Analysis for Talana Cronje

Sean van der Merwe

2019/09/23

### Links

[Consultation Unit](https://www.ufs.ac.za/natagri/departments-and-divisions/mathematical-statistics-and-actuarial-science-home/statistical-consultation-unit/statistical-consultation-unit)

[Consultant](https://seanvdm.co.za)

[Frequently asked questions](https://seanvdm.co.za/post/faqconsult/)

### Introduction and disclaimer

The data collected pertains to a specific set of conditions, and we should not try to extend the results beyond that setting without seriously considering and accounting for any systematic differences between that setting and any broader setting.

The data analysed here is inherently random. If the data collection were to be repeated then the results will differ. We will try to estimate the extent to which this will be the case, but those estimates are themselves uncertain.

While [the computer software used](https://en.wikipedia.org/wiki/R_(programming_language)) is tried and tested, the analysis involves multiple human elements. Both the client and statistician may have introduced human error at various stages of the research process.

**Thus, no guarantee can ultimately be given on the correctness of any findings below.**

Links are given occasionally to sources explaining the statistical concepts addressed. These are colloquial sources to aid understanding, not reference sources. However, most of these sources link to reference sources should the researcher require academic references for specific topics.

### Basic data properties

The significance level chosen is 0.05. [This means that we only look at results where the p-value is less than](https://en.wikipedia.org/wiki/P-value) 0.05, as other results could easily just be chance variation.

We separate the the factors that are presumed to cause differences in response (Cultivar and Herbicide) from the measurable outcomes that are presumed to be affected by the factors (TuberNumber, TuberMass, HaulmNumEmerged, HaulmNumAlive, HeightMin, HeightMax, AvgHeight, DM, NumPlants).

### Multivariate analysis

When tests are done on multiple factors or multiple outcomes then it increases the false positive rate. We normally allow a small posibility for false positives in order to maintain a reasonable power to detect true positives; but repeated testing increases this rate to the point were false positives are virtually guaranteed, which is not acceptable. Two approaches are used to combat this problem.

The first and best is to do a global test of general significance. The second is [we adjust the probabilities to account for the fact that we are doing multiple tests.](https://en.wikipedia.org/wiki/Multiple_comparisons_problem) According to the [Holm-Bonferroni method](https://en.wikipedia.org/wiki/Holm%E2%80%93Bonferroni_method) we compare the smallest p-value to /(number of tests)=0.0125, the second smallest to /(number of tests - 1), and so on.

As a starting point, we convert each explanatory variable into a binary (zero-one) variable based on whether it has a value of zero or larger than zero. Then we can calculate the average rate of response for each treatment within cultivar, and overall.

Overall response to treatment for cultivar Mondial where 0 is no tubers and 1 is at least some tubers.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Control | T0.20x | T0.40x | T0.60x | T0.80x |
| TuberNumber | 1 | 0.25 | 0.00 | 0 | 0.0 |
| TuberMass | 1 | 0.25 | 0.00 | 0 | 0.0 |
| HaulmNumEmerged | 1 | 0.75 | 0.25 | 0 | 0.5 |
| HaulmNumAlive | 1 | 0.75 | 0.25 | 0 | 0.5 |
| HeightMin | 1 | 0.75 | 0.25 | 0 | 0.5 |
| HeightMax | 1 | 0.75 | 0.25 | 0 | 0.5 |
| AvgHeight | 1 | 0.75 | 0.25 | 0 | 0.5 |
| DM | 1 | 0.75 | 0.25 | 0 | 0.5 |
| NumPlants | 1 | 0.75 | 0.25 | 0 | 0.5 |

Overall response to treatment for cultivar Sifra where 0 is no tubers and 1 is at least some tubers.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Control | T0.20x | T0.40x | T0.60x | T0.80x |
| TuberNumber | 1 | 0.00 | 0.25 | 0.25 | 0.00 |
| TuberMass | 1 | 0.00 | 0.25 | 0.25 | 0.00 |
| HaulmNumEmerged | 1 | 0.75 | 0.25 | 0.25 | 0.25 |
| HaulmNumAlive | 1 | 0.50 | 0.25 | 0.25 | 0.25 |
| HeightMin | 1 | 0.75 | 0.25 | 0.25 | 0.25 |
| HeightMax | 1 | 0.75 | 0.25 | 0.25 | 0.25 |
| AvgHeight | 1 | 0.75 | 0.25 | 0.25 | 0.25 |
| DM | 1 | 0.75 | 0.25 | 0.25 | 0.00 |
| NumPlants | 1 | 0.75 | 0.25 | 0.25 | 0.25 |

