Conference on Applied Statistics in Agriculture 2000 - 12th Annual Conference Proceedings

PERFORMANCE OF THE EXACT AND CHI-SQUARE TESTS ON SPARSE CONTINGENCY TABLES

D. G. Renter

J. J. Higgins

J. M. Sargeant

Recommended Citation

Renter, D. G.; Higgins, J. J.; and Sargeant, J. M. (2000). "PERFORMANCE OF THE EXACT AND CHI-SQUARE TESTS ON SPARSE CONTINGENCY TABLES," *Conference on Applied Statistics in Agriculture*. https://doi.org/10.4148/2475-7772.1253

PERFORMANCE OF THE EXACT & CHI-SQUARE TESTS ON SPARSE CONTINGENCY TABLES

D.G. Renter,[1] J.J. Higgins,[2] J.M. Sargeant[1]

1) Food Animal Health & Management Center, College of Veterinary Medicine, Kansas State University
2) Department of Statistics, College of Arts & Sciences, Kansas State University

ABSTRACT: A cross-sectional observational study design was used to determine the prevalence of *Escherichia coli* O157:H7 in wild deer feces. Samples were voluntarily submitted at a number of different locations. In order to determine if the proportions of *E. coli* O157:H7 positive samples submitted were equal for each of the 26 locations, a 26 by 2 contingency table was analyzed. There were only four *E. coli* O157:H7 positive samples, which resulted in a sparse table. It is possible to obtain statistically significant results in sparse tables using Fisher's exact test, whereas the chi-square test is generally unreliable in such situations. Thus, Fisher's exact test should be considered when small expected cell counts bring into question the validity of the chi-square test. However, the statistical conclusions based on either the exact test or an asymptotic chi-square test are shown to vary drastically by slight alterations in the distribution of non-empty cells. Therefore, a different statistical conclusion very easily could have been reached if a volunteer had submitted a sample at a different location. In addition, we show that the computational times for exact tests in SAS® can be an applicational limitation.

1. INTRODUCTION

In the fall of 1998 deer fecal samples were tested for *Escherichia coli* O157:H7; a bacterial organism capable of causing severe human illness. Volunteer deer hunters in Southeastern Nebraska collected the fecal samples from their harvested deer. The hunters submitted samples at one of the local check stations, which are used by the Nebraska Game and Parks Commission for registering harvested deer. There were only 4 samples positive for *E. coli* O157:H7 of the 1,608 samples. Therefore, the prevalence estimate for *E. coli* O157:H7 in this study population was 0.25 % (Standard Deviation: 0.12%). One hundred and seventy-two of the 1,608 samples submitted could not be identified as to the location of check station submission. Two of the check stations on the periphery of the study area had no sample submissions during the study period. Therefore, 1,436 samples (4 positives) were submitted at 26 check stations. Although the primary objective of the original project was to determine the prevalence of *E. coli* O157:H7 in wild deer, a secondary objective was to determine if the proportions of positive samples submitted were equal for each of the stations. The distribution of positive and negative samples submitted at the 26 stations (numbered by alphabetical order) is shown in Table 1.

To test for a difference in proportions the SAS® PROC FREQ command was used to generate a Fisher's exact test on the data (Figure 1). The corresponding p-value for this test was not significant (0.166) so we concluded that the proportion of positive samples submitted was not significantly different between check stations. The asymptotic chi-square statistic, which would not be appropriate with this sparse data set, was nearly significant (P = 0.062) at the 5% level.

2. PURPOSES AND METHODS

There are three purposes of this study. One is to demonstrate that Fisher's exact test can produce statistically significant results in extremely sparse tables. We would hope that this would encourage agricultural researchers to consider this test when the assumptions of the chi-square test are inappropriate. The second purpose is to investigate the sensitivity of the exact test when the pattern of responses is slightly changed in sparse tables. This is particularly relevant for this context. Hunters were not required to check their deer at any *specific* station so the distribution of positive submissions could have easily been slightly different. The third purpose is to compare p-values for the exact test to p-values for the chi-square test in order to quantify the degree to which the chi-square approximation is inappropriate for sparse table. We conclude with a comment on the computational efficiencies of two statistical packages for conducting exact tests.

