Data Science Career Track

The Art of Statistics, Chapter 10: Answering questions and claiming discoveries

Take-Away Notes

When hypothesis testing in statistics, we test a **Null Hypothesis (H0)** against an **Alternative Hypothesis (HA)**.

- **The *p*-value** of some observed data is a measure of the incompatibility between that observed data and the Null. The lower the *p*-value, the lower the probability of the observed data *on the assumption of the Null*; hence, since we definitely saw the observed data, the lower the *p*-value, the lower the probability of the Null.
- The *p*-value does *not* measure the probability that the studied hypothesis is true or the probability that the data were produced by randomness alone.
- Typically, thresholds of 0.05 and 0.01 are the **significance levels** (the levels such that if the *p*-value is below them, the Null is rejected).
- *p*-values relate to confidence intervals systematically: if for a given value *x*, the 95% confidence interval excludes *x*, we can reject the null hypothesis of *x* at $p < 0.05$.
- There is a degree of arbitrariness in our selection of a significance level. We need to be careful to avoid misleading conclusions that a result is 'insignificant' if its *p*-value is (for example) 0.05000001.

There are two main types of errors that can arise when doing hypothesis testing.

Truth	**Decision**	**Decision**
	Not rejecting the Null	**Rejecting the Null**
Null	Correct	Type 1 Error (mistakenly

		rejecting a true Null)
Alternative	Type 2 Error (failing to reject a false Null)	Correct

The Type 1 Error rate is known as the **size of a test** and is signaled by the symbol α. The **power of a test** is the probability of correctly rejecting a false Null, and is denoted (1 - β).

Sequential testing can give rise to spurious 'significant' results and occurs when a statistical test is repeatedly carried out on accumulating data, thus inflating the chance of a Type 1 Error.

- The Sequential Probability Ratio Test (SPRT) was devised to counter the risks of sequential testing. SPRT monitors the accumulating evidence about deviations, and can at any time be compared with simple thresholds; as soon as one of those thresholds is crossed, an alert is triggered.