



College of Medicine

# **Services offered**

- Study design
- Sample size & Power calculations
- Data management
- Statistical analysis/interpretation
- Manuscript review
- Grant proposal review

#### No cost to COM researchers...but acknowledgement, please

More info: College of Medicine web page  $\rightarrow$  Research  $\rightarrow$  COM research  $\rightarrow$  Research community  $\rightarrow$  Research Centres/Units/Facilities





## Objectives...

- To provide some guidance for how to best work with a statistician
- To provide some suggestions for data entry that can make your research life happier
- To provide an overview of sample size calculation

Possible (often sub-optimal!) approaches:

1) Helper  $\rightarrow$  technician-type role, "just the p-value please!"

2)"Data-Blesser"  $\rightarrow$  curb-side advice; no "hands on" involvement

3) Archaeologist  $\rightarrow$  "my other statistician stopped returning my e-mails..."

### 4) Leader $\rightarrow$ lack of substantive expertise. Best approach $\rightarrow$ collaborative!

Kirk RE. Statistical consulting in a university: dealing with people and other challenges. American Statistician 1991 45(1):28-34. Leoutsakos, J. Working with statistician [Internet]. Baltimore (MD): Johns Hopkins University, School of Medicine; 2010.Available from http://www.hopkinsmedicine.org/psychiatry/bayview/research/WorkingwithaStatistician.pdf

### Four key visits:

- 1. Design stage
  - Come **early** with a specific question
  - Know how your question fits into the literature
  - Think about a manageable research approach
  - Be flexible, be clear
  - Sample size calculation is often an early step
    - May be undertaken based on published information in the literature or from a preliminary pilot

### 2. Data entry

Spreadsheet format

- Entry/Coding
- "Test drive" your entry

### 3. Analysis

- Not a "magic-wand, insert-data-here-out-comes-p-values-here" process → Both an art and a science
- Potential problems: unexpected data distributions, violation of assumptions → modified analysis plan
- May require multiple meetings during this stage

Good analysis may take weeks to months!
(Give yourself and your analyst as much time as possible!!!)

- 4. Publication/Presentation
  - "How do I say this?" → assistance with writing statistical section and results presentation
  - Reviewers' comments  $\rightarrow$  revision is the norm!

### So, what kind of SHAPE is your data in?

- **S** single spreadsheet
- H horizontal entry
- A aggregated categories
- P personal de-identification
- **E** error examination

### **SHAPE – Single Spreadsheet**

- Best to include all information for all subjects on one spreadsheet if possible/practical
- Excel is adequate
- Avoid "hiding" fields during entry or minimized row heights/column widths → better to use Excel's freeze panes or split window functions

### **SHAPE – Single Spreadsheet**

FILE	HOME	r INSERT	PAGE LAYO	DUT FORMULA	S DATA REVIEW	VIEW	DEVELOPER Nuance PDF		Booki - E	xcel							? 📧 🗕 🗗 🗙 Bryce, Rhonda 🕶 💭
Normal E7	Page Break Preview Workbook	Page Custo Layout View Views	$\begin{bmatrix} & & \\ & $	✓ Formula Bar es ✓ Headings Show	Zoom 100% Zoom 1 Selectic Zoom	to New Window	Arrange Freeze All Panes Unhide Unlock all Children Freeze For Keep the top	View Side by Side     Synchronous Scrolling     Reset Window Position     was and columns to scroll     entire worksheet.     Row     row visible while scrolling     rest of the worksheet.     Polumn	Switch Windows *	Macros Macros							
	A		В	С	D		Keep the firs	st column visible while rough the rest of the worksho	eet. G	н	I	J	К	L	М	Ν	0
1	ld	Ag	e	Sex	SurgDate		DCDate	LOS	Гх								
2		1	24	1	1/1	0/2015	1/15/20	15 5		1							
3		2	52	0	1/1	2/2015	1/14/20	15 2		0							
4		3	34	0	1/1	3/2015	1/16/20	15 3		0							
5		4	32	1	1/1	3/2015	1/14/20	15 1		1							
6		5	57	1	1/1	5/2015	1/17/20	15 2		1							
7		6															
8		7															
9		8															
10		9															
11		10															
12																	
13																	
14																	
15																	
16																	
17																	
18																	
19																	
20																	
21																	
-	•	Sheet1	+								: 4						
READY	<b>a</b>														E	≣ ■ ■ -	

