Correlation

- Analyze the relationship, if any, between variables x and y by examining the linear strength between them
- Correlation is a numerical measure of this linear relationship
  - can be negative (as x increases y decreases) or positive (as x increases y increases)
- Give an example of two variables that could be negatively correlated
- Give an example of two variables that could be positively correlated

Correlation (cont')

- Linear relationship of x, y pairs can be seen on a scatterplot
- Correlations: $x, Y_1, Y_2, Y_3, Y_4$
  
  \[
  \begin{array}{c|cccc}
  & Y_1 & Y_2 & Y_3 & Y_4 \\
  \hline
  x & -0.059 & 0.987 & -0.987 & 0.295 \\
  \end{array}
  \]

Correlation (cont')

- Think of the scatterplot as a grid with quadrants split up by the mean of x and the mean of y
  - Next take each x point and convert it to a z score, repeat with the y's
  - Finally the sum of the $z_x z_y$'s will be used to calculate the correlation
  - In this case $\sum z_x z_y$ is positive

Correlation (cont')

- In this case $\sum z_x z_y$ would be close to 0
- In this case $\sum z_x z_y$ would be negative

Correlation (cont')

- Pearson's sample correlation coefficient is
  \[
  r = \frac{\sum z_x z_y}{n-1}
  \]
  - The value of r does not depend on the unit of measurement for either x nor y
  - The value of r does not change if x and y are interchanged
  - The value of r is always between -1 and +1. The closer to ±1 the stronger the linear relationship: (in absolute value) 0 to <0.5 is weak; 0.5 to <0.8 is moderate; 0.8 to 1.0 is strong
  - r will only be exactly equal to +1 (or -1) when the points lie in a straight line
  - the measure of r is the extent to which x and y are linearly related
Correlation (cont')

Example: Cereal data (cont') Calculate the correlation between sugar content and rating. Suppose that \( \sum x^2 \) is given as \(-57.735\)

\[
\frac{\sum xy}{n} - \frac{\sum x \cdot \sum y}{n^2} = \frac{-57.735}{n}
\]

Does this linear relationship seem strong, moderate or weak?

Pearson correlation of rating and sugars = \(-0.760\)

Linear Relationships

Example: The data below are airfares ($) and distance (miles) to various US cities from Baltimore, Maryland.

<table>
<thead>
<tr>
<th>Destination</th>
<th>Distance</th>
<th>Airfare</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlanta</td>
<td>576</td>
<td>178</td>
</tr>
<tr>
<td>Boston</td>
<td>370</td>
<td>138</td>
</tr>
<tr>
<td>Chicago</td>
<td>612</td>
<td>94</td>
</tr>
<tr>
<td>Dallas</td>
<td>512</td>
<td>158</td>
</tr>
<tr>
<td>Detroit</td>
<td>1409</td>
<td>138</td>
</tr>
<tr>
<td>Denver</td>
<td>1502</td>
<td>258</td>
</tr>
</tbody>
</table>

Descriptive Statistics: Distance, Airfare

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>N*</th>
<th>Mean</th>
<th>SE Mean</th>
<th>StDev</th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>12</td>
<td>0</td>
<td>713</td>
<td>116</td>
<td>403</td>
<td>189</td>
<td>380</td>
<td>985</td>
<td>1502</td>
<td>1502</td>
</tr>
<tr>
<td>Airfare</td>
<td>12</td>
<td>0</td>
<td>166.9</td>
<td>17.2</td>
<td>59.5</td>
<td>94.0</td>
<td>108.0</td>
<td>168.0</td>
<td>195.5</td>
<td>278.0</td>
</tr>
</tbody>
</table>

Linear Relationships (cont')

• Until now we have described data using statistics such as the sample mean.

The Fitted Regression Line

• Suppose we have \( n \) pairs \((x, y)\)
• If a scatterplot of the data suggests a general linear trend, it would be reasonable to fit a line to the data.
• The question is which is the best line?

