

Statistics Ain't Nothing But Numbers: Using Excel to Understand Data in New and Exciting Ways

Michael T. Braun, PhD

Yu-Ling Chiu, PhD

Children and Family Research Center

School of Social Work

University of Illinois at Urbana-Champaign

Statistics...

- Is a mindset!
- Uses basic math (addition, subtraction, multiplication, division)
- Is the combination of thinking about data and using software tools



Sample Parent Survey

1. Casework ID:

2. Client: ID:

3. What is your gender: Male Female

4. What is your current age? _____

5. How good is your caseworker at:

	Poor	Fair	Good	Very good	Excellent
Making you feel at ease?	1	2	3	4	5
Fully understanding your worries?	1	2	3	4	5
Really listening?	1	2	3	4	5

6. Did your caseworker refer you to other services you needed: Y N

Imagine for a moment...

- After gathering data from some parents, what kinds of details would you report about this survey?



Imagine for a moment...

- You also ask parents if they were offered any services, and you want to know if women were offered services more frequently than men
- What would you compare?



Imagine for a moment...

- Next, you wonder if being offered services is associated with higher ratings of workers.
- How might you look to see if this was true?

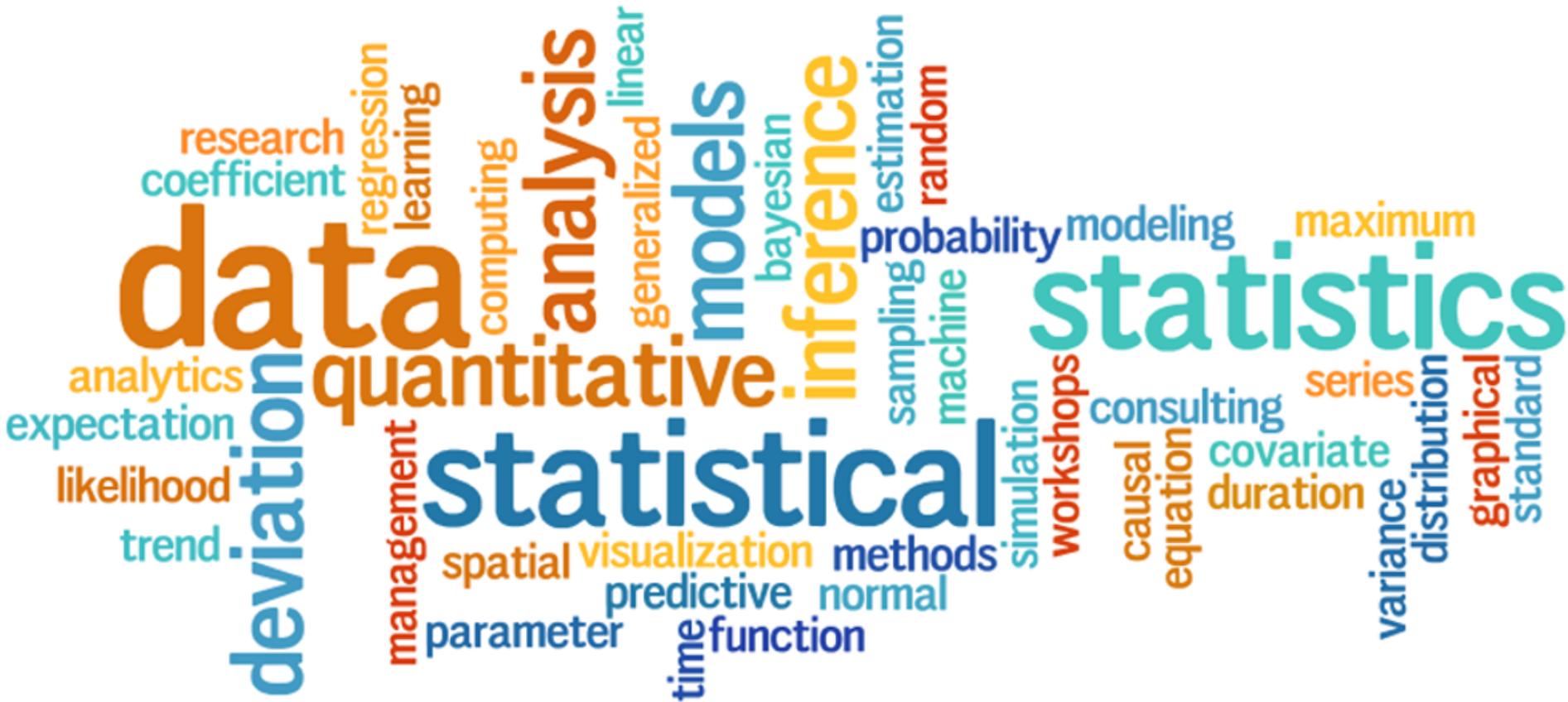


Imagine for a moment...

