Even with the continuity correction there will be some doubt about the adequacy of the $\chi^2$ approximation when the frequencies are particularly small. An exact test was suggested almost simultaneously in the mid-1930s by R. A. Fisher, J. O. Irwin and F. Yates. It consists in calculating the exact probabilities of the possible tables described in the previous subsection. The probability of a table with frequencies

| | | |
|---|---|---|
| a | b | $r_1$ |
| c | d | $r_2$ |
| $c_1$ | $c_2$ | N |

is given by the formula

$$P[\, a \mid r_1, r_2, c_1, c_2 \,] = \frac{r_1!\, r_2!\, s_1!\, s_1!}{N!\, a!\, b!\, c!\, d!} \qquad (4.25)$$

This is, in fact, the probability of the observed cell frequencies ***conditional*** on the observed marginal totals, under the null hypothesis of no association between the row and column classifications. Given any observed table, the probabilities of all tables with the same marginal totals can be calculated, and the P value for the significance test calculated by summation. Example 4.14 illustrates the calculations and some of he difficulties of interpretation which may arise. The data in Table 4.6, due to M. Hellman, are discussed by Yates (1934).

Table 4.6 Data on malocclusion of teeth in infants (Yates, 1934)

| | Infants with | | |
|---|---|---|---|
| | Normal teeth | Malocclusion | Total |
| Breast-fed | 4 | 16 | 20 |
| Bottle-ed | 1 | 21 | 22 |
| Total | 5 | 37 | 42 |

There are six possible tables with the same marginal totals as those observed. since neither a nor c (in the notation given above) can fall below 0 or exceed 5, the smallest marginal total in the table. The cell frequencies in each of these tables are shown in Table 4.7. Below them are shown the probabilities of these tables, calculated under the null hypothesis.

Table 4.7 Cell frequencies in tables with the same marginal totals as those in Table 4.6

| | | | | | |
|---|---|---|---|---|---|
| 0 20 \| 20 | 1 19 \| 20 | 2 18 \| 20 | 3 17 \| 20 | 4 16 \| 20 | 5 15 \| 20 |
| 5 17 \| 22 | 4 18 \| 22 | 3 19 \| 22 | 2 20 \| 22 | 1 21 \| 22 | 0 22 \| 22 |
| 5 37 \| 42 | S 37 \| 42 | 5 37 \| 42 | 5 37 \| 42 | 5 37 \| 42 | 5 37 \| 42 |
| a 0 | 1 | 2 | 3 | 4 | 5 |
| $P_a$ 0.0310 | 0.1720 | 0.3440 | 0.3096 | 0.1253 | 0.0182 |

The Probabilities of the various tables are calculated in the following way: the probability that a = 0 is, from (4.25),

$$P_0 = \frac{20!\, 22!\, 5!\, 37!}{42!\, 0!\, 20!\, 5!\, 7!} = 0.03096.$$

Tables of log factorials (Fisher and Yates, 1963, Table XXX) are often useful for this calculation, and many scientific calculators have a factorial key (although it may only function correctly for integers less than 70). Alternatively the expression for $P_0$ can be calculated without factorials by repeated multiplication and division after cancelling common factors:

$$P_0 = \frac{22 \times 21 \times 20 \times 19 \times 18}{42 \times 41 \times 40 \times 39 \times 38} = 0.03096.$$

The probabilities for a = 1, 2, . . ., 5 can be obtained in succession. Thus,

$$P_1 = \frac{5 \times 20}{1 \times 18} \times P_0$$

$$P_2 = \frac{4 \times 19}{2 \times 19} \times P_1, \text{ etc.}$$

The results are shown above.

[Note from JH: The 5 tables from the tea-tasting experiment with to the 2x2 tables with all marginal totals = 4 are another example of this *hypergeometric* distribution]

This is the complete *conditional distribution* for the observed marginal totals, and the probabilities sum to unity as would be expected. Note the importance of carrying enough significant digits in the first probability to be calculated; the above calculations were carried out with more decimal places than recorded by retaining each probability in the calculator for the next stage. The observed table has a probability of 0.1253. To assess its significance we could measure the extent to which it falls into the tail of the distribution by calculating the probability of that table or of one more extreme. For a one-sided test the procedure clearly gives P = 0.1253 + 0.0182 = 0.1435. The result is not significant at even the 10% level.

