

Statistical tests

TP-Transformer

Operations

```
# Read operation data
tp_op_data <- read.csv('results_TP_aggregate_op.csv',
                      header=FALSE, sep = ',')
names(tp_op_data) <- c('Distances', 'Matched', 'Differences')
tp_op_data$Matched[tp_op_data$Matched == 0] <- -1

# Linear model for operations
attach(tp_op_data)
tp_op_model <- lm(Differences ~ 1 + Distances + Matched + Distances*Matched)
summary(tp_op_model)

##
## Call:
## lm(formula = Differences ~ 1 + Distances + Matched + Distances *
##     Matched)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.294  -4.587  -1.972   1.846  83.635
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.219587   0.005945   878.0  <2e-16 ***
## Distances       0.426533   0.001284   332.2  <2e-16 ***
## Matched        -0.877403   0.005945  -147.6  <2e-16 ***
## Distances:Matched 0.245936   0.001284   191.6  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.138 on 9209636 degrees of freedom
## Multiple R-squared:  0.01234,    Adjusted R-squared:  0.01234
## F-statistic: 3.835e+04 on 3 and 9209636 DF,  p-value: < 2.2e-16

# Remove from memory
detach(tp_op_data)
rm(tp_op_data)
```

Digits

```
# Read x data
tp_x_data <- read.csv('results_TP_aggregate_x.csv',
                     header=FALSE, sep = ',')
names(tp_x_data) <- c('Distances', 'Matched', 'Differences')
tp_x_data$Matched[tp_x_data$Matched == 0] <- -1

# Linear model for digits
attach(tp_x_data)
```

```
tp_x_model <- lm(Differences ~ 1 + Distances + Matched + Distances*Matched)
summary(tp_x_model)
```

```
##
## Call:
## lm(formula = Differences ~ 1 + Distances + Matched + Distances *
##     Matched)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.9780 -1.9877 -0.3706  1.7539  6.3653
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.5184663   0.0018100   838.9  <2e-16 ***
## Distances      0.3114875   0.0003251   958.1  <2e-16 ***
## Matched       -1.0887822   0.0018100  -601.5  <2e-16 ***
## Distances:Matched 0.2121703   0.0003251   652.6  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.357 on 15936300 degrees of freedom
## Multiple R-squared:  0.05618,    Adjusted R-squared:  0.05618
## F-statistic: 3.162e+05 on 3 and 15936300 DF,  p-value: < 2.2e-16
```

```
# Remove from memory
```

```
detach(tp_x_data)
rm(tp_x_data)
```

Constituents

```
# Read constituent data
tp_con_data <- read.csv('results_TP_aggregate_con.csv',
                        header=FALSE, sep = ',')
names(tp_con_data) <- c('Distances', 'Matched', 'Differences')
tp_con_data$Matched[tp_con_data$Matched == 0] <- -1

# Linear model for constituents
attach(tp_con_data)
tp_con_model <- lm(Differences ~ 1 + Distances + Matched + Distances*Matched)
summary(tp_con_model)
```

```
##
## Call:
## lm(formula = Differences ~ 1 + Distances + Matched + Distances *
##     Matched)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.003  -4.649  -1.920   2.012  83.593
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.7289351   0.0027671  2070.4  <2e-16 ***
```

```
## Distances          0.2497248  0.0005947  419.9   <2e-16 ***
## Matched            -1.1323435  0.0027671 -409.2   <2e-16 ***
## Distances:Matched  0.2218165  0.0005947  373.0   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.924 on 36516532 degrees of freedom
## Multiple R-squared:  0.01057,    Adjusted R-squared:  0.01057
## F-statistic: 1.3e+05 on 3 and 36516532 DF,  p-value: < 2.2e-16

# Remove from memory
detach(tp_con_data)
rm(tp_con_data)
```

Standard Transformer

Operations

```
# Read operation data
tf_op_data <- read.csv('results_TF_aggregate_op.csv',
                      header=FALSE, sep = ',')
names(tf_op_data) <- c('Distances', 'Matched', 'Differences')
tf_op_data$Matched[tf_op_data$Matched == 0] <- -1

# Linear model for operations
attach(tf_op_data)
tf_op_model <- lm(Differences ~ 1 + Distances + Matched + Distances*Matched)
summary(tf_op_model)

##
## Call:
## lm(formula = Differences ~ 1 + Distances + Matched + Distances *
##     Matched)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.187  -4.541  -1.920   1.872   83.618
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.583178   0.005728   800.2   <2e-16 ***
## Distances       0.694400   0.001459   476.0   <2e-16 ***
## Matched      -0.936812   0.005728  -163.6   <2e-16 ***
## Distances:Matched 0.319338   0.001459   218.9   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.089 on 9209636 degrees of freedom
## Multiple R-squared:  0.02415,    Adjusted R-squared:  0.02415
## F-statistic: 7.598e+04 on 3 and 9209636 DF,  p-value: < 2.2e-16

# Remove from memory
detach(tf_op_data)
rm(tf_op_data)
```

Digits

```
# Read x data
tf_x_data <- read.csv('results_TF_aggregate_x.csv',
                     header=FALSE, sep = ',')
names(tf_x_data) <- c('Distances', 'Matched', 'Differences')
tf_x_data$Matched[tf_x_data$Matched == 0] <- -1

# Linear model for digits
attach(tf_x_data)
tf_x_model <- lm(Differences ~ 1 + Distances + Matched + Distances*Matched)
summary(tf_x_model)

##
## Call:
## lm(formula = Differences ~ 1 + Distances + Matched + Distances *
##     Matched)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.4249 -1.9876 -0.3778  1.7484  6.5726
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.5093758   0.0018832   801.5  <2e-16 ***
## Distances       0.2649003   0.0002878   920.5  <2e-16 ***
## Matched        -1.0693612   0.0018832  -567.9  <2e-16 ***
## Distances:Matched 0.1763851   0.0002878   612.9  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.363 on 15936300 degrees of freedom
## Multiple R-squared:  0.05189,    Adjusted R-squared:  0.05189
## F-statistic: 2.907e+05 on 3 and 15936300 DF,  p-value: < 2.2e-16

# Remove from memory
detach(tf_x_data)
rm(tf_x_data)
```

Constituents

```
# Read constituent data
tf_con_data <- read.csv('results_TF_aggregate_con.csv',
                       header=FALSE, sep = ',')
names(tf_con_data) <- c('Distances', 'Matched', 'Differences')
tf_con_data$Matched[tf_con_data$Matched == 0] <- -1

# Linear model for constituents
attach(tf_con_data)
tf_con_model <- lm(Differences ~ 1 + Distances + Matched + Distances*Matched)
summary(tf_con_model)

##
## Call:
## lm(formula = Differences ~ 1 + Distances + Matched + Distances *
```

```

##      Matched)
##
## Residuals:
##      Min        1Q      Median        3Q        Max
## -11.830   -4.632   -1.841    2.015   83.983
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.6880276  0.0024037  2366.3  <2e-16 ***
## Distances      0.2745511  0.0005205   527.5  <2e-16 ***
## Matched       -0.9171896  0.0024037  -381.6  <2e-16 ***
## Distances:Matched 0.1734084  0.0005205   333.2  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.909 on 36516532 degrees of freedom
## Multiple R-squared:  0.01433,    Adjusted R-squared:  0.01433
## F-statistic: 1.769e+05 on 3 and 36516532 DF,  p-value: < 2.2e-16
# Remove from memory
detach(tf_con_data)
rm(tf_con_data)

```