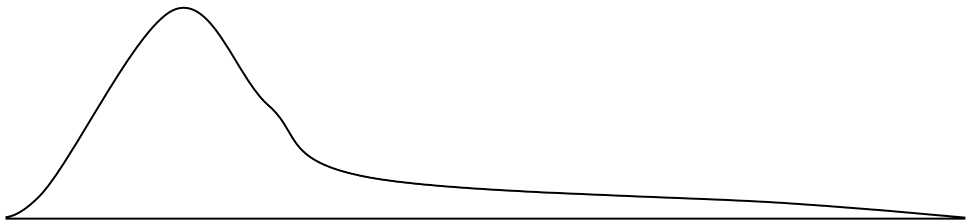
<u>X^2 (Chi-Square)</u>

- one of the most versitile statistics there is
- X^2 is a **<u>skewed</u>** distribution



- There is a different X^2 distribution for every number of degrees of freedom
- The X^2 distribution is Table 7 in the back of your book.

One use for X^2 is testing standard deviations.

- This is most often used in quality control situations in industry.

<u>QUESTION</u>:

- Is the standard deviation too large?
- Is the data too spread out?

<u>HYPOTHESES</u>:

H_0: The standard deviation is close to what it should be.

H_1: The standard deviation is too big. (It is significantly larger than it should be.)

<u>CRITICAL VALUE</u>:

- df = n – 1
- Look up α in the column at the top.

<u>FORMULA</u>:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

Important—this test is ***<u>NOT</u>*** built into the TI-83. You MUST do it with the formula.

- σ is what the standard deviation **should** be.
- ***s*** is what the standard deviation actually is in your sample.

<u>Example</u>:

Bags of Fritos® are supposed to have an average weight of 5.75 ounces. An acceptable standard deviation is .05 ounces.

Suppose a sample of 6 bags of Fritos® finds a standard deviation of .08 ounces. Is this unacceptably large? (Use α = .05)

- df = 6 – 1 = 5
- Critical: $\chi^2 = 11.07$
- Test:

$$\chi^2 = \frac{(5).08^2}{.05^2} = 12.8$$

(Note that the mean is irrelevant in the problem.)

- This is significant.

Example:
A wire manufacturer wants its finished product to be within a certain tolerance. For this to happen, the standard deviation should be less than 2.4 microns. Suppose a sample of size 20 finds the standard deviation is 3.1 microns. Do they need to adjust the machinery? Use α = .01

- df = 20 – 1 = 19
- Critical: $\chi^2 = 36.19$
- Test:

$$\chi^2 = \frac{(19)3.1^2}{2.4^2} = 31.6997$$

This is **NOT** significant. They don't need to adjust the machines.

REVIEW

When checking out, customers prefer consistent service—rather than lines that move at different speeds. A discount store company finds that in the past the average wait to check out has been 249 seconds, with a standard deviation of 46 seconds. They try a new check-out method at 12 different check lanes and find that the standard deviation with the new method is 54 seconds. Does this mean the new method has a significantly bigger variation in wait time? Do a standard deviation χ^2 test at the 10% level of significance.

Remember
CRITICAL VALUE (table):
d.f. = n – 1
So 11 d.f. and α = .10
$\chi^2(11,.100) = 17.28$
TEST STATISTIC (calculator):

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

$$\chi^2 = \frac{11 \bullet 54^2}{46^2}$$

$\chi^2 = 15.16$

Categorical Chi-Square Test
(a.k.a. "Goodness of Fit" Test)
QUESTION:

- Is the distribution of data into various categories different from what is expected?
- Key idea—you have **qualitative** data (characteristics) that can be divided into **more than 2** categories.
- You're comparing what the distribution in different categories **should** be with what it actually is in your sample.

HYPOTHESES:
H_1: The distribution is significantly different from what is expected.
H_0: The distribution is not significantly different from what is expected.
CRITICAL VALUE:

- df = k – 1
- one less than the number of categories

TEST STATISTIC:

- This test is on the TI-84, but **not** on the TI-83.
- Unless you have a TI-84, you will need to use the formula.

Formula:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

For each category:

- Subtract observed value (what it is in your sample) minus expected value (what it should be).
- Square the difference.
- Divide the square by the expected value.

Add up the answers for all categories.

