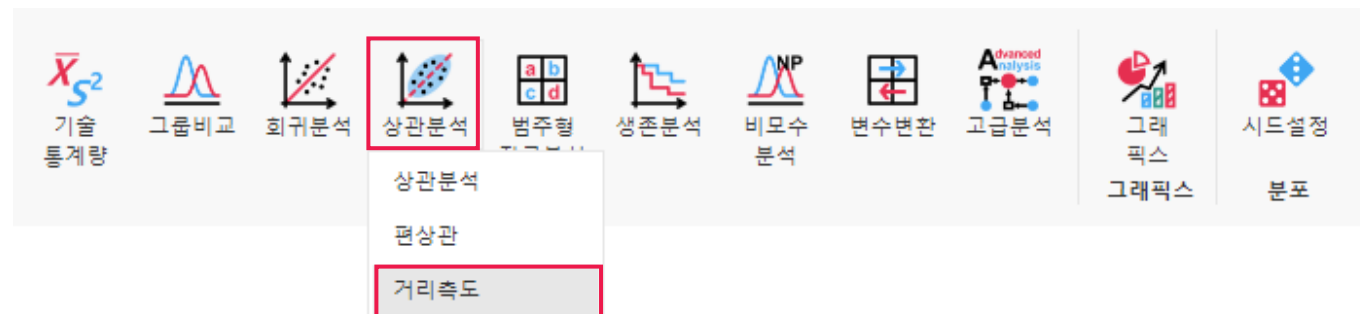


# 거리측도

거리측도(Distance Measure)는 두 개 이상의 대응관측치나 대응변수 사이의 비유사성(Dis-similarity) 정도를 거리(Distance)를 이용하여 측정하는 통계분석 모듈입니다. 거리는 입력변수가 연속형인지, 이분형인지에 따라 다양한 방식으로 계산할 수 있으며, 변수 간 계산과 관측치 간 계산을 수행할 수 있습니다. 계산된 거리를 저장하여 분류분석 등에서 활용할 수 있습니다.

## 메뉴 호출하기

- 상관분석 > 거리측도



• 변수설정 탭

거리측도

변수설정 거리측도

데이터

전체변수

id  
bweight  
lowbw  
gestwks  
preterm  
matage  
hyp  
sex

① 연속형변수(선택-2개이상가능)

>  
<

② ☐ 연속형변수 표준화

③ 이분형변수(선택-2개이상가능)

>  
<

④ 거리계산

☒ 변수간 ☐ 관측치간

⑤ 저장경로

⑥ 로그변환시 사용될 밑

☒ e ☐ 2 ☐ 10

도움말 재설정

메뉴 요소	설명
① 연속형 변수	거리를 계산하고자 하는 연속형 변수들을 전체변수로부터 선택합니다. 반드시 2개 이상의 연속형 변수가 선택되어야 합니다.
② 연속형 변수 표준화	연속형 변수에 선택된 변수들을 표준정규분포를 따르도록 표준화한 뒤, 거리를 계산합니다.
③ 이분형 변수	거리를 계산하고자 하는 이분형 변수들을 전체변수로부터 선택합니다. 반드시 2개 이상의 이분형 변수가 선택되어야 하며, 수치형으로 코딩한 변수만 지정할 수 있습니다.
④ 거리계산	거리를 계산하는 기준으로, 다음 2가지 옵션 중 1개를 선택할 수 있습니다. <ul style="list-style-type: none"> <li>변수간 (Default) : 지정된 변수들 (열 별 개체들) 간의 거리를 계산합니다.</li> <li>관측치간 : 엑셀 시트상에 입력된 행 별 개체들 간의 거리를 계산합니다.</li> </ul>
⑤ 거리계산 > 저장경로	[거리계산]-'관측치간'을 선택하는 경우 활성화됩니다. 관측치간 거리를 계산하는 경우, 출력되는 결과가 크기 때문에, Rex output이 아닌 사용자의 컴퓨터 내 지정된 저장 경로에 .csv 파일로 저장합니다.
⑥ 로그변환 시 사용될 밑	연속형 변수의 거리 계산 과정에서 변수의 로그변환에 사용될 밑(base)으로, 다음 3가지 옵션 중 1개를 선택할 수 있습니다. <ul style="list-style-type: none"> <li>e (Default) : 자연상수로, 이 경우 자연로그가 적용됩니다.</li> <li>2</li> <li>10 : 이 경우, 상용로그가 적용됩니다.</li> </ul>