- Finally, you are curious to know if parent age is associated with ratings of workers.
- How might you look at this?



Thinking about these things...



...is the conceptual work of statistics.



Statistics...

- Provide us with ways to describe data
- Help us tell if there are differences between groups of things, even when we can't assess every single thing
- Allow us to explain why differences occur and predict what will happen in the future using what has happened in the past



Some Key Terms

- We gather a "sample" (the people who answered are survey) to try to answer questions about the "population" (everyone in the group we care about, whether they answered the survey or not)
- That's easier when we have a larger sample as we have more "power" to infer from the sample to the population



Some Key Terms

- But because cannot measure everyone (most of the time), we use a statistical "test" to figure out how much we can trust that our sample will be like the population
- And we can use those same tests to figure out if two groups (like men and women) are the same or are "significantly different," which generally means the odds that they are the same is less than 5%

Types of Data

- There are two main types of data to know...
 1. Categorical: Markers of groups/categories
 2. Continuous: Anything on a range



Continuous or Categorical?

- Date/Time
- Number of Children
- Sex
- Adopted/Reunified
- Age
- Height
- Weight
- Race
- Region
- Distance from home of origin
- Yes/No
- Amount of Agreement—
Strongly Disagree to
Strongly Agree



Sample Parent Survey

1. Casework ID:

2. Client: ID:

3. What is your gender: Male Female

4. What is your current age? _____

5. How good is your caseworker at:

	Poor	Fair	Good	Very good	Excellent
Making you feel at ease?	1	2	3	4	5
Fully understanding your worries?	1	2	3	4	5
Really listening?	1	2	3	4	5

6. Did your caseworker refer you to other services you needed: Y N

Microsoft Excel

- A powerful and useful tool
- Google Sheets also works, if you don't have access to Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	69	79	8	58	37	67	48	20	62	50	99	50	53	8	23	32	68	23	6	44	5	37	51
2	70	25	22	81	84	14	57	74	15	67	92	5	65	44	45	88	29	95	6	91	36	86	15
3	94	40	78	32	9	8	59	65	72	8	98	7	50	34	78	92	55	8	68	25	45	43	99
4	14	66	95	20	95	71	69	8	40	95	53	59	51	96	43	29	77	92	67	88	67	44	68
5	35	86	9	44	42	60	17	85	60	94	65	21	39	78	96	82	90	16	77	7	48	97	27
6	89	73	28	71	32	1	21	35	21	54	40	16	9	33	69	59	17	31	70	42	45	56	16
7	74	6	39	13	62	58	30	99	13	50	55	42	90	60	81	21	98	92	30	11	77	76	27
8	66	87	84	73	25	95	84	14	65	80	34	7	48	99	97	73	70	36	14	20	80	41	64
9	14	30	41	41	30	77	72	34	52	28	38	41	24	36	76	69	23	44	82	79	54	20	87
10	40	71	79	51	60	51	19	57	27	45	82	68	73	22	48	54	94	19	24	61	46	69	13
11	71	28	29	58	7	62	84	1	31	73	94	42	23	26	10	51	34	50	54	95	6	81	88
12	71	14	40	24	35	19	10	13	10	45	29	53	67	67	78	14	91	85	49	68	65	41	63
13	28	79	23	16	45	73	8	5	4	46	15	44	1	73	8	25	50	51	22	92	89	89	62
14	20	72	68	55	53	26	85	54	71	10	32	86	30	70	59	2	50	2	93	38	10	57	38
15	53	15	28	93	18	78	80	11	57	5	80	96	50	50	67	62	65	30	6	52	8	17	24
16	77	17	46	9	46	11	56	28	89	77	16	90	4	24	36	79	13	91	46	59	88	40	92
17	97	77	52	81	50	92	44	78	13	97	72	26	6	75	71	21	90	53	19	40	33	52	90
18	10	66	82	11	68	40	11	49	83	100	50	82	52	78	67	55	16	47	47	57	26	62	67
19	31	70	2	63	80	41	83	47	44	12	98	17	99	27	21	42	52	44	86	74	10	59	77
20	79	93	68	15	45	97	72	90	91	56	68	74	49	87	27	51	44	40	18	86	23	17	89
21	27	81	63	60	32	54	93	52	20	39	96	43	40	100	59	51	26	94	92	17	27	4	28
22	41	80	72	85	75	73	44	45	42	28	94	74	8	80	81	79	15	95	73	51	79	56	70
23	9	83	46	87	86	83	99	90	5	17	15	8	58	29	2	58	85	7	27	44	6	34	42
24	18	73	17	3	56	7	14	27	29	68	12	48	24	89	62	54	39	34	10	71	1	13	61