Example Airfare (cont')

– We can see from the scatterplot that greater distance is associated with higher airfare.
– In other words airports that tend to be further from Baltimore than tend to be more expensive airfare.
– To decide on the best fitting line, we use the least-squares method to fit the least squares (regression) line.
Equation of the Regression Line

• RECALL: \( y = mx + b \)
• In statistics we call this \( \hat{y} = a + bx \)
  where
  \( y \) is the dependent variable
  \( x \) is the independent variable
  \( a \) is the y-intercept
  \( b \) is the slope of the line

\[
\hat{y} = a + bx
\]

\[
\hat{y} = \frac{\sum y - b \sum x}{n}
\]

\[
\sum x^2 - \frac{\sum x \sum y}{n}
\]

Equation of the Regression Line (cont’)

• Example: Airfare (cont’)

Regression Analysis: Airfare versus Distance

The regression equation is

Airfare = $83.27 + 0.117 \text{ Distance}$

Predictor Coef SE Coef T      P
Constant     83.27 22.95  3.63  0.005
Distance   0.11738 0.02832  4.14  0.002

S = 37.8270   R-Sq = 63.2%   R-Sq(adj) = 59.5%

Analysis of Variance

Source          DF     SS     MS      F      P
Regression       1  24574  24574 17.17  0.002
Residual Error  10  14309   1431
Total           11  38883

Equation of the Regression Line (cont’)

• When we write the least squares regression equation we use the following notation:

\[
\hat{y} = 83.27 + 0.117x
\]

– \( b \) expresses the rate of change of \( y \) with respect to \( x \)
  • For every one mile increase in distance, airfare will go up by an additional 0.117 dollars.
  • We could actually describe this as for a 100 mile increase in distance airfare rises by $11.70
– \( a \) expresses where the regression line will hit the y axis
  • It may or may not be interpretable, depends on the context
  • In this case does an airfare of $83.27 when distance traveled is 0 miles make sense?

Equation of the Regression Line (cont’)

• Predict the airfare for a city that is 576 miles away. If you look at the original data set, Atlanta's distance was 576 miles and the airfare was $178

\[
\hat{y} = 83.27 + 0.117x
\]

\[
= 83.27 + 0.11738(576)
\]

\[
= $150.88 \text{ (watch units!)}
\]

• Calculate the corresponding residual
  – HOLD that thought
  \[
  \text{Residual} = 178 - 150.88 = 27.12
  \]

Equation of the Regression Line (cont’)

• It is important to only make predictions for values that are within our sampled range of \( x \) data
• Extrapolation beyond the scope of our sampled data is dangerous because we do not know what happens to the relationship between \( x \) (distance) and \( y \) (airfare) outside this range
• In other words, this line may not continue on with the same slope forever
Equation of the Regression Line (cont’)

- Predict the airfare for a city that is 2842 miles away from Baltimore. Does this seem like a legitimate prediction? Explain.
  \[ 83.27 + 0.11738(2842) = 416.86 \]
  - This does not seem like a legitimate prediction because our sample range of data goes from 189 to 1502 miles
  - No making predictions outside our sampled range of data!
  - This city (San Francisco) falls outside of this range
  - NOTE: The actual airfare for this city was $198

Equation of the Regression Line (cont’)

- We can predict Y for X that are “reasonable” (within the range of modeled X values)
  - We should not predict Y for X values that are “not reasonable” (outside the range of modeled X values)

Residuals

- For each observed x value (x_i) there is a predicted y value (\( \hat{y} \)) based on the regression equation
  \[ \hat{y} = a + bx \]

- Also associated with each (x_i, y_i) there is a residual
  - the vertical distance between each predicted y (\( \hat{y} \)) and observed y
  - Residual = y_i - \( \hat{y} \)

- When we add up all the residuals they sum to 0

Residuals (cont’)

- Which city has the largest (in absolute value) residual? Quantify this value.
  - HINT: look at the scatter plot. How can you tell?
  St. Louis because it lies the furthest (vertically) from the regression line
  \[ \hat{y} = 83.27 + 0.11738(737) = 169.78 \]
  Residual = 98 – 169.78 = -$71.78

Residuals (cont’)

