

Pytania do egzaminu ustnego z Wielowymiarowej Analizy Danych

1. Na czym polega istota testowania hipotez dotyczących wektora średnich w ujęciu wielowymiarowym i w jaki sposób podejście to rozszerza klasyczne metody jednowymiarowe?
2. Jakie są kluczowe założenia testów wielowymiarowych i dlaczego ich spełnienie jest konieczne dla poprawności wnioskowania?
3. Na czym opiera się statystyka Hotellinga T^2 i w jaki sposób uogólnia ona test t-Studenta na przypadek wielowymiarowy?
4. Podaj wzory statystyk Wilksa, Pillai'a, Lawleya-Hotellinga i Roya w kontekście testowania różnic między grupami.
5. Dlaczego testy wielowymiarowe są szczególnie wrażliwe na obecność obserwacji odstających i jak wpływa to na wyniki analiz?
6. Jak przebiega konstrukcja zmiennych kanonicznych w analizie kanonicznej i w jaki sposób dobierane są wektory wag maksymalizujące korelację między liniowymi kombinacjami dwóch zbiorów zmiennych?
7. W jaki sposób interpretować pierwszą parę zmiennych kanonicznych w kontekście zależności między dwoma zestawami cech oraz jak ocenić, które zmienne oryginalne mają największy udział w tej relacji?
8. Jakie znaczenie mają wartości własne w analizie kanonicznej?
9. W jaki sposób przeprowadzane jest testowanie istotności kolejnych korelacji kanonicznych oraz jak interpretować statystyki i wartości p w kontekście odrzucania hipotez dotyczących braku zależności między zbiorami?
10. Jak analizować macierze korelacji między zmiennymi oryginalnymi a zmiennymi kanonicznymi (*loadings*)?
11. Podaj przykład modelu CCA z wagami, ładunkami i przedstaw jego interpretację.
12. Jakie są podstawowe założenia analizy czynnikowej?
13. W jaki sposób model analizy czynnikowej rozdziela wariancję zmiennych obserwowalnych na część wspólną i specyficzną oraz jak interpretować te komponenty w kontekście jakości dopasowania modelu?
14. Jaką rolę pełnią ładunki czynnikowe w interpretacji czynników i w jaki sposób identyfikujemy grupy zmiennych tworzących wspólny czynnik?

15. Na czym polega rotacja czynników i dlaczego stosowanie rotacji ortogonalnych oraz ukośnych prowadzi do różnych typów interpretacji? Podaj przykłady różnic z punktu widzenia struktury czynników.
16. W jaki sposób oceniać liczbę czynników w modelu, wykorzystując informacje o wartościach własnych, wykresie osypiska oraz kryterium udziału wariancji? Jakie konsekwencje ma przyjęcie zbyt dużej lub zbyt małe liczby czynników?
17. W jaki sposób zasoby zmienności wspólnej (*communality*) odzwierciedlają stopień, w jakim dana zmienna jest reprezentowana przez czynniki, oraz jak analizować niskie i wysokie wartości w interpretacji wyników?
18. Na czym polega przypadek Haywooda w analizie czynnikowej?
19. Jakie są kluczowe różnice między eksploracyjną analizą czynnikową (EFA) a analizą czynnikową potwierdzającą (CFA)?
20. Jak definiujemy model SEM i co oznaczają poszczególne jego komponenty, takie jak zmienne ukryte, zmienne obserwowałe, ścieżki i błędy pomiarowe?
21. Czym się różni modelowanie ścieżek *path model* od modelu SEM?
22. Czym jest macierz implikowana na podstawie modelu SEM/CFA i do czego służy?
23. Podaj przykłady trzech wskaźników dopasowania modelu przedstawionych w wykładzie i jak interpretować ich wartości w kontekście oceny jakości modelu strukturalnego?
24. Jaka jest różnica pomiędzy CB-SEM i PLS-SEM?
25. Czy się różnią zmienne formatywne od reflektywnych?
26. Na czym polega spójność wewnętrzna i jak jest mierzona?
27. Jakie są główne cele analizy głównych składowych?
28. Jak interpretować wartości własne oraz wektory własne w kontekście analizy PCA?
29. Jakie znasz kryteria doboru liczby składowych głównych?
30. Czym różni się ICA od PCA i co oznacza operacja *whiteningu*?
31. Jak interpretować otrzymane komponenty niezależne w ICA?
32. Na czym polega idea klasycznego skalowania wielowymiarowego?
33. Jakie są różnice między metrycznym skalowaniem wielowymiarowym a niemetrycznym skalowaniem wielowymiarowym?
34. W jaki sposób t-SNE przekształca odległości między obserwacjami w rozkłady prawdopodobieństwa?
35. Najważniejsze kroki w zastosowaniu t-SNE.
36. Na czym polega algorytm UMAP i jakie są jego główne zalety w porównaniu do innych metod redukcji wymiarowości?
37. Na czym polegają metody hierarchiczne analizy skupień i jak interpretuje się dendrogram?
38. Jakie są różnice między metodami aglomeracyjnymi a deglomeracyjnymi w analizie skupień?
39. Podaj trzy przykłady miar odległości i metod aglomeracji w analizie skupień.
40. Jak się definiuje podział twardy?
41. Jak się definiuje podział rozmyty?
42. Opisz przegieg grupowania metodą k-średnich.
43. Czym różnią się metody k-średnich, k-median i k-medoid?
44. Opisz algorytm DBSCAN.

45. Opisz algorytm OPTICS.
46. Opisz algorytm HDBSCAN.
47. Opisz ideę GMM.
48. Do czego służy analiza korespondencji i czym jest tabela kontyngencji?
49. Opisz miary jakości odtworzenia przestrzeni czynnikowej w analizie korespondencji.
50. Jaki jest cel stosowania analizy log-liniowej?
51. Do czego służy algorytm Iterative Proportional Fitting (IPF)?
52. Czym się różnią zera strukturalne od zer próbkkowych w tabeli kontyngencji? Podaj po dwa przykłady każdego z nich.