2 X 2 Contingency Tables

jdt

12/9/2020

Contents

In this document we will consider a very simple case where we have two discrete binary variables X and Y. Several examples will be considered.

Theory - 2x2 Table

In this section, we consider a special case of the general $r \times c$ tables where both X and Y are binary random variables ($r = 2$ and $c = 2$). In addition, we will consider the circumstance when one of the categorical variables, say X, is not random (e.g., X specifies gender). In this situation, the conditional probability of category $Y = j$ given $X = i$, $\pi_{j|i}$, is the parameter of interest. Consider the 2×2 table where one is interested in comparing $\Pr[Y = 1 \mid X = 1] = \pi_1 = \pi_{1|1}$ and $\Pr[Y = 1 \mid X = 2] = \pi_2 = \pi_{1|2}$ when $Y = 1$ is an event of interest The response variable Y is statistically independent of the row classification, X, when $\pi_1 - \pi_2 = 0$. This concept of computing this difference works well when $r = 2$ but doesn't when $r > 2$. In which case, the following ratios are commonly used.

Relative Risk

The relative risk is

$$RR = \frac{\Pr[[Y = 1 \mid X = 1]}{\Pr[Y = 1 \mid X = 2]} = \frac{\pi_1}{\pi_2}. \tag{1}$$

Odds Ratio

The odds of an success is

$$\text{Odds of success} = \frac{\pi}{(1-\pi)}. \tag{2}$$

The odds ratio of an success for the two rows defined by X is

$$OR = \theta = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} \tag{3}$$

when X is not random. The odds ratio is

$$OR = \theta = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}. \tag{4}$$

when X and Y are random, Note: In the conditional case, independence implies that RR and $OR = 1$.

Inference for 2 $\times$ 2 Tables

Suppose one has the following table where the row variable X is the random assignment of subject to either the control (placebo) or treatment (active) groups and Y denotes whether or not there is a "favorable (f)" or "unfavorable (u)" outcome.

	f	u	Total
active	n_{11}	n_{12}	n_{1+}
placebo	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	n

If one was interested in testing the null hypothesis that there is no association between the treatment and the outcome of the treatment and the marginal totals are fixed then it follows that,

$$\Pr[n_{ij}] = \frac{n_{1+}!\ n_{2+}!\ n_{+1}!\ n_{+2}!}{n!\ n_{11}!\ n_{12}!\ n_{21}!\ n_{22}!},$$

and

$$E(n_{ij} \mid H_0) = \frac{n_{i+}\ n_{+j}}{n} = m_{ij}, \quad V(n_{ij} \mid H_0) = \frac{n_{1+}\ n_{2+}\ n_{+1}\ n_{+2}}{n^2(n-1)} = v_{ij}.$$

When the total sample size n is sufficiently large, n_{11} is a sufficient statistic and has an approximate normal distribution from which one has,

$$Q = \frac{(n_{11} - m_{11})^2}{v_{11}} \sim \chi^2(df = 1),$$

and

$$Q_p = \sum_{i=1}^{2}\sum_{j=1}^{2}(n_{ij} - m_{ij})^2/m_{ij} = \frac{n}{n-1}\ Q.$$

It can be shown that the Pearson correlation coefficient, $\hat{\rho}$ is related to Q_p by

$$\hat{\rho} = [n_{1+}\ n_{2+}/n_{+1}\ n_{+2}]^{1/2}\ (\hat{\pi}_1 - \hat{\pi}_2) = \sqrt{Q_p/n}.$$

Inference for Difference in Proportions

Suppose one wants to test the hypothesis that the probability of a favorable outcome given the active treatment, Pr[f | active] $= \pi_{1|1} = \pi_{11}/\pi_{1+} = \pi_1$, is the same as the probability of having a favorable outcome using the placebo, Pr[f | placebo] $= \pi_{1|2} = \pi_{21}/\pi_{2+} = \pi_2$. This hypothesis is denoted as $H_0 : \pi_1 = \pi_2$. Define $\hat{\pi}_1 = n_{11}/n_{1+}$ and $\hat{\pi}_2 = n_{21}/n_{2+}$ in which case it follows that $E[\hat{\pi}_1 - \hat{\pi}_2] = \pi_1 - \pi_2$ and $Var[\hat{\pi}_1 - \hat{\pi}_2] = \pi_1(1-\pi_1)/n_{1+} + \pi_2(1-\pi_2)/n_{2+}$. Using an unbiased estimate of $Var[\hat{\pi}_1 - \hat{\pi}_2]$ given by

$$v_d = \frac{\hat{\pi}_1(1-\hat{\pi}_1)}{(n_{1+}-1)} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{(n_{2+}-1)}$$

allows one to define a 100(1-α)% confidence interval for $(\pi_1 - \pi_2)$ as

$$(\hat{\pi}_1 - \hat{\pi}_2) \pm \{z_{\alpha/2}\ \sqrt{v_d}\}$$

or

$$(\hat{\pi}_1 - \hat{\pi}_2) \pm \{z_{\alpha/2}\ \sqrt{v_d} + [1/2(1/n_{1+} + 1/n_{2+})]\}.$$

Inference for Odds Ratio and Relative Risk

Relative Risk

From equation (1) the relative risk is given as $RR = \pi_1/\pi_2$ where $\Pr[Y = 1 \mid X = 1] = \pi_1 = \pi_{1|1}$ and $\Pr[Y = 1 \mid X = 2] = \pi_2 = \pi_{1|2}$. An estimate for the relative risk is, $\widehat{rr} = \hat{\pi}_1/\hat{\pi}_2$. The asymptotic properties for the log of the relative risk are easier to derive than for the relative risk, in which case, the estimated standard error for the relative risk, $\log(rr)$, is

$$\hat{\sigma}_{log(rr)} = [\frac{(1-\hat{\pi}_1)}{\hat{\pi}_1\ n_{1+}} + \frac{(1-\hat{\pi}_2)}{\hat{\pi}_2\ n_{2+}}]^{1/2}.$$

The Wald confidence interval is

$$log(rr) \pm z_{\alpha/2}\ \hat{\sigma}_{log(rr)}.$$

Odds Ratio

From equation (4) the odds ratio is $OR = \theta = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}$. An estimate for the odds ratio is

$$\widehat{OR} = \hat{\theta} = \frac{n_{11}/n_{12}}{n_{21}/n_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}}.$$

Note, since $\theta = \infty$ if either n_{12} or n_{21} equal zero [this can happen with positive probability]. An alternative estimate for the odds ratio is given by

$$\tilde{\theta} = \frac{(n_{11}+0.5)(n_{22}+0.5)}{(n_{12}+0.5)(n_{21}+0.5)}.$$

$\hat{\theta}$ and $\tilde{\theta}$ have the same asymptotic distribution but neither are well behaved for small n. As in the relative risk, the log of the odds ratio has better asymptotic properties. The estimated standard error for the log odds ratio is

$$\hat{\sigma}_{log(\hat{\theta})} = [\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}]^{1/2}.$$

The Wald confidence interval for $log(\theta)$ is

$$log(\hat{\theta}) \pm z_{\alpha/2}\ \hat{\sigma}_{log(\hat{\theta})}.$$

Note: the computation and asymptotic normality of the log odds and log relative risk follow from the Delta method [1]

[1] Agresti (edition 2) pages 73-77.

