Note: In this problem set, expressions in green cells match corresponding expressions in the text answers.

```
Clear["Global`*"]
```

I think I will have to bring in my normal table in case I need to use it.

```
(* α is level of significance;
cvm is degrees of freedom; 100000 degrees==∞ *)
α = {0.05, 0.025, 0.010, 0.005, 0.001}
cvm = {{1, 6.31, 12.7, 31.8, 63.7, 318.3},
   {2, 2.92, 4.30, 6.96, 9.92, 22.3},
   {3, 2.35, 3.18, 4.54, 5.84, 10.2}, {4, 2.13, 2.78, 3.75,
    4.60, 7.17}, {5, 2.02, 2.57, 3.36, 4.03, 5.89},
   {6, 1.94, 2.45, 3.14, 3.71, 5.21}, {7, 1.89, 2.36, 3.00,
    3.50, 4.79}, {8, 1.86, 2.31, 2.90, 3.36, 4.50},
   {9, 1.83, 2.26, 2.82, 3.25, 4.30}, {10, 1.81, 2.23,
    2.76, 3.17, 4.14}, {11, 1.80, 2.20, 2.72, 3.11, 4.02},
   {12, 1.78, 2.18, 2.68, 3.05, 3.93}, {13, 1.77, 2.16,
    2.65, 3.01, 3.85}, {14, 1.76, 2.14, 2.62, 2.98, 3.79},
   {15, 1.75, 2.13, 2.60, 2.95, 3.73}, {16, 1.75, 2.12,
    2.58, 2.92, 3.69}, {17, 1.74, 2.11, 2.57, 2.90, 3.65},
   {18, 1.73, 2.10, 2.55, 2.88, 3.61}, {19, 1.73, 2.09,
    2.54, 2.86, 3.58}, {20, 1.72, 2.09, 2.53, 2.85, 3.55},
   {22, 1.72, 2.07, 2.51, 2.82, 3.50}, {24, 1.71, 2.06,
    2.49, 2.80, 3.47}, {26, 1.71, 2.06, 2.48, 2.78, 3.43},
   {28, 1.70, 2.05, 2.47, 2.76, 3.41}, {30, 1.70, 2.04,
    2.46, 2.75, 3.39}, {40, 1.68, 2.02, 2.42, 2.70, 3.31},
   {50, 1.68, 2.01, 2.40, 2.68, 3.26}, {100, 1.66, 1.98,
    2.36, 2.63, 3.17}, {200, 1.65, 1.97, 2.35, 2.60, 3.13},
   {100 000, 1.65, 1.96, 2.33, 2.58, 3.09}};

critCVM =
 Interpolation[Flatten[Table[{{cvm[[i, 1]], α[[j]]}, cvm[[i, j + 1]]},
    {j, 5}, {i, Length[cvm]}], 1]]
```

```
{0.05, 0.025, 0.01, 0.005, 0.001}
```

```
InterpolatingFunction[ Domain {{1., 1.00×10^5}, {0.001, 0.05}}
                       Output scalar ]
```

Below is the ChiSquare table for z.

```
α = {0.05, 0.025, 0.010, 0.005}
cxm = {{1, 3.84, 5.02, 6.63, 7.88}, {2, 5.99, 7.38, 9.21, 10.60},
   {3, 7.81, 9.35, 11.34, 12.84}, {4, 9.49, 11.14, 13.28, 14.86},
   {5, 11.07, 12.83, 15.09, 16.75}, {6, 12.59, 14.45, 16.81, 18.55},
   {7, 14.07, 16.01, 18.48, 20.28}, {8, 15.51, 17.53, 20.09, 21.95},
   {9, 16.92, 19.02, 21.67, 23.59}, {10, 18.31, 20.48, 23.21, 25.19},
   {11, 19.68, 21.92, 24.72, 26.76}, {12, 21.03, 23.34, 26.22, 28.30},
   {13, 22.36, 24.74, 27.69, 29.82}, {14, 23.68, 26.12, 29.14, 31.32},
   {15, 25.00, 27.49, 30.58, 32.80}, {16, 26.30, 28.85, 32.00, 34.27},
   {17, 27.59, 30.19, 33.41, 35.72}, {18, 28.87, 31.53, 34.81, 37.16},
   {19, 30.14, 32.85, 36.19, 38.58}, {20, 31.41, 34.17, 37.57, 40.00},
   {21, 32.7, 35.5, 38.9, 41.4}, {22, 33.9, 36.8, 40.3, 42.8},
   {23, 35.2, 38.1, 41.6, 44.2}, {24, 36.4, 39.4, 43.0, 45.6},
   {25, 37.7, 40.6, 44.3, 46.9}, {26, 38.9, 41.9, 45.6, 48.3},
   {27, 40.1, 43.2, 47.0, 49.6}, {28, 41.3, 44.5, 48.3, 51.0},
   {29, 42.6, 45.7, 49.6, 52.3}, {30, 43.8, 47.0, 50.9, 53.7},
   {40, 55.8, 59.3, 63.7, 66.8}, {50, 67.5, 71.4, 76.2, 79.5},
   {60, 79.1, 83.3, 88.4, 92.0}, {70, 90.5, 95.0, 100.4, 104.2},
   {80, 101.9, 106.6, 112.3, 116.3}, {90, 113.1, 118.1, 124.1, 128.3},
   {100, 124.3, 129.6, 135.8, 140.2}, {200, 1/2 (Sqrt[199 - 1] + 1.64)^2,
    1/2 (Sqrt[199 - 1] + 1.96)^2, 1/2 (Sqrt[199 - 1] + 2.33)^2, 1/2 (Sqrt[199 - 1] + 2.58)^2}};
(*in case degrees of freedom goes above 199,
the applicable number can be substituted in to replace
 199 above in the last line, with the understanding
 that the values in the last line are approximate.*)

critCXM =
 Interpolation[Flatten[Table[{{cxm[[i, 1]], α[[j]]}, cxm[[i, j + 1]]},
    {j, 4}, {i, Length[cxm]}], 1]]
```

```
{0.05, 0.025, 0.01, 0.005}
```

```
InterpolatingFunction[ Domain {{1., 200}, {0.005, 0.05}}  Output scalar ]
```

