FOCUS on Field Epidemiology

CONTRIBUTORS

Authors:

> **Meredith Anderson, MPH**

> **Amy Nelson, PhD, MPH**

> **FOCUS Workgroup***

Reviewers:

> **FOCUS Workgroup***

Production Editors:

> **Tara P. Rybka, MPH**

> **Lorraine Alexander, DrPH**

> **Rachel Wilfert, MPH**

Editor in chief:

> **Pia D.M. MacDonald, PhD, MPH**

* All members of the FOCUS Workgroup are named on the last page of this issue.

UNC SCHOOL OF PUBLIC HEALTH

NORTH CAROLINA CENTER FOR PUBLIC HEALTH PREPAREDNESS

Data Analysis: Simple Statistical Tests

It's the middle of summer, prime time for swimming, and your local hospital reports several children with *Escherichia coli* O157:H7 infection. A preliminary investigation shows that many of these children recently swam in a local lake. The lake has been closed so the health department can conduct an investigation. Now you need to find out whether swimming in the lake is significantly associated with *E. coli* infection, so you'll know whether the lake is the real culprit in the outbreak. This situation occurred in Washington in 1999, and it's one example of the importance of good data analysis, including statistical testing, in an outbreak investigation. (1)

The major steps in basic data analysis are: cleaning data, coding and conducting descriptive analyses; calculating estimates (with confidence intervals); calculating measures of association (with confidence intervals); and statistical testing.

In earlier issues of FOCUS we discussed data cleaning, coding and descriptive analysis, as well as calculation of estimates (risk and odds) and measures of association (risk ratios and odds ratios). In this issue, we will discuss confidence intervals and p-values, and introduce some basic statistical tests, including chi square and ANOVA.

Before starting any data analysis, it is important to know what types of variables you are working with. The types of variables tell you which estimates you can calculate, and later, which types of statistical tests you

should use. Remember, continuous variables are numeric (such as age in years or weight) while categorical variables (whether yes or no, male or female, or something else entirely) are just what the name says—categories.

For continuous variables, we generally calculate measures such as the mean (average), median and standard deviation to describe what the variable looks like. We use means and medians in public health all the time; for example, we talk about the mean age of people infected with *E. coli*, or the median number of household contacts for case-patients with chicken pox.

In a field investigation, you are often interested in dichotomous or binary (2-level) categorical variables that represent an exposure (ate potato salad or did not eat potato salad) or an outcome (had *Salmonella* infection or did not have *Salmonella* infection). For categorical variables, we cannot calculate the mean or median, but we can calculate risk. Remember, risk is the number of people who develop a disease among all the people at risk of developing the disease during a given time period. For example, in a group of firefighters, we might talk about the risk of developing respiratory disease during the month following an episode of severe smoke inhalation.

Measures of Association

We usually collect information on exposure and disease because we want to compare two or more groups of
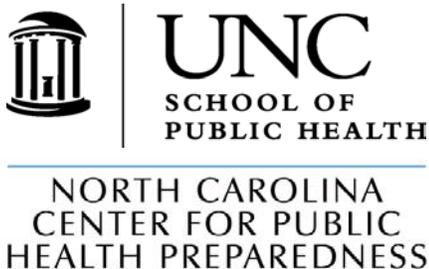
people. To do this, we calculate **measures of association**. Measures of association tell us the strength of the association between two variables, such as an exposure and a disease. The two measures of association that we use most often are the relative risk, or risk ratio (RR), and the odds ratio (OR). The decision to calculate an RR or an OR depends on the study design (see box below). We interpret the RR and OR as follows:

RR or OR = 1: exposure has no association with disease

RR or OR > 1: exposure may be positively associated with disease

RR or OR < 1: exposure may be negatively associated with disease

This is a good time to discuss one of our favorite analysis tools—the 2x2 table. You should be familiar with 2x2 tables from previous FOCUS issues. They are commonly used with dichotomous variables to compare groups of people. The table has one dichotomous variable along the rows and another dichotomous variable along the columns. This set-up is useful because we usually are interested in determining the association between a dichotomous exposure and a dichotomous outcome. For example, the exposure might be eating salsa at a restaurant (or not eating salsa), and the outcome might be Hepatitis A (or no Hepatitis A). Table 1 displays data from a case-control study conducted in Pennsylvania in 2003. (2)

So what measure of association can we get from this 2x2 table? Since we do not know the total population at risk (everyone who ate salsa), we cannot determine the risk of illness among the exposed and unexposed groups. That means we should not use the risk ratio. Instead, we will calculate the odds ratio, which is exactly what the study

---

Table 1. Sample 2x2 table for Hepatitis A at Restaurant A.

