

# feature\_\_significance

*Benjamin Hillmann*

*October 19, 2015*

```
infiles = c('genes_CEFITH_standardized.txt', 'RDP_family_AH_standardized.txt',
            'RDP_genus_AH_standardized.txt', 'OTU_AH_standardized.txt')

preprocess_mat = function(infile) {
  df = load_data_table(infile)
  mat = t(df[, -1])
  colnames(mat) = df[, 1]
  return(mat)
}

otu_mat = preprocess_mat(infiles[4])
genes_mat = preprocess_mat(infiles[1])

log_transform = function(x) {
  x = x[!is.na(x)]
  constant = min(x[x > 0])/2.
  x = x + constant
  x = x/sum(x)
  log(x)
}

prot_000000000877 = log_transform(genes_mat[, 'prot_000000000877'])
otu_11 = log_transform(otu_mat[, 'OTU_11'])

mapping_maria = load_data_table('mapping_maria_BL.txt')

dependent = as.numeric(mapping_maria$Persistence)
lm_prot_000000000877 = lm(dependent~prot_000000000877)
lm_otu_11 = lm(dependent~otu_11)

summary(lm_prot_000000000877)
```

```
##
## Call:
## lm(formula = dependent ~ prot_000000000877)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.72078 -0.21300  0.02914  0.27922  0.66513
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.34870    0.24607   1.417 0.171846
## prot_000000000877 -0.26202    0.06032  -4.344 0.000315 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.3619 on 20 degrees of freedom
## Multiple R-squared: 0.4854, Adjusted R-squared: 0.4597
## F-statistic: 18.87 on 1 and 20 DF, p-value: 0.0003151
```

```
summary(lm_otu_11)
```

```
##
## Call:
## lm(formula = dependent ~ otu_11)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4631 -0.3159 -0.1248  0.3582  1.1098
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.73439    0.27737   2.648  0.0154 *
## otu_11       -0.15411    0.06385  -2.414  0.0255 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.444 on 20 degrees of freedom
## Multiple R-squared: 0.2256, Adjusted R-squared: 0.1868
## F-statistic: 5.825 on 1 and 20 DF, p-value: 0.02551
```

```
coxtest(lm_prot_000000000877, lm_otu_11)
```

```
## Cox test
##
## Model 1: dependent ~ prot_000000000877
## Model 2: dependent ~ otu_11
##              Estimate Std. Error  z value Pr(>|z|)
## fitted(M1) ~ M2  -2.0702    0.85506  -2.4212  0.01547 *
## fitted(M2) ~ M1  -6.9702    0.68932 -10.1118 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
jtest(lm_prot_000000000877, lm_otu_11)
```

```
## J test
##
## Model 1: dependent ~ prot_000000000877
## Model 2: dependent ~ otu_11
##              Estimate Std. Error t value Pr(>|t|)
## M1 + fitted(M2)  0.53412    0.35168  1.5188 0.145285
## M2 + fitted(M1)  0.86657    0.23973  3.6148 0.001845 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```