STT 31
11-30-07

1. **Chi-square test of goodness of fit.**

a. A discrete production process has been brought under statistical control with the following probabilities for crystals produced in batch. A random sample of 30 parts is selected and sorted.

type	perfect	fine	standard	reject	
probability	0.14	0.21	0.58	0.07	(total one)
expected	4.2	6.3	17.4	2.1	(total 30)
observed	6	5	12	7	(total 30)

Chi-square statistic, the sum of $[(O - E)^2 / E]$ over four cells, is 14.1489. DF = 4 – 1 = 3 free parameters. pSIG = P(chi-square DF 3 > 14.1489) = 0.00271 (Table VII gives .005 to the right of 12.83 under chi-square with 3 DF). So it is exceedingly rare to get this much disagreement between model and data just by chance. This is exceedingly strong evidence against the hypothesis that the process is operating in the specified probability regime. Bringing and keeping a process under control involves consistently taking such measurements and acting upon the results. EXPECTED COUNTS 4.1 AND 2.1 ABOVE VIOLATE THE "ALL EXPECTED COUNTS AT LEAST 5" RULE. So we are not in conformity with proper practice. In this case one might overlook the 4.2 but combine "standard" with "fine" and do the chi-square test with DF = 3 – 1 = 2. You might try this. Does it affect the conclusion?

b. For the above, suppose a process monitor is directed to send an alert if the process rejects a test of the null hypothesis that the process remains in control, with alpha = 0.01. What action is taken in this case?

2. **Chi-square test of no difference.** A new process is being tested against the old process. Independent samples of parts produced by each process give

	best	avg	worst
old	10	16	6
new	6	6	4

Under the hypothesis that the probabilities of each of the 3 categories are the same for old and new we'd estimate those shared probabilities as 16/48, 22/48, 10/48 respectively. We've really only estimated two things because these three must total 1. The full model has 4 probabilities since each row has two degrees of freedom. So the DF is 4 – 2 = 2. This may also be calculated (r-1)(c-1) = (2-1)(3-1) = 2 where r = 2 rows and c = 3 columns. "Expected" counts project these shared 3 category probabilities over 32 old and 16 new giving

	best	avg	worst
old	16/48 32	22/48 32	10/48 32
new	16/48 16	22/48 16	10/48 16

The chi-square calculates to 0.69 with DF = 2. pSIG = P(chi-square DF 2 > 0.69) = 0.7. No evidence against H0. Note that the lower right expected count is < 5 but won't affect the conclusion.

ARE MENDEL'S DATA TOO GOOD TO BE TRUE?

In 1902, two years after Mendel's work was rediscovered, W. F. R. Weldon suspected that Mendel's results were very close to expected values and tested this suspicion with Pearson's newly developed χ^2 test. He concluded that Mendel's observed ratios were astonishingly close to his expectations (Weldon, 1902). Weldon's analysis created a brief controversy and was quickly forgotten (Mangello, 1998). The next statistician to question the proximity of Mendel's results to expected values was Ronald A. Fisher (1936) who published a now-famous paper in which he closely examined Mendel's paper and reconstructed the thought process of the experiments. Fisher's analysis is careful and thorough and reveals his admiration for Mendel's work. However, his paper is best known for its conclusion, the same one that Weldon had arrived at 32 years earlier, that Mendel's results were consistently so close to expected ratios that the validity of those results must be questioned. Fisher's work spawned a series of papers dealing with this issue. Citations of these papers can be found in several reviews (Edwards, 1986; Piegorsch, 1986; Di Trocchio, 1991; Weiling, 1991; Nissani, 1994; Orel, 1996). Unfortunately, all this effort has failed to yield a definitive solution: according to Nissani (1994, p. 182), "the subject remains every bit as controversial today as it was in 1936."

