Exercise 1.b)

Following the exact same steps as in exercise 1. a), we get the following values:

$$P(Agatha) = \frac{200}{1'000'000} = 0.0002,$$
$$P(Christie) = \frac{300}{1'000'000} = 0.0003.$$

With the Null-Hypothesis we get the mean value

$$\mu_0 = p_0 = P(Agatha) \cdot P(Christie) = 6 \cdot 10^{-8}.$$

Directly computation from the data gives the sample mean and sample variance

$$\bar{X} = p = P(Agatha\ Christie) = \frac{6}{1'000'000} = 6 \cdot 10^{-6},$$
$$s^2 = p \cdot (1-p) \approx p = 6 \cdot 10^{-6}.$$

Therefore, we get a t-value of

$$t_{obs} = \frac{\bar{X} - \mu_0}{\sqrt{\frac{s^2}{n}}} \approx 2.425$$

From the normal table, with a significance level of 1% and dof[1] $= \infty$, we can get the critical value $t_{lim} = 2.576$. The fact that $t_{obs} < t_{lim}$ means that the Null-Hypothesis is not rejected. Therefore, the words *Agatha* and *Christie* appear independently in the text.

[1] dof = degrees of freedom