Example:

A teacher wants different types of work to count toward the final grade as follows:

Daily Work ➔ 25%
Tests ➔ 50%
Project ➔ 15%
Class Part. ➔ 10%

When points for the term are figured, the actual number of points in each category is:

Daily Work ➔ 175
Tests ➔ 380
Project ➔ 100
Class Part. ➔ 75

TOTAL POINTS = 730

Critical Value:

There are 4 categories, so we have 3 degrees of freedom.
Since no level of significance is given, let's use $\alpha = .10$.
$\chi^2(3,.10) = 6.25$

Test Statistic:

To solve the problem we first compute the expected values:
(Take each % times 730 total points.)

	O	E
Daily	175	182.5
Tests	380	365
Project	100	109.5
Class	75	73

Using the formula we get .31 + .62 + .82 + .05 = **1.80**

Comparison:

- 1.80 < 6.25
- The distribution is **NOT** significantly different from what was expected.

If you have a TI-84, here's what you do …

Enter the numbers

- Go to STAT → EDIT
- Type the **observed** values in L1.
- Type the **expected** values in L2. (You can just take each percent times the total.)
- 2nd / MODE (QUIT)

Do the test

- Go to STAT → TESTS
- Choose choice "D" (you may want to use the up arrow)… χ^2GOF-Test
- Hit ENTER repeatedly. (It doesn't actually matter what you put on the "df" line.)
- In the read-out what you care about is χ^2.

EXAMPLE
You think your friend is cheating at cards, so you keep track of which suit all the cards that are played in a hand are. It turns out to be:

- ♦ → 4
- ♥ → 2
- ♣ → 13
- ♠ → 1

You'd normally expect that 25% of all cards would be of each suit. At the .01 level of significance, is this distribution significantly different than should be expected?

Critical Value

- X^2(3 d.f., .01) = 11.34

Test Statistic

	O	E	(O-E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
♦	4	5	-1	1	.2
♥	2	5	-3	9	1.8
♣	13	5	8	64	12.8
♠	1	5	-4	16	3.2

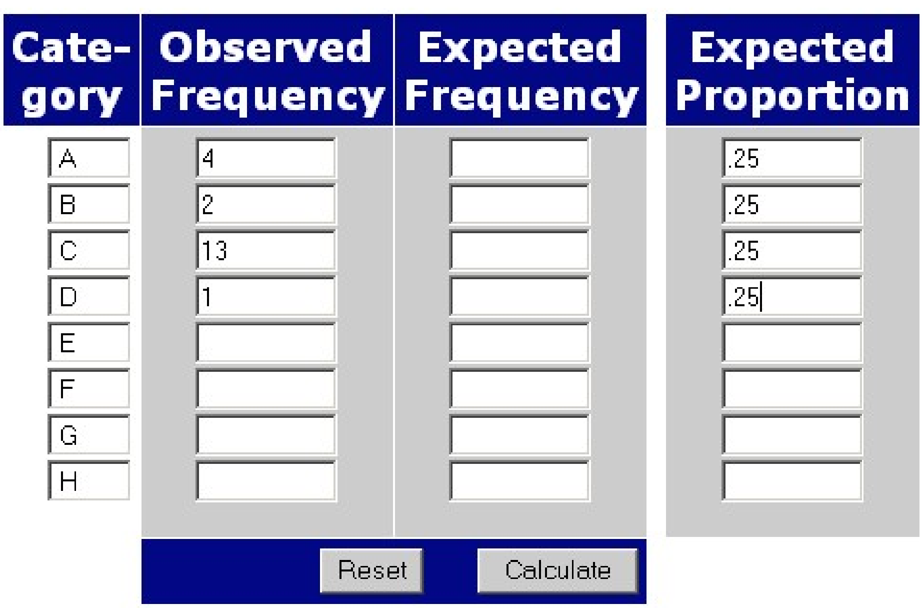
$$\Sigma \frac{(O-E)^2}{E} = 18$$

RESULT:

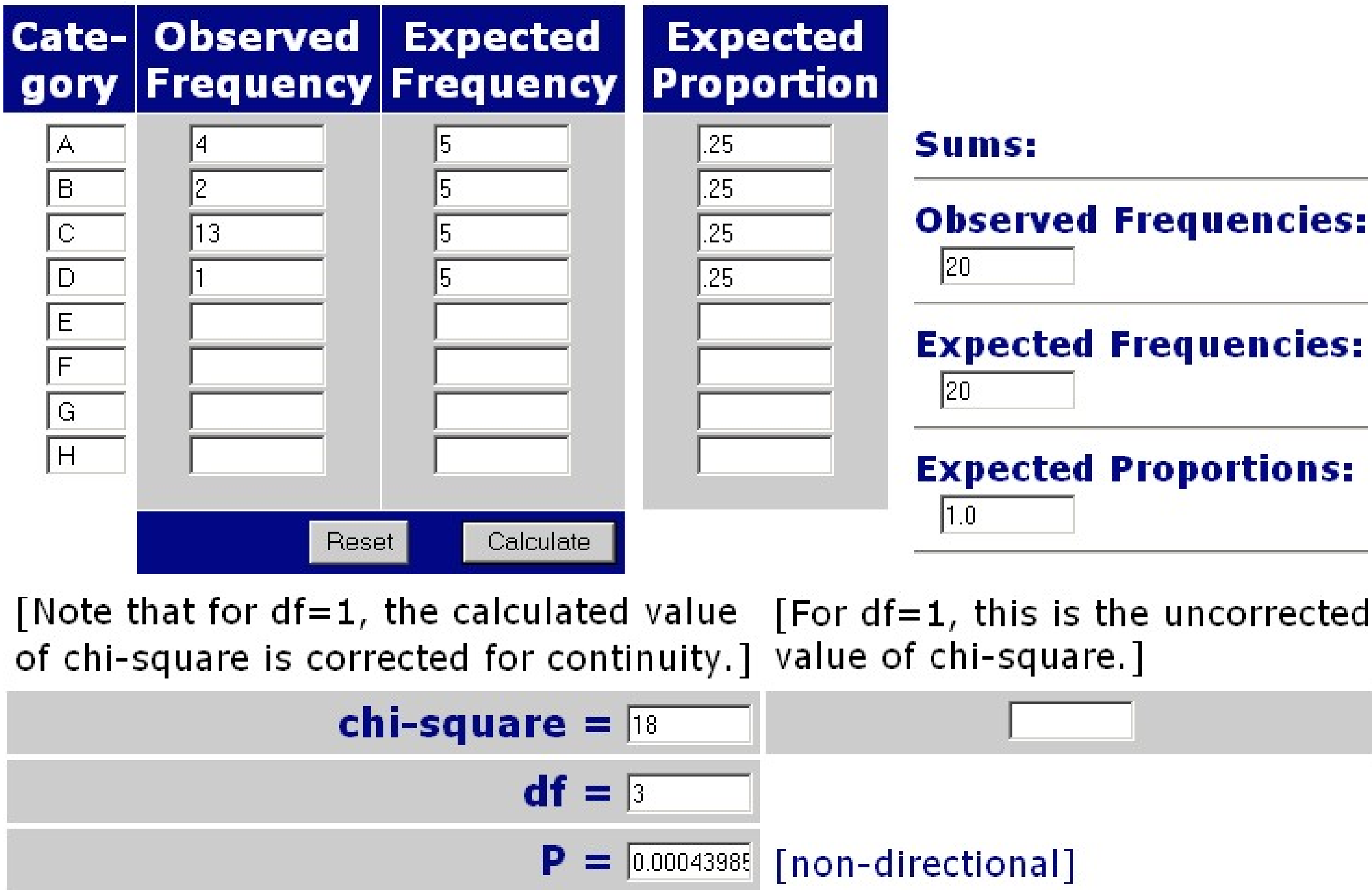
- 18 > 11.34
- Significant

You can do categorical X^2 tests online. One website with a chi-square calculator is http://faculty.vassar.edu/lowry/csfit.html

- This is linked from the class web page.
- You may want to use it for your project, if you don't have a TI-84.
- You will need to enter the observed values and **either** the expected values or the expected proportions.
- For instance, the card problem would look like:



- When you hit CALCULATE, chi-square is calculated, together with the degrees of freedom and other information.
- The results of the card problem look like this:



- If you need to do a categorical chi-square test for your project, this is the suggested method.

