SAMPLE SIZE REQUIREMENTS AND THEIR RELATIONSHIP TO THE PARAMETERIZATION OF THE PLANNED HYPOTHESIS TEST

James R. Chromy
Research Triangle Institute, Research Triangle Park, NC 27709

KEY WORDS: Binomial variables, relative risk, logit transformation

1. Introduction

With the advent of specialized software for analysis of survey data including the estimation of appropriate sampling error estimates, these data are more frequently subjected to more sophisticated analyses. For binomial data, this often includes the use of log-linear and logistic regression models. In this paper, we investigate how the statement of the study hypothesis, the assumptions about the design scenario, and the actual survey outcomes can influence the achieved power.

The discussion is limited to simple binomial variables. It relies on large sample normal approximations; at the extremes, exact tests may be more appropriate. We further limit the discussion to one-sided tests for simplicity of presentation. We believe this is an important issue, because the actual analytic strength of the study design can differ from the plan when assumptions about the level of the binomial estimates used in the design planning scenario are wrong. The method of stating the alternative hypothesis also influences the direction and magnitude of the effects on the ultimate power for planned comparisons.

2. Single Domain Cases

The simplest example of the alternate statement of study hypothesis involves inference about a single study domain. A common way of stating the null and alternative hypotheses is:

$$H_o: P - P_o = 0$$

with the alternative

$$H_1: P - P_o \geq \Delta_{11}$$

where Δ_{11} defines the alternative in terms of an absolute difference in the population parameter, P, from a hypothesized value, P_o. For a specified sample size, significance level, and power level, Δ_{11} is sometimes also called the detectable difference. If your client is more interested in relative change, you might also state the null and alternative hypotheses as:

$$H_o: \frac{P}{P_o} = 1$$

with the alternative

$$H_1: \frac{P}{P_o} \geq 1 + \Delta_{12}.$$

In this case, Δ_{12} is the hypothetical relative detectable difference.

The consulting statistician generally works with a client to establish a scenario that sets the value or values of P_o that are likely to be of most interest to the client. The client may also have some preferences about the way the hypothesis is stated: i.e., in terms of an absolute difference or a relative difference. We consider ten scenarios defined by five levels of P_o as shown in Table 1.

Tables 2 through 6 show the parameters of the alternative hypotheses when the researcher decides that he or she is also interested in testing for other values of P_o. This can occur because of poor planning, a change in perspective, or the fact that the survey has multiple purposes and the scenario used in planning the required sample size pertained to only one of those purposes. The one sample case is fairly simple and generally well understood. When stating the hypothesis in terms of an absolute difference, it is safest to choose a scenario with P_o at or near 0.50; then regardless of the final P_o of most interest, the sample size will be adequate to meet the stated power requirements. The risk in this behavior is that the sample size (and data collection costs) will be unnecessarily high if the actual P_o of most interest is well away from the mid-range of possible values. When stating the alternative hypothesis in terms of a relative change, the safe policy is to find the smallest P_o of potential interest and design for that scenario. For larger values of P_o, the power will exceed the specified level. The risk with this behavior is that the power drops off very rapidly if the researcher should become interested in values of P_o lower than those used in the design planning scenario.

3. Hypothesis Tests Comparing Two Domains

Two sample comparisons are routinely treated as a hypothesis about the difference between domain proportions or rates. Some researchers, however, prefer to think in terms of relative risk which leads to hypotheses about the ratio of proportions or rates. Many researchers anticipate the use of log-linear or logit modeling and would like to see the hypothesis stated in terms of their model parameters. For this paper, we consider four forms of the null and alternative hypotheses.

The first form is the one most usually applied by statisticians in determining the required sample size, namely:

$$H_o: P_1 = P_2$$

with the one-sided alternative

$$H_1: P_1 - P_2 \geq \Delta_{21}$$

Table 1. Design Planning Scenarios for the One Domain Problem

Scenario Number	Hypothesized population proportion, P_o	Absolute difference, Δ_{11} (Alternative 1)	Equivalent relative difference, Δ_{12} (Alternative 2)	Required sample size
1a	0.05	.05	1.0000	150
1b	0.10	.05	0.5000	253
1c	0.20	.05	0.2500	419
1d	0.50	.05	0.1000	617
1e	0.80	.05	0.0625	368

Table 2. Power to Reject the Null Hypothesis Under Design Planning Scenario 1a
($P_o = 0.05,\ \Delta_{11} = 0.05,\ \Delta_{12} = 1,\ n = 150$)