### **SHAPE – Single Spreadsheet**

🗶 i 🔓	17-1	(≃ - 🙎   =								E	Book1 -	Microsoft	Excel non-co	ommercia	l use							- 0 -	x
File	Ho	ome Inse	rt Pag	ge Layou	t Formula	as D	Data Re	view	View	Develop	er											ه 🕜 🗆 ا	di XX
Norma	Page Layout	Page Break Preview Vorkbook Vie	Custom Views	Full Screen	✓ Ruler ✓ Gridline	▼ Fo	ormula Bar leadings	Zoom	100% Zo Zoom	pom to lection	New Windo	Arrange All	Freeze Panes +	Split Hide Unhide	View Sid	e by Side nous Scrolling indow Position	Save Workspace	Switch e Windows •	Macros Macros				
	A1		. (=	fx							1												~
	Δ	B	C		D F	:	F	G	н		1	1	V	W	X	v	7	۵۵	ΔB	AC	۸D	ΔF	
1	0		U U					0				-			~		~	~~~	~			~~	
2		•																					
3																							
4																							_
5																							
6																							
7																							
8																							
9																							
10																							
11																							
12																							
13																							Ŧ
96																							
97																							
98																							
99																							
100																							
101																							
102																							
103																							
104																							
105																							_
106																							=
107							_					_											•
14 4 3	N S	neet1 / She	eet2 / S	heet3 🤇	2							►	4								1007		
Ready		_	_	_		-	_		ra				-	-	_	_	-	_	_		100% (-)-	V	÷
-					e	<b>(#</b>	1	4			P	X							é •	ې 🚯 🕪	J 🗊 🦽 🛛	10:11 Al	M 112

- Variable names on first line across columns
  - Give short, clear variable names (e.g. NumHosp)
  - No \*&%\$@!^ symbols or spaces! (But "\_" is okay)
  - Should start with a letter
  - Color-code groups of headings if helpful but NOT to convey information
- One variable per column (e.g. not sex/age)
- One line per subject *per outcome measure* If same person is measured repeatedly for the outcome of interest, their results will likely show some similarity that is due to their own unique self.





#### 30 subjects as random sample

#### 6 subjects with repeated measures (long format)

ID	Sex	Age	Doses	NumPVC
1	1	55	1	1
2	1	76	3	9
3	0	24	2	3
4	0	67	2	7
5	0	76	4	13
6	1	43	3	5
7	0	52	5	9
8	0	61	4	7
9	1	34	5	11
10	1	54	4	17
11	1	77	1	3
12	1	64	5	13
13	0	43	3	13
14	0	66	1	11
15	1	36	5	15
16	0	75	2	5
17	0	53	4	15
18	0	23	2	9
19	1	56	5	17
20	1	77	3	7
21	0	32	1	5
22	0	41	1	9
23	1	51	5	19
24	0	43	4	9
25	1	65	2	11
26	0	88	4	11
27	1	48	1	7
28	1	67	3	11
29	1	69	2	13
30	1	50	3	15

	ID	Sex	Age	🕽 Dose	NumPVC
	1	1	D	1	1
	1	1	43	2	3
	1	1	43	3	5
		1	43	4	7
Enter time	1	1	43	5	9
or data if	2	0	56	1	3
of uate fi	2	0	56	2	5
possible as	2	0	56	3	7
timo noint	2	0	56	4	9
time point	2	0	56	5	11
may be	3	1	52	1	5
	3	1	52	2	7
uneveniy	3	1	52	3	9
spaced!	3	1	52	4	11
•	3	1	52	5	13
	 4	0	49	1	7
	4	0	49	2	9
	 4	0	49	3	11
	4	0	49	4	13
	4	0	49	5	15
	 5	0	60	1	9
	 5	0	60	2	11
	 5	0	60	3	13
	 5	0	60	4	15
	 5	0	60	5	1/
	 6	1	48	1	11
	 о С	1	48	2	13
	 0	1	48 40	3	15
	 0	1	4ð 40	4 F	1/
	σ	T	4ð	Э	19

- The point: recognize that utilizing the correlation of repeated outcome measurements within an individual matters!
- Similar principle applies to other clusters (e.g. families, physician practices, etc.) as well as repeated subjects.
- Matched analyses uses this concept as well, including pre-post assessments. Be sure to tell your analyst if data is matched!

