Some Tips in Reading and Using OCHNA Data Reports

The following pages describe some of the images, tables and terms you will see in OCHNA reports. This short guide will help you to better understand how to use the data reports put out by OCHNA and become a more savvy data user.

Response Tables

Response tables show the weighted and unweighted number of responses to a particular question in the survey. The "unweighted count" refers to the actual number of responses recorded. For example, in the table below, 147 interviewees answered yes to a particular survey question. A weighted count refers to the estimated number of people in the population that are represented by actual response. This is the number shown in the table as the "population estimate." So in the table below, an estimated 39,801 Orange County residents would have answered yes to the same question if we were able to interview them all. The population estimate is compared against the total population of Orange County to derive the percentages shown.

For information on how the population estimate was derived, click here.

Sample Response Table		
YES	Population Estimate	39,801
	Percent	**20.9%**
	Unweighted Count	147
NO	Population Estimate	150,628
	Percent	**79.1%**
	Unweighted Count	509
Total	Population Estimate	190,429
	Percent	**100.0%**
	Unweighted Count	656

Estimated number in the Orange County population

Percentage reflects the estimated number in the Orange County population

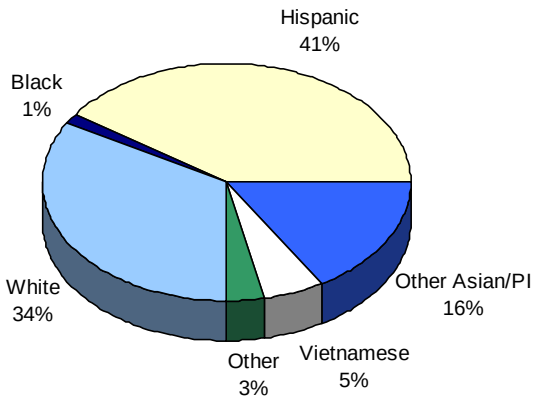
Actual number of responses to a question.

Chart Type 1: Of those Affected or Not Affected

Data can be presented and compared in different ways— in essence, from different points of view— depending on whether we decide to look at rows or columns. The charts throughout the data report are named in subtle yet systematic ways, so that that savvy data user will be able to easily interpret and understand exactly which viewpoint the chart takes. For example, the chart below specifically examines the number of those affected (i.e., who have never had a mammogram) who are of a certain condition or circumstance (i.e., their race/ethnicity). Therefore, the chart would be correctly read, "*41% of women who have never had a mammogram are Hispanic.*" It is important to note that this chart does *not* say that Hispanic women are more likely to have never had a mammogram; larger populations in the county as a whole will often have a larger piece of the pie, in part, because there are more of them.

These charts can be helpful to those who are interested in developing programs or studies that address a specific problem or circumstance.

SAMPLE CHART 1:
Race/Ethnicity of the xx.x% Women Who Have Never Had a Mammogram



Chi-square= 131.864, p-value ≤ 0.000

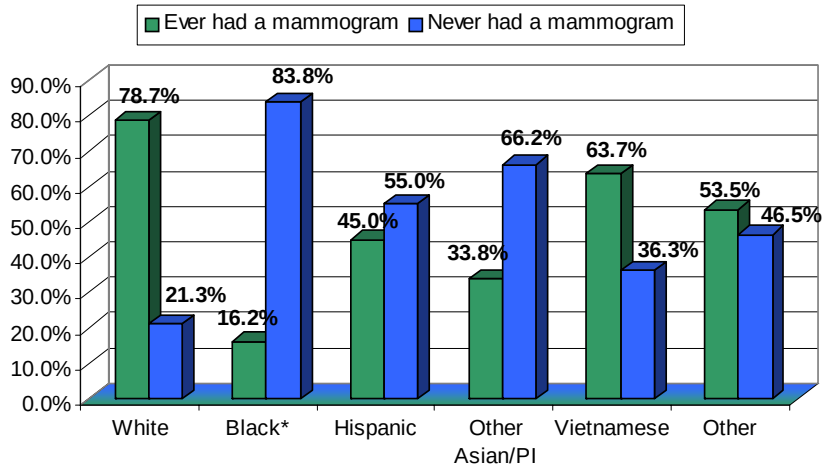
To find out what Chi-square and p-value mean, click here.