Example 1 - Asprin use and heart attacks (myocardial infaraction - MI)

R

```
MI <- matrix(c(189, 104, 10845, 10933), nrow = 2)
dimnames(MI) <- list("Group" = c("Placebo","Aspirin"), "MI" = c("Yes","No"))
MI
```

```
##          MI
## Group     Yes    No
##   Placebo 189 10845
##   Aspirin 104 10933
```

Complete the table with marginal totals and cell probabilities

```
addmargins(MI)
```

```
##          MI
## Group     Yes    No   Sum
##   Placebo 189 10845 11034
##   Aspirin 104 10933 11037
##   Sum     293 21778 22071
```

```
prop.table(MI, 1)
```

```
##          MI
## Group             Yes        No
##   Placebo 0.01712887 0.9828711
##   Aspirin 0.00942285 0.9905771
```

From the table, we estimate the probability of having MI while taking asprin is 0.0094 whereas the probability of having MI when not taking the asprin is 0.017. Neither of these probabilities are large but can we determine if they are statistical different. A board test would be to determine if the two variables Group and MI are associated using the Pearson chi-square type goodness-of-fit approach.

```
chisq.test(MI)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  MI
## X-squared = 24.429, df = 1, p-value = 7.71e-07
```

Since the p-value is very small, one can reject the hyporthesis that the two variables are independent upon one another. Let's consider the problem of testing to see if the two probabilities differ from zero.

```
prop.test(MI)
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  MI
## X-squared = 24.429, df = 1, p-value = 7.71e-07
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.004597134 0.010814914
## sample estimates:
```

```
##     prop 1     prop 2
## 0.01712887 0.00942285
```

```
p.out=prop.test(MI)
# difference in proportions
p.out$estimate[1] - p.out$estimate[2]
```

```
##      prop 1
## 0.007706024
```

In this case a better statistic is the relative risk given below

```
prop.out = prop.table(MI, margin = 1)
# relative risk of placebo vs. aspirin
prop.out[1,1]/prop.out[2,1]
```

```
## [1] 1.817802
```

The relative risk is 1.817 which means that those taking the placebo have about a 82% greater liklehood of having MI when compared with those taking the asprin.

Another statistic that is commonly used is the Odds Ratio.

Odds Ratio

```
library(epitools)
oddsratio.fisher(MI)
```

```
## $data
##          MI
## Group     Yes    No Total
##   Placebo 189 10845 11034
##   Aspirin 104 10933 11037
##   Total   293 21778 22071
##
## $measure
##          odds ratio with 95% C.I.
## Group     estimate    lower    upper
##   Placebo 1.000000       NA       NA
##   Aspirin 1.831993 1.432396 2.353927
##
## $p.value
##          two-sided
## Group       midp.exact fisher.exact   chi.square
##   Placebo           NA           NA           NA
##   Aspirin 4.989646e-07 5.032836e-07 5.691897e-07
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Conditional MLE & exact CI from 'fisher.test'"
```

```
oddsratio.wald(MI)    #large sample size procedure
```

```
## $data
##          MI
## Group     Yes    No Total
##   Placebo 189 10845 11034
```

```
##   Aspirin 104 10933 11037
##   Total   293 21778 22071
##
## $measure
##          odds ratio with 95% C.I.
## Group     estimate    lower   upper
##   Placebo 1.000000       NA      NA
##   Aspirin 1.832054 1.440042 2.33078
##
## $p.value
##          two-sided
## Group       midp.exact fisher.exact   chi.square
##   Placebo           NA           NA           NA
##   Aspirin 4.989646e-07 5.032836e-07 5.691897e-07
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

```
riskratio(MI)
```

```
## $data
##          MI
## Group     Yes    No Total
##   Placebo 189 10845 11034
##   Aspirin 104 10933 11037
##   Total   293 21778 22071
##
## $measure
##          risk ratio with 95% C.I.
## Group     estimate    lower    upper
##   Placebo  1.00000       NA       NA
##   Aspirin  1.00784 1.004759 1.010931
##
## $p.value
##          two-sided
## Group       midp.exact fisher.exact   chi.square
##   Placebo           NA           NA           NA
##   Aspirin 4.989646e-07 5.032836e-07 5.691897e-07
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

```
riskratio.wald(MI)    #large sample size procedure
```

```
## $data
##          MI
## Group     Yes    No Total
##   Placebo 189 10845 11034
##   Aspirin 104 10933 11037
```

```
##   Total   293 21778 22071
##
## $measure
##          risk ratio with 95% C.I.
## Group     estimate    lower    upper
##   Placebo  1.00000       NA       NA
##   Aspirin  1.00784 1.004759 1.010931
##
## $p.value
##          two-sided
## Group       midp.exact fisher.exact   chi.square
##   Placebo           NA           NA           NA
##   Aspirin 4.989646e-07 5.032836e-07 5.691897e-07
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

SAS

Code