Actual P_o of most interest	True P given the alternative		Power for:	
	$P_o + \Delta_{11}$	$P_o(1 + \Delta_{12})$	Alternative 11	Alternative 12
0.05	0.1000	0.1000	0.8013	0.8013
0.10	0.1500	0.2000	0.6304	0.9797
0.20	0.2500	0.4000	0.4580	1.0000
0.50	0.5500	1.0000	0.3364	1.0000
0.80	0.8500	1.6000	0.4492	1.0000

Table 3. Power to Reject the Null Hypothesis Under Design Planning Scenario 1b
($P_o = 0.10,\ \Delta_{11} = 0.05,\ \Delta_{12} = 0.5,\ n = 253$)

Actual P_o of most interest	True P given the alternative		Power for:	
	$P_o + \Delta_{11}$	$P_o(1 + \Delta_{12})$	Alternative 11	Alternative 12
0.05	0.1000	0.0750	0.9273	0.5519
0.10	0.1500	0.1500	0.8010	0.8010
0.20	0.2500	0.3000	0.6244	0.9844
0.50	0.5500	0.7500	0.4782	1.0000
0.80	0.8500	Exceeds 1	0.6497	n.a.

Table 4. Power to Reject the Null Hypothesis Under Design Planning Scenario 1c
($P_o = 0.20,\ \Delta_{11} = 0.05,\ \Delta_{12} = 0.25,\ n = 419$)

Actual P_o of most interest	True P given the alternative		Power for:	
	$P_o + \Delta_{11}$	$P_o(1 + \Delta_{12})$	Alternative 11	Alternative 12
0.05	0.1000	0.0625	0.9867	0.3661
0.10	0.1500	0.1250	0.9311	0.5204
0.20	0.2500	0.2500	0.8007	0.8007
0.50	0.5500	0.6250	0.6569	0.9998
0.80	0.8500	1.0000	0.8470	1.0000

Table 5. Power to Reject the Null Hypothesis Under Design Planning Scenario 1d
($P_o = 0.50,\ \Delta_{11} = 0.05,\ \Delta_{12} = 0.10,\ n = 617$)

Actual P_o of most interest	True P given the alternative		Power for:	
	$P_o + \Delta_{11}$	$P_o(1 + \Delta_{12})$	Alternative 11	Alternative 12
0.05	0.1000	0.0550	0.9984	0.2174
0.10	0.1500	0.1100	0.9820	0.2462
0.20	0.2500	0.2200	0.9113	0.3548
0.50	0.5500	0.5500	0.8004	0.8004
0.80	0.8500	0.8800	0.9490	0.9999

where P_1 and P_2 represent the population proportions for domains 1 and 2, respectively, and Δ_{21} is the hypothesized difference that the client wishes to detect

. The second form of the hypotheses is stated in terms of relative values of the domain proportions (sometimes referred to as "relative risk"). Its general form is

$$H_o: \frac{P_1}{P_2} = 1$$

with the one-sided alternative

$$H_1: \frac{P_1}{P_2} \geq 1 + \Delta_{22}.$$

If the client researcher wishes to fit log-linear models to the data, he or she may wish to state the hypothesis in terms of a model parameter.

$$H_o: \ln\frac{P_1}{P_2} = 0 \text{ with the one-sided alternative}$$

$$H_1: \ln\frac{P_1}{P_2} \geq \Delta_{23}.$$

Since Δ_{22} is a direct transformation of Δ_{23} that does not depend on the level of P_2, either hypothesis leads to equivalent results as is shown in the tables.

The client researcher may plan to analyze the data using a logit transformation model and may also wish to state the hypothesis in terms of those parameters. With the common model reparameterization which makes domain 2 the reference set, the model can be written as

$$\ln\frac{P_i}{(1 - P_i)} = \alpha + \beta X_i + \epsilon_i$$

where X_i is set to one for domain 1 and zero for domain 2. The relevant model parameter, β , can be represented in a design planning hypothesis test as::

$$H_o: \ln\frac{P_1/(1-P_1)}{P_2/(1-P_2)} = 0$$

with the one-sided alternative

$$H_1: \ln\frac{P_1/(1-P_1)}{P_2/(1-P_2)} \geq \Delta_{24}$$

As was the case with the log-linear model above, tests stated in terms of the odds ratio rather than the log odds ratio would yield equivalent results in this analysis. Kendall and Buckland (1971) note the regrettable use of the term "relative risk" for the odds ratio; this interpretation can lead to errors in the interpretation of logistic regression parameter estimates.

