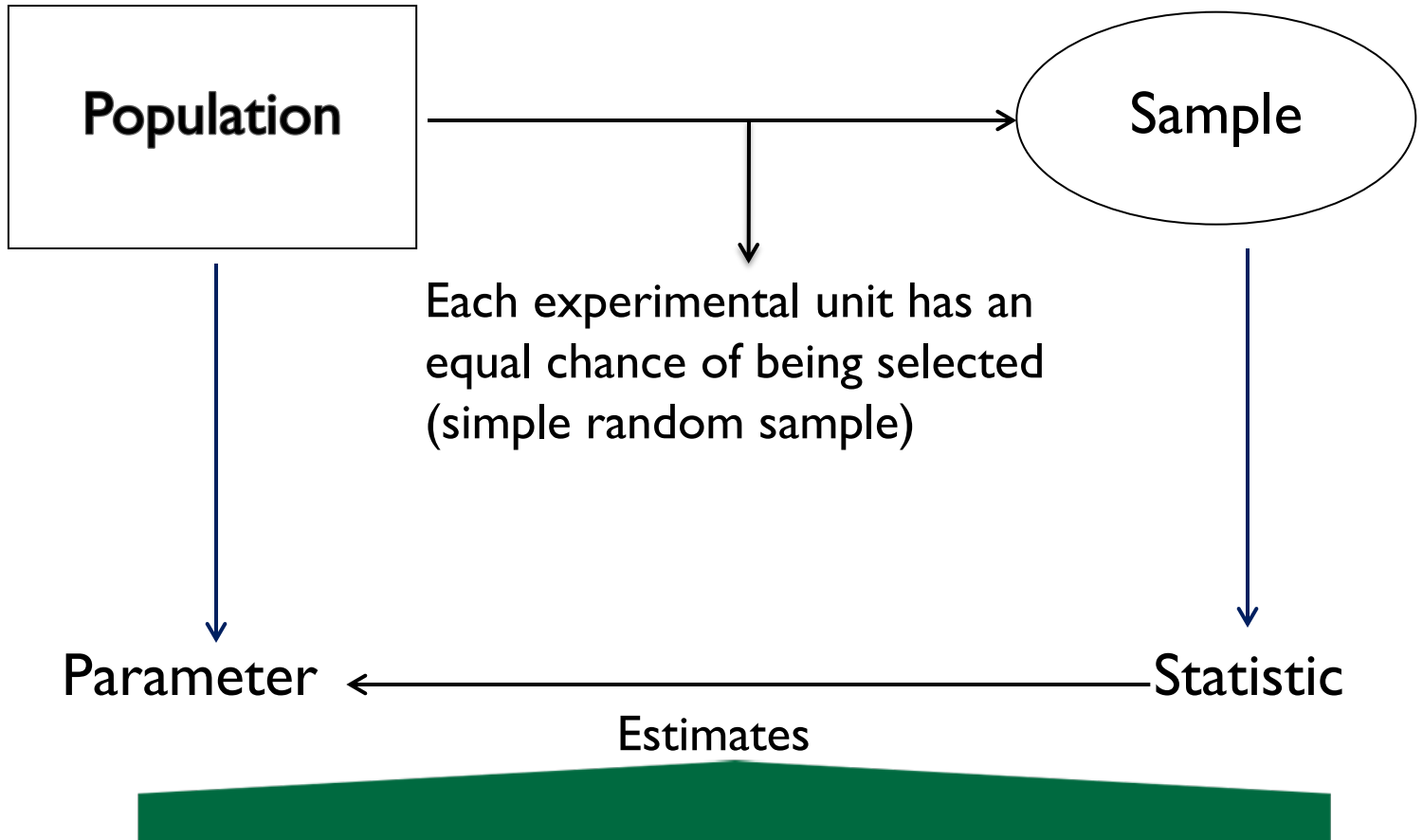
Introduction

- Choosing an appropriate statistical technique depends on



Population and Sample

The basic idea of statistics is simple: you want to extrapolate from the data you have collected to make general conclusions.