<u>Matrix Chi-Square Test</u>
(a.k.a. "Independence" Test)

- Compares two qualitative variables.
- <u>QUESTION</u>: Does the distribution of one variable change from one value to the other variable to another.
- We use this when we have two different qualitative variables.
- The information is generally arranged in a table.

<u>For example:</u>
Suppose in a TV class there were students at all 5 centers, in the following distribution:

	Male	Female
Algona	5	7
E'burg	3	2
E'ville	4	4
Spenc.	4	7
S.L.	3	3

Does the distribution of men and women vary significantly by center?

- Information for a matrix X^2 test is generally presented in a table (matrix).
- Our question essentially is—is the distribution of the columns different from row to row in the table?

<u>Hypotheses</u>:

- H_1: The variables are independent (there is not a significant difference in the distribution from row to row.)
- H_0: There is a significant difference in the distribution from row to row. (The variables are dependent.)

Critical Value:

- d.f. = (R – 1)(C – 1)
  - one less than the number of rows
    TIMES
  - one less than the number of columns
- Look up d.f. and α in the χ^2 table.

Test Statistic:

- In theory you can calculate χ^2 with the same formula used for the categorical test:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

- To find the "E's" (which you won't have to do) you'd total the rows and columns of the table and then

$$E = \frac{RT \times CT}{GT}$$

  - Row Total X Column Total ÷ Grand Total
- Realistically, no one does it that way these days.\
- If you have a TI-83 or
  TI-84, this **<u>is</u>** the χ^2`-Test` built into the "Tests" menu.
- To do this …
  1. Enter the observed matrix as [A] in the MATRIX menu.
     - Press MATRX or 2nd and x^{-1}, depending on which TI-83 you have.
     - Choose "EDIT"
     - Choose matrix [A]
     - Type the number of rows and columns, pressing ENTER after each.
     - Enter each number, going across each row, and hitting ENTER after each.
  2. Press 2nd and MODE to QUIT back to a blank screen.
  3. Go to STAT, then TESTS, and choose χ^2`-Test`.
  4. Make sure it says [A] and [B] as the observed and expected matrices. If it does just hit ENTER three times.
  5. The read-out will give you χ^2 and the degrees of freedom.

Back to our example:

	Male	Female
Algona	5	7
E'burg	3	2
E'ville	4	4
Spenc.	4	7
S.L.	3	3

Critical Value:

- Since there's no α given in the problem, let's use α = .05
- There are 5 rows and 2 columns, so we have (4)(1)=4 df
- $\chi^2(4,.05) = 9.49$

MATRX → EDIT → [A]

Enter the rows and columns

5 ENTER 2 ENTER

Type the numbers, going across and hitting ENTER after each. The screen will look like this:

```
MATRIX[A]  5 X2
[ 5   7        ]
[ 3   2        ]
[ 4   4        ]
[ 4   7        ]
[ 3   3        ]
```

2nd → MODE to QUIT
STAT → TESTS→
X²-Test (Choice C)
The screen looks like this:

```
X²-Test
 Observed:[A]
 Expected:[B]
 Calculate Draw
```

Press ENTER until you get the answer:

```
X²-Test
 X²=.9794466403
 P=.9128963711
 df=4
```

RESULT

- .979 < 9.49
- NOT significant

If you don't have a TI-83 or TI-84, you can find an online calculator for the matrix X^2-test at
http://www.georgetown.edu/faculty/ballc/webtools/web_chi.html

- This is also linked from the class web page.
- It first asks for the dimensions (rows and columns)
- For the problem we just did, it looks like this:

Web Chi Square Calculator

Chi Square Tutorial | Run a demo | Jeff's demo | Begin chi square

This page allows you to perform the chi square test for statistical significance.
and columns for your test, and a table will be generated in which you can enter
information, select one of the options above.