For a two-sided test the other tail of the distribution must be taken into account, and here some ambiguity arises. Many authors advocate that the one-tailed P value should be doubled. In the present example, the one-tailed test gave P = 0.1435 and the two-tailed test would give P = 0.2870. An alternative approach is to calculate P as the total probability of tables, in either tail, which are at least as extreme as that observed in the sense of having a probability at least as small. In the present example we should have

$$P = 0.1253 + 0.0182 + 0.0310 = 0.1745.$$

The first procedure is probably to be preferred on the grounds that a significant result is interpreted as strong evidence for a difference in the *observed direction,* and there is some merit in controlling the chance probability of such a result to no more than half the two-sided significance level. The tables of Finney *et al.* (1963) enable one-sided tests at various significance levels to be made without computation provided the frequencies are not too great.

To calculate the **mid-P** value only half the probability of the observed table is included and we have

$$\text{mid-P} = 0.5(0.1253) + 0.0182 = 0.0808$$

as the one-sided value, and the two-sided value may be obtained by doubling this to give 1617.

The results of applying the exact test in this example may be compared with those obtained by the $\chi^2$ test with Yates's correction. We find $X^2$ = 2 39 (P = 0.12) without correction and $X^2_C$ = 1.14 (P = 0.29) with correction. The probability level of 0.29 for $X^2_C$ agrees well with the two-sided value 0 29 from the exact test, and the probability level of 0.12 for $X^2$ is a fair approximation to the exact mid-P value of 0.16.

Cochran (1954) recommends the use of the exact test, in preference to the $\chi^2$ test with continuity correction, (i) if N < 20, or (ii) ir 20 < N c 40 and the smallest expected value is less than 5. With modern scientific calculators and statistical software the exact test is much easier to calculate than previously and should be used for any table with an expected value less than 5.

The exact test and therefore the $\chi^2$ test with Yates's correction for continuity have been criticized over the last 50 years on the grounds that they are conservative in the sense that a result significant at, say, the 5% level will be found in less than 5% of hypothetical repeated random samples from a population in which the null hypothesis is true. This feature was discussed in §4.7 and it was remarked that the problem was a consequence of the discrete nature of the data and causes no difficulty if the precise level of P is stated. Another source of criticism has been that the tests are conditional on the observed margins, which frequently would not all be fixed. For example, in Example 4.14 one could imagine repetitions of sampling in which 20 breast-fed infants were compared with 22 bottle-fed infants but in many of these samples the number of infants with normal teeth would differ from 5. The conditional argument is that, whatever inference can be made about .the association between breast-feeding and tooth decay, it has to be made within the context that exactly five children had normal teeth. If this number had been different then the inference would have been made in this different context, but that is irrelevant to inferences that can be made when there are five children with normal teeth. Therefore, we do not accept the various arguments that have been put forward for rejecting the exact test based on consideration of possible samples with different totals in one of the margins. The issues were discussed by Yates 1984) and in the ensuing discussion, and by Barnard (1989) and Upton (1992), .and we will not pursue this point further. Nevertheless, the exact test and the corrected $\chi^2$ test have the undesirable feature that the average value of the significance level, when the null hypothesis is true, exceeds 0.5. The mid-P value avoids this problem, and so is more appropriate when combining results from several studies (see §4.7).

As for a single proportion, the mid-P value corresponds to an uncorrected $\chi^2$ test, whilst the exact P value corresponds to the corrected $\chi^2$ test. The confidence limits for the difference, ratio or odds ratio of two proportions based on the standard errors given by (4.14), (4.17) or (4.19) respectively are all approximate and the approximate values will be suspect if one or more of the frequencies in the 2 x 2 table are small. Various methods have been put forward to give improved limits but all of these involve iterations and are tedious to carry out on a calculator. The odds ratio is the easiest case. Apart from exact limits, which involve an excessive amount of calculation, the most satisfactory limits are those of Cornfield ( 1956); see Example 16.1 and Breslow and Day (1980, §4.3) or Fleiss ( 1981, §5.6). For the ratio of two proportions a method was given by Koopman (1984) and Miettinen and Nurrninen (1985) which can be programmed fairly readily. The confidence interval produced gives a good approximation to the required confidence coefficient, but the two tail probabilities are unequal due to skewness. Gart and Nam (1988) gave a correction for skewness but this is tedious to calculate. For the difference of two proportions a method was given by Mee (1984) and Miettinen and Nurminen (1985). This involves more calculation than for the ratio limits, and again there could be a problem due to skewness (Gart and Nam, 1990).

