

Open-source and free software for in-class online surveys and data analysis

Prof. Enrique Frio

UPOU, UMUC

Free software

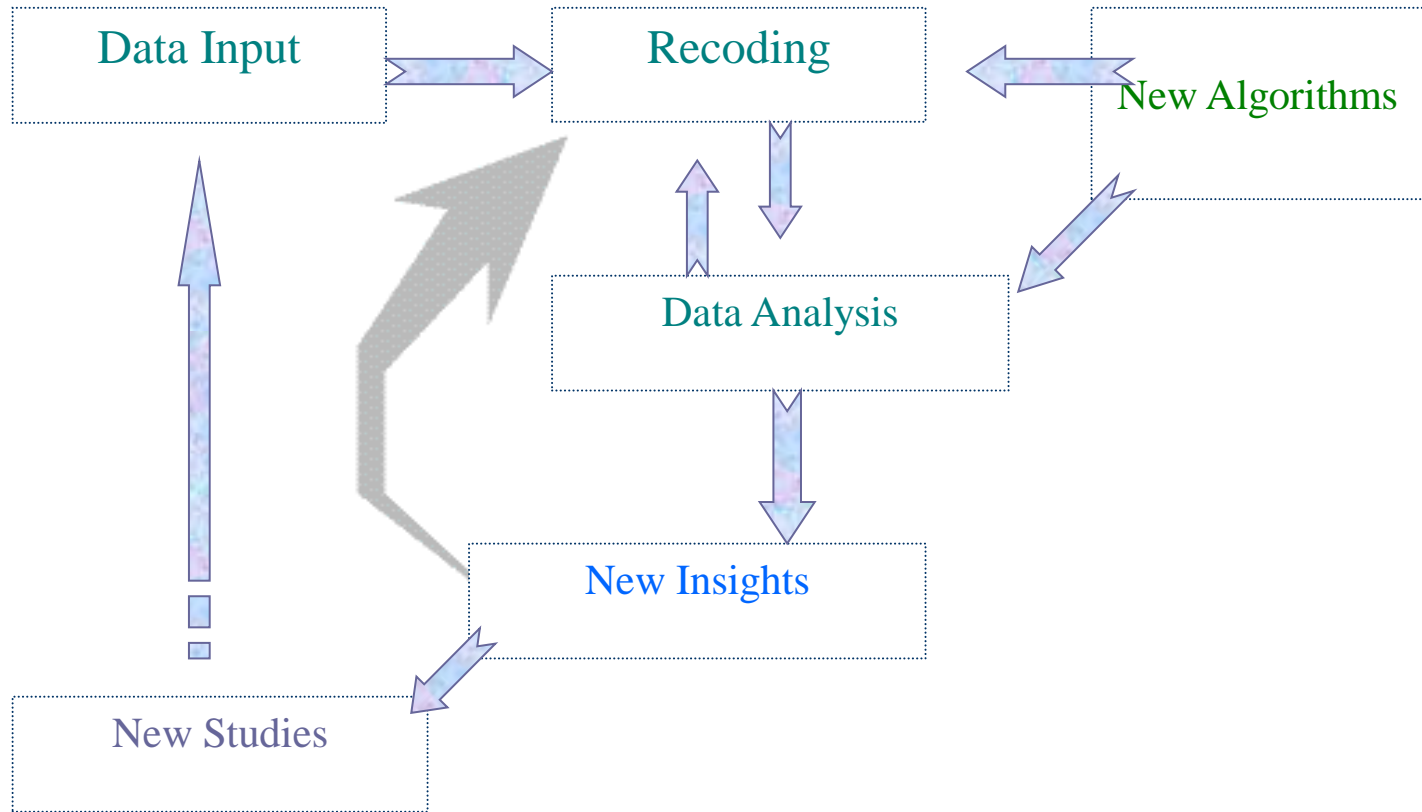
- ◆ PSPP – a clone of SPSS
 - Open source, free (GNU license)
 - Data collation, descriptive statistics
 - Hypothesis testing
 - Statistical analysis

- ◆ Kwiksurveys.com
 - Free questionnaire site
 - Questionnaire design
 - Questionnaire survey implementation
 - Automated data gathering of answers to the questionnaire

Questionnaires

- ◆ Types of answers to questions that we want from respondents:
 - Knowledge
 - Attitude
 - Behavior
 - Opinion
 - Beliefs
 - Suggestions
- + Demographic data
- ◆ Open-ended vs. closed questions

Data Analysis Cycle



Frio, E.2012.