```
ods graphics on;
title 'Example 1';
title2 'Aspirin use and Myocardinal Infarction';
data aspirin;
 input group $ disease $ count @@;
 datalines;
 placebo yes 189    placebo no 10845
 aspirin yes 104    aspirin no 10933
 ;
 proc freq data = aspirin order=data; weight count;
       tables group*disease / chisq oddsratio relrisk riskdiff nocol nocum;
run;
```

Output

Example 1

Aspirin use and Myocardinal Infarction

The FREQ Procedure

Table of group by disease			
group	disease		
	yes	no	Total
placebo	189 0.86 1.71	10845 49.14 98.29	11034 49.99
aspirin	104 0.47 0.94	10933 49.54 99.06	11037 50.01
Total	293 1.33	21778 98.67	22071 100.00

Statistic	DF	Value	Prob
Chi-Square	1	25.0139	<.0001
Likelihood Ratio Chi-Square	1	25.3720	<.0001
Continuity Adj. Chi-Square	1	24.4291	<.0001
Mantel-Haenszel Chi-Square	1	25.0128	<.0001
Phi Coefficient		0.0337	
Contingency Coefficient		0.0336	
Cramer's V		0.0337	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	189
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	<.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Column 1 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.0171	0.0012	0.0147	0.0195	0.0148	0.0197
Row 2	0.0094	0.0009	0.0076	0.0112	0.0077	0.0114
Total	0.0133	0.0008	0.0118	0.0148	0.0118	0.0149
Difference	0.0077	0.0015	0.0047	0.0107		
Difference is (Row 1 - Row 2)						

Column 2 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.9829	0.0012	0.9805	0.9853	0.9803	0.9852
Row 2	0.9906	0.0009	0.9888	0.9924	0.9886	0.9923
Total	0.9867	0.0008	0.9852	0.9882	0.9851	0.9882
Difference	−0.0077	0.0015	−0.0107	−0.0047		
Difference is (Row 1 - Row 2)						

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.8321	1.4400	2.3308
Relative Risk (Column 1)	1.8178	1.4330	2.3059
Relative Risk (Column 2)	0.9922	0.9892	0.9953

Theory - Diagnostic Tests

Binary or Diagnostic Tests

One of the applications of 2×2 contingency tables is found in the diagnostic testing literature[2]. The material given in this section has been taken from M. S. Pepe's text, "The Statistical Evaluation of Medical Tests for Classification and Prediction". Suppose that a diagnostic test Y is binary where $Y = 1$ if the test indicates a disease and $Y = 0$ if the test indicates the absence of a disease. Let the random variable D indicate the true disease state, that is, $D = 1$ if the subject has the disease and $D = 0$ if the subject does not have the disease. The possible results are given in the classification table,

	D = 0	D = 1
Y = 0	True negative (TN)	False negative (FN)
Y = 1	False positive (FP)	True positive (TP)

A test can produce errors of two types:

$$\text{False positive fraction} = \text{FPF} = \Pr[Y = 1 \mid D = 0] \tag{5}$$
$$\text{False negative fraction} = \text{FNF} = \Pr[Y = 0 \mid D = 1]. \tag{6}$$

Additional notation is often given as

$$\begin{aligned} \text{test sensitivity} &= \text{TPF} = \Pr[Y = 1 \mid D = 1] \\ \text{test specificity} &= 1 - \text{FPF} = \Pr[Y = 0 \mid D = 0] \\ \text{disease prevalence} &= \rho = \Pr[D = 1]. \end{aligned}$$

Note: an ideal test would have TPF = 1 and FPF = 0 whereas a worthless test would have TPF = FPF that is, $\Pr[Y = 1 \mid D = 1] = \Pr[Y = 1 \mid D = 0]$. For this reason one can plot the pair (FPF, TPF) on the usual (x, y) axis. Since these values are probabilities, the pair is constrained to lie in the box with vertexes $(0, 0), (0, 1), (1, 0), (1, 1)$ with the ideal test lying on the point $(0, 1)$ and the point for the worthless test lying on the diagonal line connecting $(0, 0)$ with $(1, 1)$.

The probability of misclassification, given by

$$\Pr[Y \neq D] = \rho \text{ (1 - TPF)} + (1 - \rho) \text{ (FPF)}$$

is highly dependent upon the disease prevalence ρ.

Predictive Values

A commonly used probability for evaluating a test is its predictive probability of a correct decision, given by

$$\text{positive predictive value} = \text{PPV} = \Pr[D = 1 \mid Y = 1] \tag{7}$$
$$\text{negative predictive value} = \text{NPV} = \Pr[D = 0 \mid Y = 0]. \tag{8}$$

A perfect test would have PPV = 1 and NPV = 1, whereas a worthless test would not provide any additional information over what is already known in the population. That is,

$$\text{PPV} = \Pr[D = 1 \mid Y = 1] = \Pr[D = 1] = \rho$$

and

$$\text{NPV} = \Pr[D = 0 \mid Y = 0] = \Pr[D = 0] = (1 - \rho).$$

[2]Although the material found in this section is commonly used when creating or evaluating screening or diagnostic tests, such as pap smears, PSA levels, HIV, mammograms. It has been a topic of great interest since the early months in 2020 with the onset of Covid-19 and the presence of SARS coV-2 virus or the presence of anti-bodies to the infection caused by this pathogen. In fact, we have all been forced to learn and practice critical steps in the control or mitigation of infectious diseases and pandemic outbreaks.

One can derive the following using Bayes formula when the probability of a positive test is given by $\tau = \Pr[Y = 1]$:

$$\begin{aligned} \tau &= \rho \text{ TPF} + (1-\rho) \text{ FPF} \\ \text{PPV} &= \rho \text{ TPF}/[\rho \text{ TPF} + (1-\rho) \text{ FPF}] = \rho \text{ TPF}/\tau \\ \text{NPV} &= (1-\rho) \text{ (1 - FPF)}/[(1-\rho) \text{ (1 - FPF)} + \rho \text{ (1 - TPF)}] \end{aligned}$$

and

$$\begin{aligned} \text{TPF} &= \tau \text{ PPV}/[\tau \text{ PPV} + (1-\tau) \text{ (1 - NPV)}] \\ \text{FPF} &= \tau \text{ (1 - PPV)}/[(\tau \text{ (1 - PPV)} + (1-\tau) \text{ NPV}] \\ \rho &= \tau \text{ PPV} + (1-\tau) \text{ (1 - NPV)}. \end{aligned}$$

Example

Consider the example where the probabilities are assumed to be known.

	D = 0	D = 1	
Y = 0	.223	.142	.365
Y = 1	.078	.556	.634
	.301	.698	1.00

From which one has

$$\begin{aligned} \text{TPF} &= 0.797, \quad \text{FPF} = 0.259, \quad \rho = 0.698 \\ \text{PPV} &= 0.877, \quad \text{NPV} = 0.611, \quad \tau = 0.634. \end{aligned}$$

In the next section, a graphical method for summarizing the above probabilities is given. The curve is called the Receiver Operating Curve (ROC).

Example 2 - Diagnostic Tests

R

A diagnostic test is said to have high accuracy if it achieves a high overall proportion of correct diagnoses. There are actually two aspects of accuracy— namely, the proportion of patients that the diagnostic test correctly identifies as having the disease of interest (the sensitivity of the test) and the proportion of patients that the test correctly identifies as not having the disease (the specificity of the test). The calculation of both assumes that we have a 'gold standard' diagnosis against which to evaluate the performance of our diagnostic test. For example, after patients for whom we have the results of the diagnostic test die, they are examined by a pathologist and given a 'true' diagnosis. In the account that follows, we shall assume that we are assessing how well our diagnostic test predicts this true diagnosis and conveniently ignore the complications that may arise if the diagnosis against which the test is evaluated is itself fallible.

Consider the following table

