



BIOMETRICS INFORMATION

(You're 95% likely to need this information)

PAMPHLET NO. # 41

DATE: September 18, 1992

SUBJECT: Power Analysis and Sample Size Determination for Contingency Table Tests

Statistical power is a valuable tool for estimating sample size. It helps us to ensure that the sample size is large enough that the smallest difference of biological importance can be detected with a reasonable degree of certainty (Nemec 1991). This pamphlet discusses the computation of power for contingency table tests. The SAS programs for performing power analysis and sample size calculation are included.

A contingency table test is a common method for analyzing categorical data. One of its applications is to test whether two or more classifications are independent of one another (see Biometrics Information Pamphlet No. 36 for a discussion on contingency table tests). As an example, suppose a survey was carried out in a district to assess the relationship between stream conditions and the slope on which streams are located. 100 streams were evaluated; 49 streams were located on a steep slope (over 20% gradient), 51 streams were located on a shallow slope (less than or equal to 20% gradient). The habitat quality of each stream was classified as GOOD, AVERAGE, or LOW according to fish abundance and the volume of large organic debris found along the stream. Table 1 below displays the hypothetical data for this survey.

Table 1: Observed counts for the stream assessment survey

Slope	<u>Habitat Quality</u>			Row Total
	GOOD	AVERAGE	LOW	
STEEP	26	13	10	49
SHALLOW	20	23	8	51
Column Total	46	36	18	100

Based on the collected data, we want to test whether habitat quality is independent of slope. That is, whether the proportion of streams in each of the three habitat quality levels are the same for steep slope and shallow slope streams.

To test the hypothesis that habitat quality is independent of slope, we can compute a χ^2 statistic for the contingency table:

$$\chi^2 = \sum_{\text{all cells}} \frac{(O - E)^2}{E} \quad (1)$$

where O is the observed count and E is the estimated expected count under the null hypothesis of independence.

The estimated expected count of a cell is the product of the corresponding marginal totals¹ divided by the total count. For example, the upper left cell, corresponding to GOOD habitat quality and STEEP slope has an estimated count:

$$E = \frac{49 \times 46}{100} = 22.54$$

The estimated expected counts for the other cells of the contingency table is given in Table 2.

Table 2: Expected counts based on Table 1

Slope	<u>Habitat Quality</u>		
	GOOD	AVERAGE	LOW
STEEP	22.54	17.64	8.82
SHALLOW	23.46	18.36	9.18

Therefore, the χ^2 statistic for this example is:

$$\chi^2 = \frac{(26-22.54)^2}{22.54} + \frac{(13-17.64)^2}{17.64} + \frac{(10-8.82)^2}{8.82} + \frac{(20-23.46)^2}{23.46} + \frac{(23-18.36)^2}{18.36} + \frac{(8-9.18)^2}{9.18} = 3.744$$

The degrees of freedom are (see Biometrics Information Pamphlet No. 21):

$$df = (\text{no. of slope conditions} - 1)(\text{no. of habitat levels} - 1) = (2-1)(3-1) = 2$$

With two degrees of freedom, a χ^2 value of 3.744 has a p-value of 0.154. We can therefore conclude that there is no evidence that habitat quality is dependent on slope.

This χ^2 test can be performed with the following SAS program:

SAS program 1: To perform contingency table test:

```
DATA EG1;
  LENGTH SLOPE $7 QUALITY $7;
  DO SLOPE = 'STEEP' , 'SHALLOW';
    DO QUALITY = 'GOOD' , 'AVERAGE' , 'LOW';
      INPUT COUNT@@;
      OUTPUT;
    END;
  END;
CARDS;
  26 13 10
  20 23 8
;
PROC FREQ DATA = EG1;
  TABLES SLOPE*QUALITY/NOCOL NOPERCENT CELLCHI2 CHISQ;
  WEIGHT COUNT;
RUN;
```

¹Marginal totals are the row and column totals as indicated in Table 1.

The TABLES statement requests a table of SLOPE by QUALITY. The first two options suppress the printing of column percentages and cell percentages in each cell. The option CELLCHI2 specifies the printing of each cell's contribution to the total χ^2 ; The option CHISQ requests the χ^2 test of independence with various statistics. The SAS output from this program is shown below. Note that the habitat levels have been re-ordered alphabetically and that the χ^2 statistic and p-value are consistent with our calculations.