SPSS vs. PSPP

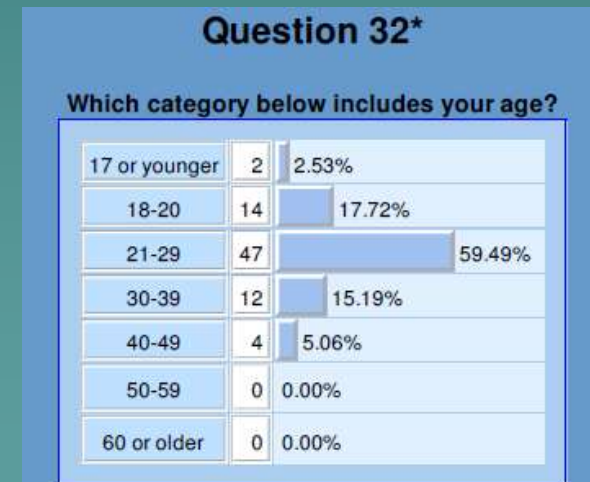
- ◆ Commercial software - IBM's SPSS -
 - *the* leading standard software for survey data analysis.
- ◆ "PSPP" – "clone" of SPSS
 - from the GNU foundation
 - new open-source program
 - allows free and unlimited data analysis capabilities on most platforms (Unix, PC, Mac).

Kwiksurveys.com

- ◆ Kwiksurveys.com = free website for quick questionnaire generation
 - fast and efficient alternative to time-consuming manual aspects of questionnaire design and implementation.
- ◆ Survey questionnaire design:
 - question generation
 - variables + variable values
 - online questionnaire implementation
 - E-mail invites or cookie-tracked links
 - automated data gathering
 - preliminary quick views / graphs of results

Kwiksurveys.com

- ◆ Questionnaire validation
 - Force respondents to answer all questions
 - * asterisks indicate required questions
- ◆ Survey results - quick view
 - Bar graphs
 - Counts
 - Percentages



Open-ended questions

• 1. Do you have a Facebook account? If not, why not?

Yes

If not, why?

[+ Add Question](#)

• 2. What is the highest level of school you have completed or the highest degree you have received?

High school degree or equivalent (e.g., GED)

Some college but no degree

Associate degree

Bachelor degree

Graduate degree

[+ Add Question](#)

• 3. Which of the following categories best describes your employment status?

