# Homework 3

October 12, 2021

Jose Carlos Munoz

Problem 2)

The First thing we have to do is to Create 4 differnet contingency tables for the 4 different possible splits

$$
\begin{aligned}
\text{Age} = &
\begin{array}{|c|c|c|c|c|}
\hline
\text{Age} & \text{None} & \text{Soft} & \text{Hard} & \text{Total} \\ \hline
\text{Young} & 4 & 2 & 2 & 8 \\
\text{Pre} & 5 & 2 & 1 & 8 \\
\text{Pres} & 6 & 1 & 1 & 8 \\ \hline
\Sigma & 15 & 5 & 4 & 24 \\ \hline
\end{array} \\
\text{Spectable Prescription} = &
\begin{array}{|c|c|c|c|c|}
\hline
\text{Spect} & \text{None} & \text{Soft} & \text{Hard} & \\ \hline
\text{M} & 7 & 2 & 3 & 12 \\
\text{H} & 8 & 3 & 1 & 12 \\ \hline
\Sigma & 15 & 5 & 4 & 24 \\ \hline
\end{array} \\
\text{Astigmatism} = &
\begin{array}{|c|c|c|c|c|}
\hline
\text{Ast} & \text{None} & \text{Soft} & \text{Hard} & \\ \hline
\text{Y} & 8 & 0 & 4 & 12 \\
\text{N} & 4 & 5 & 0 & 12 \\ \hline
\Sigma & 15 & 5 & 4 & 24 \\ \hline
\end{array} \\
\text{Tear Production Rate} = &
\begin{array}{|c|c|c|c|c|}
\hline
\text{TRB} & \text{None} & \text{Soft} & \text{Hard} & \\ \hline
\text{N} & 3 & 5 & 4 & 12 \\
\text{R} & 12 & 0 & 0 & 12 \\ \hline
\Sigma & 15 & 5 & 4 & 24 \\ \hline
\end{array}
\end{aligned}
\tag{1}
$$

From these table we are now going to calculate the expected Table if we assume that the events are independent

$$
\begin{aligned}
\text{Age}_{expected} = &
\begin{array}{|c|c|c|c|}
\hline
\text{Age} & \text{None} & \text{Soft} & \text{Hard} \\ \hline
\text{Young} & 5 & \frac{5}{3} & \frac{4}{3} \\
\text{Pre} & 5 & \frac{5}{3} & \frac{4}{3} \\
\text{Pres} & 5 & \frac{5}{3} & \frac{4}{3} \\ \hline
\end{array} \\
\text{Spectable Prescription}_{expected} = &
\begin{array}{|c|c|c|c|}
\hline
\text{Spect} & \text{None} & \text{Soft} & \text{Hard} \\ \hline
\text{M} & 7.5 & 2.5 & 2 \\
\text{H} & 7.5 & 2.5 & 2 \\ \hline
\end{array} \\
\text{Astigmatism}_{expected} = &
\begin{array}{|c|c|c|c|}
\hline
\text{Ast} & \text{None} & \text{Soft} & \text{Hard} \\ \hline
\text{Y} & 7.5 & 2.5 & 2 \\
\text{N} & 7.5 & 2.5 & 2 \\ \hline
\end{array} \\
\text{Tear Production Rate}_{expected} = &
\begin{array}{|c|c|c|c|}
\hline
\text{TRB} & \text{None} & \text{Soft} & \text{Hard} \\ \hline
\text{N} & 7.5 & 2.5 & 2 \\
\text{R} & 7.5 & 2.5 & 2 \\ \hline
\end{array}
\end{aligned}
\tag{2}
$$

Now that we have both observed and expected tables, we can now find the $\mathcal{X}^2$. To find it, we

must first calculate each of the corresponding observed and expected value, $\frac{(observed-Expected)^2}{Expected}$

$$\begin{aligned}
\text{Age} &= \left|\begin{array}{ccc} \frac{1}{5} & \frac{1}{5} & \frac{1}{3} \\ 0 & \frac{1}{15} & \frac{1}{12} \\ \frac{1}{5} & \frac{4}{15} & \frac{1}{12} \end{array}\right| \\
\text{Spectable Prescription} &= \left|\begin{array}{ccc} \frac{1}{30} & \frac{1}{10} & \frac{1}{2} \\ \frac{1}{30} & \frac{1}{10} & \frac{1}{2} \end{array}\right| \\
\text{Astigmatism} &= \left|\begin{array}{ccc} \frac{1}{30} & 2\frac{1}{2} & 2 \\ \frac{1}{30} & 2\frac{1}{2} & 2 \end{array}\right| \\
\text{Tear Production Rate} &= \left|\begin{array}{ccc} 2\frac{7}{10} & 2\frac{1}{2} & 2 \\ 2\frac{7}{10} & 2\frac{1}{2} & 2 \end{array}\right|
\end{aligned} \tag{3}$$

Once all values are found, we add them all up and get our $\mathcal{X}^2$.

$$\begin{aligned}
\mathcal{X}^2_{age} &= 3\frac{2}{15} \\
\mathcal{X}^2_{SpPr} &= 1\frac{8}{30} \\
\mathcal{X}^2_{As} &= 9\frac{2}{30} \\
\mathcal{X}^2_{TRB} &= 14\frac{4}{10}
\end{aligned} \tag{4}$$

Once we founde these values we find the degrees of freedom for alll of them and see what is th corresponding critical $\mathcal{X}^2$ value. Age is the only one with a different degree of freedom , the other attributes have the same degree of freedom. For age, the Degree of freedom is 4, the rest have a degree of freedom of 2. So for Age its Critical $\mathcal{X}^2$ value is 9.488 and the rest are 5.991. We can now say that we are able to split this node because of all the $\mathcal{X}^2$ values, 2 are higher than the critical value. Which means we can split from either of these. We split in the Tear Production attribute becuse it has the highest $\mathcal{X}^2$ value from either options.

Problem 3