Graphing Multiple Confidence Intervals

Metta Kuning*[a], Don McNeil [b]

[a] Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Thailand.
[b] Department of Statistics, Division of Economic and Financial Studies**,** Macquarie University, Australia.
*Author for correspondence; e-mail: kmetta@bunga.pn.psu.ac.th

Abstract

This study proposed a graphical method for displaying informative confidence intervals for any data set involving a categorical determinant, an outcome that could be either continuous or categorical, and an optional covariate. For continuous outcomes the graph is similar to the conventional plot of confidence intervals for individual means, but focuses on specified natural contrasts. The analogous graph for categorical outcomes involves graphing confidence intervals for specified natural odds ratios. In each case the method can be extended to include a covariate, and the extent of confounding can be illustrated on the graph. The confidence intervals are adjusted for multiplicity.

_____

Keywords: ANOVA plots, confidence intervals, graphical techniques, Mantel-Haenszel methods, multiplicity, odds ratio plots.

1. Introduction

Aided by improvements in the quality and availability of relevant computer software, graphical methods are now routinely used in statistical practice, both for preliminary and exploratory data analysis and for reporting results from studies. For preliminary analysis, box plots and scatterplot matrices are particularly useful for displaying sets of continuous variables, while clustered or divided bar charts are commonly used for graphing categorical data. To illustrate conclusions, confidence intervals can be graphed.

This article focused on investigation of a simple method for graphing multiple confidence intervals for comparing several groups of data where the outcome variable is either continuous or categorical. Manifestations already exist in (a) plots of confidence intervals of means in the analysis of variance and (b) meta-analysis plots for graphing sets of odds ratios. The aim of the study is to put these graphs into a more general paradigm in which further issues, such as multiple comparisons and model comparisons, can be dealt with straightforwardly.

2. Comparing Means

When comparing population means as in one-way analysis of variance, it is conventional to plot confidence intervals for individual means based on pooled standard deviation, s. Figure 1 (upper panel) shows a typical example of this graph. In this case the data comprise weight gains among three groups of anorexic girls given different treatments [1], and summary statistics are given in Table 1. The statistic for testing the null hypothesis in this case is $F_{2,69} = 5.429$ giving the p-value 0.0065. The horizontal lines denote 95% confidence intervals for the individual population means μ_j Denoting by $t_{\alpha/2,\nu}$ the critical value of the t distribution with two-tailed area α and ν degrees of freedom, the formula is

$$\overline{y}_j \pm t_{0.025,\nu}\, s\Big/ \sqrt{n_j}\ . \tag{1}$$

If the objective is to compare the population means, it could be argued that a more informative plot is that shown in the middle panel of Figure 1. This plot displays simultaneous confidence intervals for the contrasts between each mean and the overall population mean μ, and also uses oblique lines to join pairs of means that are not statistically different according to Tukey's multiple comparison method with $\alpha = 0.05$ [2]. Denoting the total number of observations by n and the number of samples by c, a confidence interval for the contrast $\mu_j - \mu$ is given by the formula

$$\overline{y}_j - \overline{y} \pm t_{0.025/(c-1),\nu}\, s\sqrt{1/n_j - 1/n}\ . \tag{2}$$

The vertical line denotes the overall sample mean and the confidence interval for each contrast is centered at the corresponding mean and adjusted for multiplicity by increasing its probability content from 0.95 to 1 $c \square 0.05/($

Treatment	Size (n_j)	Mean ($\bar{y}_j$)	St. Dev. (s_j)
Cognitive behavioural	29	3.007	7.309
Control	26	$\square 0.450$	7.989
Family therapy	17	7.271	7.161
Pooled	72	2.765	7.529

Table 1: Summary statistics for weight gains (pounds) of anorexic girls.
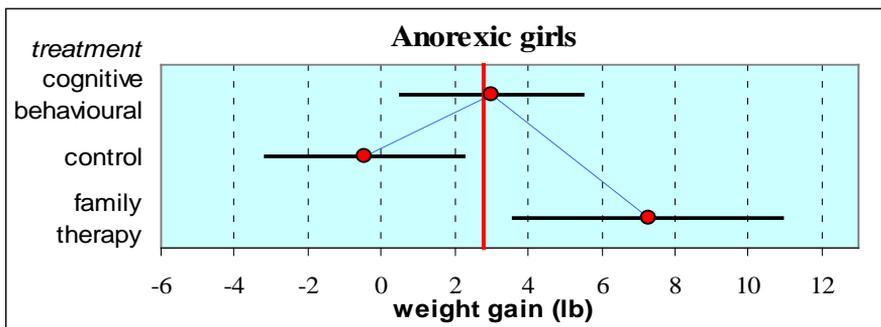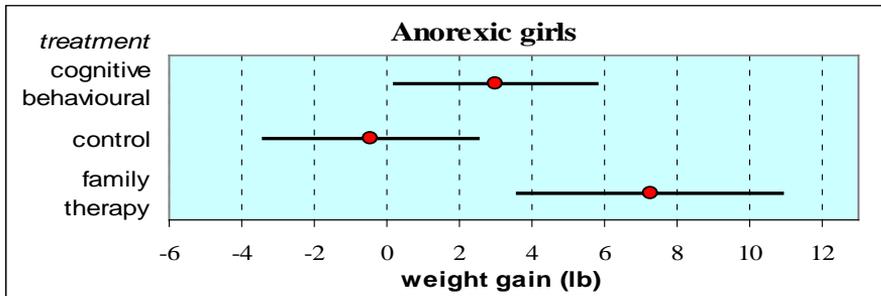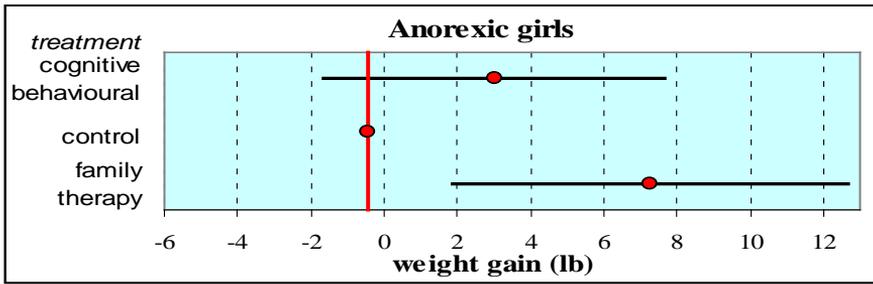
Figure 1: Confidence intervals for individual population means (upper panel), contrasts with overall mean (middle panel), and comparisons with control group (bottom panel) for weight gains of three groups of anorexic girls.