Analysis of Data

Descriptive Statistics

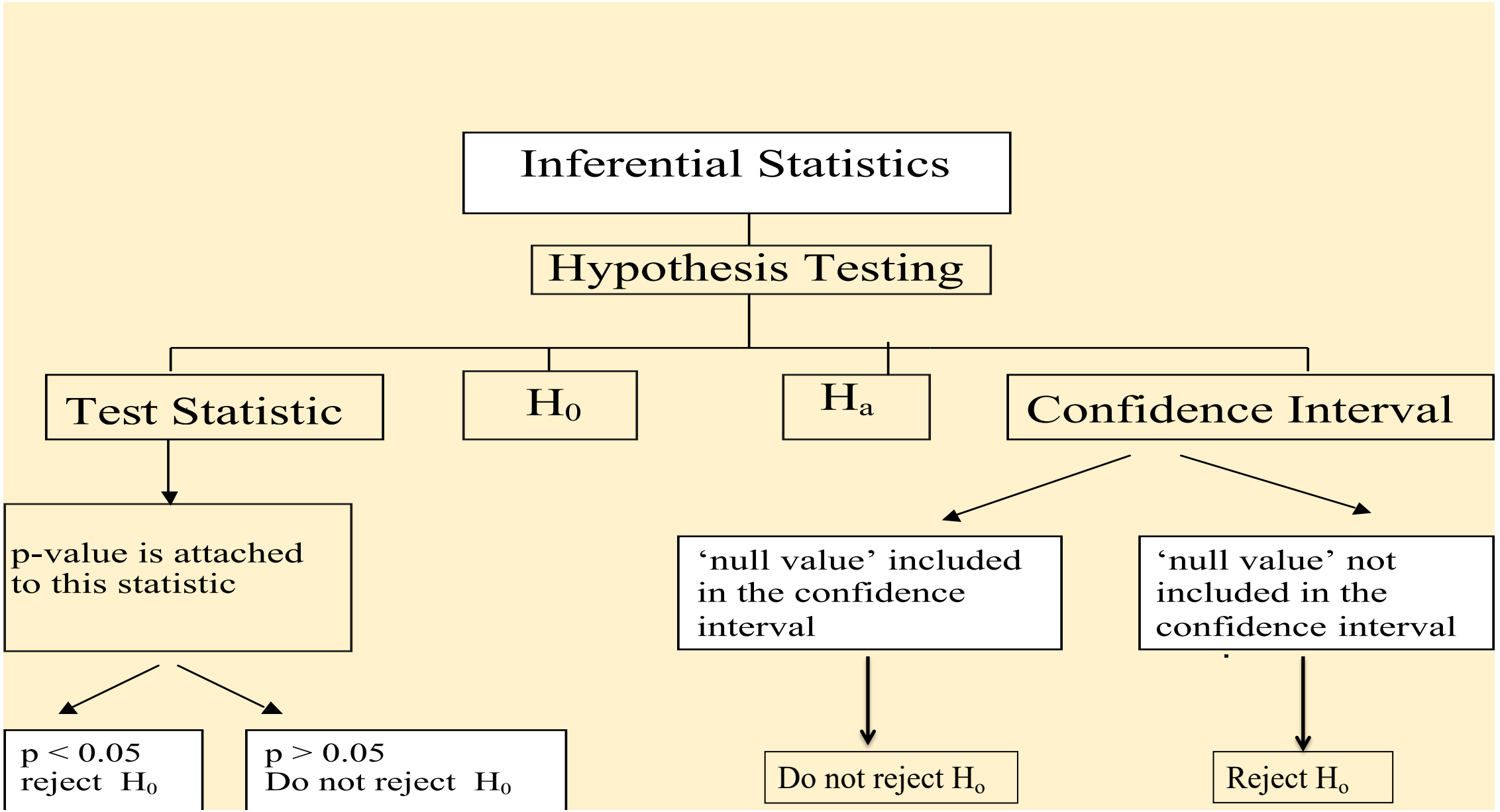
- Deals with processing data without attempting to draw any inferences from it (tables, graphs, etc.)

if you wish to organize and summarize existing data

Statistical Inference

- Concerns with Hypothesis Testing and Estimation

if you wish to generalize beyond existing data.


Inferential Statistics
Hypothesis Testing
Test Statistic
H_0
H_a
Confidence Interval
p-value is attached to this statistic
'null value' included in the confidence interval
'null value' not included in the confidence interval
p < 0.05 reject H_0
p > 0.05 Do not reject H_0
Do not reject H_o
Reject H_o

T-test

- Paired t-test

- Independent samples t-test

Purpose of t test

The t test is used for two different purposes :

(1) to estimate and test hypotheses about population variances

and

(2) to estimate and test hypotheses about population means.

- NOTE : WE ARE CONCERNED HERE WITH THE LATER USE.

Paired t-test

- **Example : Paired Samples**
- Each of 11 patients was administered a treatment on different occasions. Each patient first received the placebo. One month later they received a new drug. Blood pressure was recorded similarly on these two occasions. The data presented are systolic blood pressures.
    - The test of significance asks the following question: On the basis of these experimental results, is there any evidence of a difference in mean systolic blood pressure during the use of these two drugs?