In order to investigate how slight changes in submissions would have affected statistical conclusions, a series of sparse 26×2 tables were generated with the same number of positive responses/nonempty cells (4) and negative responses (1432) as the original data. The proportions were changed, but the total number submitted at each station (row totals) were held constant. Each new table was obtained by moving *just one positive response* from the original table. The move was made either to a station with a small number of submissions, or to one with a large number of submissions. The moves created one of three patterns: at most one positive response per station, two positive responses at one station and 1 at two other stations, or two positive responses at two stations. This allowed us to compare the p-values/conclusions for the chi-square and the Fisher's exact tests under different scenarios, and to investigate how the total number of submissions for each station (row total) and the distribution of non-empty cells affect the statistical outcome. The asymptotic p-values for the chi-square test and the p-values for Fisher's exact test were obtained using the SAS® PROC FREQ command. Tables 1A - 3B illustrate 6 example tables and the corresponding analyzes. The scenarios in A tables have the positive sample moved to a station with a small number of submissions and the scenarios displayed in B tables have the positive moved to a station with a large number of submissions. These tables show only rows with non-empty cells or changes from original table (refer to Table 1 for additional rows).

3. CONCLUSIONS

Two of the 6 contingency tables yielded significant results at the 5% level (2A with $p = .037$ and 3A with $p = .0073$), and another (1A with $p = .077$) yielded a significant result at the 10% level. These were the tables in which one positive response was moved to a station with a small number of submissions. In all these cases the p-value decreased over the value of $p = .166$ for the original table. When a move was made to a station with a large number of submissions, the p-values increased substantially over the p-value for the original table (1B with $p = .455$, 2B with $p = .586$, 3B with $p = .347$). In addition, moving a positive response from a station with a small number of submissions to one with a large number of submissions (or vice versa) resulted at times in a drastic change in the statistical outcome (e.g. 2A to 2B). P-values did not seem to be greatly affected by whether or not the response pattern had only one positive per station or more than one positive for some stations. In general, the percent positives per station is a more important factor in determining a significant difference than the absolute number of positives.

These results point to caution when analyzing sparse tables. The fact that changing the submission location of only one of the four positives can lead to a different statistical conclusion, is very noteworthy considering the method of data collection. With observational data such as these, the number of observations (submissions) per location is not under the control of the investigator. Therefore, the distribution of the total observations, as well as the distribution of outcomes can influence the conclusions. As our results show, a single variation in submission when data are sparse could change the marginal totals enough to substantially change significance of the results. In addition using the station location as a proxy variable for the deer location could result in misclassifications, which could limit our power to detect a difference should one exist.

There appears to be no consistent pattern regarding the p-values for the chi-square test in comparison to those of Fisher's exact test. In 5 of the 6 cases considered, the chi-square p-value was smaller than the p-value for Fisher's exact test. In table 1A this difference was enough that a conclusion at the 5% level would be different using the two tests. However, in Table 1B and other situations that we investigated, Fisher's exact test had a smaller p-value than the chi-square test. Thus, we are not able to say whether one test is more or less conservative than the other in sparse tables. However, there is enough difference between p-values to confirm the well-known caution against using the chi-square test with sparse tables.

4. COMPUTATIONS

In the process of investigating different distributions of the data, we became acutely aware of the computational limitations of using exact tests in SAS®. When we attempted to explore scenarios where additional positive or non-empty cells were included in the data, the time to compute an exact p-value increased drastically to the point where comparing several different scenarios became impractical. On the other hand, StatXact® statistical software, which uses fast algorithms to compute exact p-values, performed analyses very quickly. Displayed in Figure 2 is a brief comparison of computational times for exact tests using both SAS® & StatXact®. All analyses were done on a personal computer with a Pentium® lll Processor at 500 MHz; 128MB, 100MHzSDRAM. The times given for SAS® analyses were generated using PC SAS® version 6.12 on the system specified above.

5. SUMMARY

Fisher's exact test is useful for analyzing sparse contingency tables when small expected cell counts may invalidate the use of the chi-square test. However, the researcher must be cautious in interpreting results. In the context of present study, the method of sample collection resulted in the possibility that the distribution of sample submissions could have varied slightly. With very few non-empty cells and unequal row totals in sparse contingency tables such as these, slight variations in the distribution of non-empty cells can have a profound impact on statistical conclusions. The statistical conclusion from the original data table could have been significantly different had just one of the volunteers collecting samples submitted a positive sample at a different location. In addition, we have shown that there are definite computational limitations in analyzing large contingency tables using exact tests in SAS® and suggest StatXact® as a possible alternative for conducting exact tests.

The original study was funded by
United States Department of Agriculture Special Grant #96-34359-2593.

TABLE 1. Original distribution of samples submitted at check stations.