Comparing the upper two panels of Figure 1, it is clear that the two sets of confidence intervals are quite similar. In general, focusing on a contrast with the overall mean rather than an individual mean has the effect of shrinking the interval by the factor $\sqrt{(1-n_j/n)}$, whereas allowing for multiplicity inflates the interval by a factor that increases with the number of groups.

When one of the groups is a control group, as in this example, it is more appropriate to show the contrast between each group and the control group. This could be achieved by centering the vertical line at the mean of the control group instead of the overall mean and computing the corresponding confidence intervals, as shown in the bottom panel of Figure 1. If c is the number of samples and k is the index of the control group, the formula for each confidence interval is

$$\bar{y}_j - \bar{y}_k \pm t_{0.025/(c-1),v} s \sqrt{1/n_j + 1/n_k} \ . \tag{3}$$

The graph can be extended to show simultaneous confidence intervals for contrasts in more general situations involving additional covariates, such as two-way analysis of variance, simply by using regression analysis to adjust the contrasts for the covariates.

3. Combining Odds Ratios

For categorical outcome variables the statistics of interest are proportions rather than means, and odds ratios are used to measure associations. In the simplest case when each variable is binary, the data can be stored as a 2×2 contingency table of frequencies a, b, c and d (reading across rows or columns), and the association can be summarized by an odds ratio. The estimated odds ratio is

$$w = \frac{ad}{bc}.$$ (4)

when

	Disease status	
	D	ND
E	a	b
NE	c	d

Exposure status

Using the asymptotic formula (see, for example, Breslow & Day [3])

$$SE(\ln w) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}},$$ (5)

a 95% confidence interval for the population odds ratio is thus

$$w \exp(\pm 1.96\ SE(\ln w)),$$ (6)

and the association can be graphed simply as a point denoting the estimated odds ratio enclosed in an interval denoting the confidence interval.

This graphical method extends to situations involving a stratification variable, where it is commonly called a meta-analysis plot (see, for example, Sutton et al. [4]). This plot shows confidence intervals for the individual odds ratios together with a confidence interval for the overall odds ratio, which can be obtained by adjusting for the stratification variable as follows. If a_k, b_k, c_k and d_k denote the frequencies in stratum k ($k = 1, 2, …, m$) and $n_k = a_k + b_k + c_k + d_k$, a robust estimate of the strata-adjusted odds ratio [5] is

$$\tilde{w} = \frac{\sum a_k d_k / n_k}{\sum b_k c_k / n_k},$$ (7)

and an asymptotic formula for the standard error of its logarithm is [6]

$$SE(\ln \tilde{w}) = \sqrt{\frac{\sum P_k R_k}{2\left(\sum R_k\right)^2} + \frac{\sum \left(P_k S_k + Q_k R_k\right)}{2\left(\sum R_k\right)\left(\sum S_k\right)} + \frac{\sum Q_k S_k}{2\left(\sum S_k\right)^2}},$$ (8)

where $P_k = (a_k + d_k)/n_k$, $Q_i = (b_k + c_k)/n_k$, $R_k = (a_k\, d_k)/n_k$, and $S_k = (b_k\, c_k)/n_k$.

Since the *m* stratum-specific confidence intervals are viewed simultaneously, they need to be corrected for multiplicity, and this could be achieved by increasing the probability content of each from 0.95 to $1\square0.05/m$.

As an illustration, Table 2 cites some data on admission rates for men and women applicants to six graduate departments at Berkeley in 1973 [7]. These data are often used as an example of confounding (or "Simpson's paradox"), because the crude odds ratio (ignoring the stratification variable) is 1.84, suggesting discrimination against women, whereas the odds ratio after adjusting for department is 0.90. Figure 2 shows a corresponding plot of individual odds ratios (lower panel), together with the adjusted and crude overall odds ratios (upper panel), highlighting the confounding effect.