```
Liver_scan <- matrix(c(231, 27, 32, 54), nrow = 2)
dimnames(Liver_scan) <- list("Test" = c("positive","negative"), "Disease" = c("Yes","No"))
Liver_scan
```

```
##           Disease
## Test       Yes No
##   positive 231 32
##   negative  27 54
```

```
addmargins(Liver_scan)
```

```
##           Disease
## Test       Yes No Sum
##   positive 231 32 263
##   negative  27 54  81
##   Sum      258 86 344
```

```
prob.out = prop.table(Liver_scan, 2)
prob.out
```

```
##           Disease
## Test             Yes       No
##   positive 0.8953488 0.372093
##   negative 0.1046512 0.627907
```

```
sensitivity = prob.out[1,1]
sensitivity
```

```
## [1] 0.8953488
```

```
specificity = prob.out[2,2]
specificity
```

```
## [1] 0.627907
```

The sensitivity and specificity are the characteristics of the test (often determined in a laboratory setting). What we want to determine is what are the operating characteristics of the test. These values are dependent (highly) upon how prevalent the disease is within the population that is being tested. These can be described as

Positive predictive value (PPV) = probability that a patient with a positive liver scan truly has a liver abnormality

Negative predictive value (NPV) = probability that a patient with a negative liver scan does not have liver abnormality

These probabilities can be written as

```
prev = .05
positive_test = sensitivity*prev + (1 - specificity)*(1 - prev)
negative_test = (1-sensitivity)*prev + specificity*(1-prev)
PPV = (sensitivity*prev)/positive_test
NPV = ((1 - specificity)*(1 - prev))/negative_test
#Prevalence = p[Disease]
prev
```

```
## [1] 0.05
```

```
#Positive Test = true positive + false negative
positive_test
```

```
## [1] 0.3982558
```

```
#True positive test = P[D=yes | T=positive]
PPV
```

```
## [1] 0.1124088
```

```
#False positive test
1 - PPV
```

```
## [1] 0.8875912
```

```
#True negative test = P[D=no | T=negative]
NPV
```

```
## [1] 0.5874396
```

```
#False negative test
1 - NPV
```

```
## [1] 0.4125604
```

SAS

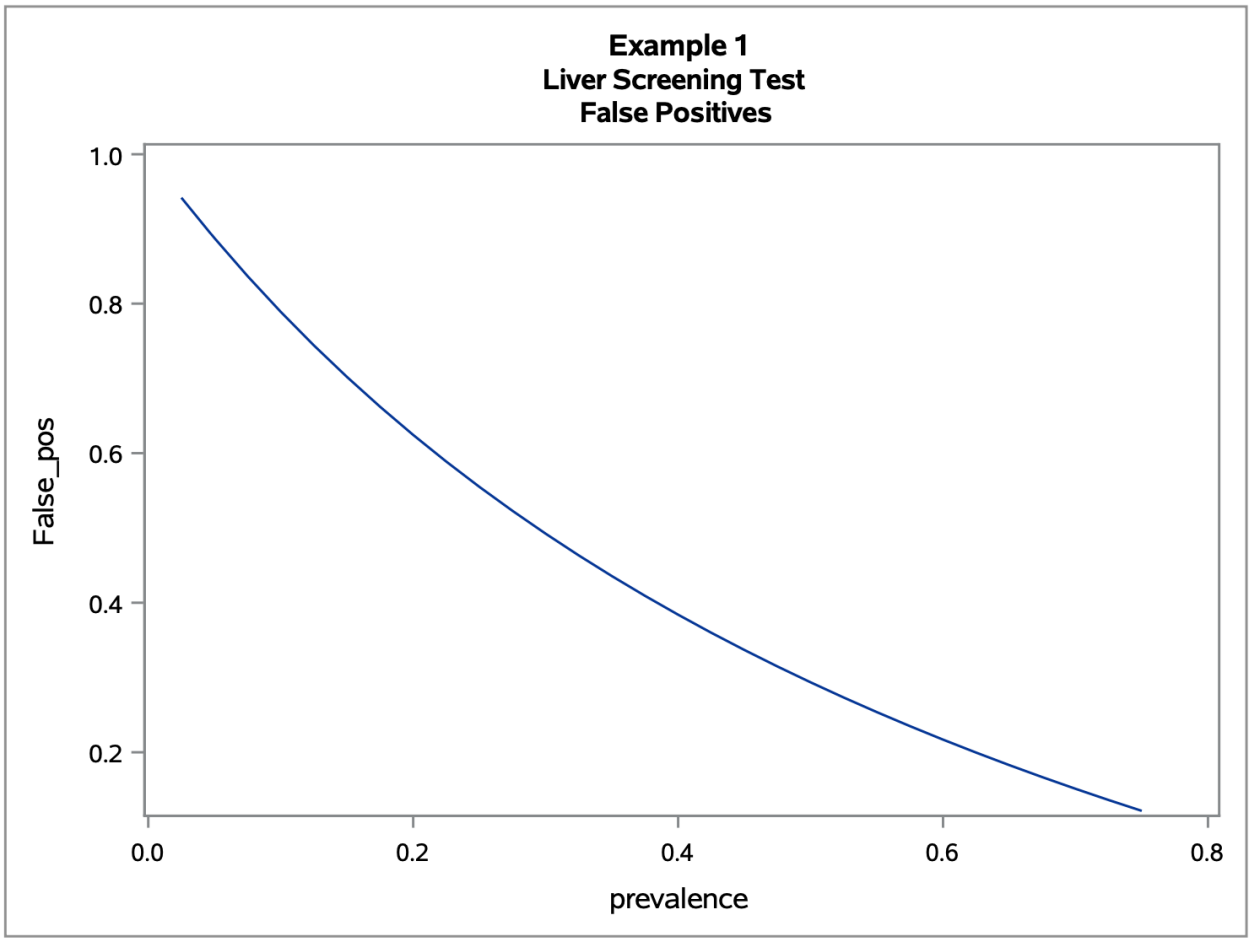
Code

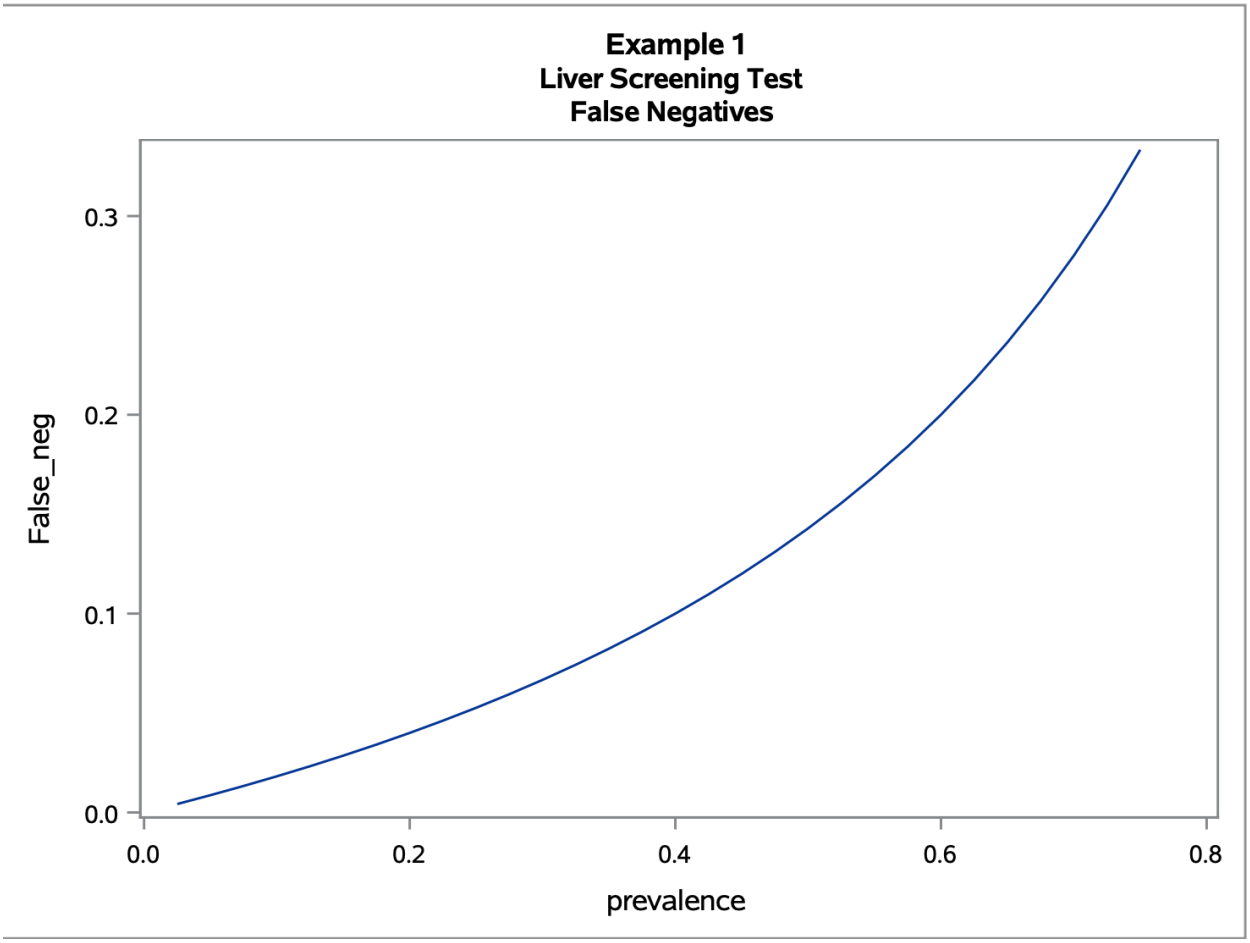
```
title2 'Liver Screening Test';
data liver;
   input test $ Disease $ count @@;
   datalines;
   positive  yes 231      positive no 32
   negative yes  27      negative no 54
;
proc freq data=liver order=data; weight count;
       tables test*disease /  norow nocum nopercent;
run;

