

Probeaufgaben

Hinweis: Die Probeaufgaben beinhalten von der TK bereitgestellte fiktive Probedaten. Obwohl es sich hier um fiktive Probedaten handelt, **basieren diese auf realen Datensätzen**. Vor diesem Hintergrund sind diese Daten genauso wie später bei der Leistungserbringung **wie echte Versichertendaten zu behandeln** und **dürfen weder an Dritte weitergegeben noch in KI-Clouddienste hochgeladen werden!**

Ziel: Mit den Probeaufgaben sollen für eine Berichts- und Reporterstellung wesentliche Fähigkeiten im Hinblick auf den Umgang mit und die Auswertung von Daten realitätsnah getestet werden. Die Lösungen sind mit dem Angebot einzureichen, andernfalls erfolgt ein Ausschluss des Angebots.

Ablauf: Von der TK bereitgestellte fiktive Probedaten (Datentabellen mit a) Berufshistorien sowie b) gemeldeten AUs zu etwa 5.000 Personen zu den Kalenderjahren 2022, 2023 sowie 2024), die in vergleichbarer Form auch für die regulären Auswertungen genutzt werden, sollen im Rahmen der Aufgabe geprüft und ggf. bereinigt, aufbereitet sowie schließlich für die drei Jahre ausgewertet werden, wobei vorgegebene Kennwerte berechnet werden müssen.

Die Datentabellen enthalten in der ersten Spalte jeweils eine Variable mit einer fortlaufenden Nummerierung aller Tabellenzeilen innerhalb der Jahresdaten, die von Ihnen ggf. optional für Verweise auf Auffälligkeiten in den Datentabellen genutzt werden können.

a) realitätsnahe Berufshistorien (Dateien mld2022, mld2023 sowie mld2024)

Tabellen mit den Merkmalen: Versichertenpseudonym [ID], Von- und Bis-Datum des jeweiligen Tätigkeitsstatus [VON, BIS], Geschlecht [GESCHL], Geburtsjahr [GEBJAHR], Kreiskennung des Wohnortes – entspricht den ersten 5 Stellen des Gemeindegeschlüssel [KREIS], Kennung Berufstätig (=1) oder ALG-I-Bezug (=3) [PERSGRP].

b) realitätsnahe Daten zu AU-Fällen (Datei au2022, au2023 sowie au2024)

Tabelle mit den Merkmalen: Versichertenpseudonym [ID], Von- und Bis-Datum der AU [AU_VON, AU_BIS], ICD-10-Diagnose [ICD]

Aufgaben

Auf Basis der bereitgestellten fiktiven Probedaten sollen nachfolgend typische Kennzahlen zu Arbeitsunfähigkeiten berechnet werden. Zuvor sollen die Daten einer Prüfung unterzogen und anschließend für die Auswertung aufbereitet werden. Eine Nachlieferung von Daten ist grundsätzlich ausgeschlossen – Sie können bei den Analysen insofern ausschließlich auf die bereitgestellten Daten zurückgreifen und müssen ggf. Strategien für den Umgang mit fehlerhaft oder inkonsistent erscheinenden Daten entwickeln, um auf Basis der bereitgestellten Daten möglichst exakte Kennzahlen und inhaltlich aussagefähige Ergebnisse zu ermitteln.

A) Unterziehen Sie zunächst alle Tabellen einer separaten Prüfung. Beschreiben Sie ggf. festgestellte Unvollständigkeiten, Inkonsistenzen oder andere Auffälligkeiten und den von Ihnen im Rahmen von weiteren Auswertungen geplanten Umgang damit sowie die von Ihnen ggf. schon vor einer weiteren Aufbereitung an den zunächst separat betrachteten Datentabellen vorgenommenen Korrekturen (bei einzelnen und nicht anderweitig einfach selektierbaren Auffälligkeiten [$n < 10$], bitte pro Datentabelle möglichst die Zeilennummer [obs_m bzw. obs_a] nennen).

B) Skizzieren Sie die Grundsätze und Ihr Vorgehen bei der weiteren Aufbereitung der Daten zu Berufshistorien und Arbeitsunfähigkeiten sprachlich und beschreiben Sie die Struktur und Eigenschaften der final aufbereiteten Daten, auf die von Ihnen später im Rahmen der Ermittlung von Kennzahlen zu Arbeitsunfähigkeiten zurückgegriffen wird (eine Bereitstellung der Syntax bzw. der verwendeten Programme ist nicht erforderlich). Weisen Sie ggf. auf Auffälligkeiten bei der gemeinsamen Betrachtung der Daten aus den beiden Tabellen hin und erläutern Sie ihren Umgang mit diesen Auffälligkeiten.

C) Ermitteln und nennen Sie – bezogen auf Berufstätige, und dabei separat für die drei Kalenderjahre 2022, 2023 sowie 2024 – die folgenden Kennzahlen basierend auf den bereitgestellten und von Ihnen zuvor sinnhaft aufbereiteten Daten (bitte gewählte Berechnungsweisen kurz erläutern):

1. abs. Anzahl der Berufstätigen
2. abs. Anzahl der Versicherungstage
3. abs. Anzahl der AU-Fälle
4. abs. Anzahl der AU-Tage
5. AU-Fälle je Versicherungsjahr (Versicherungsjahr stets a 365 Tage)
6. AU-Tage je Versicherungsjahr (Versicherungsjahr stets a 365 Tage)
7. AU-Tage je Versicherungsjahr mit Diagnosen aus dem ICD-10-Kapitel „Krankheiten des Atmungssystems“
8. Krankenstand
9. abs. Anzahl der AU-Tage, die nicht auf gesetzliche Feiertage (Feiertage bundesweit oder in einzelnen Bundesländern gültig) und nicht auf Wochenenden (Sa, So) entfallen
10. Krankenstand an regulären Arbeitstagen (zur Abgrenzung von Arbeitstagen vgl. 9.)
11. Ermitteln Sie separat für alle Einträge in den drei Tabellen mld2022 bis mld2024 jeweils zu allen IDs kleiner oder gleich 30 die Anzahl der Berufstätigkeitstage, die nicht auf Feiertage oder Wochenenden fallen (vgl. 9.)

D) Ermitteln Sie die unter 1. bis 10. bei C) gelisteten Kennzahlen in einer Auswertung beschränkt auf Berufstätige im Alter von 15 bis 64 Jahren, wobei die Ergebnisse zu den unter 5. bis 8. sowie 10. genannten Kennzahlen für die drei Jahre hier in geschlechts- und altersstandardisierter Form präsentiert werden sollen. Verwenden Sie für die Standardisierung die nachfolgend gelisteten Angaben zu Erwerbstätigen in Deutschland im Jahr 2010 als Standardpopulation.

E) Beschreiben Sie eine Datenkonstellation, in der eine direkte Standardisierung gemäß der hier vorgegebenen Standardpopulation grundsätzlich nicht sinnvoll möglich ist. Skizzieren Sie zwei denkbare Ansätze/Lösungswege mit ihren Vor- und Nachteilen, wie bei entsprechenden Datenkonstellationen dennoch „standardisierte Ergebnisse“ errechnet werden könnten.

Standardpopulation: Erwerbstätige Deutschland 2010

Altersgruppe	Personenzahl (Tsd.)	
	Männer	Frauen
15–19	664	499
20–24	1.643	1.484
25–29	1.975	1.768
30–34	2.111	1.730
35–39	2.286	1.895
40–44	3.131	2.674
45–49	3.060	2.686
50–54	2.568	2.341
55–59	2.106	1.798
60–64	1.094	759