		Outcome		
		Hepatitis A	No Hepatitis A	Total
Exposure	Ate salsa	218	45	263
	Did not eat salsa	21	85	106
	Total	239	130	369

authors did. They found an odds ratio of 19.6*, meaning that the odds of getting Hepatitis A among people who ate salsa were 19.6 times as high as among people who did not eat salsa.

$$*OR = \frac{ad}{bc} = \frac{(218)(85)}{(45)(21)} = 19.6$$

Confidence Intervals

How do we know whether an odds ratio of 19.6 is meaningful in our investigation? We can start by calculating the **confidence interval (CI)** around the odds ratio. When we calculate an estimate (like risk or odds), or a measure of association (like a risk ratio or an odds ratio), that number (in this case 19.6) is called a point estimate. The confidence interval of a point estimate describes the precision of the estimate. It represents a range of values on either side of the estimate. The narrower the confidence interval, the more precise the point estimate. (3) Your point estimate will usually be the middle value of your confidence interval.

An analogy can help explain the confidence interval. Say we have a large bag of 500 red, green, and blue marbles. We are interested in knowing the percentage of green marbles, but we do not have the time to count every marble. So we shake up the bag and select 50 marbles to give us an idea, or an estimate, of the percentage of green marbles in the bag. In our sample of 50 marbles, we find 15 green marbles, 10 red marbles, and 25 blue marbles.

Based on this sample, we can conclude that 30% (15 out of 50) of the marbles in the bag are green. In this example, 30% is the point estimate. Do we feel confident in stating that 30% of the marbles are green? We might have some uncertainty about this statement, since there is a chance that the actual percent of green marbles in the entire bag is higher or lower than 30%. In other words, our sample of 50 marbles may not accurately reflect the actual distribution of marbles in the whole bag of 500 marbles. One way to determine the degree of our uncertainty is to calculate a confidence interval.

Risk ratio or odds ratio?

The risk ratio, or relative risk, is used when we look at a population and compare the outcomes of those who were exposed to something to the outcomes of those who were not exposed. When we conduct a cohort study, we can calculate risk ratios.

However, the case-control study design does not allow us to calculate risk ratios, because the entire population at risk is not included in the study. That's why we use odds ratios for case-control studies. An odds ratio is the odds of exposure among cases divided by the odds of exposure among controls, and it provides a rough estimate of the risk ratio.

So remember, for a cohort study, calculate a risk ratio, and for a case-control study, calculate an odds ratio. For more information about calculating risk ratios and odds ratios, take a look at *FOCUS Volume 3, Issues 1 and 2*.

Calculating Confidence Intervals

So how do you calculate a confidence interval? It's certainly possible to do so by hand. Most of the time, however, we use statistical programs such as Epi Info, SAS, STATA, SPSS, or Episheet to do the calculations. The default is usually a 95% confidence interval, but this can be adjusted to 90%, 99%, or any other level depending on the desired level of precision.

For those interested in calculating confidence intervals by hand, the following resource may be helpful:

Giesecke, J. *Modern Infectious Disease Epidemiology*. 2nd Ed. London: Arnold Publishing; 2002.

The most commonly used confidence interval is the 95% interval. When we use a 95% confidence interval, we conclude that our estimated range has a 95% chance of containing the true population value (e.g., the true percentage of green marbles in our bag). Let's assume that the 95% confidence interval is 17-43%.

How do we interpret this? Well, we estimated that 30% of the marbles are green, and the confidence interval tells us that the true percentage of green marbles in the bag is most likely between 17 and 43%. However, there is a 5% chance that this range (17-43%) does not contain the true percentage of green marbles.

