

Assignments for session on “Non- Linear Models”

a. What are the assumptions of ANOVA, test it out?

Following assumptions are made for using the ANOVA test

Each group sample is drawn from a normally distributed population

All populations have a common variance

All samples are drawn independently of each other

Within each sample, the observations are sampled randomly and independently of each other

Factor effects are additive

Validating normality assumption

```
x <- yeastdata$nuc
```

```
h <- hist(x, breaks = 10, col = "green", xlab = "Score of discriminant analysis of nuclear localization signals
```

```
of nuclear and non-nuclear proteins", main = "Histogram of Score with normal curve")
```

```
xfit <- seq(min(x), max(x), length= 40)
```

```
yfit <- dnorm(xfit, mean = mean(x), sd = sd(x))
```

```
yfit <- yfit*diff(h$mids[1:2])* length(x))
```

```
lines(xfit, yfit , col = "yellow", lwd = 2)
```

Validating the Skewness or kurtosis and variances

```
library(psych)
```

```
describe(yeastdata)
```

Validating the outliers

```
boxplot(yeastdata)
```

b. Why ANOVA test? Is there any other way to answer the above question?

Well, ANOVA test ensures uniform treatment to the whole set of data for chosen level of significance and minimizes the risk of Type 1 error.

Regression analysis is the another method which could be deployed for the above question.