Employed, working 1-39 hours per week

Employed, working 40 or more hours per week

Not employed, looking for work

Not employed, NOT looking for work

Retired

Disabled, not able to work

- “Please specify”
box – forces the
respondent to
answer

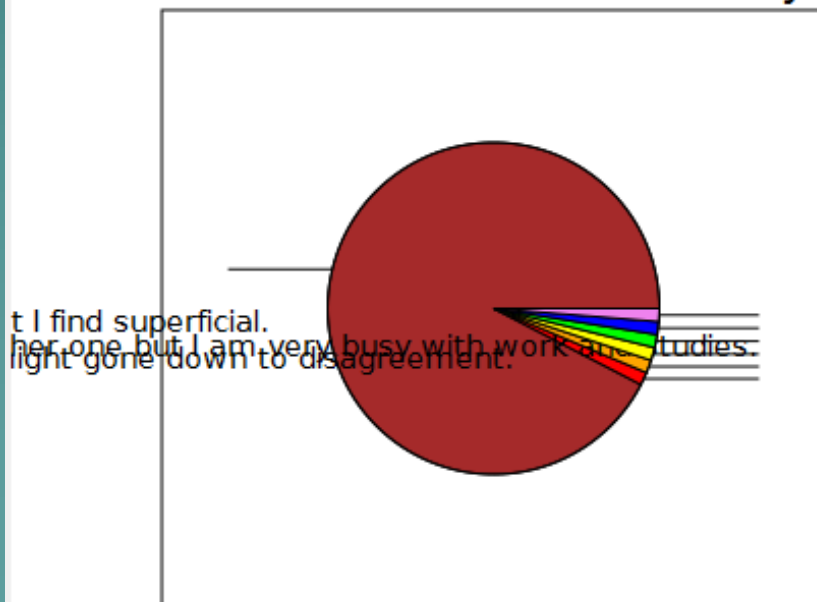
- More useful data
is obtained

Open-ended questions

No Facebook account - why

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	Honestly to avoid jealousy in my relationship with my boyfriend because if we dislike who we have accepted as a friend it might gone down to disagreement.	1	1.27	1.27	92.41
	I deleted mine when I was in undergrad studies because I blamed it for the reason I failed a major course. I wanted to create another one but I am very busy with work and studies.	1	1.27	1.27	93.67
	I find it consuming and addictive that the tendency to lag becomes too often.	1	1.27	1.27	94.94
	I personally think its dumb and not for me. I find it too superficial and people use it for various things that I find superficial.	1	1.27	1.27	96.20
	My girlfriend likes to ask a lot of questions about it.	1	1.27	1.27	97.47
	wasting time	1	1.27	1.27	98.73
<i>Total</i>		79	100.0	100.0	

No Facebook account - why



Open-ended questions

Question 27*

Which of the following sectors is your main occupation in?

Administration / Administrative	5	6.33%
Engineering	1	1.27%
Government	5	6.33%
Housewife/ Homemaker	0	0.00%
IT/ Computer-related	13	16.46%
Layman/ Minister/ Religious	0	0.00%
Medical/ Nursing	0	0.00%
Military / Police / Law Enforcement	0	0.00%
Music/ Arts/ Entertainment	6	7.59%
NGO	2	2.53%
Other Non-profit /Volunteer work	0	0.00%
Science	0	0.00%
Student	24	30.38%
Teaching/ Education	5	6.33%
Other (please specify)	18	22.78%

ID	View Survey	Email	First Name	Last Name	Other (please specify)
9139709	View				Writing
9141043	View				I'm a secretary in a review center
9141926	View				Real Estate Service
9142209	View				BPO
9144293	View				Garments (Fashion Design)
9149116	View				Contact Center/outsourcing
9156677	View				Training and Organizational Development
9159956	View				BPO
9171841	View				BPO
9203614	View				Human Resource
9256053	View				Entrepreneur
9258634	View				Graphics
9277975	View				performing artist
9294166	View				Testing/QC
9303691	View				BPO
9306357	View				BPO
9313104	View				BPO
9327584	View				Business

- ◆ **Occupation** – “Other” generated 18 other responses – very important!

Variable and data views in PSPP

STS_Social_Networks_Survey_PSPPcoded_SAV.sav [DataSet1] — PSPPIRE Data Editor

File Edit View Data Transform Analyze Utilities Windows Help

Open... Save Go To Case... Variables... Find... Insert Cases Insert Variable Split File... Weight Cases... Value Labels