Example from: Colton T. Statistics in medicine. Boston: Little, Brown and Company; 1974. Chapter 4, Inference on means; p.99-150. As

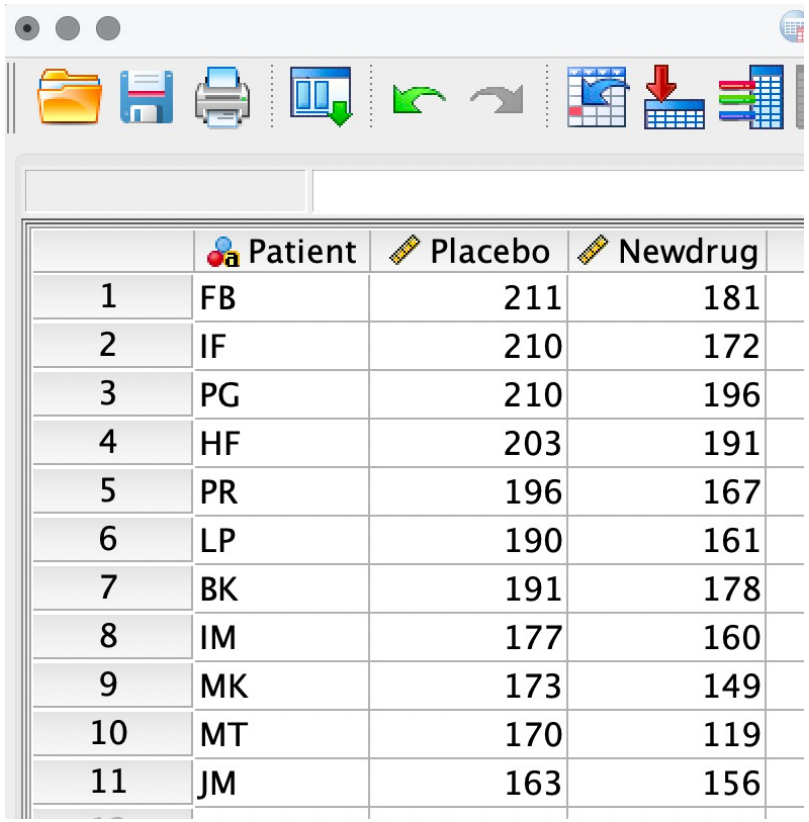
Paired t-test

Null and Alternative hypotheses:

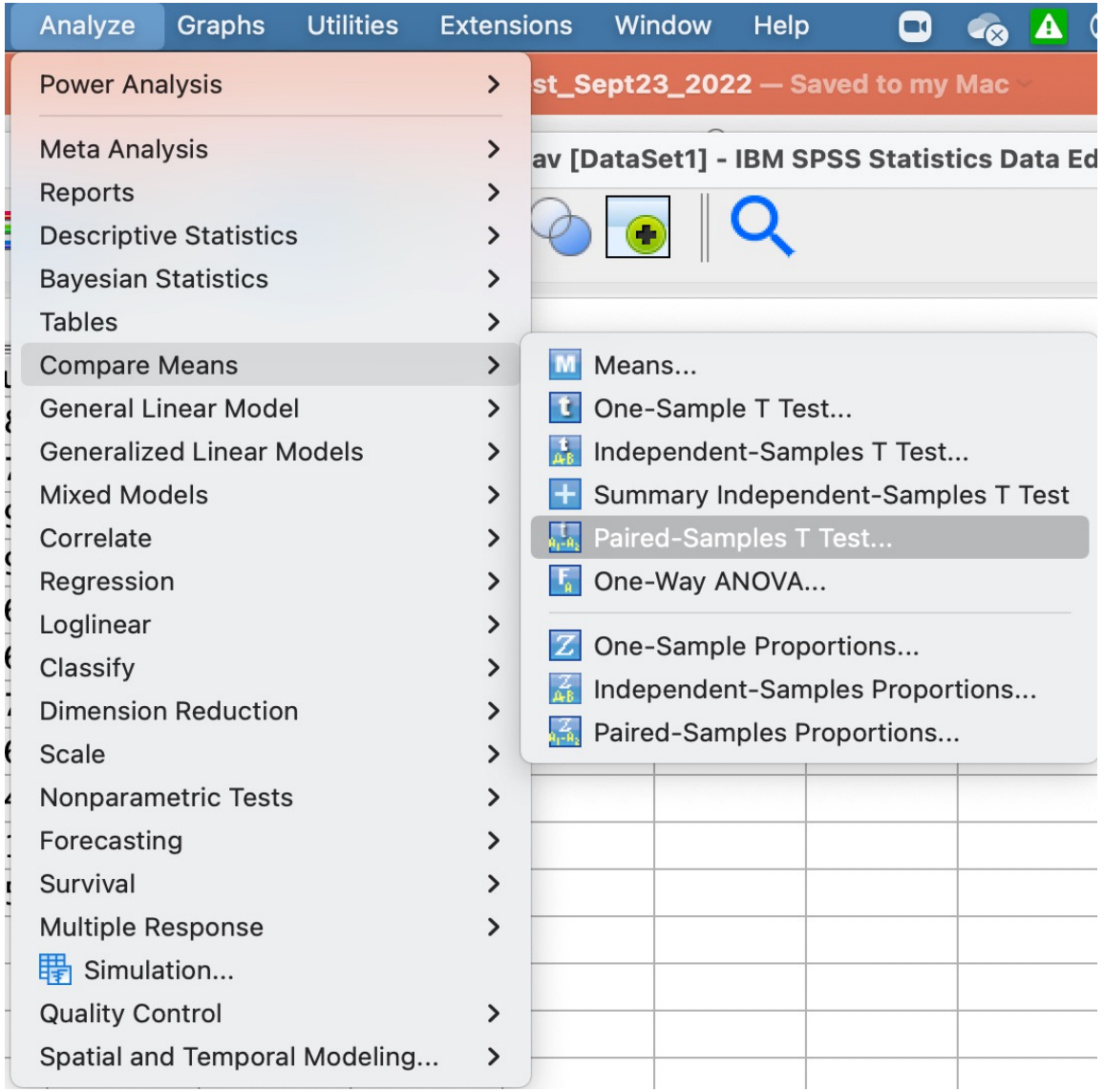
$H_0 : \mu_{SBPplacebo} = \mu_{SBPDrug}$ or $H_0 : d=0$

vs

$H_A : d \neq 0$



	Patient	Placebo	Newdrug
1	FB	211	181
2	IF	210	172
3	PG	210	196
4	HF	203	191
5	PR	196	167
6	LP	190	161
7	BK	191	178
8	IM	177	160
9	MK	173	149
10	MT	170	119
11	JM	163	156



SPSS COMMANDS FOR PAIRED t-test

Analyze

Compare Means

Paired-Samples T Test

Paired t-test

Paired Samples Test

		Paired Differences					t	df	Significance	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			One-Sided p	Two-Sided p
Pair 1	Placebo Placebo – Newdrug Intervention	24.000	13.092	3.947	15.205	32.795	6.080	10	<.001	<.001

Example – Independent samples

A psychologist measured reaction time (in seconds) of 18 individuals who were not given any stimulus and 16 individuals who were given an alcoholic stimulus. The data are given below. Do the data indicate that the reaction time is increased significantly by alcoholic stimulus?

Reatime	Group	Reatime	Group
3.00	1	5.00	2
2.00	1	4.00	2
1.00	1	4.25	2
4.00	1	4.00	2
1.50	1	3.75	2
2.00	1	3.00	2
1.00	1	4.50	2
2.25	1	2.75	2
2.50	1	2.50	2
2.25	1	3.25	2
1.75	1	3.80	2
1.50	1	4.20	2
2.00	1	4.60	2
1.75	1	4.50	2
1.00	1	4.45	2
1.50	1	4.00	2
2.00	1		
2.25	1		

Independent samples T-test

Null and Alternative hypotheses:

$H_0: \mu_1 = \mu_2$

OR

$H_0: \mu_1 - \mu_2 = 0$

Fundamental Assumptions

Population assumptions:

1. All populations are normally distributed.
2. All populations have the same variance.

Sampling assumptions :

1. The samples are independent of each other.
2. Each sample is obtained by using random sampling.
3. The outcome will be most interpretable if the data are measured on an interval or ratio scale/continuous.

Independent samples

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
REATIME	1	18	1.9583	.7389	.1742
	2	16	3.9094	.7076	.1769

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
REATIME	Equal variances assumed	.040	.843	-7.839	32	.000	-1.9510	.2489	-2.4580	-1.4441
	Equal variances not assumed			-7.859	31.805	.000	-1.9510	.2482	-2.4568	-1.4453

Hypotheses

H_0 (var) **:** Variances of two populations are equal, i.e $\sigma_1^2 = \sigma_2^2$. (NOTE: This is not our primary hypothesis)

$p > 0.05 \Rightarrow$ Do not reject H_0 (var), data are consistent with the null hypothesis. Variances of two populations are equal.