Overall response to treatment for cultivar Innovator where 0 is no tubers and 1 is at least some tubers.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Control | T0.20x | T0.40x | T0.60x | T0.80x |
| TuberNumber | 1 | 0.25 | 0 | 0.25 | 0 |
| TuberMass | 1 | 0.25 | 0 | 0.25 | 0 |
| HaulmNumEmerged | 1 | 1.00 | 1 | 0.50 | 0 |
| HaulmNumAlive | 1 | 1.00 | 1 | 0.50 | 0 |
| HeightMin | 1 | 1.00 | 1 | 0.50 | 0 |
| HeightMax | 1 | 1.00 | 1 | 0.50 | 0 |
| AvgHeight | 1 | 1.00 | 1 | 0.50 | 0 |
| DM | 1 | 1.00 | 1 | 0.33 | 0 |
| NumPlants | 1 | 1.00 | 1 | 0.50 | 0 |

Overall response to treatment for cultivar Lanorma where 0 is no tubers and 1 is at least some tubers.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Control | T0.20x | T0.40x | T0.60x | T0.80x |
| TuberNumber | 0.75 | 0.50 | 0 | 0.00 | 0.25 |
| TuberMass | 0.75 | 0.50 | 0 | 0.00 | 0.25 |
| HaulmNumEmerged | 1.00 | 0.75 | 0 | 0.25 | 0.25 |
| HaulmNumAlive | 1.00 | 0.75 | 0 | 0.25 | 0.25 |
| HeightMin | 1.00 | 0.75 | 0 | 0.25 | 0.25 |
| HeightMax | 1.00 | 0.75 | 0 | 0.25 | 0.25 |
| AvgHeight | 1.00 | 0.75 | 0 | 0.25 | 0.25 |
| DM | 1.00 | 0.67 | 0 | 0.25 | 0.25 |
| NumPlants | 1.00 | 0.75 | 0 | 0.25 | 0.25 |

Overall response to treatment for all cultivars where 0 is no tubers and 1 is at least some tubers.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Control | T0.20x | T0.40x | T0.60x | T0.80x |
| TuberNumber | 0.94 | 0.25 | 0.06 | 0.12 | 0.06 |
| TuberMass | 0.94 | 0.25 | 0.06 | 0.12 | 0.06 |
| HaulmNumEmerged | 1.00 | 0.81 | 0.38 | 0.25 | 0.25 |
| HaulmNumAlive | 1.00 | 0.75 | 0.38 | 0.25 | 0.25 |
| HeightMin | 1.00 | 0.81 | 0.38 | 0.25 | 0.25 |
| HeightMax | 1.00 | 0.81 | 0.38 | 0.25 | 0.25 |
| AvgHeight | 1.00 | 0.81 | 0.38 | 0.25 | 0.25 |
| DM | 1.00 | 0.80 | 0.29 | 0.20 | 0.20 |
| NumPlants | 1.00 | 0.81 | 0.38 | 0.25 | 0.25 |

###### Due to the extreme dependence between the measurements, we cannot tests the measurements jointly for interaction with the treatment.

We consider the measurements one at a time, remembering that many are directly dependent on each other.

### Univariate testing

In cases where a variable is numeric we often use [Analysis of Variance (ANOVA)](https://en.wikipedia.org/wiki/Analysis_of_variance) to check whether there is a difference in numeric response between groups.

However, the ANOVA assumes that the unexplained variation follows a bell-shaped density pattern. This assumption does not hold at all for this data, inflating the probability of a false positive result greatly. Sometimes the shape can be corrected using a [power transformation](https://en.wikipedia.org/wiki/Power_transform) but not in this case.

An alternative is a non-parametric procedure such as [the Kruskal-Wallis approach](https://en.wikipedia.org/wiki/Kruskal%E2%80%93Wallis_one-way_analysis_of_variance). It is similar in principle, but safer (though less powerful).