We continue to work with a one-sided test, a significance level of 0.05, and a planned power of 0.80 at the design scenario level of the population values. All of the above null and alternative hypothesis can be made equivalent for a particular scenario and would lead to the same required sample size. We state this scenario in terms of the hypothesized lower domain value, P_2, and the choice of the alternative hypothesis as indicated by positive values of Δ_{21}, Δ_{22}, Δ_{23}, or Δ_{24}. Table 7 shows the planning scenarios covered.

Tables 8 through 13 show the values of P_1 under each alternative hypothesis formulation for selected values of P_2. In the two sample case, the level of the estimates is part of the scenario, but can be stated in words without referring to the assumed levels: e.g., the difference, the relative risk, the log-linear model parameter, or the log odds ratio. As formulated here for a one-sided test, P_2 is treated as the hypothetically lower value; in fact, good target values for P_2 may not be known for the domains of interest in advance of the study.

Certain conclusions can be drawn from examining the data in Tables 8 through 13. The behavior of the power for tests of an absolute difference, Δ_{11}, follows the behavior in the one-sample case. It is safer to choose a scenario in the mid-range of possible values of P_2; this generates the largest sample size and provides more than adequate power if the design scenario assumptions are wrong. The next two tests are equivalent. The power associated with failure of the design scenario follows the behavior of the relative difference in the one-sample case. The power for logit model (log odds model) has somewhat of an inverse relationship to the behavior of the power for the absolute differences; it has its highest power in the mid-range and drops off at both extremes.

4. Implications

While the results of this examination may be intuitively obvious to many statisticians, the whole area of the interpretation of model parameters and the testing of hypotheses about those parameters remains murky for many general researchers and needs to be addressed directly by the consulting statistician. This is important as a means of avoiding disappointment with a planned study after the data are obtained. It is also important that the sampling statistician become aware of the actual planned analysis so that the hypotheses testing formulations used in developing the required sample size are consistent with those that will be used by the analyst. This requires effective communication between the sampling statistician and the researcher analyst.

References

Kendall, Maurice G., and William R. Buckland. 1971. *A Dictionary of Statistical Terms*. New York, NY: Hafner.

Table 6. Power to Reject the Null Hypothesis Under Design Planning Scenario 1e
(P_o = 0.80, Δ_{11} = 0.05, Δ_{12} = 0.0625, n =368)

Actual P_o of most interest	True P given the alternative		Power for:	
	$P_o + \Delta_{11}$	$P_o(1 + \Delta_{12})$	Alternative 11	Alternative 12
0.05	0.1000	0.0531	0.9774	0.1598
0.10	0.1500	0.1063	0.9039	0.1477
0.20	0.2500	0.2125	0.7566	0.1671
0.50	0.5500	0.5313	0.6082	0.3270
0.80	0.8500	0.8500	0.8005	0.8005

Table 7. Design Planning Scenarios for the Two Domain Problem

Scenario Number	Assumed P_1	Absolute Difference Δ_{21}	Equivalent alternatives for assumed P_1			Sample Size
			Δ_{22}	Δ_{23}	Δ_{24}	
2a	0.025	0.05	2.0000	1.09861	1.15126	238
2b	0.050	0.05	1.0000	0.69315	0.74721	343
2c	0.100	0.05	0.5000	0.40547	0.46262	540
2d	0.200	0.05	0.2500	0.22314	0.28768	862
2e	0.500	0.05	0.1000	0.09531	0.20067	1233
2f	0.800	0.05	0.0625	0.06062	0.34831	714

Table 8. Power to Reject Null Hypothesis Under Design Planning Scenario 2a
(P_2 = .025, n = 238)

Actual P_2	Value of P_1 given the alternative				Power to detect alternative			
	21	22	23	24	21	22	23	24
0.0250	0.0750	0.0750	0.0750	0.0750	0.8060	0.8060	0.8060	0.8060
0.0500	0.1000	0.1500	0.1500	0.1427	0.6656	0.9783	0.9783	0.9644
0.1000	0.1500	0.3000	0.3000	0.2600	0.5017	1.0000	1.0000	0.9985
0.2000	0.2500	0.6000	0.6000	0.4415	0.3671	1.0000	1.0000	1.0000
0.5000	0.5500	> 1	> 1	0.7597	0.2900	n.a.	n.a.	1.0000
0.8000	0.8500	> 1	>1	0.9267	0.4168	n.a.	n.a.	0.9923