**Notes by JH**

- The word "exact" means that the p-values are calculated using a finite discrete reference distribution -- the hypergeometric distribution (cousin of the binomial) rather than using large-sample approximations. It doesn't mean that it is the *correct test.* [see comment by A&B in their section dealing with Mid-P values]. While greater accuracy is always desirable, this particular test uses a 'conditional' approach that not all statisticians agree with. Moreover, compared with some unconditional competitors, the test is somewhat conservative, and thus less powerful, particularly if sample sizes are very small.

- Fisher's exact test is usually used just as a test*; if one is interested in the difference $\Delta = \pi_1 - \pi_2$ , the conditional approach does not yield a corresponding confidence interval for $\Delta$. [it does provide one for the comparative odds ratio parameter $\psi = \frac{1-\pi_1}{\pi_1} \div \frac{1-\pi_2}{\pi_2}$ ]
- Thus, one can find anomalous situations where the (conditional) test provides P>0.05 making the difference 'not statistically significant', whereas the large-sample (unconditional) CI for $\Delta$, computed as $p_1 - p_2 \pm zSE(p_1 - p_2)$, does not overlap 0, and so would indicate that the difference is 'statistically significant'. [* see the Breslow and Day text Vol I , §4.2, for CI's for $\psi$ derived from the conditional distribution]

- See letter from Begin & Hanley re 1/20 mortality with pentamidine vs 5/20 with Trimethoprim-Sulfamethoxazole in pts c Pneumocystis carinii Preumonia-Annals Int Med 106 474 1987.

- Miettinen's test-based method of forming CI's, while it can have some drawbacks, keeps the correspondence between test and CI and avoids such anomalies.

- This illustrates one important point about parameters related to binary data -- with means of interval data, we typically deal just with differences*; however, with binary data, we often switch between differences and ratios, either because the design of the study forces us to use odds ratios (case-control studies), or because the most readily available regression software uses a ratio (i.e. logistic regression for odds ratios) or because one is easier to explain that the other, or because one has a more natural interpretation (e.g. in assessing the cost per life saved of a more expensive and more efficacious management modality, it is the difference in, rather than the ratio of, mortality rates that comes into the calculation). [* the sampling variability of the estimated ratios of means of interval data is also more difficult to calculate accurately].

- Two versions of an unconditional test for the $H_0$: $\pi_1 = \pi_2$ are available: Liddell; Suissa and Shuster;

## Abstract

**Background** Some claim that food sensitivities can best be identified by intradermal injection of extracts of the suspected allergens to reproduce the associated symptoms. A different dose of an offending allergen is thought to "neutralize" the reaction.

**Methods** To assess the validity of symptom provocation, we performed a double-blind study that was carried out in the offices of seven physicians who were proponents of this technique and experienced in its use. Eighteen patients were tested in 20 sessions (two patients were tested twice) by the same technician, using the same extracts (at the same dilutions with the same saline diluent) as those previously thought to provoke symptoms during unblinded testing. At each session three injections of extract and nine of diluent were given in random sequence. The symptoms evaluated included nasal stuffiness, dry mouth, nausea, fatigue, headache, and feelings of disorientation or depression. No patient had a history of asthma or anaphylaxis.

**Results** The responses of the patients to the active and control injections were indistinguishable, as was the incidence of positive responses: 27 percent of the active injections (16 of 60) were judged by the patients to be the active substance, as were 24 percent of the control injections (44 of 180). Neutralizing doses given by some of the physicians to treat the symptoms after a response were equally efficacious whether the injection was of the suspected allergen or saline. The rate of judging injections as active remained relatively constant within the experimental sessions, with no major change in the response rate due to neutralization or habituation.

**Conclusions** When the provocation of symptoms to identify food sensitivities is evaluated under double-blind conditions, this type of testing, as well as the treatments based on "neutralizing" such reactions, appears to lack scientific validity. The frequency of positive responses to the injected extracts appears to be the result of suggestion and chance


---

† Calculated according to Fisher's exact test, which assumes that the hypothesized direction of effect is the same as the direction of effect in the data. Therefore, when the effect is opposite to the hypothesis, as it is for the data below those of Patient 9, the P value computed is testing the null hypothesis that the results obtained were due to change as compared with the possibility that the patients were more likely to judge a placebo injection as active than an active injection.