In epidemiology we are usually comfortable with this 5% chance of error, which is why we commonly use the 95% confidence interval. However, if we want less chance of error, we might calculate a 99% confidence interval, which has only a 1% chance of error. This is a trade-off, since with a 99% confidence interval the estimated range will be wider than with a 95% confidence interval. In fact, with a 99% confidence interval, our estimate of the percentage of green marbles is 13-47%. That's a pretty wide range! On the other hand, if we were willing to accept a 10% chance of error, we can calculate a 90% confidence interval (and in this case, the percentage of green marbles will be 19-41%).

Ideally, we would like a very narrow confidence interval, which would indicate that our estimate is very precise. One way to get a more precise estimate is to take a larger sample. If we had taken 100 marbles (instead of 50) from our bag and found 30 green marbles, the point estimate would still be 30%, but the 95% confidence interval would be a range of 21-39% (instead of our original range of 17-43%). If we had sampled 200 marbles and found 60 green marbles, the point estimate would be 30%, and with a 95% confidence interval the range would be 24-36%. You can see that the confidence interval becomes narrower as the sample size increases.

Let's go back to our example of Hepatitis A in the Pennsylvania restaurant for one final review of confidence intervals. The odds ratio was 19.6, and the 95% confidence interval for this estimate was 11.0-34.9. This means there was a 95% chance that the range 11.0-34.9 contained the true odds ratio of Hepatitis A among people who ate salsa compared with people who did not eat salsa. Remember that an odds ratio of 1 means that there is no difference between the two groups, while an odds ratio greater than 1 indicates a greater risk among the exposed group. The lower bound of the confidence interval was 11.0, which is greater than 1. That means we can conclude that the people who ate salsa were truly more likely to become ill than the people who did not eat salsa.

It's necessary to include confidence intervals with your point estimates. That way you can give a sense of the precision of your estimates. Here are two examples:

- In an outbreak of gastrointestinal illness at two primary schools in Italy, investigators reported that children who ate a cold salad of corn and tuna had 6.19 times the risk of becoming ill of children who did not eat salad (95% confidence interval: 4.81-7.98). (4)
- In a community-wide outbreak of pertussis in Oregon in 2003, case-patients had 6.4 times the odds of living with a 6-10 year-old child than controls (95% confidence interval: 1.8-23.4). (5)

In both of these examples, one can conclude that there was an association between exposure and disease.

Resources for further study:

- Washington State Department of Health. Guidelines for Using Confidence Intervals for Public Health Assessment. http://www.doh.wa.gov/Data/Guidelines/ConfIntguide.htm
- Swinscow TDV. Chapter 8: The Chi-Square Test. *Statistics at Square One. 9th Ed*. BMJ Publishing Group; 1997. http://bmj.bmjjournals.com/collections/statsbk/8.shtml
- Simple Interactive Statistical Analysis http://home.clara.net/sisa/two2hlp.htm

Analysis of Categorical Data

You have calculated a measure of association (a risk ratio or odds ratio), and a confidence interval for a range of values around the point estimate. Now you want to use a formal statistical test to determine whether the results are statistically significant. Here, we will focus on the statistical tests that are used most often in field epidemiology. The first is the chi-square test.

Chi-Square Statistics

As noted earlier, a common analysis in epidemiology involves dichotomous variables, and uses a 2x2 table. We want to know if Disease X occurs as much among people belonging to Group A as it does among people belonging to Group B. In epidemiology, we often put people into groups based on their exposure to some disease risk factor.

To determine whether those persons who were exposed have more illness than those not exposed, we perform a test of the association between exposure and disease in the two groups. Let's use a hypothetical example to illustrate this. Let's assume there was an outbreak of *Salmonella* on a cruise ship, and investigators conducted a retrospective cohort study to determine the source of the outbreak. They interviewed all 300 people on the cruise and found that 60 had symptoms consistent with *Salmonella* (Table 2a). Questionnaires indicated that many of the case-patients ate tomatoes from the salad bar. Table 2a shows the number of people who did and did not eat tomatoes from the salad bar.

Table 2a. Sample Cohort study: Exposure to tomatoes and *Salmonella* infection

Salmonella?