- Which city has the largest predicted (fitted) value (\( \hat{y} \))? Quantify this value.
  - HINT: look at the scatter plot. How can you tell?
  Denver because it is the observation with the largest distance and therefore predicted value
  \[ \hat{y} = 83.27 + 0.11738(1502) = 259.57 \]
  - NOTE: If the slope was negative the largest predicted value would be the observation with the smallest x.
Residuals (cont')

Regression Analysis: Airfare versus Distance

<table>
<thead>
<tr>
<th>Obs</th>
<th>Distance</th>
<th>Airfare</th>
<th>Fit</th>
<th>SE Fit</th>
<th>Residual</th>
<th>St Resid</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>576</td>
<td>178.0</td>
<td>150.9</td>
<td>27.1</td>
<td>27.1</td>
<td>0.75</td>
</tr>
<tr>
<td>2</td>
<td>370</td>
<td>138.0</td>
<td>126.7</td>
<td>14.6</td>
<td>11.3</td>
<td>0.32</td>
</tr>
<tr>
<td>3</td>
<td>612</td>
<td>94.0</td>
<td>155.1</td>
<td>11.3</td>
<td>-61.1</td>
<td>-1.69</td>
</tr>
<tr>
<td>4</td>
<td>1216</td>
<td>278.0</td>
<td>226.0</td>
<td>18.0</td>
<td>52.0</td>
<td>1.56</td>
</tr>
<tr>
<td>5</td>
<td>409</td>
<td>158.0</td>
<td>131.3</td>
<td>13.9</td>
<td>26.7</td>
<td>0.74</td>
</tr>
<tr>
<td>6</td>
<td>1502</td>
<td>238.0</td>
<td>239.6</td>
<td>24.9</td>
<td>-1.6</td>
<td>-0.05</td>
</tr>
<tr>
<td>7</td>
<td>944</td>
<td>148.0</td>
<td>194.3</td>
<td>12.8</td>
<td>3.7</td>
<td>0.10</td>
</tr>
<tr>
<td>8</td>
<td>958</td>
<td>188.0</td>
<td>200.4</td>
<td>13.8</td>
<td>-12.4</td>
<td>-0.35</td>
</tr>
<tr>
<td>9</td>
<td>189</td>
<td>184.3</td>
<td>105.5</td>
<td>14.6</td>
<td>-7.5</td>
<td>-0.23</td>
</tr>
<tr>
<td>10</td>
<td>170.2</td>
<td>170.2</td>
<td>170.2</td>
<td>3.4</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>218</td>
<td>138.1</td>
<td>107.9</td>
<td>19.6</td>
<td>30.1</td>
<td>0.90</td>
</tr>
<tr>
<td>12</td>
<td>731</td>
<td>98.0</td>
<td>140.8</td>
<td>10.9</td>
<td>-71.8</td>
<td>-1.98</td>
</tr>
</tbody>
</table>

Residual Plots

• How can we check to make sure that it is reasonable to fit a line to the data?
  – Good: Look at a scatter plot
  – Better: Look at a plot of the residuals vs. the predicted values

• The residual plot is a useful tool that extends from simple linear regression to multiple regression
  – A scatter plot only works for simple regression when you have one independent variable
  – It is often the case that you would like to predict Y with more than one independent variable

Residual Plots (cont')

• We can use a residual plot to check for linearity or other potential problems:
  • PLOT:
    residuals = (yi – ŷi) vs. predicted (fitted) = ŷi
    – WANT: residual plot to dance around 0 (not much difference between predicted and actual Yi's)
    – WANT: no specific pattern

Variability in Regression

• Consider our airfare example
• The dependent variable, airfare, varies from airport to airport, regardless of distance
  – A statistical measure of the total variability in airfare is called sums of squares total
  \[ SS(TO) = \sum (y_i - \bar{y})^2 \]
Variability in Regression (cont')

Regression Analysis: Airfare versus Distance

The regression equation is

Airfare = 83.3 + 0.117 Distance

Predictor  Coef  SE Coef  T     P
Constant    83.27  22.95  3.63  0.005
Distance   0.11738  0.02832  4.14  0.002

