韓國數學敎育學會誌 <數學敎育>
1981. 6. Vol. XIX. No. 2

Journal of the Korea Society of Mathematical Education
Jun. 1981. Vol. XIX. No. 2

An Approximation for Calculating Sample Sizes for Comparing Independent Propotions in case of $p_1 \leqq 0.2$

Kang Sup Lee

Dankook University, Seoul, Korea

1. Introduction and Summary.

Let p_1 and p_2 are proportion of binomially distributed populations. And consider the problem of determining the sample sizes required to compare two indepecent probabilities p_1 and p_2. $H_0 : p_1 = p_2$ is the null hypothesis at significance level α against the alternative $H_1 : p_1 < p_2$ with power β. Define Φ to be the cumulative standard normal distribution and define $\Phi(z_r) = r$.

According to Casagrande, Pike and Smith (CPS) [1], there are some approximations for calculating sample sizes for comparing independent proportions in case of $n_1 = n_2 = n$:

(i) The "arcsin formula" as given, for example, in Cochran and Cox,

$$n = \frac{(z_{1-\alpha} + z_\beta)^2}{2(\arcsin\sqrt{p_2} - \arcsin\sqrt{p_1})^2} \tag{1}$$

(ii) The "uncorrected χ^2 formula" as given, for example, in Fleiss,

$$n = \frac{(z_{1-\alpha}\sqrt{2\bar{p}\bar{q}} + z_\beta\sqrt{p_1 q_1 + p_2 q_2})^2}{\delta^2} \tag{2}$$

where $\bar{p} = (p_1 + p_2)/2$, $\bar{q} = 1 - \bar{p}$ and $\delta = p_2 - p_1$.

(iii) The "corrected χ^2 method" given by Kramer and Greenhouse,

$$n = \frac{m}{4\delta^2}\left(1 + \sqrt{1 + \frac{8\delta}{m}}\right)^2 \tag{3}$$

where $m = (z_{1-\alpha}\sqrt{2\bar{p}\bar{q}} + z_\beta\sqrt{p_1 q_1 + p_2 q_2})^2$.

Exact sample sizes for $\alpha = 0.05$ and $\alpha = 0.01$ are calculated by Haseman [3].
CPS derived the corrected formula

$$n = \frac{m}{4\delta^2}\left(1 + \sqrt{1 + \frac{4\delta}{m}}\right)^2 \tag{4}$$

from the comparison ofexact values with those of three approximations.

In general case of $n_1 \neq n_2$, let $r = n_2/n_1$. According to Fleiss, Tytun and Ury (FTU) [2], Ury obtained the following formula by modifying the CPS's,

(5) $$n_1=\frac{m'}{4}\left(1+\sqrt{1+\frac{2(r+1)}{rm'\delta}}\right)^2$$

as the approximate sample size from the first population, where

(6) $$m'=\frac{1}{r_6\delta^2}(z_{1-\alpha}\sqrt{(r+1)\bar{p}\bar{q}}+z_\beta\sqrt{(rp_1q_1+p_2q_2)})^2$$

$\bar{p}=\frac{p_1+rp_2}{r+1}$ and $\bar{q}=1-\bar{p}$.

And FTU derived the simple approximation

(7) $$n_1=m'+\frac{r+1}{r\delta}.$$

Also, Ury and Fleiss [4] derived the approximation using Yates' correction,

(8) $$n_1=\frac{m'}{4}(1+\sqrt{1+2\omega})^2$$

where $\omega=\frac{\delta}{(z_\alpha+z_\beta)^2\bar{p}\bar{q}}$, $\bar{p}=\frac{p_1+rp_2}{r+1}$ and $\bar{q}=1-\bar{p}$.

2. The Proposed Approximation.

1) Case of $r\neq 1$.

FTU's formula is obtained by using approximation

(9) $$\sqrt{1+\frac{1}{x}}\approx 1+\frac{1}{2x}.$$

But there is a bias since $1+\frac{1}{2x}>\sqrt{1+\frac{1}{x}}$ for $x>0$. Therefore we want to reduce the bias. So using

$$\sqrt{1+\frac{1}{x}}\approx 1+\frac{1}{3x},$$

we proposed another approximation

$$n_1=\frac{m'}{4}\left(1+\sqrt{1+\frac{2(r+1)}{rm'\delta}}\right)^2$$
$$=\frac{m'}{4}\left(2+\frac{2(r+1)}{rm'\delta}+2\sqrt{1+\frac{2(r+1)}{rm'\delta}}\right)$$
$$\approx m'+\frac{5}{6}\times\frac{r+1}{r\delta}.$$

2) Case of $r=1$.

