Exercise 2.a)

The given observed data is

Observed	*Woman*	not *Woman*	total
you	350	650	1′000
not *you*	19′650	39′350	59′000
total	20′000	40′000	60′000

For each value X_i in the table we compute the probability as

$$p_i = \frac{X_i}{n} = \frac{X_i}{60'000}.$$

Writing this back to a table we get

Probabilities	*Woman*	not *Woman*	total
you	0.00583	0.01083	0.01667
not *you*	0.3275	0.65583	0.98333
total	0.33333	0.66	1

We get the expected value for each table cell by calculating

$$p_{total} \cdot n_{total},$$

where p_{total} is the total probability of the word *you* appearing (0.01667) or not appearing (0.98333), and n_{total} is the total number of words appearing in texts written by a woman (20′000) or not (40′000) – depending on which entry of the table is being calculated.

Expected	*Woman*	not *Woman*	total
you	$0.01667 \cdot 20'000 \approx 333$	$0.01667 \cdot 40'000 \approx 667$	1′000
not *you*	$0.98333 \cdot 20'000 \approx 19'667$	$0.98333 \cdot 40'000 \approx 39'333$	59′000
total	20′000	40′000	60′000

The next step is to compute χ^2 as

$$\chi^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i},$$

where O_i is the observed and E_i is the expected number of a word appearing in a text. Therefore, we have

$$\begin{aligned}\chi^2 &= \frac{(350-333)^2}{333} + \frac{(650-667)^2}{667} + \frac{(19'650-19'667)^2}{19'667} + \frac{(39'350-39'333)^2}{39'333} \\ &\approx 1.32319.\end{aligned}$$

For a significance level of 1% and dof[1] = 1, the critical value for χ^2 is 6.635. Because $\chi^2 \approx 1.32319 < 6.635$, it's clear that the Null-Hypothesis is accepted and that there's no correlation between the term *you* and female authors.

[1]dof = degrees of freedom