Dept.	men admitted	women admitted	men not admitted	women not admitted
A	512	89	313	19
B	353	17	207	8
C	120	202	205	391
D	138	131	279	244
E	53	94	138	299
F	22	24	351	317

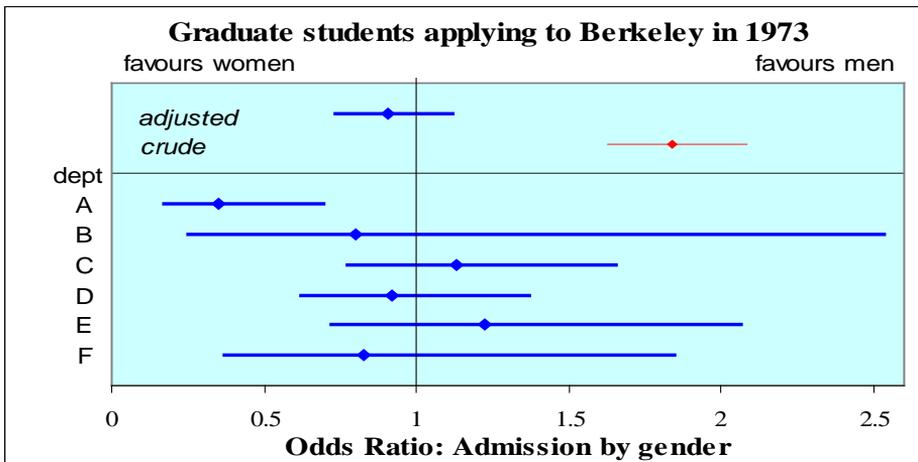**Table 2:** Admissions to six graduate programs at UC Berkeley in 1973.

Figure 2: Odds ratio plot showing 1973 Berkeley graduate school admissions.

Since the Mantel-Haenszel formula assumes a common odds ratio within each stratum of the population, it is desirable to assess this assumption using a homogeneity test. Using an idea attributed by Bartlett to Fisher (see Agresti [8], Breslow & Day [3]) defined a chi-squared goodness-of-fit statistic with $m \square 1$ degrees of freedom as follows. The expected counts for each 2×2 subtable have odds ratio $\tilde{w}$ and the same row and column totals as the observed subtable, and are thus computed by solving a quadratic equation. The chi-squared statistic is then obtained by computing a component for each cell of the form (*observed* $\square$ *expected*)^2/*expected* and summing the components from all 4*s* cells. For the 1973 Berkeley graduate admissions, where $m = 6$, this test statistic is 18.83 with 5 degrees of freedom (*p*-value 0.002) so the homogeneity assumption is not tenable, and it is clear from Figure 2 that Department A was anomalous. However, confounding remains evident when the data are regraphed as in Figure 2 with Department A omitted.

4. Graphing *r*×*c* Tables

The odds ratio confidence interval plot can be extended to categorical variables with more than two categories, using the following rationale. The conventional (Pearson) chi-squared statistic for testing independence takes the form

$$X = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(a_{ij} - \hat{a}_{ij}\right)^2}{\hat{a}_{ij}}, \tag{9}$$

where a_{ij} is the count in cell (*i, j*) and $\hat{a}_{ij}$ is its expected value under the independence assumption. Since $\hat{a}_{ij} = (a_{ij} + b_{ij})(a_{ij} + c_{ij})/n$, where

$$b_{ij} = \sum_{i=1}^{r} a_{ij} - a_{ij}, \ c_{ij} = \sum_{j=1}^{c} a_{ij} - a_{ij}, \ n = \sum_{i=1}^{r} \sum_{j=1}^{c} a_{ij},$$

Equation (9) may be written in the form

$$X = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(b_{ij}c_{ij}\right)^2 \left(w_{ij} - 1\right)^2}{n\left(a_{ij} + b_{ij}\right)\left(a_{ij} + c_{ij}\right)}, \tag{10}$$

where $d_{ij} = n - a_{ij} - b_{ij} - c_{ij}$ and

$$w_{ij} = \frac{a_{ij}d_{ij}}{b_{ij}c_{ij}}. \tag{11}$$

Note that if r and c are both 2, w_{11} reduces to the conventional odds ratio given in Equation (4). In general, w_{ij} are the odds ratios obtained by collapsing the $r \times c$ table into a set of 2×2 tables around each pivotal cell having count a_{ij}, giving "floating" odds ratios like those suggested by Bithell [9] and Easton et al [10]. Given that the overall measure of association X is a weighted sum of squares of differences between these odds ratios and their null values, they form a basis for visualizing the nature of the association. Each w_{ij} is simply based on a 2×2 table of counts, so the formula for its standard error is the same as that used for binary data. If none of the counts is zero, the asymptotic formula (5) could be used, giving

$$SE\ (\ln w_{ij}) = \sqrt{\frac{1}{a_{ij}} + \frac{1}{b_{ij}} + \frac{1}{c_{ij}} + \frac{1}{d_{ij}}}\ . \qquad (12)$$

If both r and c exceed 2 it is informative to display all rc odds ratios, even though only $(r-1)(c-1)$ of them are independent. To allow for multiplicity each confidence interval could be inflated so that its probability content is 1 $ - 0.05/\{r(c-1)\}$.

As a further illustration, Table 3 gives data on eye and hair colour collected by Snee [12] from 592 students at the University of Delaware, and used to illustrate a graphical method for displaying results based on a simplex confidence region plot. In this case the chi-squared test for independence is highly statistically significant ($X = 138.3$ with 9 degrees of freedom, $p < 0.0001$).

	Hair colour			
Eyes	black	brown	red	blond
brown	68	119	26	7
blue	20	84	17	94
hazel	15	54	14	10
green	5	29	14	16

Table 3: Hair colour and eye colour for 592 students from the University of Delaware.