3. If 100 flips of a coin result in 40 head and 60 tails, can we assert on the 5% level that the coin is fair?

I notice that χ^2 is the basis for the solution by the text of this problem. In this instance the text considers that there is only one degree of freedom, n-1 =2-1, so that c-Chi is equal to

```
cx = critCXM[1, 0.05]
```

```
3.84
```

Using numbered line (1) on p. 1097, the test will look like the following, where K is the number of samples, b_j is the proposed sample result, and e_j is the theoretical result, looking altogether like

$\chi_0{}^2 = \text{Sum}\left[\frac{(b_j - e_j)^2}{e_j}, \{j, 1, K\}\right]$

$$\frac{(40-50)^2}{50} + \frac{(60-50)^2}{50}$$

4

In this case K = 2, and the green cell matches the text answer for the sum described. Because 4 > 3.84, it means that the test fails. In order to see what the limit is, I would have to do

$$\text{Solve}\left[\frac{(n-50)^2}{50} + \frac{(100-n-50)^2}{5} == 3.84, n\right]$$

```
{{n → 45.8221}, {n → 54.1779}}
```

It is convenient in this case that both limits are given by the same **Solve** expression.

5. Can you claim, on a 5% level, that a die is fair if 60 trials give 1, . . . , 6 with absolute frequencies 10, 13, 9, 11, 9, 8?

```
cx = critCXM[5, 0.05]
```

```
11.07
```

```
tris = {10, 13, 9, 11, 9, 8}
{10, 13, 9, 11, 9, 8}
```

```
Total[tris]
60
```

$$\text{Sum}\left[\frac{(n-10)^2}{10}, \{n, \{10, 13, 9, 11, 9, 8\}\}\right]$$

$$\frac{8}{5}$$

The green cells above match the text answers for c and for $\chi_0{}^2$. However, the text answer, while correct in judging the die to be fair, has a typo for the less-than symbol.

7. If a service station had served 60, 49, 56, 46, 68, 39 cars from Monday through Friday between 1 p.m. and 2 p.m., can one claim on a 5% level that the differences are due to randomness? First guess. Then calculate.

First an observation on the service station operating days, which must include either Sat or Sun in order to total six in one week.

The c value is the same as the last problem.

```
cx = critCXM[5, 0.05]
```

```
11.07
```

```
cars = {60, 49, 56, 46, 68, 39}
{60, 49, 56, 46, 68, 39}
```

```
Total[cars]
318
```

```
% / 6
53
```

The particular 2-hour time slot has a value of 53 cars for a non-random reason, or else it is random.

$$N\left[\text{Sum}\left[\frac{(n-53)^2}{53}, \{n, \{60, 49, 56, 46, 68, 39\}\}\right]\right]$$

```
10.2642
```

The c-value and $\chi_0{}^2$ value agree with the answer in the text. The c-value is the greater, therefore the differences are due to randomness. However, consider the following

$$N\left[\text{Sum}\left[\frac{(n-53)^2}{53}, \{n, \{60, 49, 56, 45, 69, 39\}\}\right]\right]$$

```
11.1321
```

Just by shifting one car from the 4th workday to the 5th, the distribution is broken, at least at the 95% significance level, and would, I guess, no longer be considered random.

9. In a table of properly rounded function values, even and odd last decimals should appear about equally often. Test this for the 90 values of $J_1[x]$ in table A1 in appendix 5.