### SHAPE – Aggregated categories

- Continuous predictors are generally stronger.
- Enter continuously but think categorically!
  - Clinical relevance
  - Statistical necessity
- Numbers please!
  - e.g. Variable: hypertension  $\rightarrow$  1=yes, 0=no
  - If categorical, commit!
  - Include a data dictionary/code sheet
    - E.g. BMDdiffFN = change in bone density between year one and year 2 at femoral neck, 0 = no change, 1 = 5% decrease and 2 = 5% increase

### SHAPE – Aggregated categories

- Missing data???
  - Ideally no empty cells → Is this zero? Missing? Oversight?
    - -Zero is an important number
    - -Truly missing  $\rightarrow$  Unavailable vs. Impossible
      - »Need to differentiate those who are truly missing data in order to evaluate possible bias
      - » Utilize one or more non-range numbers to indicate (e.g. missing, unavailable = 999; not applicable =leave blank).
    - Minimize simple oversights → decreased power

### SHAPE – Personally de-identified

- Remove ALL personally identifying data!!!!!
  - Names, PHN's, chart numbers, phone numbers
  - Check all tabs of the worksheet
  - Better yet, never enter them
    - Each subject has a unique identification number with any corresponding personal info stored elsewhere

### SHAPE – Error examination

- Known as "data cleaning"
- Dealing with impossible values, text, data rearrangement
- Time consuming, often uncertain, for analyst
- Best approach: get it right the first time!
  - A few spreadsheet tips.....

ID	Sex	Age	Doses	NumPVC
1	1	55	1	1
2	1	76	3	9
3	0	24	2	3
4	0	67	2	7
5	0	76	4	13
6	1	43	3	5
7	0	52	5	9
8	0	61	4	7
9	1	34	5	11
10	1	54	4	17
11	1	77	1	3
12	1	64	5	13
13	0		3	13
14	0	66	1	11
15	1	36	5	15
16	0	75	2	5
17	0	53	4	15
18	0	23	2	9
19	1	56	5	17
20	1	77	3	7
21	0	32	1	5
22	0	41	1	9
23	1	511	5	19
24	0	43	4	9
25	1	65	2	11
26	0	88	4	11
27	1	48	1	7
28	1	67	3	11
29	1	9	2	13
30	1	50	3	15

File	Hor	me Inse	rt Page	Layout Fo	ormulas Data	Review	View	Developer						
Ê	Cut	v -	Calibri	* 11	~ А́ А́	· = 😑 🗧	<u>۶</u> -	Wrap Text		Gene	ral	*	Σ Aut	oSum + A
Paste	Form	nat Painter	BIU	· ·	🌺 • 🛕 • 🔳			💽 Merge & Ce	nter 👻	\$ -	%,	00. 00. 0.€ 00.	Conditional Format Cell Insert Delete Format	So ar≭ Fil
	Clipboard	G.		Font	G.		Alignment	t	Ga.		Number	G.		Editing
	118	-	. (-	$f_{x}$									Highlight Cells Rules > Greater Than	
	А	В	С	D	E	F	G	Н	1	1	J	K		Q
1	ID	Sex	Age	Doses	NumPVC								10 Less Than	
2	1	1	55	1	1									
3	2	1	76	3	9								Data Bars → Between	
4	3	0	24	2	3									
5	4	0	67	2	7								Color Scales > Equal To	
6	5	0	76	4	13									
7	6	1	43	3	5								Text that Contains	
8	7	0	52	5	9								ab -	
9	8	0	61	4	7								New Rule	
10	9	1	34	5	11								Clear Rules	
11	10	1	54	4	17								Manage Bules	
12	11	1	77	1	3								<u>Duplicate Values</u>	
13	12	1	64	5	13								More Pules	
14	13	0		3	13								<u>m</u> ore Rules	
15	14	0	66	1	11									
16	15	1	36	5	15									
17	16	0	75	2	5									
18	17	0	53	4	15									
19	18	0	23	2	9									
20	19	1	56	5	17									
21	20	1	77	3	7									
22	21	0	32	1	5									
23	22	0	41	1	9									
24	23	1	511	5	19									
25	24	0	43	4	9									
14 4	▶ H She	eet1 She	et2 She	et3 / 🔁 /										
Read	y   🛅	July one	June / Diffe											
		. 😁	-			1						-		

