Remarks on Hypothesis Testing

April 3, 2020

The main point of this section is to discuss why the null hypothesis is chosen in the way it is and why we prefer to "fail to reject the null Hypothesis" rather than "accept the null Hypothesis". From the perspective of solving the problems in the book, this section doesn't give too much information, but from a perspective of using statistics to do actual data analysis it is enlightening.

1 Type II errors

As we have seen, the computation of the probability of a Type II error is generally not very straightforward, and requires additional assumption on the unknown parameter. These assumptions often seem unnatural, because by its nature, we don't know a lot about the unknown parameter. One exception is in a classification problem where there are only two possible outcomes, and we wish to determine which of the two outcomes was observed.

The Rejection Region is generally constructed with α in mind (indeed we've using z_α or something similar). So by construction we know the probability of rejecting the null Hypothesis when it is true. Because of this way of choosing RR, β, the probability of a Type II error (not rejecting H_0 when it is false) can actually be quite high. A reason for this can be seen in a simple example:

Example 1.1. *Consider a test for an unknown parameter θ, with $H_0 : \theta = 0$, $H_a : \theta > 0$, a normally distributed test statistic $\hat{\theta}$, that has mean θ and a RR $\hat{\theta} > 1$.*

If the unknown parameter were just slightly bigger than 1, then the probability $\hat{\theta}$ falls outside of RR is almost .5. So β would be almost 1/2. Which is typically an unacceptable error probability.

The above example shows the issue with "accepting the null Hypothesis". If the above experiment were run, and the test statistic lies outside of RR. Then the researcher could try to collect more samples. This deceases the standard error of $\hat{\theta}$, leading to a larger RR, say $\hat{\theta} > .5$. This would then decrease β, for an unknown parameter that happens to be just bigger than 1. The null hypothesis can also be modified and a new experiment run. The failure to reject H_0, might show that a less ambitious null hypothesis should be taken.

2 Meaning of a small α

When we're doing estimation there's always two numbers being measured. One is the size of error, something of the form $|\hat{\theta} - \theta|$. The other is the probability the estimator lies in a certain set, something of the form $\mathbb{P}(|\hat{\theta} - \theta| > k)$. These two quantities don't necessarily imply anything about each other. Rejecting the null Hypothesis, $H_0 : \theta = \theta_0$ with a very small α, doesn't mean the true parameter θ is very far away from θ_0. It just means it is extremely likely they are not equal.

If a test is done with a lot of data, then a very small standard error can be achieved, leading to large RR. So if, for example $H_0 : \mu_1 = \mu_2$ is tested against $H_0 : \mu_1 \neq \mu_2$ and the truth is $\mu_1 = \mu_2 + .1$, it can be confidently observed that they are not equal, but the .1 might not be of practical significance.

Another example can be in election polling, if with a very small α (or p-values) someone reports that a candidate will win an election, it doesn't mean they will win by a lot. It means if the group that was sampled from turns out to vote, then we can be very confident this candidate will win. On the other hand, if a large number of the supporters hear their candidate can confidently win, and don't bother voting, then the polling data becomes not very relevant.

3 Choice of H_0

When doing a 1-sided test $H_0 : \theta = \theta_0$ against $H_a : \theta < \theta_0$, it might seem weird that we don't have a hypothesis that has anything about $\theta > \theta_0$. This is because our main interest was trying to verify H_a. So $\theta > \theta_0$ is not so interesting.

We could have chosen a different null hypothesis $H_0^* : \theta > \theta_0$, but we then would have computed α as being the maximum probability of rejecting H_0 when it is true. This maximum will occur at the boundary point, θ_0. So our choice of H_0 gives the same information.

Since our main goal is to have some control on the probability we accept H_a when it is false the H_0 we use suffices.