Figure 3 shows the corresponding odds ratio plot for these data. In agreement with Snee's conclusion, it shows that the subjects with black hair tended to have brown eyes, blonds tended to have blue eyes, and that there was some tendency for subjects with red hair to have green eyes and for subjects with brown hair to have hazel-coloured eyes.
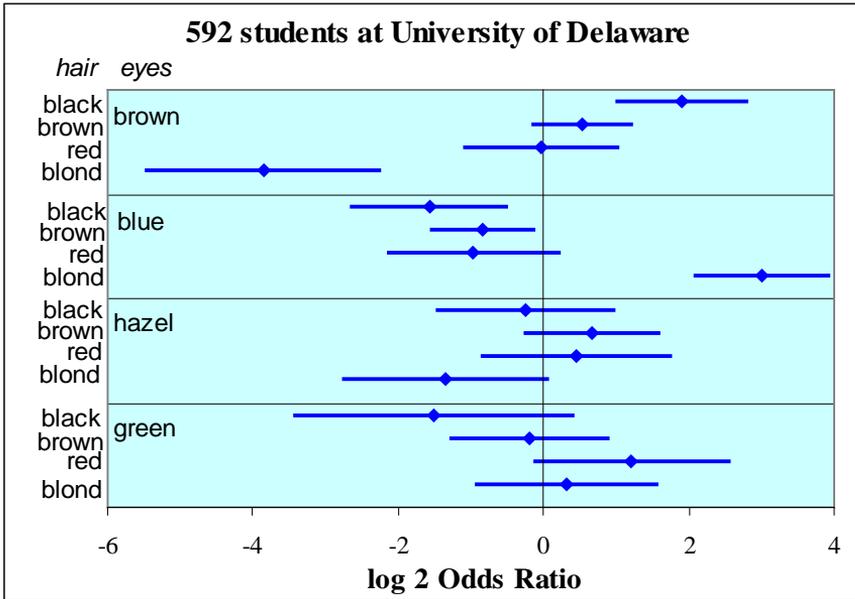
Figure 3: Odds ratio plot of Snee's hair and eye colour data.

5. Ordinal Data

In many cases categorical variables are ordinal rather than nominal, and thus Pearson's independence test has low power for detecting an association. Clayton [13] defined an ordinal odds ratio to measure the strength of association for such data, as an improvement on a chi-squared test for trend due to Armitage [14]. For an $r \times c$ table, the formula is

$$w_{ord} = \frac{\sum_{i=1}^{r-1}\sum_{j=1}^{c-1} u_{ij} A_{ij} D_{ij}}{\sum_{i=1}^{r-1}\sum_{j=1}^{c-1} u_{ij} B_{ij} C_{ij}}, \tag{13}$$

where A_{ij} is the accumulated count in cells with row indexes less than or equal to i and column indexes less than or equal to j, $B_{ij} = A_{rj} - A_{ij}$, $C_{ij} = A_{ic} - A_{ij}$, $D_{ij} = n - A_{ij} - B_{ij} - C_{ij}$, and

$$u_{ij} = \frac{\sum_j \left(a_{ij} + a_{i+1,j}\right)\sum_i \left(a_{ij} + a_{i,j+1}\right)}{n^3}. \tag{14}$$

Clayton also gave an asymptotically valid formula for the standard error of this estimate, namely

$$SE\ (\ln\ w_{ord}) = \frac{1}{\sqrt{\sum_{i=1}^{r-1}\sum_{j=1}^{c-1} u_{ij}}}\ . \tag{15}$$

The ordinal odds ratio provides a model giving expected cell counts for such data. These expected values are such that each 2×2 table, obtained from the original table by accumulating data with row and column indexes up to or above a specified pair (i, j), has odds ratio w_{ord}. As in the case of the Breslow-Day homogeneity test, these values are solutions to quadratic equations. The expected cell counts for the $r×c$ table are then obtained by differencing the expected accumulated counts, and the goodness-of-fit of the ordinal model may be assessed by constructing a chi-squared statistic based on the differences between these expected counts and the observed counts. The corresponding number of degrees of freedom is $(r-1)(c-1)-1$.

Both hair colour and eye colour could be regarded as ordinal variables, with hair colour ordered by increasing darkness from blond to red to brown to black, and eye colour ordered similarly from blue to green to hazel to brown. Applying equations (13) and (15) to the data in Table 3 with this ordering, w_{ord} = 4.462 and $SE\ (\ln\ w_{ord})$ = 0.140. Using the method described above, the resulting expected counts are given in Table 4. The chi-squared goodness-of-fit statistic is 36.40 with 8 degrees of freedom ($p < 0.0001$). While this is a sizable improvement compared to the independence model (where X = 138.3 with 9 degrees of freedom), the fit remains unacceptable.