• 거리측도 탭

거리측도

변수설정
거리측도

① 연속형변수

L<sub>p</sub> family

☒ euclidean  
☐ manhattan  
☐ chebyshev  
☐ minkowski  
power

② L<sub>1</sub> family

☐ sorensen  
☐ gower  
☐ soergel  
☐ kulczynski\_d  
☐ canberra

③ Intersection

☐ intersection  
☐ non-intersection  
☐ wavehedges  
☐ czekanowski  
☐ motyka

④ Squared\_chord

☐ fidelity  
☐ bhattacharyya  
☐ squared\_chord  
☐ squared\_euclidean

⑤ Squared L<sub>2</sub>

☐ pearson  
☐ neyman  
☐ squared\_chi

⑥ Inner Product

☐ inner\_product  
☐ harmonic\_mean  
☐ cosine  
☐ hassebrook  
☐ jaccard

⑦ Shannon's

☐ jeffreys  
☐ topsoe  
☐ jensen-shannon  
☐ jensen\_difference

⑧ Combinations

☐ taneja  
☐ kumar-johnson  
☐ avg

⑨ 이분형변수

☒ JACCARD S3  
☐ SOCKAL & MICHENER S4  
☐ SOCKAL & SNEATH S5  
☐ ROGERS & TANIMOTO S6  
☐ CZEKANOWSKI S7

☐ GOWER & LEGENDRE S9  
☐ OCHIAI S12  
☐ SOKAL & SNEATH S13  
☐ Phi of PEARSON S14  
☐ GOWER & LEGENDRE S2

도움말
재설정

확인
취소

메뉴 요소	설명
① L <sub>p</sub> family	<p>[변수설정] 탭에서 '연속형 변수'가 지정된 경우 활성화됩니다. L<sub>p</sub> Minkowski family로 다음 4가지 옵션 중 2개 이상 중복 선택이 가능합니다.</p> <ul style="list-style-type: none"> <li>euclidean : <math>d = \sqrt{\sum  P_i - Q_i ^2}</math></li> <li>manhattan : <math>d = \sum  P_i - Q_i </math></li> <li>chebyshev : <math>d = \max  P_i - Q_i </math></li> <li>minkowski : <math>d = (\sum  P_i - Q_i ^p)^{\frac{1}{p}}</math></li> </ul> <p>- power : [minkowski]를 선택한 경우 입력 칸이 활성화됩니다. minkowski power를 입력합니다. 자연수만 입력 가능하며, Default는 3입니다. power에 입력된 값은 수식 상에 <math>p</math>로 반영됩니다.</p>
② L <sub>1</sub> family	<p>[변수설정] 탭에서 '연속형 변수'가 지정된 경우 활성화됩니다. L<sub>1</sub> family로 다음의 5가지 옵션 중 2개 이상 중복 선택이 가능합니다.</p> <ul style="list-style-type: none"> <li>sorensen : <math>d = \sum  P_i - Q_i  / \sum (P_i + Q_i)</math></li> <li>gower : <math>d = (\sum  P_i - Q_i )^{-1}</math></li> <li>soergel : <math>d = \sum  P_i - Q_i  / \sum \max(P_i, Q_i)</math></li> <li>kulczynski_d : <math>d = \sum  P_i - Q_i  / \sum \min(P_i, Q_i)</math></li> <li>canberra : <math>d = \sum  P_i - Q_i  / (P_i + Q_i)</math></li> </ul>

