2. To ascertain whether a certain die was fair, 1000 rolls of the die were recorded, with the result that the numbers of times the die landed $i, i = 1, 2, 3, 4, 5, 6$ were, respectively, 158, 172, 164, 181, 160, 165. Approximate the p-value of the test that the die was fair

(a) by using the chi-square approximation, and
(b) by using a simulation.

1.

$(a)\ H_0: p_i = \frac{1}{6}$ ， $i = 1, \dots, 6$

$$T = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} = \frac{75.16 + 28.41 + 7.13 + 205.34 + 44.49 + 2.79}{166.67}$$

$= 2.179$

$p - value \approx P\{\chi_5^2 \geq 2.179\} = 0.824$

$Do\ \ not\ \ reject\ \ H_0$

$(b)\ H_0: p_i = \frac{1}{6}$ ， $i = 1, \dots, 6$

$By\ \ simulating\ \ N_i\ \ under\ \ H_0$ ，$then\ \ compute\ \ T = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i}$

$If\ \ T > 2.179$ ，$count += 1$ ，$repeat\ \ 10000\ \ times$

$So\ \ p - value = \frac{count}{10000}$ ，$via\ \ simulation$: `p-value =  0.8211`

15. Consider the following data resulting from three samples:
Compute the approximate p-value of the test that all the data come from a single probability distribution

Sample 1:	121	144	158	169	194	211	242
Sample 2:	99	128	165	193	242	265	302
Sample 3:	129	134	137	143	152	159	170

(a) by using the chi-square approximation, and
(b) by using a simulation.

2.

(a) H_0: $all\ \ data\ \ come\ \ from\ \ same\ \ distribution$

$$Under\ \ H_0\ ，\ \ n = \sum_{i=1}^{m} n_i = 21$$

$R_1 = 2 + 8 + 10 + 13 + 16 + 17 + 18.5 = 84.5$

$R_2 = 1 + 3 + 12 + 15 + 18.5 + 20 + 21 = 90.5$

$R_3 = 4 + 5 + 6 + 7 + 9 + 11 + 14 = 56$

$$R = \frac{12}{n(n+1)} \sum_{i=1}^{m} \frac{\left[R_i - \frac{n_i(n+1)}{2}\right]^2}{n_i} = \frac{12}{21 * 22 * 7}(56.25 + 182.25 + 441)$$

$= 2.5213$

$p - value \approx P\{\chi_2^2 \geq 2.5213\} = 0.2834697$ ，$so\ \ do\ \ not\ \ reject\ \ H_0$

(b) H_0: $all\ \ data\ \ come\ \ from\ \ same\ \ distribution$

$Under\ \ H_0$ ， $n = \sum_{i=1}^{m} n_i = 21$

$U_i \sim U(0,1)$ ，$i = 1, \dots 21$ ，$and\ \ rank\ \ U_i$

$R_1 = rank(U_1 \sim U_7)$

$R_2 = rank(U_8 \sim U_{14})$

$R_3 = rank(U_{15} \sim U_{21})$

$$R = \frac{12}{n(n+1)} \sum_{i=1}^{m} \frac{\left[R_i - \frac{n_i(n+1)}{2}\right]^2}{n_i}$$

$If\ \ R \geq 2.5213$ ，$count = count + 1$ ，$repeat\ \ sim\ \ times$

$p - value \approx count/sim$

Via simulation:

```
p-value =  0.2939
```