Eyes	Hair colour			
	black	brown	red	blond
brown	71.7	115.1	14.7	18.6
hazel	14.5	54.0	10.8	13.7
green	7.1	34.3	9.4	13.1
blue	14.8	82.6	36.1	81.6

Table 4: Expected frequencies for Table 3 data based on Clayton's ordinal odds ratio

Figure 4 shows the odds ratio plot for the data from Table 3 with the eye colours reordered and augmented by lines joining the odds ratios based on the model. Comparing the data with the model, too few red-headed subjects had blue eyes and too few blond-haired subjects had brown eyes.
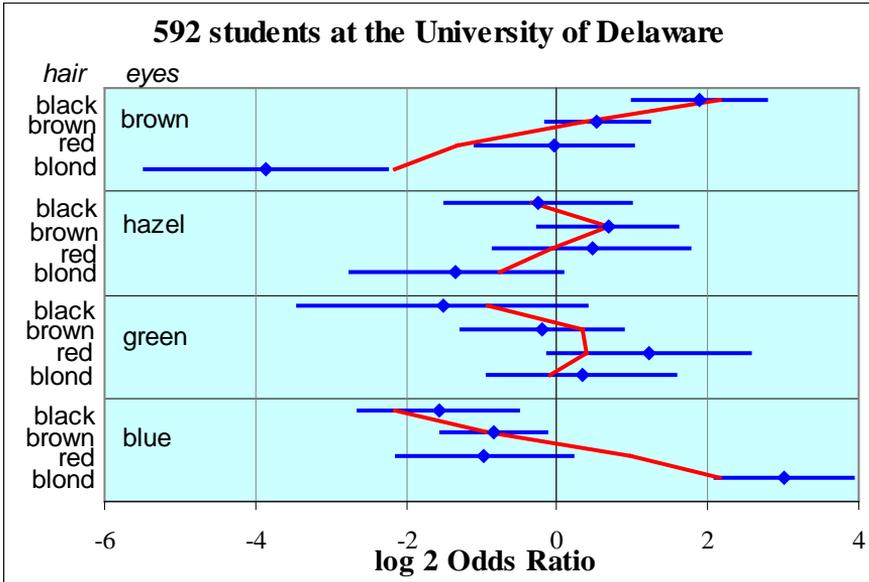
Figure 4: Odds ratio plot of hair and eye colour data with ordinal model superimposed

6. Stratified r×c Tables

Birch [15] derived an asymptotically valid chi-squared test for independence in a set of m stratified r×c tables. The formula for the test statistic is

$$X = \left(\sum \mathbf{a}_k - \sum \hat{\mathbf{a}}_k \right)^T \left(\sum \mathbf{V}_k \right)^{-1} \left(\sum \mathbf{a}_k - \sum \hat{\mathbf{a}}_k \right), \tag{16}$$

where $\mathbf{a}_k$ is a column vector with $(r-1)(m-1)$ elements of the form

$$\left(a_{11k}, a_{12k}, ..., a_{1,c-1,k}, ..., a_{r-1,c-1,k} \right), \tag{17}$$

$\hat{\mathbf{a}}_k$ its expected value, and $\mathbf{V}_k$ its null covariance matrix, given by

$$Cov\left(a_{ijk}, a_{i'j'k} \right) = \frac{a_{i+k}\left(\delta_{ii'} n_k - a_{i'+k} \right) a_{+jk}\left(\delta_{jj'} n_k - a_{+j'k} \right)}{n_k^2 \left(n_k - 1 \right)}, \tag{18}$$

where δ_{uv} is 1 if $u = v$, 0 otherwise.

Note that when $c = 2$, X reduces to the logrank test for comparing r survival curves (see, for example, Kalbfleisch & Prentice [16]). In this case the stratification variable is the failure time.

Using the Mantel-Haenszel formula (7), we may define the adjusted odds ratio associated with the cell (i, j) as

$$w_{ij} = \frac{\sum_{k=1}^{s} a_{ijk} d_{ijk} / n_k}{\sum_{k=1}^{s} b_{ijk} c_{ijk} / n_k} . \tag{19}$$

Similarly, Equation (12) provides a formula for the standard error of ln (w_{ij}).

These formulas provide a basis for plotting cell-specific adjusted odds ratios and their confidence intervals. Moreover, an $r \times c$ table of adjusted counts with cell-specific odds ratios is obtained using the method described in Section 3, yielding a homogeneity test with $(r-1)(c-1)(m-1)$ degrees of freedom.

As an illustration, Table 5 shows the frequencies of rapes and injuries in various stranger rape situations for 247 women who were assaulted, classified by whether or not a specified mode of resistance was used. In the December 1979 Newsletter of the Bureau of Social Science Research (Washington DC) Albert Biderman and Douglas Neal plotted these outcome frequencies as a set of 32 quadrantal rose plots (see Fienberg [17]), one for each combination of the four situations, the four modes of resistance, and whether or not resistance was used. More than one mode of resistance could have been used.

Wainer [18] looked at this graph and concluded that "we can see at a glance that the most dangerous situation is being taken away from a public place and that in any situation the mode of resistance chosen does not seem to have much effect". The first conclusion is also evident from the odds ratio plots shown in the upper panels of Figure 5. These plots display the odds ratios giving the strength of the association of each outcome with each location after adjusting for whether or not a particular mode of resistance was used. While they clearly show that the risk of rape and injury when taken away from a public place was relatively high for all modes of resistance, they also show relatively high risks of rape without (physical) injury for women who were asleep at home, as well as relatively low risks of rape for women who were assaulted at but not taken away from a public place.

Location	Outcome	freq	Mode of resistance tried			
			Dissuad	Scream	Flee	Fight
Asleep at home	Raped / Injured	12	8	6	0	0
	Raped / Not Injured	17	9	5	5	4
	Not raped / Injured	5	3	4	0	4
	Not raped / Not	1	1	0	0	0
Awake at home	Raped / Injured	9	8	3	3	6
	Raped / Not Injured	11	8	3	2	0
	Not raped / Injured	5	3	1	2	5
	Not raped / Not	8	5	3	1	3
Taken from public place	Raped / Injured	50	41	20	17	28
	Raped / Not Injured	22	16	5	6	7
	Not raped / Injured	11	8	9	8	7
	Not raped / Not	12	8	2	3	5
Not taken from public place	Raped / Injured	16	14	10	5	11
	Raped / Not Injured	11	8	3	1	4
	Not raped / Injured	25	9	20	12	19
	Not raped / Not	32	13	15	15	16

Table 5: Frequencies of rapes and injuries in various situations and modes of resistance for 247 women who were sexually assaulted by strangers.