proc logistic data=liver noprint;
class test disease;
model disease(event='yes')=test /outroc=rocs; freq count;
run;
data rocs;
set rocs;
sensitivity=_sensit_;
specificity=1-_1mspec_;

   do i = .025 to .75 by .025;
prevalence=i;
PPV=(sensitivity*prevalence)/((sensitivity*prevalence) +
    (1-specificity)*(1-prevalence));
NPV=(specificity*(1-prevalence)) / ((1-sensitivity)*prevalence +
    specificity*(1-prevalence));
False_pos = 1 - PPV;
False_neg = 1 - NPV;
miss_class = prevalence*(1 - sensitivity) + ( 1 - specificity)*(1 - prevalence);
   output;
   end;
drop _sensit_;
run;
data rocs;set rocs; if specificity gt 0;run;
proc print data=rocs label;
var prevalence sensitivity specificity PPV false_pos NPV false_neg miss_class;
format PPV false_pos NPV false_neg miss_class 4.2;
run;

title3 'False Positives';
proc sgplot data=rocs;
series y=False_pos x=prevalence;
run;
```

```
title3 'False Negatives';
proc sgplot data=rocs;
series y=False_neg x=prevalence;
run;

title3 'Miss Classification';
proc sgplot data=rocs;
series y=miss_class x=prevalence;
run;
```

Output

Obs	prevalence	sensitivity	specificity	PPV	False_pos	NPV	False_neg	miss_class
1	0.025	0.89535	0.62791	0.06	0.94	1.00	0.00	0.37
2	0.050	0.89535	0.62791	0.11	0.89	0.99	0.01	0.36
3	0.075	0.89535	0.62791	0.16	0.84	0.99	0.01	0.35
4	0.100	0.89535	0.62791	0.21	0.79	0.98	0.02	0.35
5	0.125	0.89535	0.62791	0.26	0.74	0.98	0.02	0.34
6	0.150	0.89535	0.62791	0.30	0.70	0.97	0.03	0.33
7	0.175	0.89535	0.62791	0.34	0.66	0.97	0.03	0.33
8	0.200	0.89535	0.62791	0.38	0.62	0.96	0.04	0.32
9	0.225	0.89535	0.62791	0.41	0.59	0.95	0.05	0.31
10	0.250	0.89535	0.62791	0.45	0.55	0.95	0.05	0.31
11	0.275	0.89535	0.62791	0.48	0.52	0.94	0.06	0.30
12	0.300	0.89535	0.62791	0.51	0.49	0.93	0.07	0.29
13	0.325	0.89535	0.62791	0.54	0.46	0.93	0.07	0.29
14	0.350	0.89535	0.62791	0.56	0.44	0.92	0.08	0.28
15	0.375	0.89535	0.62791	0.59	0.41	0.91	0.09	0.27
16	0.400	0.89535	0.62791	0.62	0.38	0.90	0.10	0.27
17	0.425	0.89535	0.62791	0.64	0.36	0.89	0.11	0.26
18	0.450	0.89535	0.62791	0.66	0.34	0.88	0.12	0.25
19	0.475	0.89535	0.62791	0.69	0.31	0.87	0.13	0.25
20	0.500	0.89535	0.62791	0.71	0.29	0.86	0.14	0.24
21	0.525	0.89535	0.62791	0.73	0.27	0.84	0.16	0.23
22	0.550	0.89535	0.62791	0.75	0.25	0.83	0.17	0.22
23	0.575	0.89535	0.62791	0.77	0.23	0.82	0.18	0.22
24	0.600	0.89535	0.62791	0.78	0.22	0.80	0.20	0.21
25	0.625	0.89535	0.62791	0.80	0.20	0.78	0.22	0.20
26	0.650	0.89535	0.62791	0.82	0.18	0.76	0.24	0.20
27	0.675	0.89535	0.62791	0.83	0.17	0.74	0.26	0.19
28	0.700	0.89535	0.62791	0.85	0.15	0.72	0.28	0.18
29	0.725	0.89535	0.62791	0.86	0.14	0.69	0.31	0.18
30	0.750	0.89535	0.62791	0.88	0.12	0.67	0.33	0.17


Example 1
Liver Screening Test
False Positives
1.0
0.8
0.6
0.4
0.2
False_pos
0.0
0.2
0.4
0.6
0.8
prevalence


Example 1
Liver Screening Test
False Negatives
False_neg
0.0
0.1
0.2
0.3
0.0
0.2
0.4
0.6
0.8
prevalence

Theory – Multiple 2 x 2 Tables

Mantel-Haenszel Test

Suppose that one has q independent 2×2 tables where one assumes that the marginal sums are fixed. These assumptions insure that one has a hypergeometric distribution. It follows that,

$$E(n_{hij} \mid H_0) = \frac{n_{hi+}\ n_{h+j}}{n_h} = m_{hij} \text{ and } V(n_{hij} \mid H_0) = \frac{n_{h1+}\ n_{h2+}\ n_{h+1}\ n_{h+2}}{n_h^2(n_h - 1)} = v_{hij}.$$

From which the Mantel-Haenszel test statistic is given by

$$Q_{MH} = \frac{[\sum_{h=1}^{q} n_{h11} - \sum_{h=1}^{q} m_{h11}]^2}{\sum_{h=1}^{q} v_{h11}}.$$

Homogeneity of Odds Ratios

The Breslow-Day statistic is given by

$$Q_{BD} = \sum_{h}^{q} \sum_{i}^{2} \sum_{j}^{2} \frac{(n_{hij} - m_{hij})^2}{m_{hij}}, \tag{9}$$

which has a asymptotic chi-square distribution with $q - 1$ degrees of freedom.

The SAS USER's guide has the following concerning the Breslow-Day procedure.

Breslow-Day Test for Homogeneity of the Odds Ratios

When you specify the CMH option, PROC FREQ computes the Breslow-Day test for stratified analysis of 2×2 tables. It tests the null hypothesis that the odds ratios for the q strata are all equal. When the null hypothesis is true, the statistic has approximately a chi-square distribution with $q - 1$ degrees of freedom. Refer to Breslow and Day (1980) and Agresti (1996).

The Breslow-Day statistic is computed as

$$Q_{BD} = \sum_{h} \frac{(n_{h11} - E(n_{h11}|OR_{MH}))^2}{var(n_{h11}|OR_{MH})}.$$

> For the Breslow-Day test to be valid, the sample size should be relatively large in each stratum, and at least 80% of the expected cell counts should be greater than 5. Note that this is a stricter sample size requirement than the requirement for the Cochran-Mantel-Haenszel test for $q \times 2 \times 2$ tables, in that each stratum sample size (not just the overall sample size) must be relatively large. Even when the Breslow-Day test is valid, it may not be very powerful against certain alternatives, as discussed in Breslow and Day (1980).

If you specify the BDT option, PROC FREQ computes the Breslow-Day test with Tarone's adjustment, which subtracts an adjustment factor from Q_{BD} to make the resulting statistic asymptotically chi-square.

$$Q_{BDT} = Q_{BD} - \frac{(\sum_h (n_{h11} - E(n_{h11}|OR_{MH})))^2}{\sum_h var(n_{h11}|OR_{MH})}$$

Refer to Tarone (1985), Jones et al. (1989), and Breslow (1996).

Example 3 - Comparing Multiple 2 x 2 Tables

R

It is not uncommon to have extra variables that can separate the original table into multiple tables. Below we consider a simple example comparing two countries preference towards soft drinks. The combined table is

```
soft_drink <- matrix(c(36, 43, 29, 44), nrow = 2)
dimnames(soft_drink) <- list("Country" = c("America","UK"), "Choice" = c("Yes","No"))
soft_drink
```

```
##          Choice
## Country   Yes No
##   America  36 29
##   UK       43 44
```

Complete the table with marginal totals and cell probabilities

```
addmargins(soft_drink)
```

```
##          Choice
## Country   Yes No Sum
##   America  36 29  65
##   UK       43 44  87
##   Sum      79 73 152
```

```
prop.table(soft_drink, 1)
```

```
##          Choice
## Country         Yes        No
##   America 0.5538462 0.4461538
##   UK      0.4942529 0.5057471
```

```
chisq.test(soft_drink)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  soft_drink
## X-squared = 0.3175, df = 1, p-value = 0.5731
```

Consider the test for proportion that favor the soft drink.

```
prop.test(soft_drink)
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  soft_drink
## X-squared = 0.3175, df = 1, p-value = 0.5731
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.1139733  0.2331599
## sample estimates:
##    prop 1    prop 2
## 0.5538462 0.4942529
```

```
p.out=prop.test(soft_drink)
# difference in proportions
p.out$estimate[1] - p.out$estimate[2]
```

```
##     prop 1
## 0.05959328
```

In this case a better statistic is the relative risk given below

```
prop.out = prop.table(soft_drink, margin = 1)
# relative risk of placebo vs. aspirin
prop.out[1,1]/prop.out[2,1]
```

```
## [1] 1.120572
```

The relative risk is 1.12 which means that American have about a 12% greater likelihood of favoring soft drink when compared with those the UK. This difference is not statistically significant.

The above table combined both males and females. Suppose we separate the table by gender and reproduce the results. Install needed library

```
library(epitools)
```

```
#Males
soft_drink_males <- matrix(c(29, 19, 6, 15), nrow = 2)
dimnames(soft_drink_males) <- list("Country" = c("America","UK"), "Choice" = c("Yes","No"))
addmargins(soft_drink_males)
```

```
##          Choice
## Country   Yes No Sum
##   America  29  6  35
##   UK       19 15  34
##   Sum      48 21  69
```

```
prop.table(soft_drink_males, 1)
```

```
##          Choice
## Country         Yes        No
##   America 0.8285714 0.1714286
##   UK      0.5588235 0.4411765
```