STATION #	# POSITIVE	# NEGATIVE	TOTAL #	% POSITIVE
1	2	55	57	3.51
2	0	15	15	0
3	0	18	18	0
4	0	81	81	0
5	0	152	152	0
6	0	80	80	0
7	1	120	121	0.83
8	0	57	57	0
9	0	71	71	0
10	0	19	19	0
11	0	74	74	0
12	1	28	29	3.45
13	0	121	121	0
14	0	39	39	0
15	0	37	37	0
16	0	10	10	0
17	0	14	14	0
18	0	67	67	0
19	0	51	51	0
20	0	66	66	0
21	0	11	11	0
22	0	58	58	0
23	0	48	48	0
24	0	39	39	0
25	0	60	60	0
26	0	41	41	0
TOTAL	**4**	**1432**	**1436**	**0.28**

FIGURE 1. Results of analysis on original distribution of samples.

Statistic	DF	Value	Prob
Chi-Square	**25**	**36.641**	**0.062**
Likelihood Ratio Chi-Square	25	17.444	0.865
Mantel-Haenszel Chi-Square	1	3.752	0.053
Fisher's Exact Test (2-Tail)			**0.166**

Sample Size = 1436
WARNING: 50% of the cells have expected counts less than 5.
Chi-Square may not be a valid test.

TABLE 1A. One positive per station - Moved 1 positive to a station (2) with a *small* number of submissions.

STATION #	# POSITIVE	# NEGATIVE	TOTAL #	% POSITIVE
1	1	56	57	1.75
2	1	14	15	6.67
7	1	120	121	0.83
12	1	28	29	3.45
TOTAL	4	1432	1436	0.28

Statistic	DF	Value	Prob
Chi-Square	25	41.694	0.019
Fisher's Exact Test (2-Tail)			0.077

TABLE 1B. One positive per station - Moved 1 positive from the station (1) with 2 positives in the original data to a station (4) with a *large* number of submissions.

STATION #	# POSITIVE	# NEGATIVE	TOTAL #	% POSITIVE
1	1	56	57	1.75
4	1	80	81	1.23
7	1	120	121	0.83
12	1	28	29	3.45
TOTAL	4	1432	1436	0.28

Statistic	DF	Value	Prob
Chi-Square	25	22.138	0.628
Fisher's Exact Test (2-Tail)			0.455

TABLE 2A. One station with two positives - Moved 1 positive from the original data (7) to a station (14) with a *small* number of submissions.

STATION #	# POSITIVE	# NEGATIVE	TOTAL #	% POSITIVE
1	2	55	57	3.51
7	0	121	121	0
12	1	28	29	3.45
14	1	38	39	2.56
TOTAL	4	1432	1436	0.28

Statistic	DF	Value	Prob
Chi-Square	25	42.897	0.014
Fisher's Exact Test (2-Tail)			0.037

TABLE 2B. One station with two positives - Moved 1 positive from the original data (12) to a station (13) with a *large* number of submissions.

STATION #	# POSITIVE	# NEGATIVE	TOTAL #	% POSITIVE
1	2	55	57	3.51
7	1	120	121	0.83
12	0	29	29	0
13	1	120	121	0.83
TOTAL	4	1432	1436	0.28

Statistic	DF	Value	Prob
Chi-Square	25	27.203	0.346
Fisher's Exact Test (2-Tail)			0.586

TABLE 3A. Two stations with two positives - Moved 1 positive from the original data (7) to a station (12) with a *small* number of submissions.

STATION #	# POSITIVE	# NEGATIVE	TOTAL #	% POSITIVE
1	2	55	57	3.51
7	0	121	121	0
12	2	27	29	6.9
TOTAL	4	1432	1436	0.28

Statistic	DF	Value	Prob
Chi-Square	25	70.908	0.001
Fisher's Exact Test (2-Tail)			7.30E -03

TABLE 3B. Two stations with two positives - Moved 1 positive from the original data (12) to a station (7) with a *large* number of submissions.

STATION #	# POSITIVE	# NEGATIVE	TOTAL #	% POSITIVE
1	2	55	57	3.51
7	2	119	121	1.65
12	0	29	29	0
TOTAL	4	1432	1436	0.28

Statistic	DF	Value	Prob
Chi-Square	25	33.153	0.127
Fisher's Exact Test (2-Tail)			0.347

FIGURE 2. Comparisons of software computation times.

	SAS®	StatXact®
	The PROCEDURE FREQ used:	
* Original Data (4 Positives)	6.04 seconds	<2 seconds
* 5 Positives*	4 minutes 37.8 seconds	<2 seconds
* 6 Positives*	1 hour 36 minutes 59 seconds	<2 seconds
* 7 Positives*	11 hours 28 minutes 10 seconds	<2 seconds
* 8 Positives*	52 hours 49 minutes 43 seconds	<2 seconds