However, the more important information, not immediately evident from the rose plots, appears in the odds ratio plots shown in the bottom panels of Figure 5. These plots clearly show that the risk of injury without rape was relatively high if the woman screamed or tried to fight or flee, and that the risk of rape and injury was higher if the woman tried to dissuade her assailant.

Next we consider some ordinal data analyzed in various different ways by Agresti [19]. These data, listed in Table 6, refer to the severity of side effects (none, slight, or moderate) resulting from four operations of increasing complexity (A, B, C, and D) for treating patients with duodenal ulcers at four hospitals. In his 1974 paper, Clayton also proposed a method for computing a single odds ratio to measure the association between two ordinal variables after adjusting for the effect of a stratification variable, and gave an asymptotic formula for the standard error of its logarithm. For the data in Table 6, this method gives $w_{ord} = 1.578$, with SE (ln w_{ord}) = 0.173 (p-value 0.0082), indicating a

substantial association between the severity of side effects and the complexity of the operation. The corresponding chi-squared goodness-of-fit statistic is 16.48 with 23 degrees of freedom, so the fit is very good.

Figure 6 shows an odds ratio plot of the association between the side effects severity and operation after adjusting for hospital, with odds ratios based on the ordinal odds ratio.
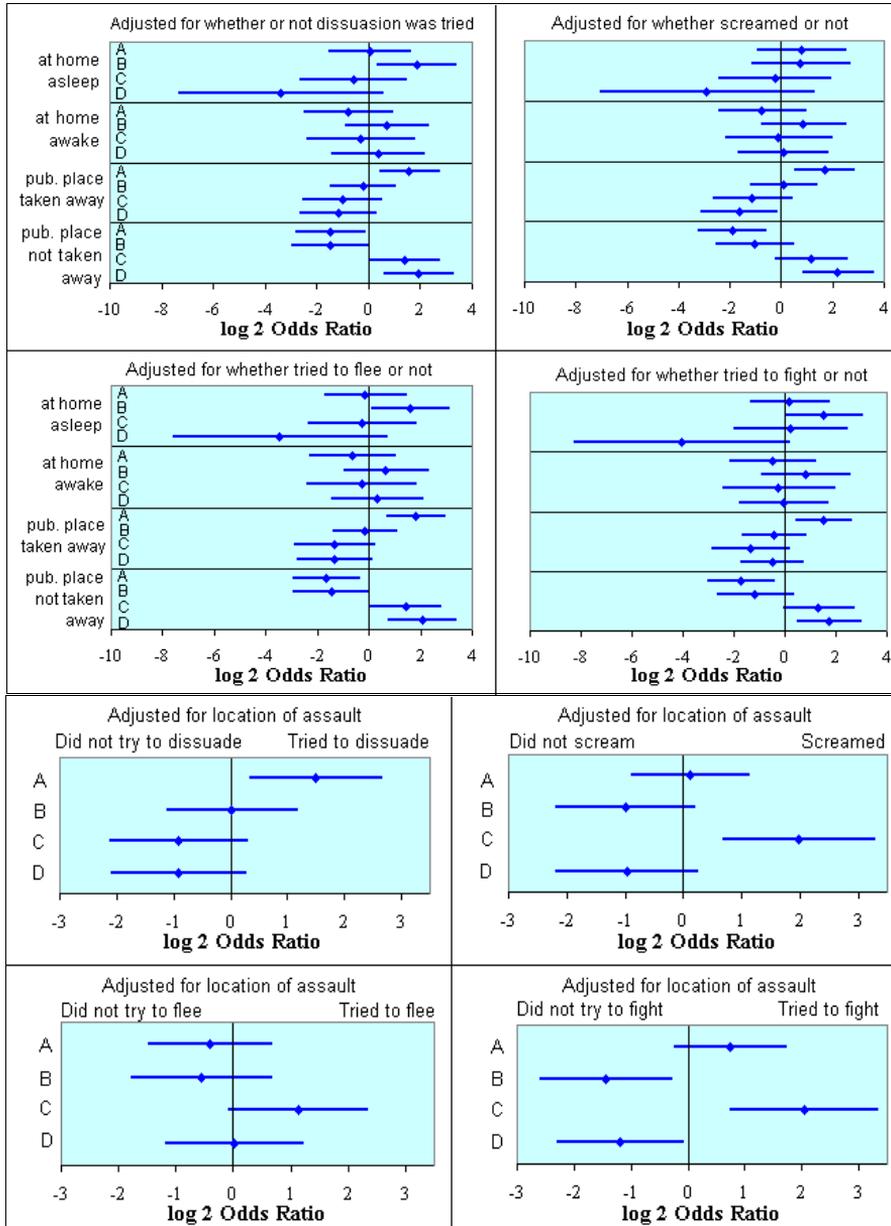
Figure 5: Odds ratio plots showing associations between assault outcome (A = raped & injured, B = raped & not injured, C = not raped & injured, D = not raped & not injured) and (a) location, adjusted for each resistance mode (top panels), and (b) whether or not a specified mode of resistance was used, adjusted for location (bottom panels)

For these data Birch's formula (16) gives X = 10.60 with 6 degrees of freedom (p = 0.10), and the corresponding homogeneity test statistic is 12.73 with 18 degrees of freedom, indicating little evidence of an overall association if the data are regarded as nominal rather than ordinal.