S = 37.8270  R-sq = 63.2%  R-sq(adj) = 58.5%

Analysis of Variance

Source     DF     SS     MS      F      P
Regression   1  24574 24574 17.17  0.002
Residual Error 10  14309 1431
Total        11  38883

The Coefficient of Determination

• The coefficient of determination is another measure of the strength of the linear relationship between X and Y
  – aka: “The proportion of the variability in Y that is explained by the linear regression of Y on X”
  – simply put this is a measure of the total variability of Y explained by X
• Denoted by \( r^2 \)

\[
 r^2 = 1 - \frac{SS(resid)}{SS(Tot)}
\]

– This is 1 – the proportion of unexplained variability = proportion of explained variability

Example: Airfare (cont')

Calculate and interpret \( r^2 \)

Only 63.2% of the total variability in airfare can be explained by a linear regression with distance.

RULE OF THUMB:  64% to 100% indicates a strong linear relationship; 25% to <64% indicates is moderate; <25% is poor.

NOTE:  \( r^2 \) close to zero does not mean that there is no relationship between X and Y, only that it is not a linear relationship.

Residual Standard Deviation

• The ‘best’ straight line is the one that minimizes the residual sums of squares
• The residual standard deviation can be used as our description of the closeness of the data points to the regression line

\[
 s_e = \sqrt{\frac{SS(resid)}{n-2}} = \sqrt{\frac{\sum(y_i - \hat{y})^2}{n-2}}
\]

– how far off predictions tend to be that are made using the regression model
– Similar idea to \( s \) (measures variability around \( \hat{y} \))
– \( s_e \) (measures variability about the regression line)
Residual Standard Deviation (cont')

• Similar interpretation to the sd from the empirical rule in ch 4.
  - 68% of our data falls within $\pm 1$ $s_e$ from the line
  - 95% of our data falls within $\pm 2$ $s_e$ from the line
• We expect most of our data to fall within $2s_e$ from the regression line
Example: Airfare (cont')

$\hat{s_e} = \sqrt{\frac{SS(resid)}{n-2}} = 37.83$
  - A typical prediction tends to be off by $37.83$
  - Most of our observed values will fall within $2(37.83) = 75.66$ from their predicted values.

Curvilinear Data

• For our method of simple linear regression to be valid it needs to be reasonable to fit a line to the data
  - If a line does not fit the data this does not mean that x is not useful for predicting y, just not in a linear relationship
• Curvilinear data can distort regression results by:
  - a fitted line that doesn't represent the data
  - the correlation is misleadingly small
  - $s_e$ is inflated

Curvilinear Data (cont')

Example: For married couples with one or more offspring, a demographic study was conducted to determine the effect of the families annual income (at marriage) on time (months) between marriage and the birth of the first child.

<table>
<thead>
<tr>
<th>Income</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>5775</td>
<td>16.20</td>
</tr>
<tr>
<td>9800</td>
<td>35.00</td>
</tr>
<tr>
<td>24210</td>
<td>20.00</td>
</tr>
<tr>
<td>13795</td>
<td>37.20</td>
</tr>
<tr>
<td>6500</td>
<td>31.75</td>
</tr>
<tr>
<td>25015</td>
<td>24.40</td>
</tr>
<tr>
<td>7400</td>
<td>31.75</td>
</tr>
<tr>
<td>4120</td>
<td>5.00</td>
</tr>
<tr>
<td>20170</td>
<td>36.00</td>
</tr>
<tr>
<td>22400</td>
<td>30.80</td>
</tr>
<tr>
<td>9340</td>
<td>30.00</td>
</tr>
<tr>
<td>15385</td>
<td>39.20</td>
</tr>
<tr>
<td>6440</td>
<td>20.00</td>
</tr>
<tr>
<td>18000</td>
<td>41.25</td>
</tr>
<tr>
<td>19625</td>
<td>44.00</td>
</tr>
<tr>
<td>13000</td>
<td>44.00</td>
</tr>
<tr>
<td>18180</td>
<td>41.25</td>
</tr>
<tr>
<td>4608</td>
<td>9.70</td>
</tr>
<tr>
<td>35.00</td>
<td>20.00</td>
</tr>
<tr>
<td>10625</td>
<td>38.20</td>
</tr>
<tr>
<td>18000</td>
<td>41.25</td>
</tr>
<tr>
<td>13500</td>
<td>44.00</td>
</tr>
<tr>
<td>38000</td>
<td>44.00</td>
</tr>
<tr>
<td>10000</td>
<td>44.00</td>
</tr>
<tr>
<td>8000</td>
<td>44.00</td>
</tr>
<tr>
<td>5400</td>
<td>9.20</td>
</tr>
<tr>
<td>15385</td>
<td>39.20</td>
</tr>
<tr>
<td>15385</td>
<td>39.20</td>
</tr>
<tr>
<td>19625</td>
<td>38.20</td>
</tr>
<tr>
<td>13000</td>
<td>44.00</td>
</tr>
</tbody>
</table>