**Notes on P-Values from Fisher's Exact Test** in previous article

Response

**Table 1: Responses of 18 Patients Forced to Decide Whether Injections Contained an Active Ingredient or Placebo**

| Pt. No* | Active Injection | | Placebo Injection | | P Value† |
|---|---|---|---|---|---|
| | resp | no resp | resp | no resp | |
| 3 | 2 | 1 | 1 | 8 | 0.13 |
| 1 | 2 | 1 | 2 | 7 | 0.24 |
| 14a | 2 | 1 | 2 | 7 | 0.24 |
| 12 | 1 | 2 | 0 | 9 | 0.25 |
| 16 | 2 | 1 | 3 | 6 | 0.36 |
| 18 | 2 | 1 | 4 | 5 | 0.50 |
| 14b | 1 | 2 | 2 | 7 | 0.87 |
| 4 | 1 | 2 | 2 | 7 | 0.87 |
| 5 | 1 | 2 | 2 | 7 | 0.87 |
| 9 | 0 | 3 | 0 | 9 | -- |
| 2a | 0 | 3 | 1 | 8 | 0.75 |
| 13 | 0 | 3 | 1 | 8 | 0.75 |
| 15 | 1 | 2 | 3 | 6 | 0.76 |
| 6 | 0 | 3 | 2 | 7 | 0.55 |
| 8 | 0 | 3 | 2 | 7 | 0.55 |
| 17 | 1 | 2 | 5 | 4 | 0.50 |
| 2b | 0 | 3 | 3 | 6 | 0.38 |
| 7 | 0 | 3 | 3 | 6 | 0.38 |
| 10 | 0 | 3 | 3 | 6 | 0.38 |
| 11 | 0 | 3 | 3 | 6 | 0.38 |

*Patients were numbered in the order they were studied.

The order in the table is related to the degree that the results agree with the hypothesis that patients could distinguish active injections from placebo injections. The results listed below those of Patient 9 do not support this hypothesis, placebo injections were identified as active at a higher rate than were true active injections. The letters a and b denote the first and second testing sessions, respectively, in Patients 2 and 14. true active injections. ID denotes intradermal, and SC subcutaneous.

The value is the P value associated with the test of whether the common odds ratio (the odds ratio for all patients) is equal to 1.0. The common odds ratio was equal to 1.13 (computed according to the Mantel-Haenszel test).

| | | + | – | Total |
|---|---|---|---|---|
| Patient no. 3 | Active Injection | 2 | 1 | 3 |

| | | | | |
|---|---|---|---|---|
| | Placebo Injection | 1 | 8 \| | 9 |
| | | 3 | 9 | |

All possible tables with a total of 3 +ve responses

| | | | | |
|---|---|---|---|---|
| | 0 3 | 1 2 | 2 1 | 3 0 |
| | 3 6 | 2 7 | 1 8 | 0 9 |
| | 9• 8• 7 | 3•3 | 2•2 | 1•1 |
| prob | –––––––– | 0.382 • --- | 0.491 • --- | 0.123 • --- |
| | 12•11•10 | 1•7 | 2•8 | 3•9 |
| | 0.382 | 0.491 | 0.123 | 0.005 |
| *(pt #)* | *(2b,7,10,11)* | *(14b, 4, 5)* | *(3)* | |
| P-Value* | 1.0 | 0.618 | 0.128 | 0.005 |

---

| | | | | |
|---|---|---|---|---|
| | | Response + | – | Total |
| Patient no. 1 | Active Injection | 2 | 1 \| | 3 |
| | Placebo Injection | 2 | 7 \| | 9 |
| | | 4 | 8 | |

All possible tables with a total of 4 +ve responses

| | | | | |
|---|---|---|---|---|
| | 0 3 | 1 2 | 2 1 | 3 0 |
| | 4 5 | 3 6 | 2 7 | 1 8 |
| | 8• 7• 6 | 3•4 | 2•3 | 1•2 |
| prob | –––––––– | 0.255 • --- | 0.510 • --- | 0.218 • --- |
| | 12•11•10 | 1•6 | 2•7 | 3•8 |
| | 0.255 | 0.510 | 0.218 | 0.018 |
| *(pt #)* | | *(15)* | *(1,14a)* | |
| P-Value | 1.0 | 0.745 | 0.236 | 0.018 |

