StatsErrorAnalysis

Timo Roettger

10/4/2020

1. Size of subset?

What is the proportion of analyzable articles relative to all available articles?

```
## [1] 0.474
```

Almost half of all articles are analyzable

2. Error Analysis

How many individual statistical results could be checked?

```
## [1] 515
```

How many of these are flagged as a statistical errors?

```
## [1] 25
```

How many statistical errors are detected that cross the alpha level (= Decision errors)?

```
## [1] 5
```

What is the proportion of errors?

```
## [1] 0.049
```

How many articles have at least one error?

```
## [1] 18
```

Proportion of articles with at least one error relative to those that statcheck has evaluated?

```
## [1] 37.957
```

Interim summary:

Statcheck was able to check 515 results. 47% of all articles contained at least one statistical result that could be automatically checked. Overall, there were 25 and 18 had at least one error. In other words, 38% of all articles (that were checkable) contained an error.

3. Type of errors (+ manually checked)

There were 11 errors that were difficult to evaluate, 2 errors that could be a rounding error after applying a one-sided test, there were 9 clear rounding errors, 2 typos in which $<$, $>$, or $=$ were errorneously used, and there was one case in which the p-value was very small and was mistakenly specified as $p = 0$.

4. Bias?

How often do errors increase or decrease the actual p-value?

```
## [1] 0.68
```

How many non-rounding errors increase or decrease the actual p-value?

```
## [1] 0.636
```

How many rounding errors increase or decrease the actual p-value?

```
## [1] 0.667
```

It seems as if two thirds of all errors lead to errorneously speficying the p-value as smaller as calculated from the test statistic.