	Yes	No	Total
Tomatoes	41	89	130
No Tomatoes	19	151	170
Total	60	240	300

To see if there is a significant difference in the amount of illness between those who ate tomatoes (41/130 or 32%) and those who did not (19/170 or 11%), one test we could conduct is a handy little statistic called χ^2 (or chi-square). In order to calculate a run-of-the mill chi-square, the following conditions must be met:

- There must be at least a total of 30 observations (people) in the table.
- Each cell must contain a count of 5 or more.

To conduct a chi-square test, we compare the observed data (from our study results) to the data we would expect to see. So how do we know what data would be expected? We need to know the size of our population, so we start with the totals from our observed data, as in Table 2b.

Table 2b. Row and column totals for tomatoes and *Salmonella* infection

Salmonella?

	Yes	No	Total
Tomatoes			130
No Tomatoes			170
Total	60	240	300

This gives us the overall distribution of people who ate tomatoes and people who became sick. Based on these distributions, we can fill in the empty cells of the table with the expected values, using the totals as weights. A computer program will calculate the expected values, but it is good to know that these numbers do not just fall out of the sky; there is actually a simple method to calculate them!

Expected Value $= \dfrac{\textit{Row Total x Column Total}}{\textit{Grand Total}}$

For the first cell, people who ate tomatoes and became ill:

Expected Value $= \dfrac{130 \times 60}{300} = 26$

We can use this formula to calculate the expected values for each of the cells, as shown in Table 2c.

Table 2c. Expected values for exposure to tomatoes

Salmonella?

	Yes	No	Total
Tomatoes	$\dfrac{130 \times 60}{300} = 26$	$\dfrac{130 \times 240}{300} = 104$	130
No Tomatoes	$\dfrac{170 \times 60}{300} = 34$	$\dfrac{170 \times 240}{300} = 136$	170
Total	60	240	300

To calculate the chi-square statistic, you use the observed values from Table 2a and the expected values that we calculated in Table 2c. You use this formula: [(Observed – Expected)^2/ Expected] for each cell in the table, as in Table 2d. Then you add these numbers together to find the chi-square statistic.

Table 2d. Expected values for exposure to tomatoes

Salmonella?

	Yes	No	Total
Tomatoes	$\frac{(41-26)^2}{26} = 8.7$	$\frac{(89-104)^2}{104} = 2.2$	130
No Tomatoes	$\frac{(19-34)^2}{34} = 6.6$	$\frac{(151-136)^2}{136} = 1.7$	170
Total	60	240	300

The chi-square (χ^2) for this example is 19.2 (8.7 + 2.2 + 6.6 + 1.7 = 19.2). What exactly does this number tell you? In general, the higher the chi-square value, the greater the likelihood that there is a statistically significant difference between the two groups you are comparing. To know for sure, though, you need to look up the p-value in a chi-square table. We will talk about the p-value and how the p-value is related to the chi-square test in the section below. First, though, let's talk about different types of chi-square tests.

Many computer programs give several types of chi-square tests. Each of these chi-square tests is best suited to certain situations.

The most commonly calculated chi-square test is Pearson's chi-square, or the uncorrected chi-square. In fact, if you see output that is simply labeled "chi-square," it is likely that it is actually Pearson's chi-square. A general rule of thumb is to use Pearson's chi square when you have a fairly large sample (>100). For a 2x2 table, the computer takes some shortcuts when calculating the chi-square, which does not work so well for smaller sample sizes but does make things faster to calculate.

The figure to the right identifies the types of tests to use in different situations. If you have a sample with less than 30 people or if one of the cells in your 2x2 tables is less than 5, you will need to use Fisher's Exact Test instead of a chi-square test. If you have matched or paired data, you should use McNemar's Test instead of a standard chi-square test (we'll talk more about McNemar's Test in the next issue of FOCUS).

Below are a few examples of studies that compared two groups using a chi-square test or Fisher's exact test. In each study, the investigators chose the type of test that best applied to the situation. Remember that the chi-square value is used to determine the corresponding p-value. Many studies, including the ones below, report only the p-value rather than the actual chi-square value.