(*1-sided, guided by $H_{alt}$: π of +ve responses with Active > π of +ve responses with Placebo)

| | | | | |
|---|---|---|---|---|
| | | Response + | – | Total |
| Patient no. 18 | Active Injection | 2 | 1 \| | 3 |
| | Placebo Injection | 4 | 5 \| | 9 |
| | | 6 | 6 | |

All possible tables with a total of 6 +ve responses

| | | | | |
|---|---|---|---|---|
| | 0 3 | 1 2 | 2 1 | 3 0 |
| | 6 3 | 5 4 | 4 5 | 3 6 |
| | 6• 5• 4 | 3•6 | 2•5 | 1•4 |
| prob | –––––––– | 0.091 • --- | 0.409 • --- | 0.409 • --- |
| | 12•11•10 | 1•4 | 2•5 | 3•6 |
| | 0.091 | 0.409 | 0.409 | 0.091 |
| *(pt #)* | | *(17)* | *(18)* | |
| P-Value | 1.0 | 0.909 | 0.500 | 0.091 |

(1-sided, as above)

---

**In Table 1, the P-values for patients below patient 9 are calculated as 1-sided, but guided by the opposite $H_{alt}$ from that used for the patients in the upper half of the table, i.e. by $H_{alt}$: π of +ve responses with Active < π of +ve responses with Placebo).**

**It appears that the authors decided the "sided-ness" of the $H_{alt}$ after observing the data!!! and that they used different $H_{alt}$ for different patients!!!**

**Politics and small sample sizes** Active Intervention and Conservation: Africa's Pachyderm Problem

Joel Berger[1] and Carol Cunningham[2] POLICY FORUM • SCIENCE • VOL. 263 • 4 Mar 1994 • pages 1241-1242

1 Ecology Evolution, and Conservation Biology Program 2 Department of Environmental and Resource Sciences University of Nevada Reno NV 89512, USA.

*The Namibian government expelled the authors form Namibia following the publication of this article; the reason given was that* ***their "data and conclusions were premature"*** *.. jh* ¶•

Since 1900 the world's population has increased from about 1.6 to over 5 billion) the U.S. population has kept pace, growing from nearly 75 to 260 million. While the expansion of humans and environmental alterations go hand in hand, it remains uncertain whether conservation programs will slow our biotic losses. Current strategies focus on solutions to problems associated with diminishing and less continuous habitats, but in the past, when habitat loss was not the issue, active intervention prevented extirpation. Here we briefly summarize intervention measures and focus on tactics for species with economically valuable body parts, particularly on the merits and pitfalls of biological strategies tried for Africa's most endangered pachyderms, rhinoceroses.

[ ... ]

Given the inadequacies of protective. legislation and enforcement, Namibia. Zimbabwe, and Swaziland are using a controversial preemptive measure, dehorning (Fig. D) with the hope that complete devaluation will buy time for implementing other protective measures (7) In Namibia and Zimbabwe, two species, black and white rhinos (Ceratotherium simum), are dehorned, a tactic resulting in ***sociological and biological uncertainty: Is poaching deterred? Can hornless mothers defend calves from dangerous predators?***

On the basis of our work in Namibia during the last 3 years (8) and comparative information from Zimbabwe, some data are available. Horns regenerate rapidly, about 8.7 cm per animal per year, so that 1 vear after dehorning the regrown mass exceeds 0.5 kg. Because poachers apparently do not prefer animals with more massive horns (8), frequent and costly horn removal may be required (9). In Zimbabwe, a population of 100 white rhinos, with at least 80 dehorned, was reduced to less than 5 animals in 18 months *(10).* These discouraging results suggest that intervention by itself is unlikely to eliminate the incentive for poaching. Nevertheless, some benefits accrue when governments, rather than poachers, practice horn harvesting, since less horn enters the black market Whether horn stockpiles may be used to enhance conservation remains controversial, but mortality risks associated with anesthesia during dehoming are low (5).