Fi	le Ho	ome Inse	rt Page	Layout Fo	ormulas Data	Review	View [	Developer						
Pas	Cut	t py *	D Z U	▼ 11	$\cdot A^{*} A^{*} \equiv $	= =   >>		/rap Text	Gen	eral	▼ ≪.000	 ≦≶	Eormat C	
- 43	🗧 🗳 For	mat Painter	BI	2 *   🖽 *				lerge & Center	r* \$	· % ,	.00 ⇒.0	Formatting ~ a	s Table ~ Sty	les -
	Clipboar	d G		Font	6	AI	ignment	_	G	Number	Gi.	St	yles	
	11	Greater Than					? X							
	A	Format cell	s that are G	REATER THA	N:			н	1.1	J	К	L	м	N
1	ID													
2	1	90			📧 with Red Borde	er	-							
3	2				Light Red	Fill with Dark R	led Text							
4	3				Green Fill	with Dark Gree	in Text							
5	4 4	-		-	Light Red	Fill								
6	5	0	76	4	13Red Borde	5								
7	6	1	43	3	5 Custom Fe	ormat								
8	7	0	52	5	9									
9	8	0	61	4	7									
10	9	1	34	5	11									
11	10	1	54	4	17									
12	11	1	77	1	3									
13	12	1	64	5	13									
14	13	0		3	13									
15	14	0	66	1	11									
16	15	1	36	5	15									
17	16	0	75	2	5									
18	17	0	53	4	15					]				
19	18	0	23	2	9									
20	19	1	56	5	17									
21	20	1	77	3	7									
22	21	0	32	1	5									
23	22	0	41	1	9									
24	23	1	511	5	19									
25	24	0	43	4	9									
14 4	► N S	neet1 She	et2 / She	et3 ⁄ 🖓 /										

File	Hor	me Inse	rt Pag	ge Layout Fo	ormulas Data	Review	View	Developer															2	<b>?</b> -
Ê	🔏 Cut	y •	Calibri	* 11	· A A =	= =	≫	Wrap Text		General		<b>v</b>	<u></u>			+			Σ AutoSu III ▼	m *	AZ7	A		
Paste	IForn	nat Painter	B I	<u>u</u> - 🖽 -	<u>∞ - A</u> - ≡			Merge & Cer	ter 🔻	\$ - %	•	.00 *-0	Formatting *	as Table *	Styles *	Insert	Delete F	ormat	🖉 Clear 🔻	F	Sort 8 Filter 1	Select *		
	Clipboard	l G		Font	Gi.		Alignmen		- Fa	Nur	ber	F2	S	styles			Cells			Edit	<b>2</b> ↓	Sort A to	z	
	A1		• (=	f <sub>x</sub> ID																	Z↓	S <u>o</u> rt Z to	A	
1	А	В	С	D	E	F	G	н			J	Κ	L	M		N	0	Р	C	L	¥†	C <u>u</u> stom S	ort	
1	ID	Sex	Age	Doses	NumPVC																7=	Eilter		<u> </u>
2	1	1	55	1	1																W.	Clear		-
3	2	1	76	3	9																7.	Reannly		
4	3	0	24	2	3																-6°4			_
5	4	0	67	2	7																			
6	5	0	76	4	13																			
7	6	1	43	3	5																			
8	7	0	52	5	9																			
9	8	0	61	4	7																			
10	9	1	34	5	11																			
11	10	1	54	4	17																			
12	11	1	77	1	3																			
13	12	1	64	5	13																			
14	13	0		3	13																			
15	14	0	66	1	11																			
16	15	1	36	5	15																			
17	16	0	75	2	5																			
18	17	0	53	4	15																			
19	18	0	23	2	9																			
20	19	1	56	5	17																			
21	20	2	77	3	7																			
22	21	0	32	1	5																			
23	22	0	41	1	9																			
24	23	1	511	5	19																			
25	24	0	43	4	9																			
14 4 1	N She	eet1 She	et2 / S	heet3 / 🞾 /						1						1								
Ready																		(	ount: 5			100% 🤆	)	$-\nabla$
8		3		0 🬔	😍 🤘	3				P			10.10			1			í • ∢	•	6	â atl	<b>P*</b>	10:20 20/11/2