Main null hypothesis : H_0 : There is no difference in the reaction time between two groups.

OR

H_0: $\mu_1 = \mu_2$

OR

H_0: $\mu_1 - \mu_2 = 0$

$p<0.05 \Rightarrow$ Reject H_0, reaction time increased significantly with a stimulus

UNIVERSITY OF SASKATCHEWAN

ANOVA

ANalysis Of VAriance (ANOVA)

- Response/ Outcome/dependent variable (continous)
- Independent/factor variable (nominal variable or ordinal variable) has more than two levels/categories/groups

Example. Mental distress score among cancer patients – divided into four groups: (i) stage 1, (ii) stage II, (iii) stage III, and (iv) stage IV. (Dependent variable: Mental distress score -continuous, and independent variable: cancer stage with four levels/groups--ordinal).

Example – One-way ANOVA

- A study was conducted to compare the effectiveness of three comprehensive therapeutic programs for the treatment of mild to moderate acne. Three methods were employed:

- Treatment 1: Old method. It involved washing twice a day, everyday with a polyethylene scrub and an abrasive soap together with the use of 20 mg of tetracycline.
- Treatment 2: Currently in use. It required use of tretinoin cream, avoidance of sun, twice daily washing with an emollient soap and water, and utilization of 250 mg of tetracycline twice daily.
- Treatment 3: New method. It required water avoidance, twice daily washing with a lipid free cleanser, and use of tretinoin and benzoyl peroxide.

Example – One-way ANOVA

These three methods were to be compared for effectiveness in reducing the number of acne lesions in patients. Thirty-five patients participated in the study. These patients were randomly split into three subgroups of size 10, 12, and 13. One group was assigned treatment I, another, treatment II; and the third, treatment III. At the end of 16 weeks, the percentage improvement in the number of lesions was noted for each patient.

Description of Completely Randomized Design:

- Collection of *N* experimental units
- Study the effects of *k* different treatments.
- Units are randomly divided into k groups
- n_1= size of group 1; n_2= size of group 2 , ... , n_k= size of group k.
- The k subgroups are viewed as constituting independent random samples
- Drawn from populations with mean responses $\mu_1, \mu_2, \ldots, \mu_k$.
- Each group receives a different treatment.

Purpose of ANOVA

Analysis of variance is used for two different purposes :

(1) to estimate and test hypotheses about population variances

and

(2) to estimate and test hypotheses about population means.

- NOTE : WE ARE CONCERNED HERE WITH THE LATER USE.

ANOVA

1. For comparing multiple populations means
2. Makes only one comparison (Tests one null hypothesis)
3. To determine if any of the population means differ from the rest

Testing the Null Hypothesis

To test the null hypothesis:

a) A test statistic.
b) Probability distribution of a test statistic must be known under the assumption that the null hypothesis is true (i.e the k population means are equal).
c) For this to occur, certain assumptions must be made about the populations from which the samples are drawn.

