1. Using the table (it is best that you confirm by calculator also) determine the t-score needed for constructing a 98% (not 95%) CI for population mean based on a sample of n = 4 the population scores follow a **normal** distribution whose mean and standard deviation are not known.

Use t-table A-98, column with Confidence Level 98% at its base; row with df = n-1 = 4-1=3. The table entry t = 4.541.

2. Refer to #1. Supposing that the sample scores are {2.33, 2.72, 2.74, 2.30} determine the CI in question.

s = 0.240052

xBAR +/- t s / $\sqrt{n}$ = $\boxed{\{1.97746,\ 3.06754\}}$

3. Refer to #1 and #2. Which of the following statements are accurate if the appropriate t score is used and the **population is perfectly normal distributed**?

P(μ in $\bar{x} \pm$ t $\frac{s}{\sqrt{n-1}}$) ~ 0.98

P(μ in $\bar{x} \pm$ t $\frac{s}{\sqrt{n}}$) ~ 0.98

P(μ in $\bar{x} \pm$ t $\frac{s}{\sqrt{n-1}}$) = 0.98 if perfect precision is extended to all calculations

$\boxed{P\left(\mu \text{ in } \bar{x} \pm t\, s / \sqrt{n}\right) = 0.98 \text{ if perfect precision is extended to all calculations}}$

4. A 98% t-based CI is prepared from a sample of n = 4 from a normal population. It takes the form [2.21, 2.36]. Which are correct?

P(μ in [2.21, 2.36]) ~ 0.98

P(μ in [2.21, 2.36]) = 0.98

Neither. The probabilities are 0 or 1, we don't know which.

5. Verify that several entries of the book's chi-square table (you choose some) agree with those from the posted chi-square table. Use your calculator to obtain answers confirming the latter. Report your comparisons here.

df	**P-value, book chi-square, table chi-square**		
6	**0.05**	**12.592**	**12.5916**
14	**0.01**	**21.141**	**29.1412**
70	**0.10**	**85.527**	**85.527**

The URL http://irapilgrim.mcn.org/men01.html links to a paper of R. A. Fisher in which you will find Fisher's views on aspects of Mendel's data.

6. Fisher gives an aggregate chi-square statistic for some of Mendel's experiments. The total DF is 64 followed by the aggregate chi-square statistic. The passage is immediately above Table VI. Locate the passage and read off the value of the aggregate chi-square statistic. Use the posted chi-square table to verify Fisher's claim that the P-value is ~ 1 by getting a good approximate answer from the table. Finish up by using your calculator to more accurately determine the P-value.

	Expectations	Chi-square	Probability of exceding deviations observed
3:1ratios Seed characters	2	0.2779	
3:1ratios Plant characters	5	1.8610	
	7	2.1389	.95
2:1ratios Seed characters	2	0.5982	
2:1ratios Plant characters	6	4.5750	
	8	5.1733	.74
Bifactorial experiment	8	2.8110	.94
Gametic ratios	15	3.6730	.9987
Trifactorial experiment	26	15.3224	.95
Total	64	29.1186	.99987
Illustration of plant variation	20	12.4870	.90
Total	84	41.6056	.99993

Leave out the bottom part of IV (my chi-square table only goes to 79 df). Look at the subtotal of df = 64. Remember, the expectation of a chi-square is equal to its df. This table refers to expectations (df). Total chi-square for that part of the table is 29.1186. Looking up 29.1186 in the body of the chi-square table (for df 64) we find closest entry 30.1729, having a P-value of

0.9999.

7. Refer to #6. What is the interpretation of P-value ~ 1? Did Mendel's data agree rather too well with the models he had for them or did the data disagree rather too strongly from Mendel's models?

It is extremely unlikely that the experiments in question would have produced so small a chi-square (i.e. that the data would have fit the models so well overall). The probability of a fit worse than what was observed is 0.9999. The data agrees with the models rather too well it would seem.

8. Confirm Fisher's P-value for the combined experiments reported in Table IV.

Fisher gets 0.99987. a little more accuracy than my table provides. Your calculator should do better.

9. Use the data of Table 26.2 of your textbook to prepare a chi-square test of the hypothesis that college (i.e. Ag, A&S, Eng., SS) is independent of outcome (i.e. empl, grad schl, other).

Let's suppose that in 26.1 the 2096 graduates are a random sample classified into a 3 by 4 table.

9a. Are the graduates classified in 26.2 really a random sample? What evidence is given for this?

It is not clear to me that this is a sensible application of chi-square. In my view this application of chi-square is simply asking how this data would be interpreted IF it were obtained as the result of sorting a random sample of 2096 undergraduates (from a larger population) into the table. The purpose would be to investigate the hypothesis that IN THE POPULATION the outcomes (employed, grad school, other) are independent of the choice of college.

You may find instances in which the population is somehow thought of as consisting of all the possible ways things might have turned out for these

people all the way through their choice of college through to the outcome. I find that view rather un-interesting, others may not.

9b. What is the population, or what are the populations being sampled?

This is touched on in 9a.

9c. In view of 9a, 9b is this a good illustration of chi-square in your opinion?

Not in my opinion.

9d. Does it seem to you that the row or column totals are fixed in advance?

It seems that they were not. The students and the processes operating in their lives appear to have sorted them into the colleges.

9e. Leaving aside the above, if we formally prepare a chi-square statistic only to illustrate the workings of the method we need the table of "expected counts." Give that table.

Each expected count is calculated

E = (row total)(column total)/(overall total)

For example, for cell "employed Agriculture" the expected count for that cell is

E = 1052 669 / 2096 = 335.777

The contribution of this cell to the overall chi-square is

$$(O - E)^2 / E = (379 - 335.777)^2 / 335.777 = 5.56389$$

9f. Are all expected counts at least 5?

Yes. The smallest E is for the cell "other social science"

$$\boxed{E = 279 \times 322 / 2096 = 42.8616}$$

9g. Are any expected counts close to 5 (or less than 5)? If so, they may be major players in whether the chi-square statistic will be "large". Check for this by determining the standardized residuals (O-E)/$\sqrt{\text{E}}$ for each of the six cells. Identify any cells with unusually small or large standardized residuals. You interpret them as z-scores. (See page 699).

For the cell with smallest E we have a standardized residual of

$$\boxed{(O - E) \Big/ \sqrt{E} = (58 - 42.8616) \Big/ \sqrt{42.8616} = 2.31}$$

While the cells with smaller E are good places to look for trouble, the observed scores O are also players. You have to check them all.

9h. Determine which type of chi-square test of the hypothesis is being advanced by the book (its name).

On pg. 95 the book offer the test of homogeneity which formally uses the same chisquare statistic and same df as the chi-square test of independence we just discussed. The test of homogeneity is used when the row totals are not random but are fixed in advance. One may apply it also if the column totals are fixed in advance.

9i. Determine the P-value. If the conditions had all been met for a proper application of the chi-square method to this data what would you be able to conclude?

As reported on page 697 the total chi-square from all 12 cells of the table is 54.51. The applicable df for this test of the hypothesis (that outcome is independent of college) is (R-1)(C-1) = (3-1)(4-1) = 6. From the chi-square table the P-value << 0.00001 (off the table). Your calculator will do better.

This data departs strongly from what would typically be seen (as measured by chi-square statistic) if the hypothesis of independence.