Chart Type 2: Within a Certain Population or Characteristic

Other charts in this data report look at those within certain populations or characteristics who are affected or not affected by a condition or circumstance. For example, the chart here looks at the percentage of those of a certain race/ethnicity who have and who have not had a mammogram as compared with others within the same race/ethnicity. Hence, the chart would be correctly read, "*55.0% of Hispanic women have never had a mammogram.*" These charts tell us which persons from different demographic categories are more likely to be affected by the condition or circumstance in question.

These charts are helpful to those examining differences among certain populations and in comparing needs among populations in order to better target services or campaigns.

SAMPLE CHART 2:
Mammogram Testing Within Each Race/Ethnicity



Population Tables

These types of population tables accompany charts and show the estimated number of Orange County residents that are affected by the question. For example, the figures on the chart below are the estimated number of women of each race/ethnicity who either had or never had a mammogram. Because we are consistently looking at the same population, but from different viewpoints (e.g., those who have never had a mammogram who are Hispanic, or Hispanic women who have never had a mammogram) the same population estimate table serves for both Chart Type 1 and Chart Type 2.

Race	Ever had a Mammogram	Never Had a Mammogram
White	499,425	135,298
Black	1,137	5,884
Hispanic	134,157	164,289
Other Asian/PI	33,506	65,632
Vietnamese	38,711	22,094
Other	15,485	13,439

Q&A: Some Common Data Questions

Where does the population estimate come from?

- The population estimates, also known as population weights, were created using a complex statistical process that "weighs" the demographics of each interviewee to determine how many people in the total population with similar demographics are represented by their survey responses. This allows for the calculation of unbiased population estimates from the sample data and for the sample responses to be more accurately representative of the entire Orange County population. The population weights were developed using 2004 population estimates for Orange County from the California Department of Finance (DOF).

Survey responses were "weighted" to more accurately reflect the demographics of Orange County.

What do chi-square and p-value mean?

- Basically, the chi-square (pronounced 'kI– skwer') number reported underneath the chart is a result of performing a chi-square test for statistical significance. The chi square test evaluates whether a relationship between two variables is significant, or probable. The chi square is only reported when the results of the test indicate that a significant (or highly probable) relationship exists. Significance testing was conducted for all of the survey questions to determine if there was a consistent difference as to how different groups, such as those characterized by race/ethnicity, income, gender, and age, responded.

- The p-value (a.k.a. probability value) indicates the significance level of a survey result. A significance level tells to what degree a result is reliable, or to what degree it may be due to chance. The most common level of reliability is 95%, meaning that the finding has a 95% chance of being true (also called a confidence interval). This is shown as a p-value of ".05," which indicates that the finding has a 5% (.05) chance of *not* being true (or of being false). This is the opposite of a 95% significance level, although it means essentially the same thing. The p-value is only reported when a relationship between two variables is significant (i.e., the p-value is less than or equal to 0.05).

A significance level tells us to what degree a result is reliable.

Why don't the unweighted counts match up with the percentages and population estimates in some tables?

- Occasionally you will see response tables where there might seem to be more or less unweighted counts (or number of responses) than are warranted by the percentage or the population estimate. This is because some respondents had a higher "weight" than others according to their demographic characteristics when the calculations were done for the population estimates. This is done so that the survey is as accurate a reflection of Orange County residents as possible. In general, these mismatches tend to occur when a survey question only had a small number of respondents.

If you come up with any additional questions, please feel free to contact OCHNA at (714) 547-3631 or staff@ochna.org.

Questions to Think About When Reviewing Data (Any data!)