F	ile Ho	me Inse	ert Pag	e Layout 🛛 F	ormulas Data	Review	View	D	eveloper								
ľ	🍋 🔏 Cut	y •	Calibri	* 11	· A A ■	= =	≫	∎ w	/rap Text		Gene	eral	٣	_ ≦\$			+
Pas	te 🛷 Forr	mat Painter	BI	<u>u</u> - 🔛 -	<u></u>	<b>= =</b>	•	•a• M	lerge & Cent	ter 👻	\$ -	%,	00. 0. <b>→</b> 0.◆ 00.	Conditional Formatting *	Format as Table - Si	Cell tyles *	In
	Clipboard	d G		Font	E.		Alignme	ent		E.		Number	Gi.	S	tyles		
	A1		• (=	∫x ID													
1	А	В	С	D	E	F	(	6	Н		I.	J	К	L	М	N	1
1	ID 🔻	Sex 🔻	Age	Doses -	NumPVC -	1											
<b>A</b> ↓	Sort Smalle	est to Larges	st		1												
Z↓	Sort Large	st to Smalles	st		9												
	Sor <u>t</u> by Co	lor			3												
w.	Clear Filter	r From "Sev"			7												
21	Eiltar by C				13												
	Number Ei	Harr			5												
	Number <u>r</u> i	itters		·	9												
	Search			2	7												
	🗹 (Sel	lect All)		_	11												
					17												
					3												
					13		_										
					13												
					11												
					5		-										
					15		_										
					9		_										
		OK		Cancel	17												
21	20	2		3	7												
22	21	0	32	1	5												
23	22	0	41	1	9												
24	23	1	511	5	19												
25	24	0	43	4	9												
14 4	▶ ₩ Sh	eet1 Sh	eet2 7 Sh	eet3 / 🔁 /													
Rea	dy 🞦																

F	ile Ho	ome Inse	ert Page	Layout F	ormulas Data	Review	View	, 6	Developer								
ľ	Cut		Calibri	* 11	· A A =	= =	≫	<b>≣</b> ∙ w	Vrap Text		Gen	eral	•			ð	+
Pas	ste 🛷 For	mat Painter	BI	· 🗄 - T	<u></u>	<b>a a b</b>		•a• N	lerge & Cent	ter 👻	\$	- % ,	00. 0. <b></b> 0.♦ 00.	Conditional Formatting * a	Format as Table - Si	Cell tyles ∓	In
	Clipboard	d G		Font	E.		Alignme	ent		Es.		Number	G.	S	tyles		
	A1		• (=	<i>f</i> <sub>∞</sub> ID													
1	A	В	С	D	E	F	(	6	Н		1	J	К	L	м		N
1	ID 💌	Sex 🔻	Age 🔻	Doses 🔻	NumPVC 🔻	1											
₽↓	Sort Small	est to Larges	t		1												
Z↓	S <u>o</u> rt Large	st to Smalles	t		9												
	Sor <u>t</u> by Co	lor			3												
w.	Clear Filter	r From "Sex"			7												
	Eilter by C	olor			13												
	Number Ei	ilter			5												
	Number Er	litters			9												
	Search			م	7		_									_	
	(Se	lect All)		-	11		_										
	0				17		_									_	
	2				3		_										
					13												
					13		_										
					11												
					5												
					15												
					9												
		OK		Tancel	17												
21	20	2	11	.:	7												
22	21	0	32	1	5												
23	22	0	41	1	9												
24	23	1	511	5	19												
25	24	0	43	4	9												
14 4	► ► Sh	eet1 Sh	eet2 / She	et3 / 🔁 /	· · ·												
Rea	dy 🛅																







