When I was reading your book *Statistical Inference.* I was confused about the concept, p-value in the chapter 8-Hypothesis Testing. Even though I can understand the definition of the valid p-value as a statistic followed by $P_\theta(p(X) \leq \alpha) \leq \alpha$ and we can use it to define the rejection region R: $p(X) \leq \alpha$, but as in other book, the p-value is just like a number to determine whether to reject the null hypothesis.

In the most case, we can calculate a p-value through the distribution when the null hypothesis is true. And as they define, p-value is the probability of observing a sample statistic that is at least as extreme as the observed sample statistic.

I can catch the point that their definition has been contained in your definition. However, it's confusing me that the p-value in their definition had to be calculated after we have seen the data. But in the exercise (8.16) of your book, we have to avoid choosing a level alpha after we've seen the data. I think it's a little bit contradictory here. Because we have to design our test before we see the data, it means the only thing I can do should be to choose a rejection region and after that, test the data to see whether to accept the null hypothesis or not. At last, we can use the alpha or the power to check the reliability of our result. But in this procedure, I can't understand how the p-value (in their definition) work and why it can explain the reliability of the result.

Thus my question is that **how the p-value (defined as the probability of observing a sample statistic that is at least as extreme as the observed sample statistic) can be a indicator of the reliability in the hypothesis test, because we have to calculate it after we've seen the data**.