• 거리측도 탭

거리측도

변수설정 [거리측도]

① 연속형변수

② L<sub>p</sub> family

☒ euclidean  
☐ manhattan  
☐ chebyshev  
☐ minkowski  
 power

③ L<sub>1</sub> family

☐ sorensen  
☐ gower  
☐ soergel  
☐ kulczynski\_d  
☐ canberra

④ Intersection

☐ intersection  
☐ non-intersection  
☐ wavehedges  
☐ czekanowski  
☐ motyka

⑤ Squared\_chord

☐ fidelity  
☐ bhattacharyya  
☐ squared\_chord  
☐ squared\_euclidean

⑥ Squared L<sub>2</sub>

☐ pearson  
☐ neyman  
☐ squared\_chi

⑦ Inner Product

☐ inner\_product  
☐ harmonic\_mean  
☐ cosine  
☐ hassebrook  
☐ jaccard

⑧ Shannon's

☐ jeffreys  
☐ topsoe  
☐ jensen-shannon  
☐ jensen\_difference

⑨ Combinations

☐ taneja  
☐ kumar-johnson  
☐ avg

⑩ 이분형변수

☒ JACCARD S3  
☐ SOCKAL & MICHENER S4  
☐ SOCKAL & SNEATH S5  
☐ ROGERS & TANIMOTO S6  
☐ CZEKANOWSKI S7  
☐ GOWER & LEGENDRE S9  
☐ OCHIAI S12  
☐ SOKAL & SNEATH S13  
☐ Phi of PEARSON S14  
☐ GOWER & LEGENDRE S2

도움말 | 재설정 | 확인 | 취소

메뉴 요소	설명
3. Intersection	<p>[변수설정] 탭에서 '연속형 변수'가 지정된 경우 활성화됩니다. Intersection family로 다음의 5가지 옵션 중 2개 이상 중복 선택이 가능합니다. <math>s</math>는 similarity measure라는 의미입니다.</p> <ul style="list-style-type: none"> <li>intersection : <math>s = \sum \min(P_i, Q_i)</math></li> <li>non-intersection : <math>d = 1 - \sum \min(P_i, Q_i)</math></li> <li>wavehedges : <math>d = \sum  P_i - Q_i  / \max(P_i, Q_i)</math></li> <li>czekanowski : <math>d = \sum  P_i - Q_i  / \sum  P_i + Q_i </math></li> <li>motyka : <math>d = \sum \min(P_i, Q_i) / (P_i + Q_i)</math></li> </ul>
④ Squared_chord	<p>[변수설정] 탭에서 '연속형 변수'가 지정된 경우 활성화됩니다. Squared-chord family로 다음 4가지 옵션 중 2개 이상 중복 선택이 가능합니다.</p> <ul style="list-style-type: none"> <li>fidelity : <math>s = \sum \sqrt{P_i * Q_i}</math></li> <li>bhattacharyya : <math>d = -\ln \sum \sqrt{P_i * Q_i}</math></li> <li>squared_chord : <math>d = \sum (\sqrt{P_i} - \sqrt{Q_i})^2</math></li> <li>squared_euclidean : <math>d = \sum (P_i - Q_i)^2</math></li> </ul>
⑤ Squared L <sub>2</sub>	<p>[변수설정] 탭에서 '연속형 변수'가 지정된 경우 활성화됩니다. Squared L<sub>2</sub> family로 다음 3가지 옵션 중 2개 이상 중복 선택이 가능합니다.</p> <ul style="list-style-type: none"> <li>pearson : <math>d = \sum ((P_i - Q_i)^2 / Q_i)</math></li> <li>neyman : <math>d = \sum ((P_i - Q_i)^2 / P_i)</math></li> <li>squared_chi : <math>d = \sum ((P_i - Q_i)^2 / (P_i + Q_i))</math></li> </ul>