- Pearson (Uncorrected) Chi-Square : A North Carolina study investigated 955 individuals referred to the Department of Health and Human Services because they were partners of someone who tested positive for HIV. The study found that the proportion of partners who got tested for HIV differed significantly by race/ethnicity (p-value <0.001). The study also found that HIV-positive rates did not differ by race/ethnicity among the 610 who were tested (p = 0.4). (6)

- Yates (Corrected) Chi-Square: In an outbreak of *Salmonella* gastroenteritis associated with eating at a restaurant, 14 of 15 ill patrons studied had eaten the Caesar salad, while 0 of 11 well patrons had eaten the salad (p-value <0.01). The dressing on the salad was made from raw eggs that were probably contaminated with *Salmonella*. (7)

- Fisher's Exact Test: A study of Group A Streptococcus (GAS) among children attending daycare found that 7 of 11 children who spent 30 or more hours per week in daycare had laboratory-confirmed GAS, while 0 of 4 children spending less than 30 hours per week in daycare had GAS (p-value <0.01). (8)

P-Values

Let's get back to our hypothetical cruise ship *Salmonella* outbreak. In our example, 32% of the people who ate tomatoes got *Salmonella* infection, compared to 11% of the people who did not eat tomatoes. (Although 32% and 11% *look* like they are different, we will need to check this with a statistical test. Often looks can be deceiving.) How do

Parade of Statistics Guys	
The right test...	*To use when....*
Pearson's chi-square (uncorrected)	Sample size >100 Expected cell counts > 10
Yates chi-square (corrected)	Sample size >30 Expected cell counts ≥ 5
Mantel-Haenszel chi-square	Sample size > 30 Variables are ordinal
Fisher exact	Sample size < 30 and/or Expected cell counts < 5

we know whether the difference between 32% and 11% is a "real" difference? In other words, how do we know that our chi-square value of 19.2 indicates a statistically significant difference? The **p-value** is our indicator!

Many statistical tests give both a numeric result (e.g., our chi-square value of 19.2) and a p-value. The p-value is a number that ranges between 0 and 1. So what does the p-value tell you? The p-value is the probability of getting the result you got, assuming that the two groups you are comparing are actually *the same.*

In other words, we start by assuming that there is no difference in outcomes between the groups (e.g., the people who ate tomatoes and those who did not). Then we look at the test statistic and p-value to see if they indicate otherwise. A low p-value means that, assuming the groups are actually the same, the probability of observing these results just by chance is very small. Therefore, you can call the difference between two groups statistically significant. A high p-value means that the two groups were not that different. A p-value of 1 means that there was no difference at all between the two groups.

Generally, if the p-value is less than 0.05, the difference observed is considered statistically significant. Although this is somewhat arbitrary, we often consider statistically significant differences to be real. If the p-value is below 0.05, we conclude that the difference observed is a true difference and did not happen by chance.

A number of statistical tests can be employed to obtain the p-value. The test you use depends on the type of data you have. In the section above, we discussed the chi-square test. If the chi-square statistic is small, this indicates that

---

A cautionary note...

Statistical tests such as the chi-square are based on several assumptions about the data, including **independence** of observations. The assumption of independence means that the value of one observation does not influence the value of another observation.

If this assumption is not actually true in your study, then it is incorrect to use the test. Situations in which the chi-square test should not be used include: repeat observations of the same group of people (e.g., pre- and post-tests) and matched pair designs in which cases and controls are matched on variables such as age and sex. These situations require special data analysis methods.

Luckily for us, in most of the situations we face in field epidemiology, the assumptions of the chi-square test are met. For example, the people who ate the potato salad were not the same people as those who did not eat the potato salad.

---

observed and expected data were not very different, and the p-value is large (>0.05, i.e. the difference was not statistically significant). If the chi-square statistic is large, this generally means that the p-value is small, and the difference could be statistically significant.

At the beginning of this issue, we mentioned an outbreak of *E. coli* O157:H7 associated with swimming in a lake. In that study, the investigators reported that "case-patients were significantly more likely than camper control subjects to have taken lake water into the mouth (p-value =0.002) and to have swallowed lake water (p-value =0.002)." (1) Because the p-values were each less than 0.05, both exposures were considered to be statistically significant risk factors.

Analysis of Continuous Data

Data do not always fit into discrete categories but continuous numeric data can also be of interest in a field investigation. You might want to compare clinical symptoms between groups of patients. For example, you might want to compare the temperature (fever) of adult patients with that of children. You could compare the average age of patients to the average age of non-patients, or the respiratory rate of those who were directly exposed to a chemical plume to the rate of those who were not exposed.