F	ile	Ho	me	Insert I	Page Layout	For	mulas	Data	Rev	view	View	Dev	eloper								
¥	A		*					Connect Properti	ions es	<b>2</b> ↓ [		Y	🛠 Clear 🔓 Reapply		-		8		<b>E</b> ?		*
Fre	om iess	From Web	From Text	From Other Sources *	Connecti	g Re ions /	efresh All <del>v</del> ◎	Edit Link	s	X↓	Sort	Filter	Madvanced	Text to Column	Remove s Duplicat	e D es Valid	ata Co lation *	onsolidate	What-If Analysis *	Group	Ung
			Get Ex	ternal Data			Con	nections			S	ort & Filt	er		e opnioi		Data Val	idation			
		16		- (=	f.,												Circle In	alid Data			
		10	-	•	,x	-		-									Circle In	alid Data		-	_
1	A		B	0	Dagag	E Num	DVC	F	G	5	н		J	ĸ	L	- 🐉	Clea <u>r</u> Va	lidation Cir	cles	Q	_
2	1	,	1	Age	1	NUM	PVC														
2	2		1	76	2	1												_			_
3	2		0	24	2	9															-
4 5	3		0	67	2	7				-											
6			0	76	4	19	1					<u> </u>									
7	6		1	/0	3	5	,					<u> </u>	-								-
8	7	,	0	52	5	9											-				_
9	8		0	61	4	7															
10	9		1	34	5	11															
11	10	n	1	54	4	17	,														
12	1	1	1	77	1	3				_											
13	1	2	1	64	5	13	1														
14	1	3	0		3	13															
15	14	4	0	66	1	11															
16	1	5	1	36	5	15	5														
17	10	5	0	75	2	5															
18	1	7	0	53	4	15	;														_
19	18	8	0	23	2	9															
20	19	9	1	56	5	17	,														
21	20	D	2	77	3	7															
22	2	1	0	32	1	5															
23	2	2	0	41	1	9															
24	2	3	1	511	5	19	)														
25	24	4	0	43	4	9															
26	2	5	1	65	2	11	L														
27	20	5	0	88	4	11	L														
28	2	7	1	48	1	7															
29	20	8	1	67	3	11															
30	29	9	1	9	2	13	3														
14	())	Sh	eet1	Sheet2	Sheet3	2/											[	4			
Rea	ady	<b>*</b>																			

🚳 😓 🚞 O 💪 😃 🗠 💽 🔀 💽

F	ile	Home	Insert	Page Layou	eloper												
彩 Fr(	A Com F	rom From	From Oth	er Existi	ing Refresh	Conned Properti	tions es	$\begin{array}{c} A \downarrow \\ Z \downarrow \\ Z \downarrow \\ A \downarrow \\ A \downarrow \\ Sort \end{array}$	Filter	🛠 Clear 🐌 Reapply	Text	to Remo	ove Da	ata Co	nsolidate	What-If	Grou
Acc	ess V	Veb Text	Sources	Connec	ctions All - 6	e Edit Lini	cs			• Advance	Colur	nns Duplic	ates Valida	ation ~		Analysis -	Ÿ
		Get B	External Data	3	Cor	nections		S	ort & Filte	er			Dat	a Tools			
		D1	<b>-</b> (0	$f_{x}$	Doses												
4	Α	В	С	D	E	F	0	G H	1	J	K	L	M	N	0	Р	0
1	ID	Sex	Age	Doses	NumPVC												
2	1	1	55	1	1												
3	2	1	76	3	9		0						2	X			
4	3	0	24	2	3			Data Validation	ו				<u> </u>				
5	4	0	67	2	7			Contract		-							
6	5	0	76	4	13			Setungs In	nput Messa	age Error	Alert						
7	6	1	43	3	5		-	Validation crit	eria								
8	7	0	52	5	9		- 11	Allow:									
9	8	0	61	4	7			Whole num	her	-	/ Ignore b	lank					
10	9	1	34	5	11		-	Whole Hum									
11	10	1	54	4	17		-	Data:									
12	11	1	77	1	3		-	between		•							
13	12	1	64	5	13		-	Minimum:									
14	13	0		3	13		-	1				<b>1</b>					
15	14	0	66	1	11		-	Maximum:									
16	15	1	36	5	15		-	2				FEE					
17	16	0	75	2	5			3									
18	17	0	53	4	15			Apply the	se change	s to all other	cells with	the same se	ttings				
19	18	0	23	2	9												
20	19	1	56	5	1/			Clear All				OK	Car	ncel			
21	20	2	77	3	7												
22	21	0	32	1	5			_	_		_	_	_				
23	22	0	41	1	9												
24	23	1	511	5	19												
25	24	0	43	4	9												
26	25	1	65	2	11		-										
27	26	0	88	4	11												
28	2/	1	48	1	/												
29	28	1	67	3	11												
30	29	Chart	Chart2	2 Chart2	13										4		
End		Sneet1	_ Sneet2	Sneet3										L		Auge	41170
EDI	e 1															Average: 5	411/6