```
chisq.test(soft_drink_males)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  soft_drink_males
## X-squared = 4.7216, df = 1, p-value = 0.02979
```

```
#oddsratio.fisher(soft_drink_males)
oddsratio.wald(soft_drink_males)  #large sample size procedure
```

```
## $data
##          Choice
## Country   Yes No Total
##   America  29  6    35
##   UK       19 15    34
##   Total    48 21    69
##
## $measure
##          odds ratio with 95% C.I.
## Country   estimate    lower    upper
```

```
##   America 1.000000       NA       NA
##   UK      3.815789 1.258159 11.57267
##
## $p.value
##          two-sided
## Country   midp.exact fisher.exact chi.square
##   America         NA           NA         NA
##   UK       0.0174903   0.01940439 0.01490867
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

```
#riskratio(soft_drink_males)
riskratio.wald(soft_drink_males)  #large sample size procedure
```

```
## $data
##          Choice
## Country   Yes No Total
##   America  29  6    35
##   UK       19 15    34
##   Total    48 21    69
##
## $measure
##          risk ratio with 95% C.I.
## Country   estimate    lower    upper
##   America 1.000000       NA       NA
##   UK      2.573529 1.132635 5.847474
##
## $p.value
##          two-sided
## Country   midp.exact fisher.exact chi.square
##   America         NA           NA         NA
##   UK       0.0174903   0.01940439 0.01490867
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

```
#Females
soft_drink_females <- matrix(c(7, 24, 23, 29), nrow = 2)
dimnames(soft_drink_females) <- list("Country" = c("America","UK"), "Choice" = c("Yes","No"))
addmargins(soft_drink_females)
```

```
##          Choice
## Country   Yes No Sum
##   America   7 23  30
##   UK       24 29  53
##   Sum      31 52  83
```

```
prop.table(soft_drink_females, 1)
```

```
##          Choice
## Country         Yes        No
##   America 0.2333333 0.7666667
##   UK      0.4528302 0.5471698
```

```
chisq.test(soft_drink_females)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  soft_drink_females
## X-squared = 3.062, df = 1, p-value = 0.08014
```

```
#oddsratio.fisher(soft_drink_males)
oddsratio.wald(soft_drink_females)
```

```
## $data
##          Choice
## Country   Yes No Total
##   America   7 23    30
##   UK       24 29    53
##   Total    31 52    83
##
## $measure
##          odds ratio with 95% C.I.
## Country    estimate     lower    upper
##   America 1.0000000        NA       NA
##   UK      0.3677536 0.1347282 1.003819
##
## $p.value
##          two-sided
## Country   midp.exact fisher.exact chi.square
##   America         NA           NA         NA
##   UK      0.05040856   0.06016876 0.04703122
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

```
#riskratio(soft_drink_males)
riskratio.wald(soft_drink_females)
```

```
## $data
##          Choice
## Country   Yes No Total
##   America   7 23    30
##   UK       24 29    53
##   Total    31 52    83
##
## $measure
##          risk ratio with 95% C.I.
## Country    estimate     lower     upper
##   America 1.0000000        NA        NA
##   UK      0.7136998 0.5210735 0.9775345
```

```
##
## $p.value
##          two-sided
## Country   midp.exact fisher.exact chi.square
##   America         NA           NA         NA
##   UK      0.05040856   0.06016876 0.04703122
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

The Mantel-Haenszel odds ratio estimates the odds ratio for association between country and choice, controlling for the possible confounding effects of the stratifying variable (gender here). Need to install a R package

```
library("lawstat")
```

```
myarray <- array(c(soft_drink_males,soft_drink_females),dim=c(2,2,2))
cmh.test(myarray)
```

```
##
##  Cochran-Mantel-Haenszel Chi-square Test
##
## data:  myarray
## CMH statistic = 0.02428, df = 1.00000, p-value = 0.87617, MH Estimate =
## 1.05388, Pooled Odd Ratio = 1.27025, Odd Ratio of level 1 = 3.81579,
## Odd Ratio of level 2 = 0.36775
```

It appears that there is a gender difference but once this varaible is accounted for, there is not a preference difference between the two countries.

SAS

Code

```
title 'Example 3';
title2 'Soft Drink Choice';
title3 ' ';

data soft;
      input gender $ country $ question $ count @@;
      datalines;
   male   American  y 29 male   American  n  6
   male   British   y 19 male   British   n 15
   female American  y  7 female American  n 23
   female British   y 24 female British   n 29
   ;
proc freq order=data;
      weight count;
          tables country*question/chisq riskdiff nocol nopercent relrisk oddsratio;
title3 'Combined Table';
 run;

proc freq order=data;
      weight count;
      tables gender*country*question /
           chisq cmh nocol nopercent relrisk oddsratio;
```

