I’ve done three tests on the length of the verdict and the judicial assistant: a linear regression, an ANCOVA and a independent t-test on a small subsection of the dataset.

## Linear regression

1. A linear regression allows me to see whether there is a relation between the length of the verdict and the experience of the judicial assistant. I control for a number of characteristics of a verdict that determine the length of the verdict, such as (1) complexity of the case (measured in number of references), (2) contextual variables related to the case (single-judge / panel, case of first instance), (3) and the outcome of the case. The experience of the judge is also added as a control.
2. Based on previous literature and practice within the Dutch judiciary, judicial assistants have the task to write the (first) draft of the verdict. The result of this regression shows what influence their experience has on the length of the verdict.

|  |  |  |  |
| --- | --- | --- | --- |
| 1. **Table 4.1. Linear regression analysis of the length of 4967 verdicts in administrative law in 2020** | | | |
| 1. **Parameter** | 1. **β** | 1. **S.E.** | 1. ***p*-value** |
| 1. Constant | 1. 15139.61 | 1. 846.20 | .00 |
| 1. Experience of judicial assistant in days | 1. .63 | 1. .08 | .00 |
| 1. Number of references to case law | 1. 511.66 | 1. 31.93 | .00 |
| 1. Number of references to law | 1. 1407.05 | 1. 54.94 | .00 |
| 1. Experience of presiding judge in days | 1. -.14 | 1. .06 | .01 |
| 1. Decision unfounded (opposite: founded) | 1. -1818.98 | 1. 424.57 | .00 |
| 1. Single judge case (opposite: panel) | 1. -9652.49 | 1. 486.66 | .00 |
| 1. Case in first instance (opposite: appeal) | 1. 1107.39 | 1. 478.65 | .02 |
| 1. R2 adjusted = .29. Experience of judges in days. The outcome of the case is unfounded; the opposite is founded (plaintiff is right). Cases with one judge are single judge cases; panel cases could have 3 or 5 judges. Cases in first instance – generally – happen at district courts; higher courts handle appeal cases. Decision, single judge and case in first instance are dummy coded variables. | | | |

1. From the results from the linear regression (table 4.1.), we see that each variable significantly explains the length of a verdict (*p <* .05). We also see the direction of the effects (does the variable cause the verdict to be shorter or longer).
2. Three interesting effects can be seen: (1) the more experienced a judicial assistant is, the longer the verdicts they work on are. For each 100 days of experience, their verdicts become approximately 63 words longer. (2) The more references to case law and law references are included in a verdict, the longer a verdict becomes. (3) Cases handled by a single judge, are significantly shorter (approximately 9652 words) than cases handled by panels. Also interesting to note is that verdicts that are unfounded, are shorter by around 1818 words. The main conclusion we can draw from this table, is that the longer a verdict is, the more experienced the assistant working on the verdict appears to be. The overall model fit was R2 = .29.

## ANCOVA

1. Also interesting to note is that verdicts that are unfounded, are shorter by around 1818 words.

## **Finding effects for one judge and two judicial assistants**

1. From the dataset, I took a subsection for one judge (A.C. Rop) that worked together with two assistants (N.S.J. Letschert, R. Stijnen) on a multiple cases at the district court of Rotterdam. This subsection was taken because it contained a judicial assistant who had an average experience of 1529 days, while the other had an average experience of 7143 days. Additionally, all cases they worked on appeared for a judge in first instance and the outcomes of the cases were equally distributed amongst both assistants. I ran a independent samples t-test on this dataset to test whether the length of their verdicts were significantly different.
2. I compared the length of 15 cases for the novice assistant (*M* = 8232 characters, *SD* = 4818), with the length of 14 cases for the experienced assistant (*M* = 22997 characters, *SD* = 17211). This demonstrated a significant difference in length, *t*(14.90) = -3.10, *p* < .01. There is a mean difference of -14765 characters, with a high effect size, *d =* 1.17. Levene’s test (equal variances) failed (*p* < *0.01*).
   1. Running this test on such a small subsection of the dataset, for one judge and two assistants that all work for the same institution, means that these results are not generalizable. Running similar tests for different institutions, levels of procedure and for panels of judges, each for one judge with two assistants, will result in a multiple testing problem.