Microsoft Excel

- First... what is a function in Excel?
- A function lets you take some data in your spreadsheet and summarize it in some way

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	69	79	8	58	37	67	48	20	62	50	99	50	53	8	23	32	68	23	6	44	5	37	51
2	70	25	22	81	84	14	57	74	15	67	92	5	65	44	45	88	29	95	6	91	36	86	15
3	94	40	78	32	9	8	59	65	72	8	98	7	50	34	78	92	55	8	68	25	45	43	99
4	14	66	95	20	95	71	69	8	40	95	53	59	51	96	43	29	77	92	67	88	67	44	68
5	35	86	9	44	42	60	17	85	60	94	65	21	39	78	96	82	90	16	77	7	48	97	27
6	89	73	28	71	32	1	21	35	21	54	40	16	9	33	69	59	17	31	70	42	45	56	16
7	74	6	39	13	62	58	30	99	13	50	55	42	90	60	81	21	98	92	30	11	77	76	27
8	66	87	84	73	25	95	84	14	65	80	34	7	48	99	97	73	70	36	14	20	80	41	64
9	14	30	41	41	30	77	72	34	52	28	38	41	24	36	76	69	23	44	82	79	54	20	87
10	40	71	79	51	60	51	19	57	27	45	82	68	73	22	48	54	94	19	24	61	46	69	13
11	71	28	29	58	7	62	84	1	31	73	94	42	23	26	10	51	34	50	54	95	6	81	88
12	71	14	40	24	35	19	10	13	10	45	29	53	67	67	78	14	91	85	49	68	65	41	63
13	28	79	23	16	45	73	8	5	4	46	15	44	1	73	8	25	50	51	22	92	89	89	62
14	20	72	68	55	53	26	85	54	71	10	32	86	30	70	59	2	50	2	93	38	10	57	38
15	53	15	28	93	18	78	80	11	57	5	80	96	50	50	67	62	65	30	6	52	8	17	24
16	77	17	46	9	46	11	56	28	89	77	16	90	4	24	36	79	13	91	46	59	88	40	92
17	97	77	52	81	50	92	44	78	13	97	72	26	6	75	71	21	90	53	19	40	33	52	90
18	10	66	82	11	68	40	11	49	83	100	50	82	52	78	67	55	16	47	47	57	26	62	67
19	31	70	2	63	80	41	83	47	44	12	98	17	99	27	21	42	52	44	86	74	10	59	77
20	79	93	68	15	45	97	72	90	91	56	68	74	49	87	27	51	44	40	18	86	23	17	89
21	27	81	63	60	32	54	93	52	20	39	96	43	40	100	59	51	26	94	92	17	27	4	28
22	41	80	72	85	75	73	44	45	42	28	94	74	8	80	81	79	15	95	73	51	79	56	70
23	9	83	46	87	86	83	99	90	5	17	15	8	58	29	2	58	85	7	27	44	6	34	42
24	18	73	17	3	56	7	14	27	29	68	12	48	24	89	62	54	39	34	10	71	1	13	61

Microsoft Excel

- A function is typed into a cell

=SUM(A2:A15)

“=” Tells Excel that what follows is a function, not just some text to display

“SUM” The name of the function. SUM in this case tells Excel to add up all values in the provided range

“(A2:A15)” This is the range of cells (From A2 down to A15) that Excel should sum