Substituting $r=1$ into (10)

(11) $$n=m'+\frac{2}{3\delta}.$$

3. Comparision of Approximations.

Now, consider the problem of estimating power for prespecified sample size. If we use the approximation (7), FTU derived

(12) $$z_\beta=\frac{\sqrt{r\delta^2n_1-(r+1)\delta}-z_{1-\alpha}\sqrt{(r+1)\bar{p}\bar{q}}}{(rp_1q_1+p_2q_2)}$$

as the approximate percentile corresponding to the actual power. If we use the approximation (10), the approximate power $\Phi(z_\beta)$ is given by

$$z_\beta=\frac{\sqrt{r\delta^2 n_1}-\frac{5}{6}(r+1)\delta-z_{1-\alpha}\sqrt{(r+1)\bar{p}\bar{q}}}{(rp_1q_1+p_2q_2)} \tag{13}$$

using (6) and (10).

Since z_β in (13) $>$ z_β in (12), the approximation (10) is more powerful. Table 1 shows the numerical example.

Table 1.

Approximate powers for detecting a difference between $p_1=0.15$ and $p_2=0.25$ using a one sided significance test with a total sample size of 360 and a significance level of 0.05.

r	Approximation (7)		Approximation (10)	
	z_β	Power	z_β	Power
0.33	0.24	0.5948	0.27	0.6064
0.50	0.49	0.6879	0.55	0.7088
1	0.60	0.7257	0.62	0.7324
2	0.41	0.6591	0.49	0.6879
3	0.19	0.5753	0.29	0.6141

Table 2 shows that the approximation denoted by (11) is more accurate than the approximation (5) for $r=1$ in case of $p_1 \leqq 0.2$. Therefore the approximation (11) is more unerring than that of (7).

Table 2.

p_1 \ δ	.05	.1	.15	.2	.25	.3	.35	.4	.45	.5	.55	.6	.65	.7
0.05	504	165	89	57	42	33	25	21	18	15	13	11	10	9
	506*	169*	93*	61*	45*	34*	27*	22*	19*	16*	13*	12	10*	9
	513	172	95	63	46	35	28	23	20	17	14	12	11	9
0.1	782	232	119	74	52	39	31	25	20	17	15	12	11	10
	781*	233*	119*	75*	53*	39*	31*	25*	20*	17*	14#	12*	10#	9#
	787	237	121	77	54	41	32	26	21	18	15	13	11	10
0.15	1024	289	142	87	60	45	34	27	23	18	16	12	11	10
	1021	289*	142*	87*	60*	44#	34*	27*	22#	18*	15#	12*	11	9#
	1027	292	144	89	61	45	35	28	23	19	16	13	11	10
0.2	1231	338	162	97	65	47	36	30	23	18	16	12	11	10
	1227#	336#	161	96	65*	47*	36*	28#	22#	18*	15#	13*	11	9#
	1233	339	163	98	66	48	37	29	23	19	16	14	11	10
0.25	1402	377	178	106	71	51	36	31	24	18	16	12	11	9
	1398#	375#	177	104#	69#	49#	37*	29#	23#	19	15#	12*	10#	9
	1404	378	179	106	71	51	38	30	24	19	16	13	11	9

	1538	408	190	111	72	53	40	31	24	18	16	12	10	—
0.3	1535	405#	188#	109#	72*	51#	38#	29#	23#	18*	15#	12*	10*	—
	1541	408	190	111	73	52	39	30	24	19	16	13	11	—
	1640	428	200	115	72	53	40	31	23	18	15	11	—	—
0.35	1638*	426#	196#	112#	73*	51#	38#	29#	22#	18*	14#	12	—	—
	1644	429	198	114	75	53	39	30	23	19	15	12	—	—
	1710	445	202	116	72	53	36	30	23	17	13	—	—	—
0.4	1706#	439#	200#	113#	73*	51#	37*	28#	22#	17*	13*	—	—	—
	1713	442	202	115	75	52	38	29	23	18	14	—	—	—
	1746	446	202	115	72	51	36	27	20	15	—	—	—	—
0.45	1741#	443#	200#	112#	72*	49#	36*	27*	20*	16*	—	—	—	—
	1747	447	202	114	73	51	37	28	21	17	—	—	—	—
	1746	445	200	111	71	47	34	25	18	—	—	—	—	—
0.5	1741#	439#	196#	109#	69#	47*	34*	25*	19*	—	—	—	—	—
	1747	442	198	111	71	48	35	26	20	—	—	—	—	—

Upper figure: Exact valuc.

Middle figure: Due to approximation (11).

Lower figure: Due to approximation (4) (or (5) for $r=1$).

*: Middle figure is more accurate than lower figure.

#: Lower figure is more accurate than middle figure.

References.

[1] Casagrande, J.T., Pike, M.C. and Smith, P.G. (1978). An improved approximate formula for calculating sample sizes for comparing two binomial distributions. *Biometrics* **34**, 483-486.

[2] Fleiss, J.L., Tytun, A. and Ury, H.K. (1980). A simple approximation for calculating sample sizes for comparing independent proportions. *Biometrics* **36**, 343-346.

[3] Haseman, J.K. (1978). Exact sample sizes for use with the Fisher-Irwin test for 2×2 tables. *Biometrics* **34**, 106-109.

[4] Ury, H.K. and Fleiss, J.L. (1980). On approximate sample sizes for comparing two independent proportions with the use of Yates' corrrection. *Biometrics* **36**, 347-351.