Table 9. Power to Reject Null Hypothesis Under Design Planning Scenario 2b
(P_2 = .050, n = 343)

Actual P_2	Value of P_1 given the alternative				Power to detect alternative			
	21	22	23	24	21	22	23	24
0.0250	0.0750	0.0500	0.0500	0.0514	0.9144	0.5313	0.5313	0.5621
0.0500	0.1000	0.1000	0.1000	0.1000	0.8009	0.8009	0.8009	0.8009
0.1000	0.1500	0.2000	0.2000	0.1900	0.6315	0.9795	0.9795	0.9570
0.2000	0.2500	0.4000	0.4000	0.3455	0.4693	1.0000	1.0000	0.9962
0.5000	0.5500	1	1	0.6786	0.3691	n.a.	n.a.	0.9992
0.8000	0.8500	> 1	>1	0.8941	0.5313	n.a.	n.a.	0.9637

Table 10. Power to Reject Null Hypothesis Under Design Planning Scenario 2c
(P_2 = .010, n = 540)

Actual P_2	Value of P_1 given the alternative				Power to detect alternative			
	21	22	23	24	21	22	23	24
0.0250	0.0750	0.0375	0.0375	0.0391	0.9838	0.3210	0.3210	0.3717
0.0500	0.1000	0.0750	0.0750	0.0771	0.9307	0.5208	0.5208	0.5727
0.1000	0.1500	0.1500	0.1500	0.1500	0.8000	0.8000	0.8000	0.8000
0.2000	0.2500	0.3000	0.3000	0.2842	0.6267	0.9848	0.9848	0.9444
0.5000	0.5500	0.7500	0.7500	0.6136	0.5001	1.0000	1.0000	0.9833
0.8000	0.8500	> 1	> 1	0.8640	0.6979	n.a.	n.a.	0.8794

Table 11. Power to Reject Null Hypothesis Under Design Planning Scenario 2d

(P_2 = .020, n = 862)

Actual P_2	Value of P_1 given the alternative				Power to detect alternative			
	21	22	23	24	21	22	23	24
0.0250	0.0750	0.0313	0.0313	0.0331	0.9992	0.1948	0.1948	=1
0.0500	0.1000	0.0625	0.0625	0.0656	0.9895	0.3019	0.3019	0.3976
0.1000	0.1500	0.1250	0.1250	0.1290	0.9329	0.4990	0.4990	0.5979
0.2000	0.2500	0.2500	0.2500	0.2500	0.8002	0.8002	0.8002	0.8002
0.5000	0.5500	0.6250	0.6250	0.5714	0.6679	0.9998	0.9998	0.9085
0.8000	0.8500	=1	=1	0.8421	0.8620	n.a.	n.a.	0.7378

Table 12. Power to Reject Null Hypothesis Under Design Planning Scenario 2e

(P_2 = .050, n = 1233)

Actual P_2	Value of P_1 given the alternative				Power to detect alternative			
	21	22	23	24	21	22	23	24
0.0250	0.0750	0.0275	0.0275	0.0304	1.0000	0.1044	0.1044	0.2033
0.0500	0.1000	0.0550	0.0550	0.0604	0.9990	0.1382	0.1382	0.3049
0.1000	0.1500	0.1100	0.1100	0.1196	0.9828	0.2018	0.2018	0.4637
0.2000	0.2500	0.2200	0.2200	0.2340	0.9083	0.3351	0.3351	0.6576
0.5000	0.5500	0.5500	0.5500	0.5500	0.8001	0.8001	0.8001	0.8001
0.8000	0.8500	0.8800	0.8800	0.8302	0.9480	0.9999	0.9999	0.6126

Table 13. Power to Reject Null Hypothesis Under Design Planning Scenario 2f

(P_2 = .080, n = 714)

Actual P_2	Value of P_1 given the alternative				Power to detect alternative			
	21	22	23	24	21	22	23	24
0.0250	0.0750	0.0266	0.0266	0.0351	0.9966	0.0723	0.0723	0.2972
0.0500	0.1000	0.0531	0.0531	0.0694	0.9744	0.0841	0.0841	0.4606
0.1000	0.1500	0.1063	0.1063	0.1360	0.8878	0.1044	0.1044	0.6787
0.2000	0.2500	0.2125	0.2125	0.2615	0.7319	0.1443	0.1443	0.8682
0.7000	0.7500	0.7438	0.7438	0.7677	0.6814	0.5793	0.5793	0.8952
0.8000	0.8500	0.8500	0.8500	0.8500	0.8004	0.8004	0.8004	0.8004