Who are they talking about – what population is being reported on?

- Is it everybody? Children too? Certain sub-populations?
- Does it include all age groups? (For example, adults 18 to 64, but not 65+?)

What geographic area are they talking about?

- The state, county, region, city, a neighborhood, an organization, a particular community?
- Is the data broken down by zip codes or by census tracts?

What is the reporting period?

- Is the data from two years ago? (Most national, state and county objective data is two years behind, so a report done in 2004 will most likely be using data from 2002 or earlier).
- Is it data gathered on a regular reporting cycle that takes a measurement at a specific point in time? Each year, month quarterly?
- Is it an average of two or more reporting years?

What type of data is being presented?

- Primary survey? What type: phone, face-to-face, mail survey?
- Objective or secondary data from city, county, state or national sources?
- Qualitative data as obtained from focus groups? (This type of data provides a sense of how a small sample of people feel about particular issues. The findings can not be generalized to the whole population, but does provide a contextual framework to use in conjunction with other data sources).

How large a population or sample are they talking about and is it large enough to support the assertions made by the data report?

- The larger the sample, the more reliable the findings will be (all other variables being equal).
- Did a State survey sample 2,500 households in the entire state and then generalize the results of that survey to the County level? What proportion of the 2,500 samples was actually from that particular County?
- Was it only 500? This may be adequate to generalize to County population as a whole, providing a smaller confidence interval (90% CI), but any attempt at breaking the data out by sub-population or by regions would not be reliable.

What is the level of accuracy of the data being presented?

- Confidence Interval (CI) – measures statistically how sure, or confident, we can be that the numbers or percentages being presented are accurate. 99% CI or 95% CI are common measures, the higher the CI the more confidence we can have in the reliability that the data being presented will be representative of the actual real number in the population surveyed.
- A p-value, or probability value, is another measure we can use to determine the of reliability of the data. A CI of 95% is the same as a p-value of .05. The lower the p-value, the more probable it is that any relationships in the data are reliable and representative.
- The larger the sample size the more reliable the data, especially when breaking the data out by various population variables such as age, race/ethnicity or gender. Is the sample size is large enough to make generalizations regarding subpopulations?

The larger the sample size, the more reliable the data.

Know the limitation of the data – all data has it's strengths and weaknesses

- **Telephone Surveys** cannot reach households without telephones, homeless populations, those who are incarcerated or institutionalized. (Note: It is estimated that 98% of Orange County households have a telephone).

- **Objective Data** (National, State and Local government based data) will most often have a lag time of 2 years from time of reported occurrence to time of public release.

- **Mail Surveys** (paper/pencil) generally have a very poor response rate, which lowers the reliability of the data. These types of surveys will often not be completed by those who have poor reading or comprehension skills.

- **Internet surveys** are limited to those who have access to a computer and are computer literate.

- **Face-to-face surveys** are extremely time and staff intensive, random selection of the population sample is more difficult.

- **Focus Group** results cannot be generalized to the population as a whole or to other sub-groups, and will only provide information about how the participants of a particular focus group, feel or think about specified issues.

OCHNA – We're Here to Help

We encourage the use of all data available to the public, but caution the user to read the details that describe the population that is being represented, the sources for the data reported, the confidence intervals or error rates being reported (how sure can you be of the accuracy of what is being reported), and how large a sample size was used to base the estimates on. The larger the sample, the better, leading to a greater CI and thus smaller rate of error (all good reports will provide this information).

We understand that this is lot of information and OCHNA is here to help make this process easier. The above suggestions are just general guidelines that will be useful in determining why there can be many different statistics reported on what initially looks like the same population. Please feel free to call our offices for questions pertaining to specific data reports or questions regarding sources for data, including OCHNA survey data.

2024 N. Broadway, Suite 100
Santa Ana, CA 92706
(714) 547-3631, fax (714) 547-3629
staff@ochna.org
http://www.ochna.org