	Hospital 1			Hospital 2			Hospital 3			Hospital 4		
	Severity of side effects											
Operation	N	S	M	N	S	M	N	S	M	N	S	M
A	23	7	2	18	6	1	8	6	3	12	9	1
B	23	10	5	18	6	2	12	4	4	15	3	2
C	20	13	5	13	13	2	11	6	2	14	8	3
D	24	10	6	9	15	2	7	7	4	13	6	4

Table 6: Cross-classification of severity of side effects by operation and hospital for 417 patients having surgery to treat duodenal ulcers.
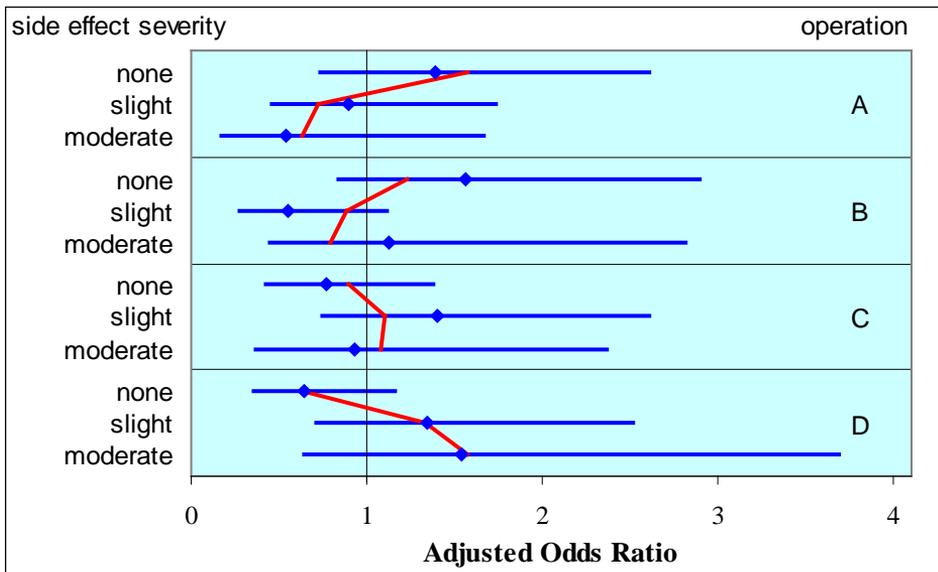

Figure 6: Odds ratio plot showing hospital-adjusted associations between side effect severity and operation, together with fitted odds ratios based on ordinal model.

The method can be used to effectively display associations between several variables. As an illustration consider some data from a 2×2×3×3×2×2 contingency table based on 5419 industrial workers cited by Andrews and Herzberg [20] and earlier analyzed by Higgins and Koch [21] and Yates [22]. The outcome variable was presence or absence of the lung disease byssinosis and the other variables were smoking history (binary), duration of employment (<10 years, 10-19 years, 20+ years), dustiness of workplace (most, less, least), gender, and race (white or non-white). Figure 7 shows an odds ratio plot of the associations between gender/race and disease presence or absence, before and after adjusting for the stratification variable obtained by taking all 18 combinations of the three exposure variables smoking, employment duration, and workplace dustiness. There is no evidence that gender and race were risk factors, although it would have appeared so if the exposure variables were not taken into account.

Figure 8 shows an odds ratio plot of the associations between the disease outcome and the exposure variables after adjusting for gender and race. The plot clearly shows that the risks of disease were greatest for smokers who had worked the longest in the dustiest workplaces.
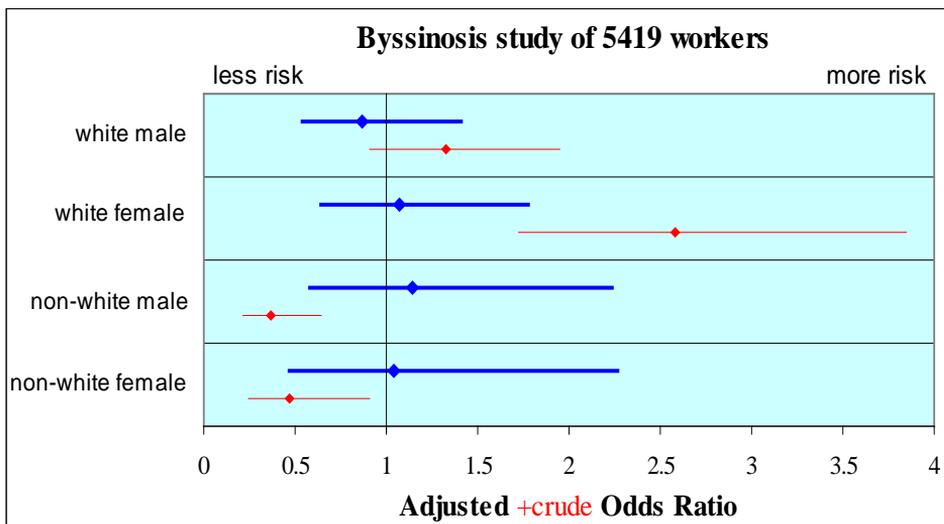

Figure 7: Odds ratio plot of byssinosis disease risk by gender and race after adjusting for exposure variables (thicker lines), with crude odds ratios (thinner lines).