Like Weldon's analysis, Fisher's was based on consistently low χ^2 values produced when he subjected Mendel's data to χ^2 tests. We present the χ^2 values and their associated probabilities with the appropriate degrees of freedom for each of Mendel's independent *Pisum* experiments in Table 2. As Edwards (1986) noted, about half of all independent experiments should yield χ^2 values with probabilities <0.5 and half with probabilities >0.5. Of Mendel's 22 experiments, only two yield χ^2 values with probabilities <0.5 and six yield χ^2 values with probabilities >0.90, indicative of the bias toward expectation in Mendel's data (Table 2).

View this table: Table 2. Results of chi-square tests for Mendel's *Pisum sativum* experiments

Experiment from Table 1	Degrees of freedom	χ^2	Probability
1.	1	0.1314	0.7169
2.	1	0.0075	0.9310
3.	1	0.1954	0.6584
4.	1	0.0318	0.8586
5.	1	0.2253	0.6350
6.	1	0.1748	0.6759
7.	1	0.3033	0.5818
8.	1	0.1735	0.6771
9.	1	0.4249	0.5145
10.	1	0.3200	0.5716
11.	1	0.8450	0.3580
12.	1	2.0000	0.1573
13.	1	0.0050	0.9436
14.	1	1.2800	0.2579
15.	1	0.1250	0.7237
16a.	3	0.4700	0.9254
16b.	8	2.8110	0.9457
17.	26	15.3224	0.9511
18.	3	0.5778	0.9015
19.	3	0.8621	0.8346
20.	3	0.6182	0.8923
21.	3	0.5306	0.9121
22.	3	1.0843	0.7809

Table 2. Results of chi-square tests for Mendel's *Pisum sativum* experiments

3. **Mendel's Data.** Let's take at face value the above chi-square values and assume that the 23 experiments are statistically independent. If so, one is entitled to form a grand chi-square statistic by totaling column three, the DF for which will be the total of column two. This is a consequence of the fact that the chi-square distribution having DF = d is actually the distribution of the sum of

squares of d independent standard normal r.v. The grand DF obtained from column two will be too large to use the chi-square tables.

The CLT comes to the rescue. For large DF = d, if the models leading to these chi-squares are all correct then

(total chi-square – d) / root(2d) ~ Z.

The above works out to (28.5193 - 67) / root(2 67) = -3.32423
so pSIG = P(Z > -3.32423) = 0.999557 ~ 1.

This means the combined data from all 23 experiments is uncomfortably close to what is expected by Mendel's theory, as measured by chi-square. There are other experiments not among these 23 which raise pSIG still higher.

Pearson invented chi-square in 1900, around the time Mendel's ideas were rediscovered and became widely known.

4. **Hardy-Weinberg Principle.** Suppose a population of breeding individuals has genotype distribution

genotype	AA	Aa	aa	
probability	p1	p2	p3	(total 1)

Ignoring sex, random mating with the parental population dying off will in one generation tend to produce a new population with

genotype	AA	Aa	aa	
probability	p^2	$2p(1-p)$	$(1-p)^2$	for p = p1 + (p2) / 2.

This is the Hardy-Weinberg Principle. It follows by simply observing that under random mating the whole business amounts to sampling two letters A or a according to p or (1-p).

Suppose we are able to genetically type a sample of 100 offspring and wish to test the hypothesis of random mating. Let's suppose the data is

genotype	AA	Aa	aa
counts	20	45	35

We form "expected counts" by figuring that if the hypothesis of random mating is correct we can estimate p by the proportion of letters A in the sample which is (2(20) + 45) / 200 = 0.425. This leads to "expected" counts

genotype	AA	Aa	aa
"expected"	$.425^2$ 100	2(.425)(.575) 100	$.575^2$ 100
i.e.	18.0625	48.875	33.0625

The chi-square statistic for a test of the hypothesis of random mating works out to 0.628593.

Calculating DF involves more than just k-1 = 3 – 1 because **we have to deduct one DF for estimating p**. So DF = 3 – 1 – 1 = 1. Therefore pSIG = P(chi-square with DF = 1 exceeds 0.629) which is rather close to 1. If anything this is unusually strong support for random mating. **Hardy-Weinberg found random mating so stable as to require outside influences such as mutation/migration etc. in order for a population to evolve.**