Fundamental Assumptions

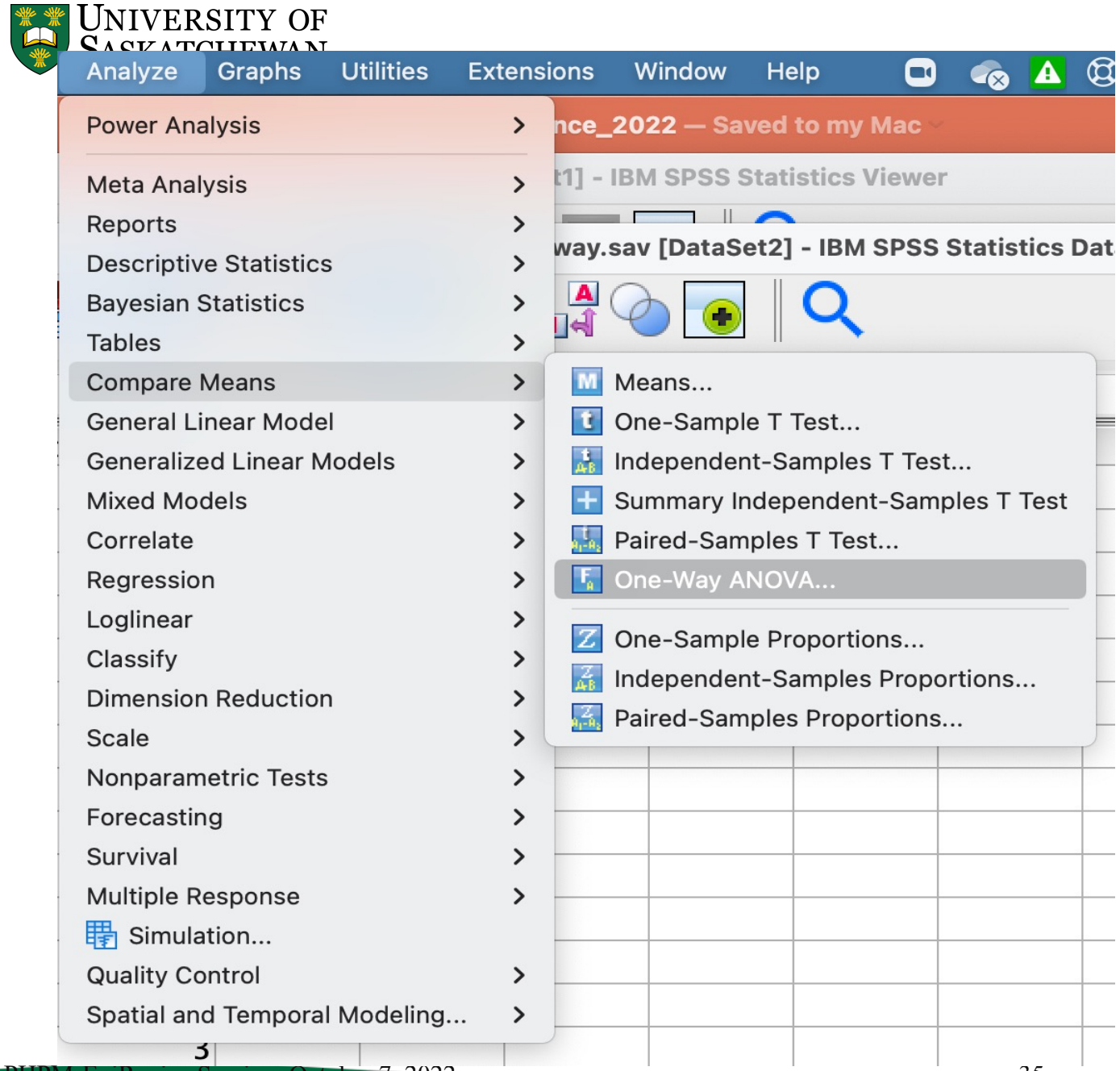
Population assumptions:

1. All populations are normally distributed.
2. All populations have the same variance.

Sampling assumptions :

1. The samples are independent of each other.
2. Each sample is obtained by using independent random sampling.
3. The outcome will be most interpretable if the data are measured on an interval or ratio scale/continuous.

	id	improv	treatment
2	2	49.40	1
3	3	50.10	1
4	4	49.80	1
5	5	50.60	1
6	6	50.80	1
7	7	47.10	1
8	8	52.50	1
9	9	49.00	1
10	10	46.70	1
11	11	68.00	2
12	12	67.00	2
13	13	70.10	2
14	14	64.50	2
15	15	68.00	2
16	16	68.30	2
17	17	71.90	2
18	18	71.50	2
19	19	69.90	2
20	20	68.90	2
21	21	67.80	2
22	22	68.90	2
23	23	67.50	3
24	24	62.50	3
25	25	64.20	3
26	26	62.50	3
27	27	63.90	3



Descriptives

percent percent improvement

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1 treatment 1	10	49.4600	1.73218	.54776	48.2209	50.6991	46.70	52.50
2 treatment 2	12	68.7333	2.00061	.57753	67.4622	70.0045	64.50	71.90
3 treatment 3	13	63.6000	2.22224	.61634	62.2571	64.9429	60.50	67.50
Total	35	61.3200	8.16008	1.37931	58.5169	64.1231	46.70	71.90

Test of Homogeneity of Variances

percent percent improvement

Levene Statistic	df1	df2	Sig.
.491	2	32	.616

This table is used to test the assumption that all populations have the same variances, i.e.