I: SESSION_ 9137810

	FBWHY	PERSONAS	SECRETVW	FBEVENTS	ValuPriv	FRNDSUGG	MOMDAD	FBPHONE	UNFRIEND
1	So that I am able to contact and be updated about my distant relatives	Strongly Agree	No, never	Yes	Important	Neutral	Mom and Dad are	No, never	Not applicable
2	Easiest Communication Method	Agree	Yes	No	Very Important	Very Unlikely	Mom only	No, never	.
3	Interaction with friends/family whom I do not see not so often	Strongly Agree	Yes	Yes	Very Important	Very Unlikely	either of them are	No, never	.
4	to find my relatives abroad and to get some news about the announcements	Strongly Agree	Yes	Yes	Very Important	Neutral	either of them are	No, never	.
5	to connect to my long lost friend and find relatives abroad	Neutral	Yes	Yes	Very Important	Neutral	Not applicable	but I withdrew it	.
6	To meet and connect friends and advertise my products	Strongly Agree	No, never	Yes	Very Important	Neutral	either of them are	No, never	.
7	Blog exposure	Agree	Yes	Yes	Very Important	Unlikely	Mom and Dad are	it is still out there	.
8	stays connected with friends	Strongly Agree	Yes	Yes	Very Important	Very Unlikely	Not applicable	No, never	No
9	communication	Strongly Agree	I won't tell!	Yes	Very Important	Unlikely	Mom only	No, never	.
10	For fun	Strongly Agree	Yes	Yes	Important	Unlikely	Dad only	No, never	.
11	To communicate with my family, esp. my children (to track what happens)	Strongly Agree	Yes	Yes	Very Important	Unlikely	either of them are	it is still out there	No
12	to know what's happening outside the house and outside the office	Strongly Agree	I won't tell!	Yes	Very Important	Very Unlikely	Mom only	but I withdrew it	.
13	to get in touch with my friends, former high school and college colleagues	Strongly Agree	Not applicable	No	Very Important	Very Unlikely	Mom and Dad are	it is still out there	No
14	To look for my long lost HS and College friends	Strongly Agree	Yes	Yes	Very Important	Very Unlikely	Mom only	it is still out there	.
15	To communicate with friends	Neutral	Yes	No	Very Important	Neutral	either of them are	No, never	Not applicable
16	Stay in touch with past classmates, colleagues, etc.	Neutral	Yes	No	Very Important	Likely	Mom and Dad are	but I withdrew it	No
17	to keep updated	Neutral	Yes	No	Very Important	Unlikely	Mom only	but I withdrew it	.
18	to connect with friends	Strongly Agree	Yes	Yes	Very Important	Very Unlikely	either of them are	but I withdrew it	.
19	Connect with friends. Read the latest from TED, BBC GMANewsTV and	Agree	No, never	Yes	Very Important	Neutral	either of them are	but I withdrew it	.
20	for business	Agree	I won't tell!	Yes	Important	Unlikely	Mom and Dad are	it is still out there	.
21	News, updates, opinions	Agree	Yes	Yes	Neutral	Unlikely	Dad only	No, never	No
22	meet old friends and relatives	Agree	No, never	Yes	Very Important	Unlikely	Mom only	No, never	.
23	to get updates from friends	Strongly Agree	No, never	No	Very Important	Unlikely	Not applicable	No, never	No
24	groups, games	Strongly Agree	Yes	Yes	Very Important	Very Unlikely	Mom and Dad are	but I withdrew it	.
25	to be connected with old and new friends	Strongly Agree	Yes	Yes	Very Important	Very Unlikely	Not applicable	No, never	No
26	to communicate with old friends	Disagree	Yes	No	Important	Unlikely	either of them are	but I withdrew it	No

PSPP – Variable view

- ◆ Allows coding of variables, variable labels, and value labels

35	FBAppls	String	14	0		None	None
36	Marstat	String	18				
37	Relign	String	26				
38	Relign2	String	59				
39	Dependts	String	11				
40	Occpn	String	31				
41	Occpn2	String	39				
42	Msgs	String	38				
43	Safer	String	17				
44	Mzuck	String	135	0		None	None
45	Gender	String	6	0		None	None
46	Age	Scientific	13	3		None	None

Variable Type

Numeric
 Comma
 Dot
 Scientific notation
 Date
 Dollar
 Custom currency
 String

Width: 13
Decimal Places: 3

OK
Cancel
Help

Recoding data and variable values in PSPP

- ◆ Recoding – done to facilitate analysis
 - combine categories of a variable in order to increase the valid counts/ frequencies
 - E.g. Combine “Strongly agree” and “Agree” into “Agree”
 - Also recode categories with very few or zero counts into missing values
 - increases the amount of expected counts to have a valid statistical analysis

Recoding data and variable values in PSP

Question 24

What is your marital status?

Never married	64	82.05%
Married	12	15.38%
Separated/Divorced	2	2.56%
Widowed	0	0.00%

Recode into Different Variables: Old and New Values

Old Value:

Value: 4

System Missing

System or User Missing

Range:

through

Range: LOWEST thru value

Range: value thru HIGHEST

All other values

New Value:

Value: 2

System Missing

Copy old values

Add

Edit

Remove

Old	New
1	1
2	2
3	SYSMIS
4	SYSMIS

Output variables are strings Width: 0

Convert numeric strings to numbers ('5' -> 5)

Continue Cancel Help

Recode into Different Variables

Variables:

Old	New
MARITAL	MARITA2

Old and New Values

Output Variable:

Name: MARITA2

Label: Marital status 2

Change

OK

Paste

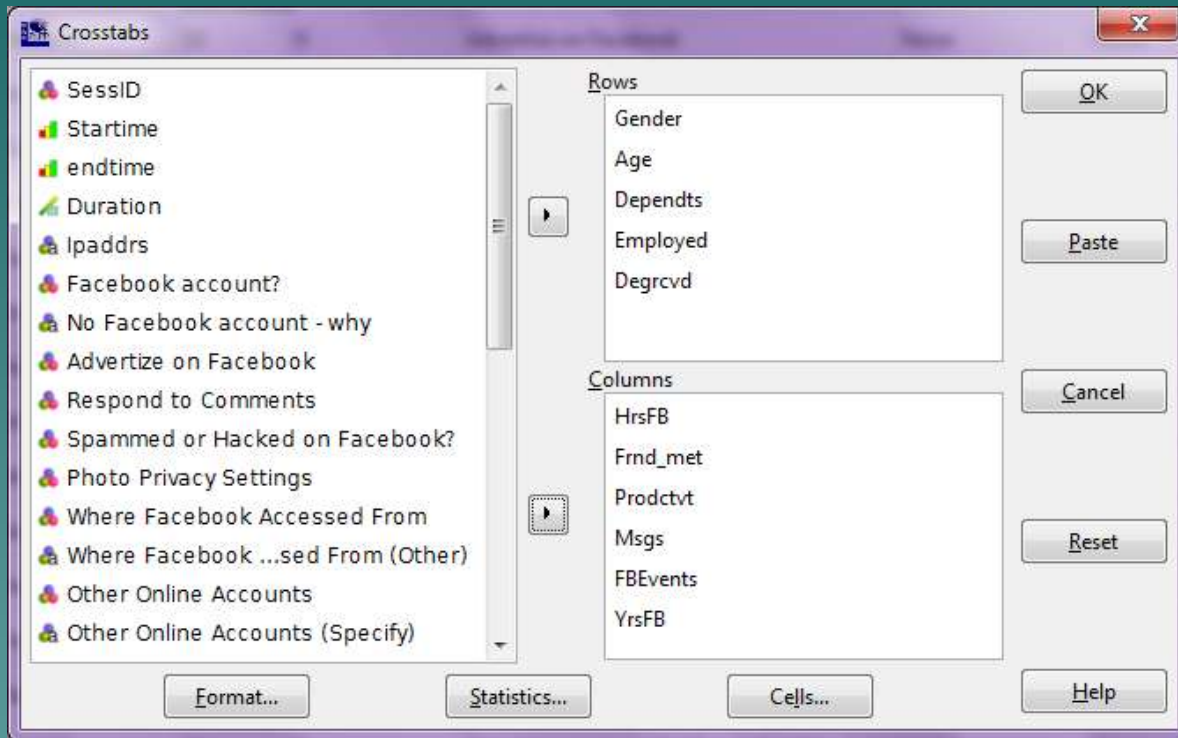
Cancel

Reset

Help

- ◆ Marital status – “Separated/divorced” and “Widowed” recoded as **System Missing (SYSMIS)** values

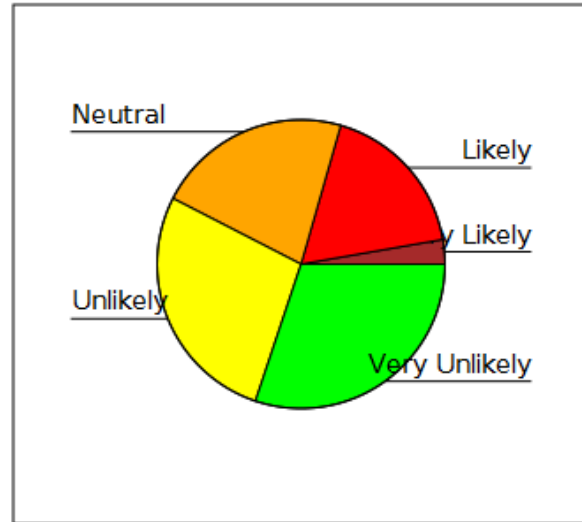
Crosstabs dialog box, PSpP



- ◆ Variables can be added to both rows & columns and analyzed in one go, much like in SPSS

Sample results

Friend suggestion - likelihood of addi



- ◆ Most (*Valid Percent = 27.40 + 30.14 = 57.54%*) are unlikely to add a person as a friend due to Facebook's "friend suggestion"

Friend suggestion - likelihood of adding as friend

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Very Likely	1	2	2.53	2.74	2.74
Likely	2	13	16.46	17.81	20.55
Neutral	3	16	20.25	21.92	42.47
Unlikely	4	20	25.32	27.40	69.86
Very Unlikely	5	22	27.85	30.14	100.00
.		6	7.59	Missing	
<i>Total</i>		79	100.0	100.0	

