Exercise 1.a)

$n = 1'000'000$

Observed	*Agatha*	not *Agatha*	total
Christie	20	97	117
not *Christie*	10	999′873	999′883
total	30	999970	1′000′000

Probabilites	*Agatha*	not *Agatha*	total
Christie	0.00002	0.000097	0.000177
not *Christie*	0.00001	0.999873	0.999883
total	0.00003	0.99997	1

Our hypothesis is that the words *Agatha* and *Christie* are independent. Following this assumption, we have

$P(\mathit{Agatha\ Christie}) = P(\mathit{Agatha}) \cdot P(\mathit{Christie}) = 0.00003 \cdot 0.000177 = 0.5 \cdot 10^{-8}$.

The direct estimation from the table above gives

$$P(\mathit{Agatha\ Christie}) = \frac{20}{1'000'000} = 0.2 \cdot 10^{-4}.$$

Based on the observed data we can build a Bernoulli model with parameters

$$\begin{aligned} \bar{x} &= p = 0.2 \cdot 10^{-4}, \\ \sigma^2 &= p\,(1-p) \approx 0.2 \cdot 10^{-4}. \end{aligned}$$

Comparing the two models we get

$$t_{obs} = \sqrt{n}\,\frac{\bar{x}-\mu}{\sigma} = 1000 \cdot \frac{0.2 \cdot 10^{-4} - 0.5 \cdot 10^{-8}}{\sqrt{0.2 \cdot 10^{-4}}} \approx 4.471.$$

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 hypothesis from above is rejected. Therefore, the words *Agatha* and *Christie* are **not** independent.

[1] dof = degrees of freedom