Table dimensions: 5 rows x 2 columns

Generate table   Reset

- Hit Generate Table to produce a table where you can type your numbers.
- You don't need to worry about the row and column labels or the title (though if you want a nice-looking presentation copy, you can fill them in).
- For our problem, it looks like this:

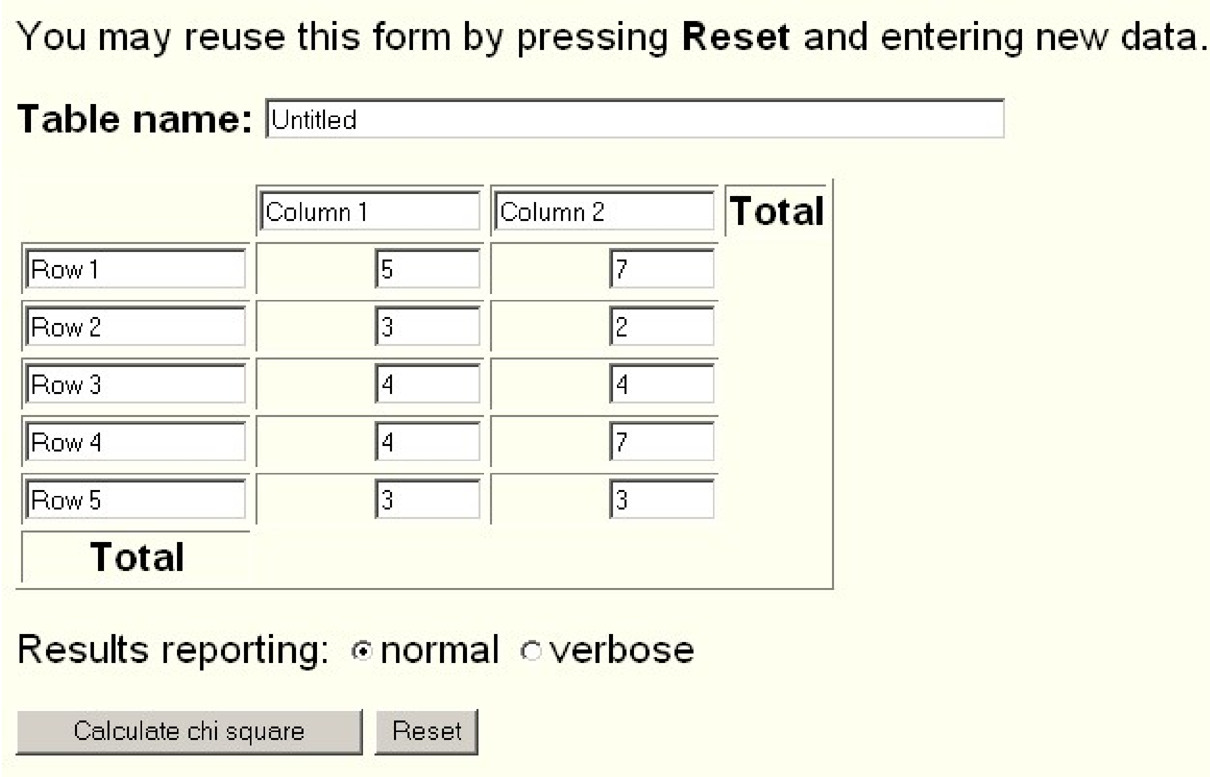

You may reuse this form by pressing **Reset** and entering new data.