It is possible to make this type of comparison through a test called **Analysis Of Variance (ANOVA).** Most of the major statistical software programs will calculate ANOVA, but the output varies slightly in different programs. Here we will discuss the output from Epi Info to introduce the ANOVA test and two other useful tests. Epi Info generates 3 pieces of information when you conduct a test using ANOVA: the ANOVA results, Bartlett's test, and the Kruskal-Wallis test. The next few paragraphs describe how to interpret all of these tests.

ANOVA

When comparing continuous variables between groups of study subjects, a test that results in a p-value is used. However, instead of using chi-square tests as we did for categorical data, we use a **t-test** (for comparing 2 groups) or an **f-test** (for comparing 3 or more groups). The ANOVA procedure uses either the t-test or the f-test, depending on the number of groups being compared. A t-test compares averages between two groups, takes into account the variability in each group, and results in a statistic (t) that has a p-value.

For example, let's say that we are testing age differences between 2 groups. If the groups have very similar average ages and a similar "spread" or distribution of age values,

the t-statistic will be relatively small and the p-value will not be significant (i.e., it will be >0.05). If the average ages of the 2 groups are quite different, the t-statistic will be larger, and the p-value will be smaller. If the p-value is less than 0.05, the groups have significantly different ages.

Bartlett's Test

In t-tests and f-tests, there is one critical assumption that you should know about. ANOVA assumes that the two groups have similar **variances**. In other words, the two groups have a similar "spread" of age values. Think of it as comparing apples to apples. As part of the ANOVA analysis, the software program conducts a separate test just to see if the variances of the two groups are comparable. This test is called Bartlett's test for **equality of variance**. The basic assumption for this test is that the variances are comparable.

If you see a p-value for the test that is larger than 0.05, which is not significant, everything is fine. You can use the results of your ANOVA. However, if Bartlett's p-value is less than 0.05, the variances in the groups are not the same and you cannot use the results of the ANOVA. In this situation there is another test that you can use in place of ANOVA: the Kruskal-Wallis test.

Kruskal-Wallis Test

In Epi Info, a third test result is given that you will only need if Bartlett's test told you that the variances were not similar enough to use ANOVA (i.e., the p-value of Bartlett's

test was less than 0.05). The Kruskal-Wallis test does not make assumptions about the variance in the data. The Kruskal-Wallis test does not test averages but examines the distribution of the values within each of the groups. The test will result in a p-value. If the p-value is larger than 0.05, then you can conclude that there is not a significant difference between the groups. If the p-value is less than 0.05, there is a significant difference between the groups. Figure 1 gives a summary of when to use each of these tests for analyzing continuous data.

Conclusion

Analysis of epidemiologic data is a vital link in implicating exposures in disease causation. You could take years' worth of coursework just to learn epidemiologic and statistical methods. However, in field epidemiology, a few tried and true calculations and tests make up the core of analytic methods. Here we have presented confidence intervals, p-values, chi-square tests, ANOVA, and their interpretations. When you master these methods, your field epi skills will be among the best!

The next issue of FOCUS will delve even further into data analysis to tackle methods that you can use to control for confounding, including matching and logistic regression.