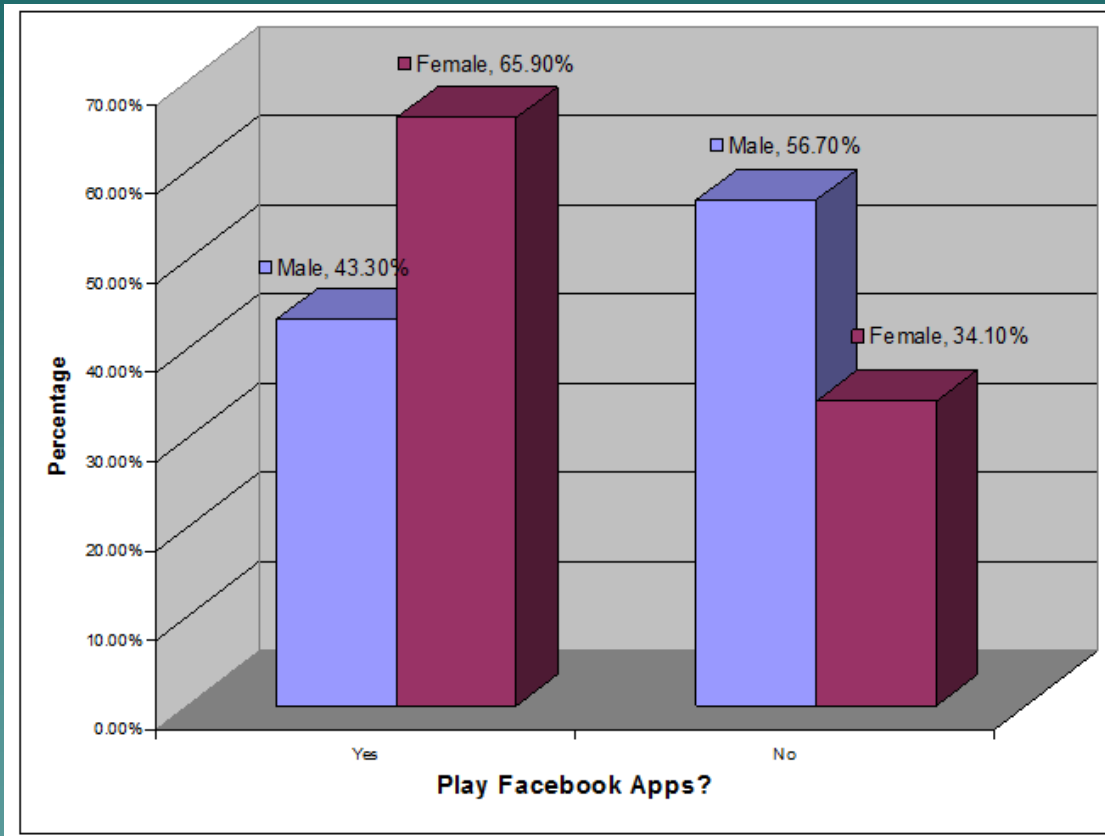
Crosstabs results

Gender	Play Facebook applications?		Total
	Yes	No	
Male	13	17	30
Expected	17	13	0
Row %	43.30%	56.70%	100.00%
Column %	31.00%	53.10%	40.50%
Total %	17.60%	23.00%	40.50%
Female	29	15	44
Expected	25	19	0
Row %	65.90%	34.10%	100.00%
Column %	69.00%	46.90%	59.50%
Total %	39.20%	20.30%	59.50%
Total	42	32	74
Total %	56.80%	43.20%	100.00%

Statistic	Value	df	Asymp. Sig. (2-tailed)
Pearson Chi-Square	3.7	1	0.05
Likelihood Ratio	3.71	1	0.05
Continuity Correction	2.84	1	0.09
Linear-by-Linear Association	3.65	1	0.06
N of Valid Cases	74		

- ◆ Crosstabulation and Chi-square results of **Gender** by **Play Facebook apps**.
- ◆ Crosstabs percentages and Expected values were labeled outside of PSPP.
- ◆ *P-value = 0.05*
- ◆ There might be an association between gender and playing Facebook applications
- ◆ Worth investigating

MS Excel graphs of Crosstabs data



- ◆ *Crosstab percentages of Gender by Play Facebook apps*
- ◆ *PSPP's graphs are not as good as those of SPSS*
- ◆ *Export data to MS Excel and generate graphs from there*

Crosstabs results

Gender * Friend suggestion - likelihood of adding as friend [count,

Gender	Friend suggestion - likelihood of adding as friend			Total
	Likely	Neutral	Unlikely	
Male	8.0	5.0	17.0	30.0
	6.2	6.6	17.3	.0
	26.7%	16.7%	56.7%	100.0%
	11.0%	6.8%	23.3%	41.1%
Female	7.0	11.0	25.0	43.0
	8.8	9.4	24.7	.0
	16.3%	25.6%	58.1%	100.0%
	9.6%	15.1%	34.2%	58.9%
Total	15.0	16.0	42.0	73.0
	20.5%	21.9%	57.5%	100.0%
	20.5%	21.9%	57.5%	100.0%

Chi-square tests.