Table name: Untitled

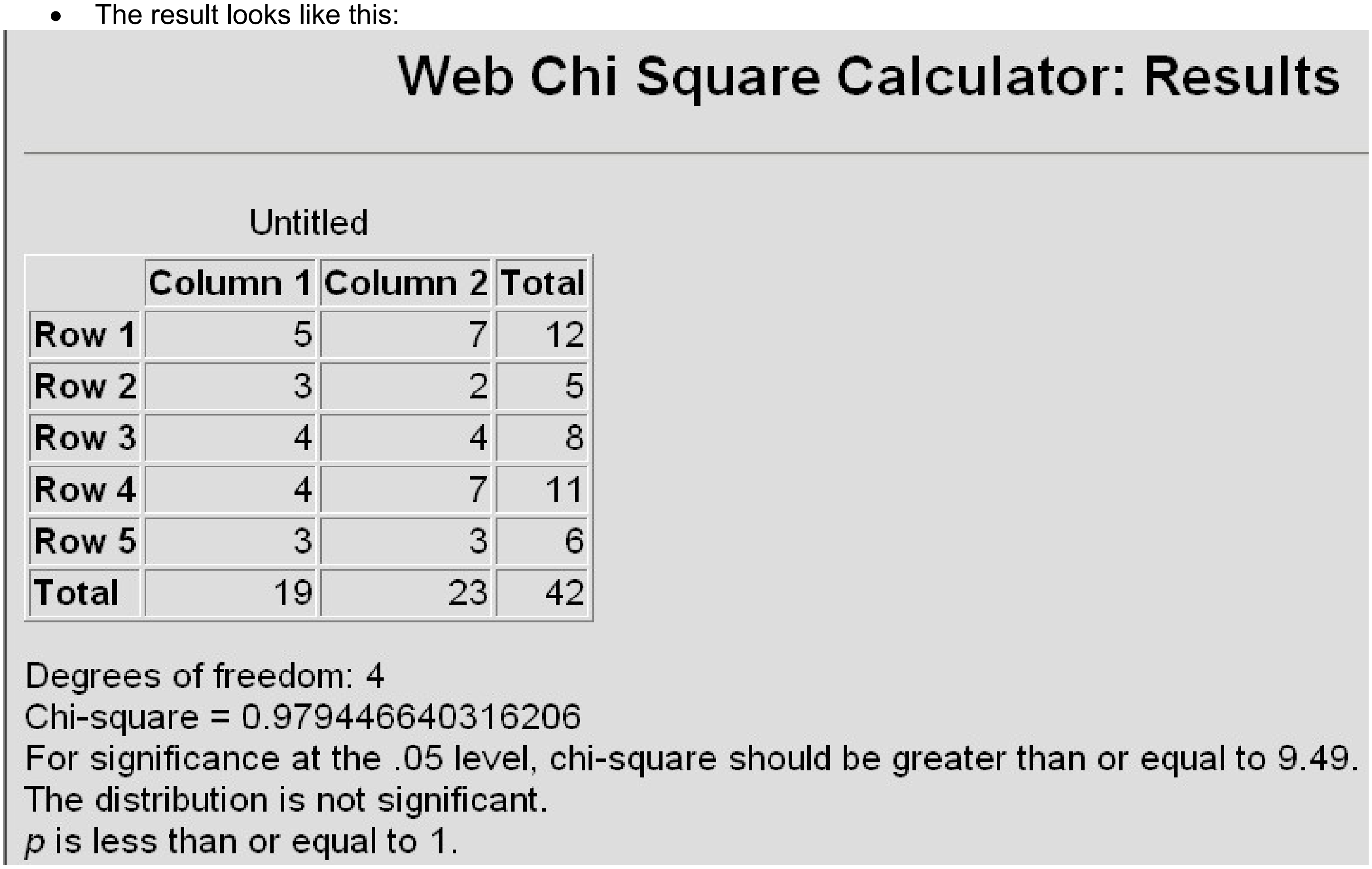
	Column 1	Column 2	**Total**
Row 1	5	7	
Row 2	3	2	
Row 3	4	4	
Row 4	4	7	
Row 5	3	3	
Total			

Results reporting: ◉ normal ○ verbose

Calculate chi square | Reset

- Press the “Calculate chi square” button to get your answer. (You will want to select “normal” unless you want to see how it was calculated.)

- The result looks like this:


Web Chi Square Calculator: Results

Untitled

	Column 1	**Column 2**	**Total**
Row 1	5	7	12
Row 2	3	2	5
Row 3	4	4	8
Row 4	4	7	11
Row 5	3	3	6
Total	19	23	42

Degrees of freedom: 4
Chi-square = 0.979446640316206
For significance at the .05 level, chi-square should be greater than or equal to 9.49.
The distribution is not significant.
p is less than or equal to 1.

(Sometimes it gives information on a critical value, and sometimes it doesn’t.)

On your final test, you won't have to calculate X^2 for a matrix chi-square test. However, you will have to do problems like this …

A college did a study where they compared the evaluations students gave their professors to the grades the students earned in the class. The results for a certain professor are given below:

Student Grade → Professor Rating ↓	A	B	C	D or F
Excellent	3	6	3	0
Average	3	2	2	1
Poor	3	1	2	2

If you did a matrix χ^2 test on this data, how many **degrees of freedom** would there be?

Use the table to find a critical value for χ^2.

If you did a matrix χ^2 test, the calculated value of χ^2 would be 5.44 . **Yes or No**—do people with different grades tend to give their professors significantly different ratings?