$H_{0(var)}$: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$

p=0.616>0.05 ⇒Do not reject $H_{0(var)}$

ANOVA

percent percent improvement

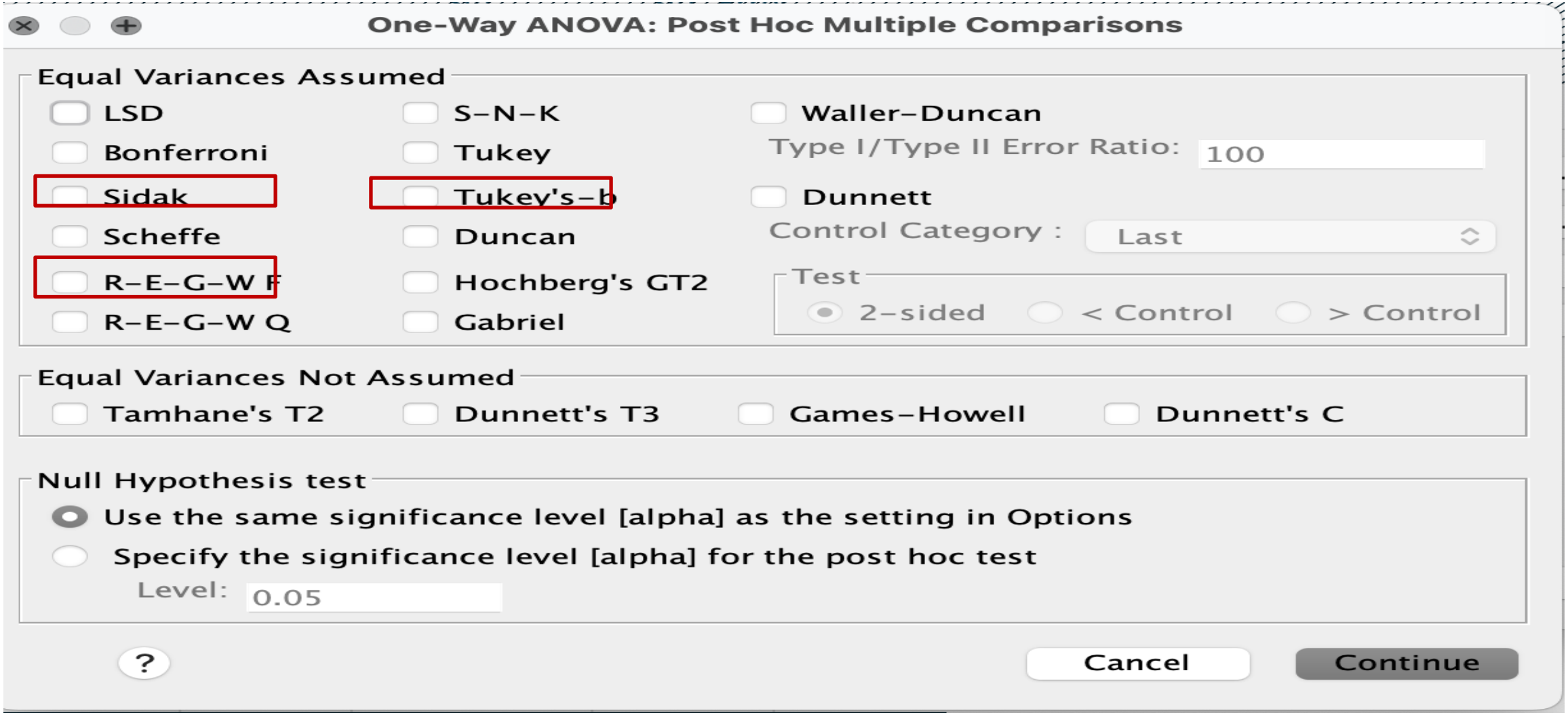
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2133.665	2	1066.833	262.019	.000
Within Groups	130.291	32	4.072		
Total	2263.956	34			

SST *k-1* *MST=SST/(k-1)* *F=MST/MSE*

TSS *SSE* *N-1* *N-k* *MSE=SSE/(N-k)*

Multiple Comparison Procedures

- Multiple comparison procedures can be used to answer the question which population means are different?
- These comparisons may be of interest to the investigator
  - before (a priori) the data were collected or
  - may arise in completely exploratory studies only after (a posteriori) the data have been examined.
- All post-hoc multiple comparison methods are essentially based on the t-test but include appropriate corrections for the fact that we are comparing more than one pair of means.



https://www.ibm.com/docs/en/spss-statistics/saas?topic=anova-one-way-post-hoc-tests

UNIVERSITY OF SASKATCHEWAN

One-way ANOVA

Multiple Comparisons

Dependent Variable: percent percent improvement

Scheffe

(I) treatmen	(J) treatmen	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1 treatment 1	2 treatment 2	-19.27333*	.86398	.000	-21.4911	-17.0556
	3 treatment 3	-14.14000*	.84874	.000	-16.3186	-11.9614
2 treatment 2	1 treatment 1	19.27333*	.86398	.000	17.0556	21.4911
	3 treatment 3	5.13333*	.80777	.000	3.0598	7.2068
3 treatment 3	1 treatment 1	14.14000*	.84874	.000	11.9614	16.3186
	2 treatment 2	-5.13333*	.80777	.000	-7.2068	-3.0598