```
title3 'Tables for each gender';
run;
```

Output

Example 3
Soft Drink Choice
Combined Table
The FREQ Procedure

Table of country by question			
country	question		
	y	n	Total
American	36 55.38	29 44.62	65
British	43 49.43	44 50.57	87
Total	79	73	152

Statistic	DF	Value	Prob
Chi-Square	1	0.5293	0.4669
Likelihood Ratio Chi-Square	1	0.5299	0.4666
Continuity Adj. Chi-Square	1	0.3175	0.5731
Mantel-Haenszel Chi-Square	1	0.5258	0.4684
Phi Coefficient		0.0590	
Contingency Coefficient		0.0589	
Cramer's V		0.0590	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	36
Left-sided Pr <= F	0.8136
Right-sided Pr >= F	0.2867
Table Probability (P)	0.1004
Two-sided Pr <= P	0.5136

Column 1 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.5538	0.0617	0.4330	0.6747	0.4253	0.6773
Row 2	0.4943	0.0536	0.3892	0.5993	0.3853	0.6036

Column 1 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Total	0.5197	0.0405	0.4403	0.5992	0.4373	0.6014
Difference	0.0596	0.0817	−0.1005	0.2197		
Difference is (Row 1 - Row 2)						

Column 2 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.4462	0.0617	0.3253	0.5670	0.3227	0.5747
Row 2	0.5057	0.0536	0.4007	0.6108	0.3964	0.6147
Total	0.4803	0.0405	0.4008	0.5597	0.3986	0.5627
Difference	−0.0596	0.0817	−0.2197	0.1005		
Difference is (Row 1 - Row 2)						

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.2702	0.6666	2.4207
Relative Risk (Column 1)	1.1206	0.8263	1.5196
Relative Risk (Column 2)	0.8822	0.6271	1.2411

Example 3
Soft Drink Choice
Tables for each gender
The FREQ Procedure

Table 1 of country by question			
Controlling for gender=male			
country	question		
	y	n	Total
American	29 82.86	6 17.14	35
British	19 55.88	15 44.12	34
Total	48	21	69

Statistic	DF	Value	Prob
Chi-Square	1	5.9272	0.0149
Likelihood Ratio Chi-Square	1	6.0690	0.0138
Continuity Adj. Chi-Square	1	4.7216	0.0298
Mantel-Haenszel Chi-Square	1	5.8413	0.0157
Phi Coefficient		0.2931	
Contingency Coefficient		0.2813	
Cramer's V		0.2931	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	29
Left-sided Pr <= F	0.9968
Right-sided Pr >= F	0.0143
Table Probability (P)	0.0112
Two-sided Pr <= P	0.0194

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	3.8158	1.2582	11.5727
Relative Risk (Column 1)	1.4827	1.0611	2.0717
Relative Risk (Column 2)	0.3886	0.1710	0.8829

Table 2 of country by question			
Controlling for gender=female			
country	question		
	y	n	Total
American	7 23.33	23 76.67	30
British	24 45.28	29 54.72	53
Total	31	52	83

Statistic	DF	Value	Prob
Chi-Square	1	3.9443	0.0470
Likelihood Ratio Chi-Square	1	4.0934	0.0431
Continuity Adj. Chi-Square	1	3.0620	0.0801
Mantel-Haenszel Chi-Square	1	3.8968	0.0484
Phi Coefficient		−0.2180	
Contingency Coefficient		0.2130	
Cramer's V		−0.2180	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	7
Left-sided Pr <= F	0.0385
Right-sided Pr >= F	0.9881
Table Probability (P)	0.0267
Two-sided Pr <= P	0.0602

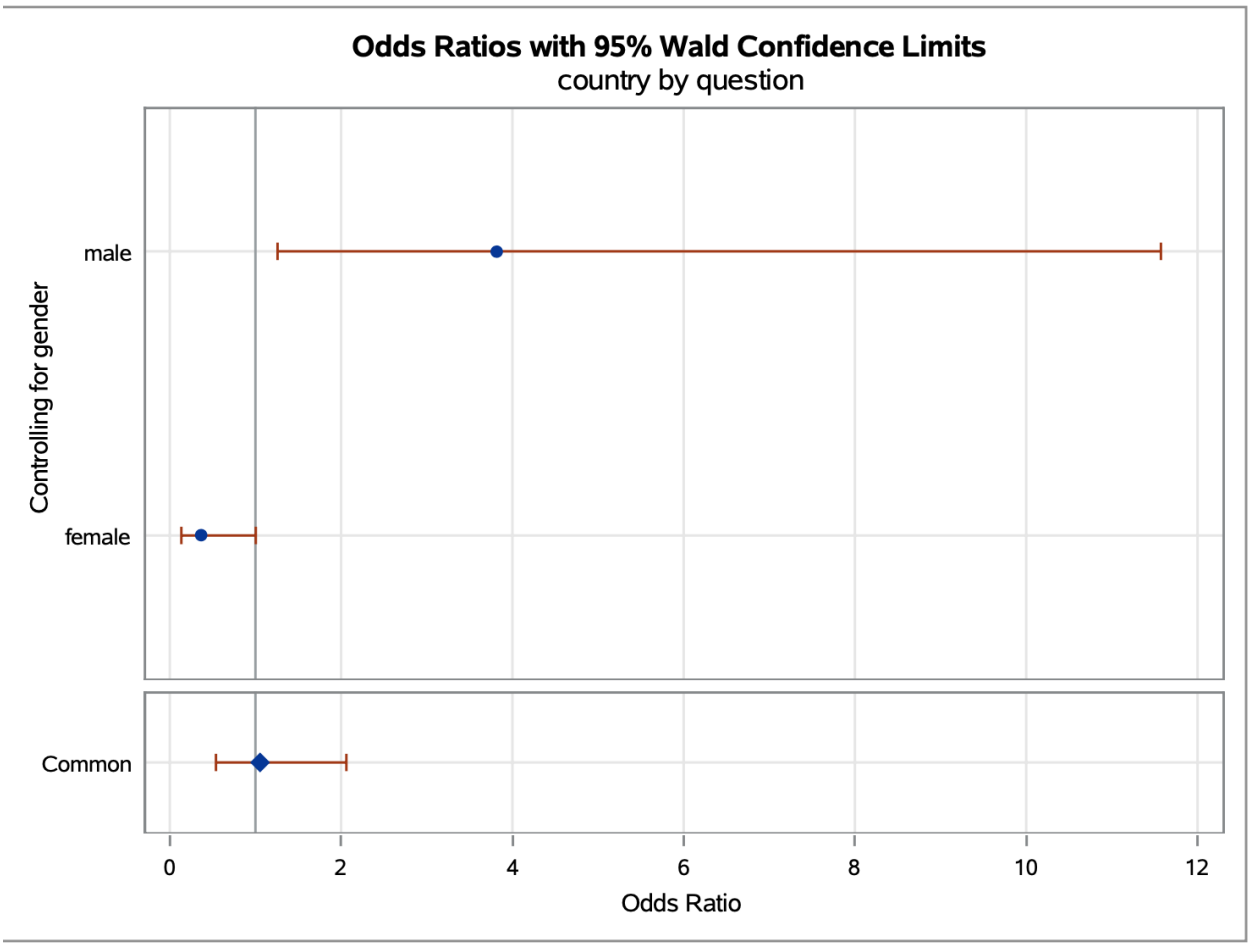
Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	0.3678	0.1347	1.0038
Relative Risk (Column 1)	0.5153	0.2526	1.0512
Relative Risk (Column 2)	1.4011	1.0230	1.9191

Example 3
Soft Drink Choice
Tables for each gender
The FREQ Procedure

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)				
Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0243	0.8762
2	Row Mean Scores Differ	1	0.0243	0.8762
3	General Association	1	0.0243	0.8762

Common Odds Ratio and Relative Risks				
Statistic	Method	Value	95% Confidence Limits	
Odds Ratio	Mantel−Haenszel	1.0539	0.5388	2.0615
	Logit	1.0545	0.5009	2.2202
Relative Risk (Column 1)	Mantel−Haenszel	1.0244	0.7441	1.4104
	Logit	1.2253	0.9051	1.6587
Relative Risk (Column 2)	Mantel−Haenszel	0.9753	0.7163	1.3278
	Logit	1.1889	0.8863	1.5948

Breslow-Day Test for Homogeneity of Odds Ratios	
Chi-Square	9.8324
DF	1
Pr > ChiSq	0.0017


Odds Ratios with 95% Wald Confidence Limits
country by question
Controlling for gender
male
female
Common
0
2
4
6
8
10
12
Odds Ratio