Statistic	Value	df	Asymp. Sig. (2-tailed)
Pearson Chi-Square	1.58	2	.45
Likelihood Ratio	1.58	2	.45
Linear-by-Linear Association	.38	1	.54
N of Valid Cases	73		

- ◆ **Gender x Friend Suggestion**
- ◆ Males and females do not differ in their likelihood of adding someone as a friend based on Facebook's "friend suggestion"
- ◆ *P-value = 0.45 (not significant)*

Crosstabs results

Employment Status * Hours per day on Facebook [count, row %, total %, expected].

Employment Status	Hours per day on Facebook			Total
	<1 hour	1-3 hours	>3 hours	
Employed working 1-39 hours per week	7.0	5.0	2.0	14.0
	7.3	4.6	2.1	.0
	50.0%	35.7%	14.3%	100.0%
	9.6%	6.8%	2.7%	19.2%
Employed working 40 or more hours per week	18.0	11.0	5.0	34.0
	17.7	11.2	5.1	.0
	52.9%	32.4%	14.7%	100.0%
	24.7%	15.1%	6.8%	46.6%
Not employed looking for work	2.0	5.0	3.0	10.0
	5.2	3.3	1.5	.0
	20.0%	50.0%	30.0%	100.0%
	2.7%	6.8%	4.1%	13.7%
Not employed NOT looking for work	11.0	3.0	1.0	15.0
	7.8	4.9	2.3	.0
	73.3%	20.0%	6.7%	100.0%
	15.1%	4.1%	1.4%	20.5%
Total	38.0	24.0	11.0	73.0
	52.1%	32.9%	15.1%	100.0%
	52.1%	32.9%	15.1%	100.0%

- ◆ **Employment status x Hours per day spent on Facebook**
- ◆ The expected counts are too low (<5 per cell) to use for interpretation in the Chi-square test
- ◆ *Sometimes, nothing can be done due to the nature of the sample itself; small values in the category pairs cannot be helped.*

PSPP - Limitations

- ◆ PSPP Output Viewer window = not amenable to highlighting and copying of contents
 - work-around = export the output to html or PDF formats = enables copying and pasting crosstabs data directly into MS Word.
- ◆ Input limitations
 - various shortcuts and keyboard input options standard in Windows applications do not work in PSPP.

PSPP - Limitations

- ◆ PSPP build notes - “Bug” list in crosstabs analysis
 - **Pearson's R** (but not Spearman) is “off a little” (Stover, 2010b).
 - This statement may be incorrect, as Pearson's R is used for linear correlation analysis of two continuous variables
 - The **Pearson chi-square** (which falls under PSPP and SPSS' Crosstabs analysis) is a different measure used for testing for relationships between two categorical (i.e., non-continuous) variables (Statsoft, 2011).
- ◆ Watch this space – development and debugging of the software are ongoing and is a continuous process

PSPP - Limitations

◆ Crosstabs output

- expected counts, row, column, and total percentages are not labeled at all.
- headings of the crosstabs output ([count, row %, column %, total %, expected]) do not correspond with the consecutive rows of the table.

◆ PSPP graphing capabilities

- useful only for basic overviews of the data
- not really suitable for presentation graphics
- *Solution*: export tabular data into graphing software (e.g., MS Excel)
- No interactive graphs capability
 - ◆ No real-time manipulation of graphs (SPSS has this capability)

Conclusions

- ◆ PSPP is a good alternative to SPSS
 - Free, open-source
 - Development and debugging ongoing
 - Online forums for questions and suggestions
 - Watch for latest versions and builds
- ◆ Kwiksurveys.com
 - Free questionnaire generation and survey implementation online
 - Automated data gathering
 - Basic data overviews (graphs and percentages of results)
 - Data exportable as CSV and numerical or text values
- ◆ Good combination for in-class surveys and data analysis