We also do [(Chi Square) tests for each binary response](https://en.wikipedia.org/wiki/Pearson%27s_chi-squared_test). We make the assumption that the expected number of responses in every treatment is the same for a single variable. We then calculate the probability of seeing differences in the categories as large or larger that what we observe under the assumption. If this probability is small we conclude that there is evidence of differences.

In the following graphs we show the median and inter-quartile range for each measure, split by cultivar. These do not indicate an statistical significance, they merely summarise the patterns in the data visually.





#### Test results

Based on the results of [the Shapiro-Wilk test for Normality](https://en.wikipedia.org/wiki/Shapiro%E2%80%93Wilk_test) below and a visual analysis of the data, it is clear that an ANOVA is not appropriate for this data as its assumptions are markedly violated.

###### Shapiro-Wilk p-values:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Mondial | Sifra | Innovator | Lanorma | All Cultivars |
| TuberNumber | 0.0000 | 0.0038 | 0.0254 | 0.0009 | 0 |
| TuberMass | 0.0000 | 0.0009 | 0.0000 | 0.0072 | 0 |
| HaulmNumEmerged | 0.4410 | 0.1033 | 0.0128 | 0.0068 | 0 |
| HaulmNumAlive | 0.4410 | 0.0061 | 0.0128 | 0.0090 | 0 |
| HeightMin | 0.3796 | 0.6412 | 0.0034 | 0.0016 | 0 |
| HeightMax | 0.0163 | 0.0646 | 0.5152 | 0.3240 | 0 |
| AvgHeight | 0.3088 | 0.0762 | 0.8512 | 0.0924 | 0 |
| DM | 0.0000 | 0.0116 | 0.0263 | 0.0008 | 0 |
| NumPlants | 0.0080 | 0.0814 | 0.0131 | 0.0770 | 0 |

Instead we consider the results of the Kruskal-Wallis tests and note that many p-values are small enough to reject the assumption of no difference in location between treatments. This implies that the treatment does have an effect on the measurements obtained.

###### Kruskal-Wallis p-values:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Mondial | Sifra | Innovator | Lanorma | All Cultivars |
| TuberNumber | 0.0025 | 0.0124 | 0.0056 | 0.0578 | 0 |
| TuberMass | 0.0025 | 0.0078 | 0.0056 | 0.0699 | 0 |
| HaulmNumEmerged | 0.0126 | 0.0176 | 0.0059 | 0.0148 | 0 |
| HaulmNumAlive | 0.0126 | 0.0236 | 0.0059 | 0.0148 | 0 |
| HeightMin | 0.0341 | 0.0948 | 0.0208 | 0.0468 | 0 |
| HeightMax | 0.0085 | 0.0183 | 0.0058 | 0.0235 | 0 |
| AvgHeight | 0.0085 | 0.0218 | 0.0071 | 0.0250 | 0 |
| DM | 0.0210 | 0.0351 | 0.0175 | 0.0566 | 0 |
| NumPlants | 0.0111 | 0.0140 | 0.0048 | 0.0172 | 0 |

The above is corroborated by the tests, although those can only be done on all the cultivars grouped together.

###### Chi-square p-values:

|  |  |
| --- | --- |
|  | x |
| TuberNumber | 0.0001 |
| TuberMass | 0.0001 |
| HaulmNumEmerged | 0.0072 |
| HaulmNumAlive | 0.0100 |
| HeightMin | 0.0072 |
| HeightMax | 0.0072 |
| AvgHeight | 0.0072 |
| DM | 0.0103 |
| NumPlants | 0.0072 |

Lastly, we give the ANOVA results for interest.

###### ANOVA p-values:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Mondial | Sifra | Innovator | Lanorma | All Cultivars |
| TuberNumber | 0.0005 | 0.0014 | 0.0000 | 0.0113 | 0 |
| TuberMass | 0.0015 | 0.0020 | 0.0000 | 0.0761 | 0 |
| HaulmNumEmerged | 0.0000 | 0.0000 | 0.0001 | 0.0084 | 0 |
| HaulmNumAlive | 0.0000 | 0.0000 | 0.0001 | 0.0080 | 0 |
| HeightMin | 0.0618 | 0.0792 | 0.0402 | 0.1556 | 0 |
| HeightMax | 0.0000 | 0.0009 | 0.0006 | 0.0092 | 0 |
| AvgHeight | 0.0000 | 0.0027 | 0.0005 | 0.0285 | 0 |
| DM | 0.0333 | 0.0000 | 0.0000 | 0.1218 | 0 |
| NumPlants | 0.0000 | 0.0000 | 0.0000 | 0.0214 | 0 |

**NB: An ANOVA or related test only tests whether an effect might be present, it does not explain what that effect might be.** To investigate the effect further we must do additional testing.