Microsoft Excel

- There are innumerable functions
- Some useful examples include

=SUM

=AVERAGE

=StDev

=CountIf

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	69	79	8	58	37	67	48	20	62	50	99	50	53	8	23	32	68	23	6	44	5	37	51
2	70	25	22	81	84	14	57	74	15	67	92	5	65	44	45	88	29	95	6	91	36	86	15
3	94	40	78	32	9	8	59	65	72	8	98	7	50	34	78	92	55	8	68	25	45	43	99
4	14	66	95	20	95	71	69	8	40	95	53	59	51	96	43	29	77	92	67	88	67	44	68
5	35	86	9	44	42	60	17	85	60	94	65	21	39	78	96	82	90	16	77	7	48	97	27
6	89	73	28	71	32	1	21	35	21	54	40	16	9	33	69	59	17	31	70	42	45	56	16
7	74	6	39	13	62	58	30	99	13	50	55	42	90	60	81	21	98	92	30	11	77	76	27
8	66	87	84	73	25	95	84	14	65	80	34	7	48	99	97	73	70	36	14	20	80	41	64
9	14	30	41	41	30	77	72	34	52	28	38	41	24	36	76	69	23	44	82	79	54	20	87
10	40	71	79	51	60	51	19	57	27	45	82	68	73	22	48	54	94	19	24	61	46	69	13
11	71	28	29	58	7	62	84	1	31	73	94	42	23	26	10	51	34	50	54	95	6	81	88
12	71	14	40	24	35	19	10	13	10	45	29	53	67	67	78	14	91	85	49	68	65	41	63
13	28	79	23	16	45	73	8	5	4	46	15	44	1	73	8	25	50	51	22	92	89	89	62
14	20	72	68	55	53	26	85	54	71	10	32	86	30	70	59	2	50	2	93	38	10	57	38
15	53	15	28	93	18	78	80	11	57	5	80	96	50	50	67	62	65	30	6	52	8	17	24
16	77	17	46	9	46	11	56	28	89	77	16	90	4	24	36	79	13	91	46	59	88	40	92
17	97	77	52	81	50	92	44	78	13	97	72	26	6	75	71	21	90	53	19	40	33	52	90
18	10	66	82	11	68	40	11	49	83	100	50	82	52	78	67	55	16	47	47	57	26	62	67
19	31	70	2	63	80	41	83	47	44	12	98	17	99	27	21	42	52	44	86	74	10	59	77
20	79	93	68	15	45	97	72	90	91	56	68	74	49	87	27	51	44	40	18	86	23	17	89
21	27	81	63	60	32	54	93	52	20	39	96	43	40	100	59	51	26	94	92	17	27	4	28
22	41	80	72	85	75	73	44	45	42	28	94	74	8	80	81	79	15	95	73	51	79	56	70
23	9	83	46	87	86	83	99	90	5	17	15	8	58	29	2	58	85	7	27	44	6	34	42
24	18	73	17	3	56	7	14	27	29	68	12	48	24	89	62	54	39	34	10	71	1	13	61

Microsoft Excel

- Best way to understand it is to use it, so let's jump into that...
- 5 minute break and then onto Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	69	79	8	58	37	67	48	20	62	50	99	50	53	8	23	32	68	23	6	44	5	37	51
2	70	25	22	81	84	14	57	74	15	67	92	5	65	44	45	88	29	95	6	91	36	86	15
3	94	40	78	32	9	8	59	65	72	8	98	7	50	34	78	92	55	8	68	25	45	43	99
4	14	66	95	20	95	71	69	8	40	95	53	59	51	96	43	29	77	92	67	88	67	44	68
5	35	86	9	44	42	60	17	85	60	94	65	21	39	78	96	82	90	16	77	7	48	97	27
6	89	73	28	71	32	1	21	35	21	54	40	16	9	33	69	59	17	31	70	42	45	56	16
7	74	6	39	13	62	58	30	99	13	50	55	42	90	60	81	21	98	92	30	11	77	76	27
8	66	87	84	73	25	95	84	14	65	80	34	7	48	99	97	73	70	36	14	20	80	41	64
9	14	30	41	41	30	77	72	34	52	28	38	41	24	36	76	69	23	44	82	79	54	20	87
10	40	71	79	51	60	51	19	57	27	45	82	68	73	22	48	54	94	19	24	61	46	69	13
11	71	28	29	58	7	62	84	1	31	73	94	42	23	26	10	51	34	50	54	95	6	81	88
12	71	14	40	24	35	19	10	13	10	45	29	53	67	67	78	14	91	85	49	68	65	41	63
13	28	79	23	16	45	73	8	5	4	46	15	44	1	73	8	25	50	51	22	92	89	89	62
14	20	72	68	55	53	26	85	54	71	10	32	86	30	70	59	2	50	2	93	38	10	57	38
15	53	15	28	93	18	78	80	11	57	5	80	96	50	50	67	62	65	30	6	52	8	17	24
16	77	17	46	9	46	11	56	28	89	77	16	90	4	24	36	79	13	91	46	59	88	40	92
17	97	77	52	81	50	92	44	78	13	97	72	26	6	75	71	21	90	53	19	40	33	52	90
18	10	66	82	11	68	40	11	49	83	100	50	82	52	78	67	55	16	47	47	57	26	62	67
19	31	70	2	63	80	41	83	47	44	12	98	17	99	27	21	42	52	44	86	74	10	59	77
20	79	93	68	15	45	97	72	90	91	56	68	74	49	87	27	51	44	40	18	86	23	17	89
21	27	81	63	60	32	54	93	52	20	39	96	43	40	100	59	51	26	94	92	17	27	4	28
22	41	80	72	85	75	73	44	45	42	28	94	74	8	80	81	79	15	95	73	51	79	56	70
23	9	83	46	87	86	83	99	90	5	17	15	8	58	29	2	58	85	7	27	44	6	34	42
24	18	73	17	3	56	7	14	27	29	68	12	48	24	89	62	54	39	34	10	71	1	13	61