Figure 1. Decision tree for analysis of continuous data.
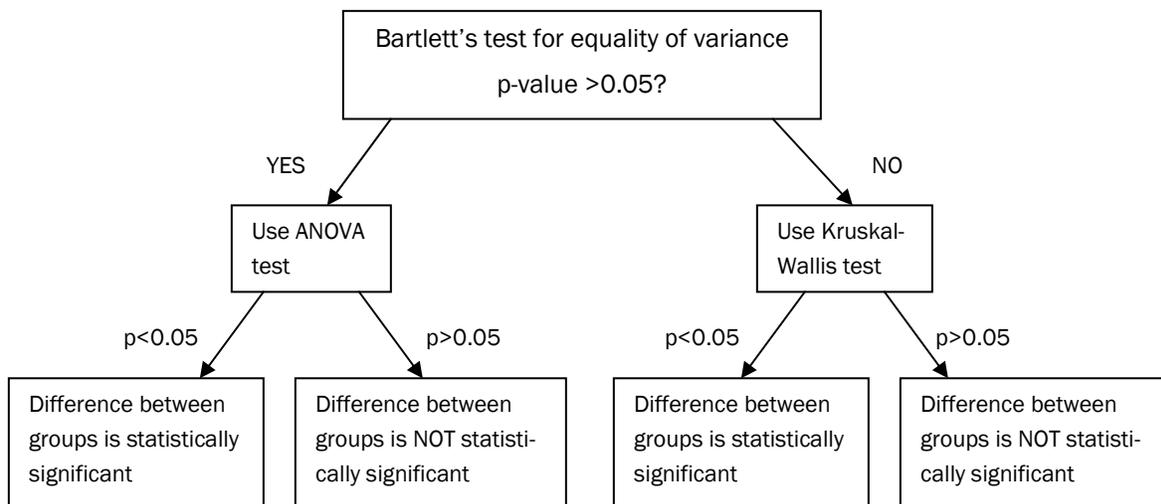
CONTACT US:

The North Carolina Center for Public Health Preparedness

The University of North Carolina at Chapel Hill

Campus Box 8165

Chapel Hill, NC 27599-8165

Phone: 919-843-5561

Fax: 919-843-5563

Email: nccphp@unc.edu

FOCUS Workgroup:

- Lorraine Alexander, DrPH
- Meredith Anderson, MPH
- David Bergmire-Sweat, MPH
- Kim Brunette, MPH
- Anjum Hajat, MPH
- Pia D.M. MacDonald, PhD, MPH
- Gloria C. Mejia, DDS, MPH
- Amy Nelson, PhD, MPH
- Tara P. Rybka, MPH
- Rachel A. Wilfert, MD, MPH

REFERENCES:

1. Bruce MG, Curtis MB, Payne MM, et al. Lake-associated outbreak of *Escherichia coli* O157:H7 in Clark County, Washington, August 1999. *Arch Pediatr Adolesc Med.* 2003;157:1016-1021.
2. Wheeler C, Vogt TM, Armstrong GL, et al. An outbreak of hepatitis A associated with green onions. *N Engl J Med*. 2005;353:890-897.
3. Gregg MB. *Field Epidemiology.* 2nd ed. New York, NY: Oxford University Press; 2002.
4. Aureli P, Fiorucci GC, Caroli D, et al. An outbreak of febrile gastroenteritis associated with corn contaminated by *Listeria monocytogenes*. *N Engl J Med.* 2000;342:1236-1241.
5. Schafer S, Gillette H, Hedberg K, Cieslak P. A community-wide pertussis outbreak: an argument for universal booster vaccination. *Arch Intern Med.* 2006;166:1317-1321.
6. Centers for Disease Control and Prevention. Partner counseling and referral services to identify persons with undiagnosed HIV --- North Carolina, 2001. *MMWR Morb Mort Wkly Rep.*2003;52:1181-1184.
7. Centers for Disease Control and Prevention. Outbreak of *Salmonella* Enteritidis infection associated with consumption of raw shell eggs, 1991. *MMWR Morb Mort Wkly Rep.* 1992;41:369-372.
8. Centers for Disease Control and Prevention. Outbreak of invasive group A streptococcus associated with varicella in a child-care center -- Boston, Massachusetts, 1997. *MMWR Morb Mort Wkly Rep.* 1997;46:944-948.

If you would like to receive electronic copies of FOCUS on Field Epidemiology, please fill out the form below:

- NAME: _____

- DEGREE (S): _____

- AFFILIATION: _____

- E-MAIL ADDRESS: _____

- May we e-mail any of your colleagues? If so, please include their e-mail addresses here:

  _____

  _____

  _____

Please fax to: (919) 919-843-5563

or mail to: North Carolina Center for Public Health Preparedness

The University of North Carolina at Chapel Hill

Campus Box 8165

Chapel Hill, NC 27599-8165

Or go online: http://www.sph.unc.edu/nccphp/focus/

UPCOMING TOPICS!

- **Advanced Data Analysis:** Methods to Control for Confounding

- Collecting Specimens in Outbreak Investigations

- **Laboratory Diagnosis: An Overview**

We are on the web!

http://www.sph.unc.edu/nccphp