

Comparison and Cross-Tab Data Analysis

Mel Langworthy
Terry Beasley

I. Comparison vs Cross-Tab

II. Response Collection – Plan of Action

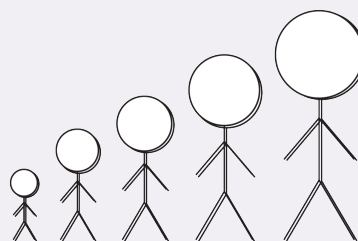
III. Comparison Reports in SurveyGizmo

IV. Cross-Tab Report – Example, Statistics, and SurveyGizmo how-to

Comparison vs Cross-Tab

Comparison	Cross-Tab
<ul style="list-style-type: none"> Simple comparison between 2 questions Ideal for demographic questions Response counts only No statistics Segments that match sample population highly recommended! 	<ul style="list-style-type: none"> Advanced comparison between 2 questions Single and multiple select questions Row, column, counts Statistics Segments that match sample population highly recommended!

Sample



Sample: Some definitions

- Population: represents your entire group of individuals for which you are trying to draw conclusions.
- Sample: a sub-group of the population, chosen using a statistically valid means, in order to represent the population as a whole.

What is sample? Why is it important?

- Your options:
 - Survey everyone
 - Survey a percentage
- Why?
 - Cost
 - Survey fatigue
 - Using a statistically valid sample is just as effective (or more effective) than trying to survey your entire population

About Sample

- A sample is statistically valid when every single person in that population has an equal chance or probability to be in a sample that you select.

What is sample size?

- How many responses do you need for your survey to be statistically accurate?
 - How accurate do you want the data to be?
 - How repeatable do you want the results to be?
 - How large is your total population?
 - What percentage of your population is represented by each segment?

Use a Sample Calculator

Determine Sample Size

Confidence Level: ☒ 95% ☐ 99%

Confidence Interval:

Population:

Sample size needed:

Or... estimate **400 responses!!!**

For Comparison Reports

- If you are segmenting your sample data for comparison, you need to make sure that the segments that you are using for comparison are the same as the segments in the represented population.
 - Ex: When comparing men and women in the US, you would need to make sure that the ratio within our survey was the same as the ratio within the larger US population.

Country	Under 15	Age 15-64	Over 65	Total
United States	1.04	1	0.75	0.97
Cuba	1.06	1	0.83	0.99

Males to # Females

For Cross-Tab Reports

- Need to ensure that the data you have (per question that you are cross-tabbing) is statistically valid for representing the larger population.
 - Survey items to consider
 - Question logic
 - Page logic
 - Percent branch action

Comparison - Compatible questions

Grouping Question types

- Radio Button
- Dropdown Menu (Single)
- Rating (Likert Scale)

Most questions can be compared

Question types that will not show data in Comparison report

- Drag and Drop Ranking
- Ranking Grid
- Continuous Sum

Cross-Tab - Compatible questions

- Radio Button
- Checkboxes
- Dropdown Menu (Single)
- Image Select (Single)
- Image Select (Multiple)
- Rating (Likert Scale)
- Radio Button Grid
- Checkbox Grid
- Custom Group Checkbox Grid
- Custom Group Radio Button Grid

Example Question

1. Midpoint County has 3 high schools; East Midway, North Central, and South Middle.
2. There are 3 family owned ice cream shops in Midpoint County, each specializing in a different flavor of ice cream.

Q. Is there a relation between the high school attended and favorite ice cream flavor?

Survey Setup

- Each high school graduates about 200 – 210 students per year (population).
- Sample size needed is 132 - 136 recent graduates per high school.
- Collect responses.
- Perform a Cross-tab analysis.
- Look for correlation between high school and favorite ice cream flavor.

Cross-Tab – Observed Values

School	Ice Cream			Total
	Vanilla	Chocolate	Strawberry	
East Midway	88 64.4% 161	32 24.2% 99	15 11.4% 32	132 100%
North Central	40 30.3% 79	28 18.7% 80	66 50.0% 141	132 100%
South Middle	35 25.7% 64	40 29.4% 79	61 44.9% 120	136 100%
Totals	160	98	142	400

Chi-Square: Ice Cream and School

Pearson Chi-Square	65.0748
Degrees of Freedom	4
p-Value	0 (< 0.05, likely correlated)

Demystifying Chi-Square Analysis

- The null hypothesis! Things are not related!
- Chi-Square Analysis tests if two variables are independent
 - Independent variables = no relationship
 - “accept the null hypothesis that variables are not related”
 - Non-significant
 - Dependent variables = related
 - “reject the null hypothesis. Variables are related”
 - Statistically significant

Calculating Chi-Square

$$\chi^2 = \frac{(\text{Observed Value} - \text{Expected Value})^2}{(\text{Expected Value})}$$

Perform calculation on every square in the cross-tab table, then sum all Chi-Square Values.

Cross-Tab – Expected Values

		What is your favorite Ice Cream flavor?			
		Vanilla	Chocolate	Strawberry	Row Totals
Where did you go to High School?	East Midway	44	44	44	132
	Row%	11.1%	11.1%	11.1%	33%
	North Central	44	44	44	132
	Row%	11.1%	11.1%	11.1%	33%
	South Middle	48	44	44	136
	Row%	11.2%	11.1%	11.1%	34%
	Column totals	136	132	132	400
	Column %	34%	33%	33%	100%

Cross-Tab – Observed values

		What is your favorite Ice Cream flavor?			
		Vanilla	Chocolate	Strawberry	Row Totals
Where did you go to High School?	East Midway	85	32	15	132
	Row%	64.4%	24.2%	11.4%	33.0%
	North Central	40	26	66	132
	Row%	30.3%	19.7%	50.0%	31.51%
	South Middle	35	40	61	136
	Row%	25.7%	29.4%	44.9%	33.56%
	Column totals	160	98	142	400
	Column %	40.0%	24.5%	35.5%	100%

Degrees of freedom

- The number of values in a statistic calculation that are free to vary.
 - Number of independent parameters

$$(\# \text{ Q1 options} - 1) + (\# \text{ Q2 options} - 1) = df$$

The p-Value

- Significance level of the test
 - Traditionally 10%, 5% or 1% (0.1, 0.05 or 0.01)
- Reflect the researchers willingness to accept a type of error, or the probability of rejecting the null hypothesis

A p-Value of 0.045 means there is 4.5% chance that the items are independent.

- Reject null hypothesis
- Statistically significant
- Dependent variables

A p-Value of 0.06 means there is a 6% chance that the items are independent.

- Accept null hypothesis
- Not significant
- Independent variables

Cross-Tab – Observed Values

School	Ice Cream			Total
	Vanilla	Chocolate	Strawberry	
East Midway	85 64.4% 161	32 24.2% 99	15 11.4% 32	132 100%
North Central	40 30.3% 76	26 19.7% 80	66 50.0% 141	132 100%
South Middle	35 25.7% 64	40 29.4% 120	61 44.9% 126	136 100%
Totals	160	98	142	400

Chi-Square: Ice Cream and School

Pearson Chi-Square	65.0748
Degrees of Freedom	4
p-Value	0 (< 0.05, likely correlated)