Curvilinear Data (cont')

• Clearly a straight line model does not accurately describe what is going on with this data.
• Does this mean there is no relationship between income and time?
  - No, just that it isn’t linear!
• A better equation would be $\hat{y} = a + b_1x + b_2x^2$

Curvilinear Data (cont')

Example: Airfare (cont')

Regression Analysis: Airfare versus Distance

The regression equation is
Airfare $= 83.3 + 0.117$ Distance

Predictor     Coef SE Coef T      P
Constant      83.27    22.95  3.63  0.005
Distance   0.11738  0.02832  4.14  0.002

$S = 37.8270$  R-$sq =$52.3%  R-$sq(adj) =$49.5%

Analysis of Variance

Source     DF     SS      MS     F      P
Regression  1   24574   24574 17.17  0.002
Residual Error  10   14309   1431
Total           11  38883

Curvilinear Data (cont')

Regression Analysis: Time versus Income

The regression equation is
Time $= 19.6 + 0.000714$ Income

Predictor     Coef SE Coef T      P
Constant      19.626      5.213  3.76  0.001
Income     0.0007138  0.0003528  2.02  0.058

$S = 10.4958$  R-$sq =$18.5%  R-$sq(adj) =$14.0%

Analysis of Variance

Source     DF     SS      MS     F      P
Regression  1   450.9  450.9  4.09  0.058
Residual Error  18  1982.9  110.2
Total           19  2433.8

Curvilinear Data (cont')

Regression Analysis: Time versus Income

The regression equation is
Time $= 19.4 + 0.000715$ Income

Predictor     Coef SE Coef T      P
Constant      19.424      5.213  3.76  0.001
Income     0.0007152  0.0003528  2.02  0.058

$S = 10.4958$  R-$sq =$18.5%  R-$sq(adj) =$14.0%

Analysis of Variance

Source     DF     SS      MS     F      P
Regression  1   450.9  450.9  4.09  0.058
Residual Error  18  1982.9  110.2
Total           19  2433.8
Curvilinear Data (cont’)

This non-linearity can be seen in the residual plot. Because of the curvature seen in the scatter plot, this plot suggests that a linear model may not be appropriate here.

Residuals Versus the Fitted Values
(response is Time)

Fitted Value
Residual

Curvilinear Data (cont’)

• Our solution would be to fit a quadratic model to address the curvature seen in the scatter plot.
  – The graph shows that visually we have a good fit with a quadratic model.
  – NOTE: Now that we have more than one independent variable this becomes a multiple regression problem.

Residuals Versus the Fitted Values
(response is Time)

Curvilinear Data (cont’)

Regression Analysis: Time versus Income, IncomeSQ

The regression equation is
Time = - 18.6 + 0.00770 Income - 0.000000 IncomeSQ

Predictor Coef SE Coef T P
Constant -18.639 4.679 -3.98 0.001
Income 0.0077004 0.0007699 10.00 0.000
IncomeSQ -0.00000025 0.00000003 -9.25 0.000

S = 4.39819   R-Sq = 86.5% R-Sq(adj) = 84.9%

Analysis of Variance
Source DF SS MS F P
Regression 2 2104.9 1052.5 54.41 0.000
Residual Error 17 328.8 19.3
Total 19 2433.8

Curvilinear Data (cont’)

Once the model is fit appropriately the residual plot looks much more random – better!
This is because the residuals all randomly disperse about the quadratic regression equation.