Figure 8: Odds ratio plot of byssinosis disease risk by smoking/employment duration/workplace dustiness after adjusting for gender/race.

7. Summary

We have shown how a graph of simultaneous confidence intervals can usefully display contrasts between means as well as odds ratios used to measure associations in contingency tables. In each case there is a natural null hypothesis: equality of population means, or unit odds ratios. If no specific alternative hypothesis is specified, the graph displays the confidence intervals for the contrasts between each outcome category and the remaining categories. The probability content of each confidence interval is increased to ensure that the overall probability content is fixed at the conventional 95% level.

The graph can be extended to more complex situations. If there is a control group, it is more appropriate to graph confidence intervals for the contrasts between each outcome category and the control. If there is a covariate, the effect of confounding can be shown by graphically comparing the confidence intervals before and after adjusting for the covariate. If the covariate is categorical, a set of confidence intervals corresponding to these categories can be graphed, after adjusting for multiplicity.

The graph is useful for preliminary data analysis, but can also be used to investigate the appropriateness of a fitted model, by including the effects corresponding to the model on the graph. A simple example of this arises when the data are ordinal, but more complex models such as multiple logistic regression and those considered by Goodman [23,24] could be graphed in this way. Using the concept of repeating small multiples [25], highly multidimensional contingency tables can be visualized concisely in various ways. For example, with six or more variables, an odds ratio plot matrix could show a two-dimensional set of plots, where each plot shows the association between two variables after adjusting for the all the remaining variables.

References

[1]   Hand, D.J., F. Daly, A.D. Lunn, K.J. McConway & E. Ostrowski, A Handbook of Small Data Sets. London: Chapman & Hall, 1994.

[2]   Kramer, C.Y., "Extension of multiple range test to group means with unequal numbers of replications," Biometrics, 1956; 12: 307-310.

[3]   Breslow, N.E. & N.E. Day., Statistical Methods in Cancer Research: Volume I - The Analysis of Case-Control Studies. Lyon: IARC, 1980.

[4]   Sutton, A.J., Abrams, K.R., and Jones, D.R., Methods for Meta-Analysis in Medical Research. New York: John Wiley & Sons, 2000.

[5]   Mantel, N., & W. Haenszel, "Statistical aspects of the analysis of data from retrospective studies of disease," Journal of the National Cancer Institute, 1959; 22: 719-748.

[6]   Robins, J., N. Breslow, & S. Greenland, "Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models," Biometrics, 1986; 42: 311-323.

[7]   Freedman, D., R. Pisani, & R. Purves, Statistics. New York: W.H. Norton & Co, 1978.

[8]   Agresti, A., An Introduction to Categorical Data Analysis. New York: John Wiley, 1996.

[9]   Bithell, J.F., "Estimation of relative risk functions," Statistics in Medicine, 1991; 10: 1745-51.

[10]   Easton, D.F., J. Peto & G.A.G. Babiker, "Floating absolute risk: an alternative to relative risk in survival and case-control analysis avoiding an arbitrary reference group," Statistics in Medicine, 1991; 10: 1025-1035.

[11]   Cleveland, W.S., The Elements of Graphing Data (revised ed). AT&T Bell Laboratories. Murray Hill, NJ, 1994.

[12]   Snee, R., "Graphical display of two-way contingency tables," The American Statistician, 1974; 28(1): 9-12.

[13]   Clayton, D.G., "Some odds ratio statistics for the analysis of ordered categorical data," Biometrika, 1974; 61: 525-531.

[14]   Armitage, P., Statistical Methods in Medical Research. Oxford: Blackwell Scientific Publications, 1971.

[15]   Birch, M.W., "The detection of partial association II: The general case," Journal of the Royal Statistical Society (B), 1965; 27: 111-124.

[16]   Kalbfleisch, J., and Prentice, R.L., The Statistical Analysis of Failure Time Data. New York: John Wiley & Sons, 1980.

[17]   Fienberg, S.E., "Perspective Canada as a Social Report," Social Indicators Research, 1975; 2: 153-174.

[18]   Wainer, H., "A rose by another name," Chance, 1995; 8(2): 46-51.

[19]   Agresti, A., Categorical Data Analysis. New York: John Wiley, 1990.

[20]   Andrews, D.F., and Herzberg, A.M., Data: A Collection of Problems from Many Fields *for the Student and Research Worker*. New York: Springer-Verlag, 1985.

[21]   Higgins, J.E., and Koch, G.G., "Variable selection and generalized chi-square analysis of categorical data applied to a large cross-sectional occupational health survey," International Statistical Review, 1977; 45: 51-62.

[22]   Yates, F., Sampling Methods for Censuses and Surveys (4th ed). London: Griffin, 1981.

[23]   Goodman, L.A., "The analysis of cross-classified data: independence, quasi-independence, and interactions in contingency tables with or without missing entries," Journal of the American Statistical Association, 1968; 63: 1091-1131.

[24]   Goodman, L.A., "The multivariate analysis of qualitative data: interaction among multiple classifications," Journal of the American Statistical Association, 1970; 65: 226-256.

[25]   Tufte, E.R., The Visual Display of Quantitative Information. Cheshire, Connecticut: Graphics Press, 1983.