- 거리측도 탭

거리측도

변수설정 [거리측도]

① 연속형변수

② L<sub>p</sub> family

☒ euclidean  
☐ manhattan  
☐ chebyshev  
☐ minkowski  
 power

③ L<sub>1</sub> family

☐ sorensen  
☐ gower  
☐ soergel  
☐ kulczynski\_d  
☐ canberra

④ Intersection

☐ intersection  
☐ non-intersection  
☐ wavehedges  
☐ czekanowski  
☐ motyka

⑤ Squared\_chord

☐ fidelity  
☐ bhattacharyya  
☐ squared\_chord  
☐ squared\_euclidean

⑥ Squared L<sub>2</sub>

☐ pearson  
☐ neyman  
☐ squared\_chi

⑦ Inner Product

☐ inner\_product  
☐ harmonic\_mean  
☐ cosine  
☐ hassebrook  
☐ jaccard

⑧ Shannon's

☐ jeffreys  
☐ topsoe  
☐ jensen-shannon  
☐ jensen\_difference

⑨ Combinations

☐ taneja  
☐ kumar-johnson  
☐ avg

⑩ 이분형변수

☒ JACCARD S3  
☐ SOCKAL & MICHENER S4  
☐ SOCKAL & SNEATH S5  
☐ ROGERS & TANIMOTO S6  
☐ CZEKANOWSKI S7  
☐ GOWER & LEGENDRE S9  
☐ OCHIAI S12  
☐ SOKAL & SNEATH S13  
☐ Phi of PEARSON S14  
☐ GOWER & LEGENDRE S2

도움말 | 재설정 | 확인 | 취소

메뉴 요소	설명
⑥ Inner Product	<p>[변수설정] 탭에서 '연속형 변수'가 지정된 경우 활성화됩니다. Inner product family로 다음 5가지 옵션 중 2개 이상 중복 선택이 가능합니다.</p> <ul style="list-style-type: none"> <li>inner_product : <math>s = \sum P_i * Q_i</math></li> <li>harmonic_mean : <math>s = 2 * \sum (P_i * Q_i) / (P_i + Q_i)</math></li> <li>cosine : <math>s = \sum (P_i * Q_i) / \sqrt{\sum P_i^2} \sqrt{\sum Q_i^2}</math></li> <li>hassebrook : <math>s = \sum (P_i * Q_i) / (\sum P_i^2 + \sum Q_i^2 - \sum (P_i * Q_i))</math></li> <li>jaccard : <math>d = 1 - \sum (P_i * Q_i) / (\sum P_i^2 + \sum Q_i^2 - \sum (P_i * Q_i))</math></li> </ul>
⑦ Shannon's entropy	<p>[변수설정] 탭에서 '연속형 변수'가 지정된 경우 활성화됩니다. Shannon's entropy family로 다음 4가지 옵션 중 2개 이상 중복 선택이 가능합니다.</p> <ul style="list-style-type: none"> <li>jeffreys : <math>d = \sum (P_i - Q_i) * \log(P_i / Q_i)</math></li> <li>topsoe : <math>d = \sum (P_i * \log(2 * P_i / P_i + Q_i) + (Q_i * \log(2 * Q_i / P_i + Q_i))</math></li> <li>jensen-shannon : <math>d = 0.5 * \sum (P_i * \log(2 * P_i / P_i + Q_i) + \sum (Q_i * \log(2 * Q_i / P_i + Q_i))</math></li> <li>jensen_difference : <math>d = \sum ((P_i * \log P_i + Q_i \log Q_i / 2) - (P_i + Q_i / 2) * \log(P_i + Q_i / 2))</math></li> </ul>

- 거리측도 탭

거리측도

변수설정 [거리측도]