#### Pairwise testing

A common approach to investigating the differences between means is [Tukey’s HSD test](https://en.wikipedia.org/wiki/Tukey%27s_range_test). However, the assumptions of this test do not hold for our data, so we will use a non-parametric equivalent. It is known as the pairwise Wilcoxon or Mann-Whitney test.

The results are not significant for any one cultivar on its own, so only the tables for the cultivars as a group are given below.

##### Pairwise comparisons using Wilcoxon rank sum test

For measurement: Count of tubers

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Control | T0.20x | T0.40x | T0.60x |
| T0.20x | 0.0000547 | NA | NA | NA |
| T0.40x | 0.0000238 | 0.8982193 | NA | NA |
| T0.60x | 0.0000498 | 1.0000000 | 1 | NA |
| T0.80x | 0.0000161 | 0.8622252 | 1 | 1 |

For measurement: Mass of tubers

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Control | T0.20x | T0.40x | T0.60x |
| T0.20x | 0.00027 | NA | NA | NA |
| T0.40x | 0.00002 | 1 | NA | NA |
| T0.60x | 0.00002 | 1 | 1 | NA |
| T0.80x | 0.00002 | 1 | 1 | 1 |

For measurement: Number of haulms that emerged from the ground

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Control | T0.20x | T0.40x | T0.60x |
| T0.20x | 0.0005535 | NA | NA | NA |
| T0.40x | 0.0000313 | 0.1932502 | NA | NA |
| T0.60x | 0.0000205 | 0.0262718 | 0.9258297 | NA |
| T0.80x | 0.0000154 | 0.0103884 | 0.9040684 | 0.9604399 |

For measurement: Number of haulms emerged from the ground that were alive

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Control | T0.20x | T0.40x | T0.60x |
| T0.20x | 0.0004356 | NA | NA | NA |
| T0.40x | 0.0000315 | 0.4403094 | NA | NA |
| T0.60x | 0.0000206 | 0.0616121 | 0.9258297 | NA |
| T0.80x | 0.0000155 | 0.0338636 | 0.9040684 | 0.9604399 |

For measurement: Maximum haulm height

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Control | T0.20x | T0.40x | T0.60x |
| T0.20x | 0.1880690 | NA | NA | NA |
| T0.40x | 0.0002941 | 0.0351566 | NA | NA |
| T0.60x | 0.0008461 | 0.0351566 | 1 | NA |
| T0.80x | 0.0002982 | 0.0222084 | 1 | 1 |

For measurement: Minimum haulm height

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Control | T0.20x | T0.40x | T0.60x |
| T0.20x | 0.0001894 | NA | NA | NA |
| T0.40x | 0.0000300 | 0.1468943 | NA | NA |
| T0.60x | 0.0000182 | 0.0339362 | 1.0000000 | NA |
| T0.80x | 0.0000113 | 0.0031949 | 0.9054491 | 1 |

For measurement: Average haulm height

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Control | T0.20x | T0.40x | T0.60x |
| T0.20x | 0.0008295 | NA | NA | NA |
| T0.40x | 0.0000322 | 0.1043146 | NA | NA |
| T0.60x | 0.0000322 | 0.0543137 | 1.00000 | NA |
| T0.80x | 0.0000116 | 0.0042827 | 0.93806 | 1 |

For measurement: Dry mass

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Control | T0.20x | T0.40x | T0.60x |
| T0.20x | 0.0070014 | NA | NA | NA |
| T0.40x | 0.0004086 | 0.0653179 | NA | NA |
| T0.60x | 0.0001801 | 0.0128688 | 1 | NA |
| T0.80x | 0.0000996 | 0.0011512 | 1 | 1 |