*. The mean difference is significant at the .05 level.

percent percent improvement

Scheffe[a,b]

treatmen	N	Subset for alpha = .05		
		1	2	3
1 treatment 1	10	49.4600		
3 treatment 3	13		63.6000	
2 treatment 2	12			68.7333
Sig.		1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 11.527.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

UNIVERSITY OF SASKATCHEWAN

Chi-Squared Test (χ^2)

UNIVERSITY OF SASKATCHEWAN

Statistical Tests

Exposure measure (independent variable)	Outcome measure (dependent variable)	Example	Test
Categorical, two groups only	Continuous	Comparing blood pressure between men and women	T-test
Categorical, two or more groups	Continuous	Comparing blood pressure between four med school classes	Analysis of Variance (ANOVA)
Categorical	Categorical	Comparing likelihood of dementia (Y/N) by sex/gender	Chi-square test (χ_2)
Any (multiple)	Categorical	Comparing likelihood of dementia (Y/N) by sex/gender *and* age	Logistic regression
Any (multiple)	Continuous	Comparing blood pressure by sex/gender and age	Linear regression

χ^2

When dependent and independent variables are categorical:

- Q.1. How do you investigate if there is an association between two categorical variables?
- ~~Q.2. How do you measure the strength of this association?~~

To answer

- Q. 1, you compute chi-square test and to answer
- ~~Q.2., you estimate odds ratio~~

Let's look at the simplest situation:

- Dependent variable and independent variable each have two categories then how do you measure the association ~~and the strength of this association~~?

		Disease		
		Diseased	Not Diseased	
Exposure	Exposed	a	b	$a + b$
	Not exposed	c	d	$c + d$
		$a + c$	$b + d$	

Chi-squares test – to measure the association

Examples: **Exposure to**: smoking (personal and second hand); grain dust, pesticides

Disease: lung cancer, chronic cough; asthma etc.

UNIVERSITY OF SASKATCHEWAN

Crosstabs

		Disease		
		Diseased	Not Diseased	
Exposure	Exposed	35	65	100
	Not exposed	15	85	100
		50	150	200

Example: Modified from Dr. Tan's handouts

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
expos * illwell	200	100.0%	0	.0%	200	100.0%

70% of diseased were exposed to the risk factor

expos * illwell Crosstabulation

			illwell		
			.00 diseased	1.00 no diseased	Total
expos	.00 exposed to a risk factor	Count	35	65	100
		% within illwell	70.0%	43.3%	50.0%
	1.00 not exposed to a risk factor	Count	15	85	100
		% within illwell	30.0%	56.7%	50.0%
Total		Count	50	150	200
		% within illwell	100.0%	100.0%	100.0%

UNIVERSITY OF SASKATCHEWAN

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	10.667[b]	1	.001		
Continuity Correction[a]	9.627	1	.002		
Likelihood Ratio	10.903	1	.001		
Fisher's Exact Test				.002	.001
Linear-by-Linear Association	10.613	1	.001		
N of Valid Cases	200				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 25.00.

UNIVERSITY OF SASKATCHEWAN

Statistical Tests

Exposure measure (independent variable)	Outcome measure (dependent variable)	Example	Test
Categorical, two groups only	Continuous	Comparing blood pressure between men and women	T-test
Categorical, two or more groups	Continuous	Comparing blood pressure between four med school classes	Analysis of Variance (ANOVA)
Categorical	Categorical	Comparing likelihood of dementia (Y/N) by sex/gender	Chi-square test (χ_2)
Any (multiple)	Categorical	Comparing likelihood of dementia (Y/N) by sex/gender *and* age	Logistic regression
Any (multiple)	Continuous	Comparing blood pressure by sex/gender and age	Linear regression

Examples from the literature

- Cohort profile: the Saskatchewan Rural Health Study—adult component | BMC Research Notes (springer.com)
- Racial discrimination and depression among on-reserve First Nations people in rural Saskatchewan | SpringerLink
- Gender Differences in the Association of Individual and Contextual Exposures with Lung Function in a Rural Canadian Population | SpringerLink
- Household Income and Psychological Distress: Exploring Women's Paid and Unpaid Work as Mediators - PMC (nih.gov)