**Chemometrics: Analysis of Variance Example**

The heights of 4 individuals of each of 3 nationalities is given below (in cm).

|  |  |  |
| --- | --- | --- |
| **American** | **Dutch** | **Irish** |
| 180 | 190 | 170 |
| 190 | 175 | 185 |
| 175 | 185 | 175 |
| 175 | 190 | 170 |

Construct a “long” data object containing the height values and nationalities.

|  |
| --- |
| > head(Data)  Height Nation  1 180 American  2 190 American  3 175 American  4 175 American  5 190 Dutch  6 175 Dutch |

The mean height of the 12 subjects is 180 cm.

The mean height for each nationality is 180cm, 185 cm and 175cm respectively.

Let us consider if nationality is a factor in heights. We can perform an Analysis of Variance procedure using the aov() function.

|  |
| --- |
| > aov(Height~Nation)  Call:  aov(formula = Height ~ Nation)  Terms:  Nation Residuals  Sum of Squares 200 450  Deg. of Freedom 2 9  Residual standard error: 7.071068  Estimated effects may be unbalanced |

|  |
| --- |
| > summary(aov(Height~Nation))  Df Sum Sq Mean Sq F value Pr(>F)  Nation 2 200 100 2 0.191  Residuals 9 450 50 |

N.B. “Nations” is the between groups sum of squares.

“Residuals” is the within groups sum of squares.

Can it be stated at a 5% significance level that there is no difference between the mean heights of these three nationalities? State your null and alternative

The null and alternative hypotheses are as follows:

*H0: The population means are equal for each of the three countries .*

*H1: The population means are not equal for each of the three countries . (At least two are different)*

We consider the p-value ( here written as Pr(>F)) which is 0.191.

We have fail to reject the null hypothesis. There is insufficient evidence to suggest that there is any difference in average heights.