For measurement: Count of plants

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Control | T0.20x | T0.40x | T0.60x |
| T0.20x | 0.0007182 | NA | NA | NA |
| T0.40x | 0.0000493 | 0.1358637 | NA | NA |
| T0.60x | 0.0000188 | 0.0143727 | 0.8416938 | NA |
| T0.80x | 0.0000127 | 0.0044471 | 0.8395696 | 0.8425297 |

We see in the tables above that the 80% treatment and sometimes the 60% treatment are better than the 20% treatment. However, we see no evidence in the data that treatment at 80% or 60% is better than 40%. *It may well be the case that higher treatments would show improved results if the experiment were expanded but this cannot be established in either direction via the above methods.*

### Regression analysis

The goal here is to determine whether there is some ideal level of treatment, based on the measurements taken. This is not a simple task as any simple model will not address the research question. Instead we fit a model that more closely resembles the true nature of the data. The current gold standard for implementing such complex models is [the STAN interface](https://mc-stan.org/). We will attempt to use this interface via [RSTAN](https://mc-stan.org/users/interfaces/rstan) to implement our models.

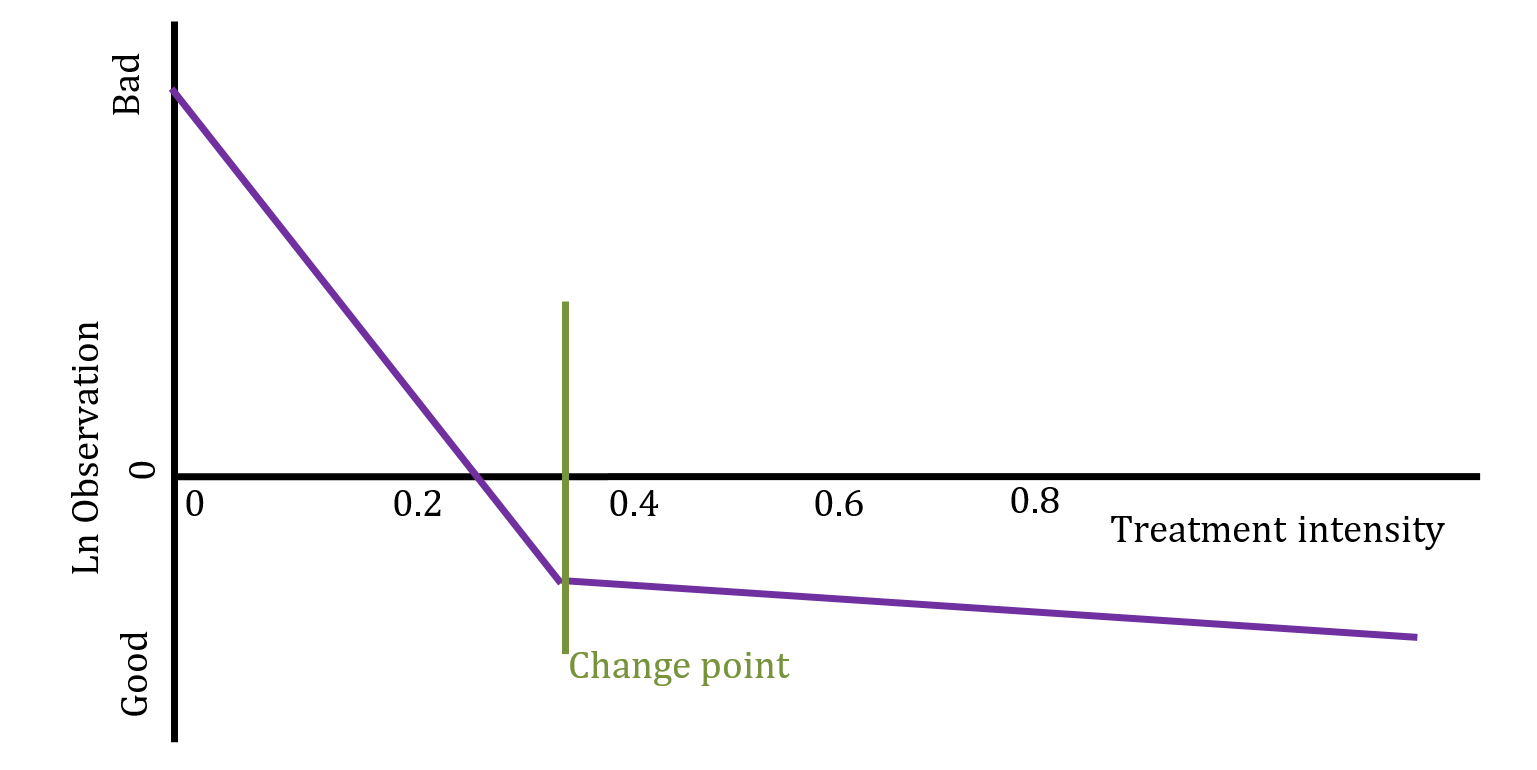
For this analysis, assume that the treatment increases in effectiveness as the dose is increased, but only up to a point, from which a higher dose no longer adds value to the enterprise. To test this we fit a model with two slopes and an unknown change point, then test whether the slopes fit our assumption. If the slopes do meet our expectations then we report the estimated change point as a proposed optimal treatment level.