Biologically, there have also been problems. Despite media attention and a bevy of allegations about the soundness of denorning ( 11 ), serious attempts to determine whether dehorning is harmful have been remiss. A lack of negative effects has been suggested because (i) horned and dehorned individuals have interacted without subsequent injury; (ii) dehorned animals have thwarted the advance of dangerous predators; (iii) feeding is normal; and (iv) dehorned mothers have given birth (12) However, most claims are anecdotal and mean little without attendant data on demographic effects. For instance, while some dehorned females give birth, it may be that these females were pregnant when first immobilized. Perhaps others have not conceived or have lost calves after birth. Without knowing more about the frequency of mortality, it seems premature to argue that dehorning is effective.

We gathered data on more than 40 known horned and hornless black rhinos in the presence and absence of dangerous carnivores in a 7,000 $km^2$ area of the northem Namib Desert and on 60 horned animals in the 22,000 $km^2$ Etosha National Park. On the basis of over 200 witnessed interactions between horned rhinos and spotted hyenas (Crocura crocura) and lions (Panthera leo) we saw no cases of predation, although mothers charged predators in about 45% of the cases. Serious interspecific aggression is not uncommon elsewhere in Africa, and calves missing ears and tails have been observed from South Africa, Kenya, Tanzania, and Namibia (13).

**To evaluate the vulnerability of dehorned rhinos to potential predators, we developed an experimental design using three regions:**

• Area A had horned animals with spotted hyenas and occasional lions

• Area B had dehorned animals lacking dangerous predators,

• Area C consisted of dehorned animals that were sympatric with hyenas only.

Populations were discrete and inhabited similar xeric landscapes that averaged less than 125 mm of precipitation annually. Area A occurred north of a country-long veterinary cordon fence, whereas animals from areas B and C occurred to the south or east, and no individuals moved between regions.

The differences in calf survivorship were remarkable. All three calves in area C died within 1 year of birth, whereas all calves survived for both dehorned females living without dangerous predators (area B; n = 3) and for horned mothers in area A (n = 4). Despite admittedly restricted samples, the differences are striking [Fisher's (3 x 2) exact test, P = 0.017; area B versus C, P = 0.05; area A versus C, P = 0.0291 ††. The data offer a first assessment of an empirically derived relation between horns and recruitment.

Our results imply that hyena predation was responsible for calf deaths, but other explanations are possible. If drought affected one area to a larger extent than the others, then calves might be more susceptible to early mortality. This possibility appears unlikely because all of westem Namibia has been experiencing drought and, on average, the desert rhinos in one area were in no poorer bodily condition than those in another. Also, the mothers who lost calves were between 15 to 25 years old, suggesting that they were not first time, inexperienced mothers (14). What seems more likely is that the drought-induced migration of more l-than 85% of the large, herbivore biomass (kudu, springbok, zebra, gemsbok, giraffe, and ostrich) resulted in hyenas preying on an alternative food, rhino neonates, when mothers with regenerating horns could not protect them.

Clearly, unpredictable events, including drought, may not be anticipated on a short-term basis. Similarly, it may not be possible to predict when govemments can no longer fund antipoaching measures, an event that may have led to the collapse of Zimbabwe's dehorned white rhinos. Nevertheless, any effective conservation actions must account for uncertainty. In the case of dehorning, additional precautions must be taken. [ ... ]

| | A | B | C | |
|---|---|---|---|---|
| survived | 4 | 3 | 0 | †† |
| died | 0 | 0 | 3 | |
| | 4 | 3 | 3 | |

B vs C

| | B | C | B | C | B | C | B | C | *tot** |
|---|---|---|---|---|---|---|---|---|---|
| survived | 3 | 0 | 2 | 1 | 1 | 2 | 0 | 3 | *3* |
| died | 0 | 3 | 1 | 2 | 2 | 1 | 3 | 3 | *3* |
| | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | |
| prob | 1/20 | | 9/20 | | 9/20 | | 1/20 | | |

A vs C

| | A | C | A | C | A | C | A | C | *tot** |
|---|---|---|---|---|---|---|---|---|---|
| survived | 4 | 0 | 3 | 1 | 2 | 2 | 1 | 3 | *4* |
| died | 0 | 3 | 1 | 2 | 2 | 1 | 3 | 3 | *3* |
| | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 3 | |
| prob | 1/35 | | 12/35 | | 18/35 | | 4/35 | | |

**¶ Do you agree??**