A 144

-

~

X		2 - 6	- 8	₹						Scrap pa	per for acad h	half pres	<ul> <li>Microsof</li> </ul>	t Excel i	non-comm	ercial use		
F	ile	Hor	me	Insert	Page Layou	it Formulas	Data	Revie	ew Viev	/ Deve	loper							
<u></u>	A						Connec	tions ies	$\begin{array}{c} A \downarrow \\ Z \downarrow \\ Z \downarrow \\ Z \downarrow \\ A \end{array}$	Y	🕻 Clear	• •			3			
Acc	ess \	Web	Text	Sources *	connec	tions All •	📾 Edit Lin	ks	Å↓ Soπ	Filter	🦅 Advanced	Colum	o Remov ns Duplica	e tes Vali	dation •	onsolidate	Analysis *	Grou
			Get Ex	ternal Data		C	onnections			Sort & Filte	r				Data Val	idation		
		<b>D1</b>		- (-	£	Deses									Data <u>v</u> al	idation		
		DI	-	• (		Doses	-					<b>v</b> 1			C <u>i</u> rcle Inv	/alid Data		
1	A		B	C	D	E	F	G	н		J	K	L	- 3	Clea <u>r</u> Va	idation Cir	cles	(
2	1	,	1	Age	1	1												
2	2		1	76	3													
4	2		0	24	2	3												
5	4		0	67	2	7												
6	5		0	76	4	13												
7	6		1	43	3	5												
8	7		0	52	5	9												
9	8		0	61 (	4	7												
10	9		1	34 🤇	5	11												
11	10	)	1	54 🤇	4	17												
12	11		1	77	1	3												
13	12		1	64 🤇	5	13												
14	13	1	0		3	13												
15	14	ļ.	0	66	1	11												
16	15	5	1	36 🤇	5	15												
17	16	5	0	75	2	5												
18	17		0	53 🤇	4	15												
19	18	3	0	23	2	9												
20	19	)	1	56 🤇	5	17												
21	20	)	2	77	3	7												
22	21		0	32	1	5										_		
23	22	2	0	41	1	9										_		
24	23	;	1	511	5	19												
25	24	•	0	43	4	9												
26	25	;	1	65	2	11					_							
27	26	;	0	88	4	11												
28	27		1	48	1	7										_		
29	28		1	67	3	11												
30	29		1	9	2	13										1.		
14 4		She	eet1 🔍	Sheet2	Sheet3										1			
Pes	CIN .																Average.	< 41176