① 연속형변수

② L<sub>p</sub> family

☒ euclidean  
☐ manhattan  
☐ chebyshev  
☐ minkowski  
 power

③ L<sub>1</sub> family

☐ sorensen  
☐ gower  
☐ soergel  
☐ kulczynski\_d  
☐ canberra

④ Intersection

☐ intersection  
☐ non-intersection  
☐ wavehedges  
☐ czekanowski  
☐ motyka

⑤ Squared\_chord

☐ fidelity  
☐ bhattacharyya  
☐ squared\_chord  
☐ squared\_euclidean

⑥ Squared L<sub>2</sub>

☐ pearson  
☐ neyman  
☐ squared\_chi

⑦ Inner Product

☐ inner\_product  
☐ harmonic\_mean  
☐ cosine  
☐ hassebrook  
☐ jaccard

⑧ Shannon's

☐ jeffreys  
☐ topsoe  
☐ jensen-shannon  
☐ jensen\_difference

⑨ Combinations

☐ taneja  
☐ kumar-johnson  
☐ avg

⑩ 이분형변수

☒ JACCARD S3  
☐ SOCKAL & MICHENER S4  
☐ SOCKAL & SNEATH S5  
☐ ROGERS & TANIMOTO S6  
☐ CZEKANOWSKI S7

☐ GOWER & LEGENDRE S9  
☐ OCHIAI S12  
☐ SOKAL & SNEATH S13  
☐ Phi of PEARSON S14  
☐ GOWER & LEGENDRE S2

도움말 | 재설정 | 확인 | 취소

메뉴 요소	설명
8. Combinations	<p>[변수설정] 탭에서 '연속형 변수'가 지정된 경우 활성화됩니다. Combinations 로 다음 3가지 옵션 중 2개 이상 중복 선택이 가능합니다.</p> <ul style="list-style-type: none"> <li>taneja : <math>d = \sum \left( (P_i + 0.5Q_i) * \log(P_i + Q_i / (2\sqrt{P_i Q_i})) \right)</math></li> <li>kumar-johnson : <math>d = \frac{\sum (P_i^2 - Q_i^2)^2}{2} * (P_i * Q_i)^{1.5}</math></li> <li>avg : <math>d = \sum  P_i - Q_i  + \max  P_i - Q_i  * 0.5</math></li> </ul>
⑨ 이분형 변수	<p>[변수설정] 탭에서 '이분형 변수'가 지정된 경우 활성화됩니다. 다음의 옵션 중 2개 이상 중복 선택이 가능합니다. 두 개의 이분형 변수의 2x2 빈도표에서 <math>n_{11} = a, n_{10} = b, n_{01} = c, n_{00} = d</math>로 정의할 때, 각 거리는 다음과 같이 계산됩니다.</p> <ul style="list-style-type: none"> <li>JACCARD S3 : <math>s = a / (a + b + c)</math></li> <li>SOCKAL &amp; MICHENER S4 : <math>s = (a + b) / (a + b + c + d)</math></li> <li>SOCKAL &amp; SNEATH S5 : <math>s = a / (a + 2(b + c))</math></li> <li>ROGERS &amp; TANIMOTO S6 : <math>s = (a + d) / (a + 2(b + c) + d)</math></li> <li>CZEKANOWSKI S7 : <math>s = 2a / (2a + b + c)</math></li> <li>GOWER &amp; LEGENDRE S9 : <math>s = (a - (b + c) + d) / (a + b + c + d)</math></li> <li>OCHIAI S12 : <math>s = a / \sqrt{(a + b)(a + c)}</math></li> <li>SOKAL &amp; SNEATH S13 : <math>s = ad / \sqrt{(a + b)(a + c)(d + b)(d + c)}</math></li> <li>Phi of PEARSON S14 : <math>s = (ad - bc) / \sqrt{(a + b)(a + c)(d + b)(d + c)}</math></li> <li>GOWER &amp; LEGENDRE S2 : <math>s = a / (a + b + c + d)</math></li> </ul>