Residuals Versus the Fitted Values
(response is Time)

Transformations

• Sometimes when you cannot meet the regression assumptions of linearity, (an others: constant variance, or normality) you can use a transformation of the dependent variable, independent variable, or both.
  – The type of transformation to use really depends on the problem with the data
  – Log variable, 1/variable, Sqrt variable are some examples of transformations that are often used, see page234 for more details.
  – If done correctly, once you transform your data you can meet regression assumptions and carry on with your analysis.

Example: Remember the highly debated presidential race of 2000? Recall that one of the highly criticized results of this election was the voting irregularity found in Palm Beach County, Florida. Democrats argued that Pat Buchanan received far too many votes, which should have been for Al Gore, all because of the confusing butterfly ballot. The variables in the file 2000pres.mtw are the 2000 election votes for each major party candidate: Gore, Bush, Buchanan, Nader, and total votes; by county in Florida.

Transformations (cont')
One of the justifications for this abnormally large number of Buchanan votes (from republicans obviously) was that Palm Beach is one of the largest counties in Florida, so there should be more votes for every candidate when compared to other counties in the state. To investigate this claim we will use regression to explore the relationship, if any, between total number of votes (i.e. a surrogate for county size) and votes for Buchanan.

First we'll start with a scatter plot of total votes by votes for Buchanan.

Palm Beach is an obvious outlier, so let's remove them from the data and calculate a regression for the rest of the counties in Florida. This way we can make a prediction of what Palm Beach should have had for Buchanan votes, based on the voting pattern in the rest of the state.

Seems like the regression results produce an ok model ($R^2 = 72\%$). Next we should check to be sure that we meet the appropriate assumptions for the residuals before we assume the model is ok.

The residual plot suggests that there are some outliers that could cause assumption problems.
Transformations (cont')

Regression Analysis: LOGBuch versus LOGTotal

The regression equation is

\[ \text{LOGBuch} = -1.09 + 0.703 \times \text{LOGTotal} \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1.09</td>
<td>0.1640</td>
<td>-6.62</td>
<td>0.000</td>
</tr>
<tr>
<td>LOGTotal</td>
<td>0.703</td>
<td>0.03610</td>
<td>19.49</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ s = 0.189026 \quad R-Sq = 85.6\% \quad R-Sq(adj) = 85.4\% \]

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>13.569</td>
<td>13.569</td>
<td>379.76</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Error</td>
<td>64</td>
<td>2.287</td>
<td>0.036</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>65</td>
<td>15.856</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The residual plot looks much better. No huge outliers, nice random pattern, suggests that we are ok with our linearity assumption.

Transformations (cont')

SO...Based on the voting pattern for the rest of Florida (excluding Palm Beach) in the 2000 election, what do we predict for number of votes for Buchanan, based on the total number of votes for Palm Beach.

Palm Beach actually cast a total of 432286 votes, which Log(432286) = 5.63577

So we would predict

\[ \text{LogBuch} = -1.09 + 0.703 (5.63577) = 2.8781 \]

Which equates to \(10^{2.8781} = 755.26\) or 756 votes. How many votes did Buchanan actually get in Palm Beach? 3407!!!

Transformations (cont')

WARNING more on this topic to come later:

In fact, a 95% prediction interval suggests that we are highly confident based on the voting pattern in the rest of Florida, Buchanan should have received between \(10^{2.4888} = 308.18\) to \(10^{2.673} = 1850.55\) votes.

Predicted Values for New Observations

<table>
<thead>
<tr>
<th>New Obs</th>
<th>Fit</th>
<th>SE Fit</th>
<th>95% CI</th>
<th>95% PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.8781</td>
<td>0.0472</td>
<td>(2.7838, 2.9723)</td>
<td>(2.4888, 3.2673)</td>
</tr>
</tbody>
</table>