Correlation_Functions

November 30, 2016

0.1 Import required Python packages

```
In [1]: from sklearn.neighbors import DistanceMetric
        import numpy as np
        import pandas as pd
```

1 Pearson Correlation

$$r = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2 \sum_i (Y_i - \bar{Y})^2}}$$

```
In [2]: def my_pearson(X, Y):
            num = sum((X - X.mean()) * (Y - Y.mean()))
            denom = np.sqrt(sum((X-X.mean())**2) * sum((Y - Y.mean())**2))
            r = num / denom
            return r

        p_dist = DistanceMetric.get_metric(my_pearson).pairwise

        def Pearson(df):
            P = pd.DataFrame(p_dist(df))
            try:
                P.index = list(df.index)
                P.index.name = df.index.names
                P.columns = list(df.index)
            except:
                pass

            P.to_csv('pearson_corr_Matrix.tsv', sep='\t')
```

2 Spearman Correlation

$$S_p(X,Y) = \frac{\sum_i (R_i - \bar{R})(Q_i - \bar{Q})}{\sqrt{\sum_i (R_i - \bar{R})^2 \sum_i (Q_i - \bar{Q})^2}}$$

```
In [3]: def Spearman(df):
            df = df.T
            S = df.corr(method='spearman')

            try:
                S.index = list(df.index)
                S.index.name = df.index.names
                S.columns = list(df.index)
            except:
                pass

            S.to_csv('spearman_corr_Matrix.tsv', sep='\t')
```

3 Czekanowski Index

$$C_z = \frac{\sum_i 2 * \min(X_i, Y_i)}{\sum_i (X_i + Y_i)}$$

```
In [4]: def my_Czek(X,Y):
            X = np.array(X)
            Y = np.array(Y)

            Cz = 2 * sum(np.minimum(X, Y)) / sum(X + Y)
            return Cz

        czek_dist = DistanceMetric.get_metric(my_Czek).pairwise

        def Czekanowski(df):
            C = pd.DataFrame(czek_dist(df))
            try:
                C.index = list(df.index)
                C.index.name = df.index.names
                C.columns = list(df.index)
            except:
                pass
            #print(C.head())
            C.to_csv('czek_corr_Matrix.tsv', sep='\t')
            return(C)
```

4 Stringent Proportional Similarity (SPS)

$$s = 1 - \frac{1}{n}\left(\sum_{\{i:X_i^2+Y_i^2\neq 0\}} \frac{|X_i^2 - Y_i^2|}{X_i^2 + Y_i^2} + \sum_{\{i:X_i^2+Y_i^2=0\}} 1\right)$$

```
In [5]: def my_sps(X, Y):
            X = np.array(X)
            Y = np.array(Y)
            id0 = np.logical_and((X != 0) , (Y != 0))
            sps = 1. - (1. / len(X)) * ( sum(np.absolute(X[id0]**2 - Y[id0]**2)/(X|
            return sps

        sps_dist = DistanceMetric.get_metric(my_sps).pairwise

        def SPS(df):
            Sp = pd.DataFrame(sps_dist(df))

            try:
                Sp.index = list(df.index)
                Sp.index.name = df.index.names
                Sp.columns = list(df.index)
            except:
                pass
            Sp.to_csv('SPS_corr_Matrix.tsv', sep='\t')
            return Sp
```