The even-odd occurrences should be just like coin flips. (My even-odd count matched the text answer's on the first try, something of a shock.)

```
lastdigits = {0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1,
   1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1,
   0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 1,
   0, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0
    0, 0, 0, 1, 0, 1, 0, 1, 1, 1};
```

```
Total[lastdigits]
48
```

```
cx = critCXM[1, 0.05]
```

```
3.84
```

$$N\left[\frac{(42-45)^2}{45}+\frac{(48-45)^2}{45}\right]$$

```
0.4
```

The text answer does not show an equation, it only advises to “accept”, which I take it means to accept the null hypothesis, meaning that the sequence of even-odd meets the randomness test. In problem 3, the coin-flipping one, the $\chi_0{}^2$ value was greater than the c-value, and it was judged that the coin was not fair. So here, with the $\chi_0{}^2$ value far smaller than the c-value, the opposite situation exists, namely that the even-odd occurrence is random.

$$\text{Solve}\left[\frac{(n-45)^2}{45}+\frac{(90-n-45)^2}{45}==3.84,\ n\right]$$

```
{{n → 35.7048}, {n → 54.2952}}
```

The following grid shows where the even-odd occurrence would go non-random.

$$\text{Grid}\left[\text{N}\left[\text{Table}\left[\left\{n, \frac{(n-45)^2}{45} + \frac{(90-n-45)^2}{45}\right\}, \{n, 35, 55\}\right]\right], \text{Frame} \to \text{All}\right]$$

35.	4.44444
36.	3.6
37.	2.84444
38.	2.17778
39.	1.6
40.	1.11111
41.	0.711111
42.	0.4
43.	0.177778
44.	0.0444444
45.	0.
46.	0.0444444
47.	0.177778
48.	0.4
49.	0.711111
50.	1.11111
51.	1.6
52.	2.17778
53.	2.84444
54.	3.6
55.	4.44444

13. Mendel's pathbreaking experiments. In a famous plant-crossing experiment, the Austrian Augustinian father Gregor Mendel (1822-1884) obtained 355 yellow and 123 green peas. Test whether this agrees with Mendel's theory according to which the ratio should be 3:1.

There are only two pea possibilities, yellow and green. Even though the probabilities for these two possibilities are different, there are only 2, and to get the number of degrees of freedom, 1 must be subtracted, leaving 1.

```
cx = critCXM[1, 0.05]
```

```
3.84
```

```
355 + 123
```

```
478
```

```
% 0.75
```

```
358.5
```

```
478 - %
```

```
119.5
```

```
N[(123 - 119.5)^2/119.5 + (355 - 358.5)^2/358.5]
```

```
0.136681
```

The green cells above match the answers in the text (the $\chi_0{}^2$ value in the text is 0.137).

15. Radioactivity. Rutherford-Geiger experiments. Using the given sample, test that the corresponding population has a Poisson distribution. x is the number of alpha particles per 7.5-s intervals observed by E. Rutherford and H. Geiger in one of their classical experiments in 1910, and a[x] is the absolute frequency (= number of time periods during which exactly x particles were observed). Use $\alpha = 5\%$.

x	0	1	2	3	4	5	6	7	8	9	10	11	12	≥ 13
a	57	203	383	525	532	408	273	139	45	27	10	4	2	0

I found an interesting reference to this situation at *https : // mathematica.stackexchange.com/questions/172473/how - to - test - if - my - results - have - a - poisson - distribution*, in the answer by iav. No observation intervals were marked by alpha particles of 13 or more in number, so I don't see the argument for including it as a trial. I believe the null observations, 57 in number, are also irrelevant.

```
aaa = {203, 383, 525, 532, 408, 273, 139, 45, 27, 10, 4, 2}
{203, 383, 525, 532, 408, 273, 139, 45, 27, 10, 4, 2}

rb = Range[12]
{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

pt2 = Table[{aaa[[n]], rb[[n]]}, {n, 12}]
{{203, 1}, {383, 2}, {525, 3}, {532, 4}, {408, 5}, {273, 6},
 {139, 7}, {45, 8}, {27, 9}, {10, 10}, {4, 11}, {2, 12}}
```

I haven't gotten comfortable with p-values. If their "intention" is to ward away a possible positive decision, they tend to be extremely tiny. The one below is healthy, and I would tend to accept it as a recommendation.

```
x = pt2[[All, 2]];
h = DistributionFitTest[x,
   PoissonDistribution[Mean@x], "HypothesisTestData"];
h["TestDataTable", All]
```

	Statistic	P-Value
Pearson χ^2	3.37903	0.641765

I do not understand how the text answer deduced that the distribution has 9 degrees of freedom.

```
cx = critCXM[9, 0.05]
16.92
```