This type of model is called a bi-phasic regression. Given an unknown changepoint , the model is formally defined as .

However, as noted previously, the observed data does not follow a Normal distribution. To bring the data in line with this assumption we make adjustments. First, we take the log of the observations. Since we can’t take the log of zero, we must treat them separately. The zeros do not occur randomly, they carry valuable information that must be incorporated into the model.

The approach implemented here is to consider zeros as values censored below a lower measurement limit. In the real experiment tubers cannot occur in non-whole numbers, but if it was possible then it would be reasonable to assume that the data generating process would want to produce numbers that are close to but not exactly zero, and certainly less than one. We assume that, based on a level of treatment, the experiment should have produced a real number of tubers greater than zero and less than one in all the cases where we observe a zero. On a log scale this implies a negative number of log tubers.

The proposed model is illustrated below:



Considering measure: Count of tubers

The p-value for the changepoint test is: 0.0037 and for the test of a flat slope after the changepoint the p-value is: 0.6383

Which means we assume there is a changepoint, with expected value 24% of recommended treatment, and 95% HPD interval of 10% to 42%.

Considering measure: Mass of tubers

The p-value for the changepoint test is: 0.006 and for the test of a flat slope after the changepoint the p-value is: 0.7757

Which means we assume there is a changepoint, with expected value 27% of recommended treatment, and 95% HPD interval of 10% to 46%.

Considering measure: Number of haulms that emerged from the ground

The p-value for the changepoint test is: 0.0583 and for the test of a flat slope after the changepoint the p-value is: 0.501

Which means there is not enough evidence in the data to suggest a changepoint followed by a flat slope.

Considering measure: Number of haulms emerged from the ground that were alive

The p-value for the changepoint test is: 0.0537 and for the test of a flat slope after the changepoint the p-value is: 0.407

Which means there is not enough evidence in the data to suggest a changepoint followed by a flat slope.

Considering measure: Maximum haulm height

The p-value for the changepoint test is: 0.164 and for the test of a flat slope after the changepoint the p-value is: 0.7747

Which means there is not enough evidence in the data to suggest a changepoint followed by a flat slope.

Considering measure: Minimum haulm height

The p-value for the changepoint test is: 0.1113 and for the test of a flat slope after the changepoint the p-value is: 0.584

Which means there is not enough evidence in the data to suggest a changepoint followed by a flat slope.

Considering measure: Average haulm height

The p-value for the changepoint test is: 0.1273 and for the test of a flat slope after the changepoint the p-value is: 0.6807

Which means there is not enough evidence in the data to suggest a changepoint followed by a flat slope.

Considering measure: Dry mass

The p-value for the changepoint test is: 0.0263 and for the test of a flat slope after the changepoint the p-value is: 0.2443

Which means we assume there is a changepoint, with expected value 34% of recommended treatment, and 95% HPD interval of 10% to 66%.

Considering measure: Count of plants

The p-value for the changepoint test is: 0.0267 and for the test of a flat slope after the changepoint the p-value is: 0.2743

Which means we assume there is a changepoint, with expected value 31% of recommended treatment, and 95% HPD interval of 10% to 59%.



### Conclusions

For a rough overall impression, I would say that a roughly 40% treatment appears optimal statistically based on the experimental results obtained.

Please contact me should any further clarity be required.