Sample Parent Survey

1. Casework ID:

2. Client: ID:

3. What is your gender: Male Female

4. What is your current age? _____

5. How good is your caseworker at:

	Poor	Fair	Good	Very good	Excellent
Making you feel at ease?	1	2	3	4	5
Fully understanding your worries?	1	2	3	4	5
Really listening?	1	2	3	4	5

6. Did your caseworker refer you to other services you needed: Y N

Descriptive Statistics in Excel

- Add the “Data Analysis” on your Excel tool bar:
File → Options → Add-Ins → select “Analysis Toolpak” → GO → Okay.
- Descriptive Statistics:
Data → Data Analysis → Descriptive Statistics →
 - Input Range: select your data including variable names on the excel sheet
 - Group by: Columns
 - Check “Labels in first row”
 - New Worksheet Ply: name your spreadsheet
 - Check “Summary statistics”

Pivot Table

- General introduction
 - Insert → PivotTable → New Worksheet.
 - Values: e.g. any dependent variables
 - Column Labels: e.g. independent variables
 - Row Labels: e.g. independent variables
 - Report Filter

Pivot Table

- Frequency: to know the distribution of each value in a variable.
 - Row Labels: e.g. the variable of “feel at ease”
 - Value: e.g. how many rate 0, 1,2,3,4,5
 - * Check the variable (“Feel at ease”) in the box. Do it twice. One for counts; one for %.
 - * Count: select ‘Value Field Setting’ → choose ‘count’ under ‘Summarize Value By’
 - * Percent (%): select ‘Value Field Setting’ → choose ‘count’ under ‘Summarize Value By’ → select ‘Show Value As’ → choose ‘% of Column Total’ under ‘Show value as’

Pivot Table

- Group Data: e.g. “Age”
 - Row Labels: “Age”
 - Value: Count of “Age”
 - Select a “Single” cell in the “Row Labels”
 - PivotTable Tools → Options → Group Field

Chi-square

- Chi-square: to investigate whether distributions of categorical variables differ from one another. The Chi-square statistic is calculated by comparing the observed accounts with expected accounts.
- Excel Pivot Table

Step1: Pivot Table -Column labels: independent variable /cause (e.g. gender)

-Row labels: dependent variable/result (e.g. "service referral")

-Values: dependent variable/result (e.g. whether service is offered) → select 'Value Field Setting' → choose 'count' under 'Summarize Value By'

-copy and paste values the pivot table twice.

Step2: Change the account to % in the Pivot (Crosstab table)

-Values: click the variable → select 'Value Field Setting' → select 'Show Value As' → choose '% of Column Total' under 'Show value as'

Step3: Chi-square calculation

-The first table will be 'observed account' table and the second table 'expected account' table.

-Calculate the 'expected account': = (Grand Total of the column*Grand Total of row)/ Grand Total of all).

-Chi-Square in excel: =chisq.test (highlight the 'observed accounts' , highlight the 'expected accounts')

- p value assessment: $p < 0.5$ means significant difference

t-Test

- *t*-Test: A *t*-test's statistical significance indicates whether or not the difference between two groups' averages (means) most likely reflects a "real" difference in the population from which the groups were sampled.
- Excel: use the function of "Data Analysis"
 - Step 1: Sort the data (e.g. Gender). Choose "expand the selection."
 - Step2: Copy and past the accounts of the variable ("Total hours hours") by gender in a new spreadsheet.
 - Step3: select 'Data' → 'Data Analysis' → 't-Test'
 - *Two-Sample Assuming Equal Variances: equal numbers of the two groups.
 - *Two-Sample Assuming Unequal Variances: unequal numbers of the two groups.
 - *Paired Two Sample for Means (Pre and Post test)
- Reference:
 1. <http://docs.statwing.com/examples-and-definitions/t-test/statistical-significance/>
 2. http://www.ruf.rice.edu/~bioslabs/Stats_tutorial/ttest17.html