SAS output from PROC FREQ:

The SAS System				
TABLE OF SLOPE BY QUALITY				
SLOPE	QUALITY			
Frequency				
Cell Chi-Square				
Row Pct	average	low	good	Total
shallow	23	8	20	51
	1.1726	0.1517	0.5103	
	45.10	15.69	39.22	
steep	13	10	26	49
	1.2205	0.1579	0.5311	
	26.53	20.41	53.06	
Total	36	18	46	100
STATISTICS FOR TABLE OF SLOPE BY QUALITY				
Statistic	DF	Value	Prob	
Chi-Square	2	3.744	0.154	
Likelihood Ratio Chi-Square	2	3.782	0.151	
Mantel-Haenszel Chi-Square	1	3.209	0.073	
Phi Coefficient		0.193		
Contingency Coefficient		0.190		
Cramer's V		0.193		
Sample Size = 100				

A non-significant result does not guarantee independence. It is possible that habitat quality and slope are dependent but the experiment is not sensitive enough to detect it. A power analysis can help us decide if the non-significant conclusion is reliable.

The power of a χ^2 test depends on its non-centrality parameter, nc (Cohen 1977, Nemec 1991). For contingency table tests, the alternative hypothesis (H_a) is that the cell frequencies are distributed in a particular way, and nc is the χ^2 statistic² evaluated under H_a . For our example,

² For log-linear models, nc is the likelihood-ratio χ^2 evaluated under the alternative hypothesis. See O'Brien 1988 for details.

the H_a of the χ^2 test is that the cell frequencies are distributed as in Table 1, and $nc = 3.744$.

The power of a test is the right tail probability under the alternative hypothesis, characterized by nc (Bergerud and Sit 1992). It can be calculated with the following SAS program:

SAS program 2: To compute power for contingency table test:

```
DATA;
  ALPHA = 0.05;  DF = 2;  NC = 3.744;
  C = CINV(1-ALPHA,DF);
  POWER = 1 - PROBCHI(C,DF,NC);
  PUT 'POWER = ' POWER;
RUN;
```

At $\alpha = 0.05$, the power of the experiment is 0.39 (this result appears in the LOG window). That is, if SLOPE and QUALITY were dependent to the extent suggested by the data, the experiment would be able to detect it 39% of the time. Such low power makes the non-significant conclusion unreliable.

Suppose the survey was only a pilot study for a larger survey. We can use the test results to estimate an adequate sample size for the new survey. For contingency table tests, the non-centrality parameter can be written as:

$$nc = d^2 N \quad (2)$$

where d is the effect size and N is the total sample size. Effect size, d , measures the degree of departure from the null hypothesis. Suppose we want to estimate the sample size required to detect a pattern similar to that shown in Table 1, then $nc = 3.744$, $N = 100$, and

$$d = \sqrt{\frac{3.744}{100}} = 0.193$$

For two-way contingency tables, d is also the phi coefficient (see SAS output). The following SAS program computes power for various N .

SAS program 3: To estimate sample size for contingency table test:

```
DATA;
  ALPHA = 0.05;  DF = 2;  D = 0.193;
  C = CINV(1-ALPHA,DF);
  DO N = 100 TO 400 BY 50;
    NC = D*D*N;
    POWER = 1-PROBCHI(C, DF, NC);
    OUTPUT;
  END;
PROC PRINT;  VAR N POWER;
RUN;
```

SAS output:

N	POWER
100	0.39
150	0.55
200	0.68
250	0.79
300	0.86
350	0.91
400	0.94

The results show that to detect a pattern similar to that displayed in Table 1, a minimum of 350 streams must be sampled for a power of at least 91%.

As an exercise, suppose we want to design an experiment to test the alternative hypothesis as shown in Table 3. How many streams should we survey to ensure the test has a power of 90%?

Table 3: Cell counts for the alternative hypothesis (H_a)

Slope	<u>Habitat Quality</u>		
	GOOD	AVERAGE	LOW
STEEP	5	7	13
SHALLOW	10	36	29

With SAS program 1, we compute that this H_a has $\chi^2 = 3.093$ and $d = 0.176$. Using SAS program 3, we calculate the power of the experiment for various sample sizes:

SAS output from SAS program 3:

<u>N</u>	<u>POWER</u>
100	0.33
150	0.47
200	0.60
250	0.70
300	0.78
350	0.85
400	0.89
450	0.93

The results indicate that to achieve a power of 90%, a sample size between 400 and 450 streams is required to detect a pattern similar to that shown in Table 3.

References

- Bergerud, W. (1989). What are degrees of freedom. *Biometrics Information Pamphlet No. 21*. B.C. Min. For., Res. Br., Victoria, B.C.
- (1992). Contingency tables and log-linear models. *Biometrics Information Pamphlet No. 36*. B.C. Min. For., Res. Br., Victoria, B.C.
- Bergerud, W. and V. Sit (1992). *Power Analysis Workshop Notes*. B.C. Min. For., Res. Br., Victoria, B.C.
- Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. Academic Press, Inc. New York.
- Nemec, A. (1991). *Power Analysis Handbook for the Design and Analysis of Forestry Trials*. Biometrics Handbook No. 2. B.C. Min. For., Res. Br., Victoria, B.C.

CONTACT: Vera Sit
356-0435