X	<b>-</b> 7 •	(= - 8	.						Scrap pap	per for acad h	half pres	<ul> <li>Microsof</li> </ul>	t Excel no	n-comme	ercial use		
Fil	e Ho	ome	Insert	Page Layou	ut Formulas	Data	Review	View	Devel	oper							
						Connec	tions A Z	A Z A	T,	K Clear				1. O			₽ E
Fron	n From	From	From Oth	er Existi	ng Refresh	en Edit Lin	Z	Sort	Filter	Ca Advanced	Text t	o Remov	e Da	ta Cor	nsolidate	What-If	Grou
Acce	ss Web	Text	Sources	Connec	tions All -		10			Advanced	Colum	ns Duplicat	tes Valida	tion ~		Analysis *	Ť
		Get Ex	cternal Data		Co	nnections			Sort & Filte	r			Data	Tools			
			<b>-</b> (0	$f_x$	5												
- 24	Α	В	С	D	E	F	G	Н	1	J	K	L	М	N	0	Р	(
20	19	1	56	5	17												
21	20	2	77	3	7												
22	21	0	32	1	5												
23	22	0	41	1	9												
24	23	1	511	5	19												
25	24	0	43	4	9												
26	25	1	65	2	11												
27	26	0	88	4	11												
28	27	1	48	1	7										- 22		
29	28	1	67	3	11				Microsoft	Excel							_
30	29	1	9	2	13												
31	30	1	50	3	15					The value	you enter	ed is not val	id.				
32				5						A user has	restricted	d values that	t can be er	tered into	this cell.		
33																	
34										<u>R</u> etr	y	Cancel	Hel	p			
35										_			_		_		
36									_			_			_		
37																	
38																	
39																	_
40																	_
41																	
42																	_
43																	
44																	
45																	
46																	
47																	
48			_														
49																	
14 4	► H S	neet1	Sheet2	Sheet3	<u></u>									1	4		
Dand																	

### SHAPE – Error avoidance/assessment

- Known as "data cleaning" Dealing with impossible values, text, data rearrangement Time consuming, often uncertain, for analyst
- Best approach: get it right the first time! A few spreadsheet tips.....
  - Conditional formatting
  - Filters
  - Data validation criteria
  - Use of spreadsheet calculation functions
    - » e.g. days between dates/years

### So, remember to get your data into SHAPE!

- **S** single spreadsheet
- H horizontal entry
- A aggregated categories
- P personal de-identification
- **E** error examination

Why do we need it? ≻To address a particular objective or hypothesis Objective: To reduce cholesterol level by an intervention

- Null hypothesis (H<sub>0</sub>): Mean cholesterol (Control) = Mean cholesterol (Intervention)
   μ (Control) - μ (Intervention) = 0
- Alternative hypothesis (H<sub>1</sub>):  $\mu$  (Control)  $\mu$  (Intervention) = 10 (=  $\delta$ )

- If the sample size is large, small value of δ can be significantly different
- If the sample size is small, large value of δ may not be significantly different
- Choose  $\delta$  based on
- What value of δ is practically important or clinically meaningful
- $\succ$  Calculate sample size based on that  $\delta$

Sample size for continuous outcome (power based)

 $n = (Z_{1-\alpha/2} + Z_{1-\beta})^2 * 2^* \sigma^2 / \delta^2$ 

 $\succ \sigma^2$  is the population variance  $\succ \delta$  is the difference we would like to detect

 $\geq \alpha$  = Probability (Type I error)

= level of significance (0.05)

- $\geq \beta$  = Probability (Type II error)
- > Power =  $1 \beta$  (0.80 or 0.90) = probability of detecting a significant difference when it exists

	Truth about the population						
Decision from the sample	H <sub>0</sub> is true	H <sub>0</sub> is false					
Fail to reject H <sub>0</sub>	Correct decision	Type II error					
Reject H <sub>0</sub>	Type I error	Correct decision					

Sample size for continuous outcome (power based)

n = 
$$(Z_{1-\alpha/2}+Z_{1-\beta})^2 * 2*\sigma^2 / \delta^2$$

- Z<sub>1-α/2</sub> is the critical value of the standard normal distribution at 1-α/2 (for α=0.05, Z<sub>1-α/2</sub> =1.96)
- $> Z_{1-\beta}$  is the critical value of the standard normal distribution at 1-β (for a power of 80%, β is 0.2 and the critical value is 0.84)

### Sample size calculation for binary outcome

Proportion of side effect of the old treatment is 0.10. Our objective is to reduce it to 0.05 by a new treatment

The required sample size for each group  $n = (Z_{1-\alpha/2}+Z_{1-\beta})^2 * (p_0(1-p_0)+p_N(1-p_N)) / (p_0-p_N)^2$ 

# Thank you!

# Questions???