

DEBRECENI EGYETEM
EGÉSZSÉGTUDOMÁNYI KAR

Egészségügyi adatok felvétele és elemzése

Biostatisztika, Epidemiológia, Egészségügyi Informatika

Gyakorlati összefoglaló
Hallgatói jegyzet

Szerkesztette:

Dr. habil. Nagy Attila Csaba

Kirilla György

Szerző:

Dr. habil. Nagy Attila Csaba



Debreceni Egyetemi Kiadó
Debrecen University Press

Lektorálta:
Dr. habil. Móré Marianna

© Debreceni Egyetemi Kiadó Debrecen University Press beleértve az egyetemi hálózaton
belüli elektronikus terjesztés jogát is

CC-BY-NC-ND-4.0

ISBN: 978-963-615-075-4

Kiadta: a Debreceni Egyetemi Kiadó, az 1795-ben alapított
Magyar Könyvkiadók és Könyvterjesztők Egyesülésének a tagja

www.dupress.hu

Felelős kiadó: Karácsony Gyöngyi

Tartalom

TARTALOM.....	4
ELŐSZÓ.....	5
EPIDEMIOLOGIAI KITEKINTÉS.....	6
Gyakorisági mutatók.....	6
Kapcsolati mutatók.....	7
Epidemiológiai vizsgálatok.....	7
Hitelesség.....	9
Szűrési tesz jellemzők.....	10
ADATGYŰJTÉS.....	11
Kérdőívekről.....	11
Kérdőívek tervezése.....	11
Kérdőívek szerkesztése.....	11
A Microsoft Forms használata.....	12
ADATOK FELVÉTELE/ADATBEVITEL.....	16
Kettős adatbevitel.....	16
Adattisztítás.....	16
Szöveg az MS-Excelben.....	17
Adatok összekötése / FKERES függvény.....	19
Adatelemzés az MS-Excelben.....	21
STATISZTIKAI PROGRAMOKRÓL.....	23
Nagy adatbázisok kezelése (KNIME).....	23
Python.....	24
ADATELEMZÉS.....	28
SZAKIRODALOM KUTATÁSA.....	30
Referenciaszerkesztő.....	32
R PROGRAM.....	35
Mintaelemszám becslés.....	37
Leíró statisztika.....	39
Egyszerű elemzés.....	46
Paraméteres próbák.....	46
Nemparaméteres próbák.....	51
Korreláció.....	53
Többszörös elemzés.....	55
ÖSSZEFOGLALÓ AZ ELEMZÉSI MÓDSZEREKRŐL.....	62
EREDMÉNYEK ÁBRÁZOLÁSA.....	63
UTÓSZÓ.....	64
FELHASZNÁLT IRODALOM.....	65
FELHASZNÁLT ADATBÁZIS.....	66
FELHASZNÁLT ADATBÁZIS.....	67
FELHASZNÁLT ADATBÁZIS (QR KÓD).....	69
FELHASZNÁLT ADATBÁZIS KÓDSZÓTÁRA.....	70
FELADATOK.....	71
MEGOLDÁSOK.....	76
MELLÉKLET.....	82
Cukorbetegség kockázat mérő FINDRISC kérdőív.....	82
Alkoholfogyasztási rendellenességek azonosítási tesztje.....	84
Fagerström-féle nikotinfüggőségi teszt.....	86
Nemzetközi Fizikai Aktivitás Kérdőív Rövidített Változata.....	87
Európai lakossági egészségfelmérés, 2019 – Kérdőív – KSH (releváns részek).....	89
Tudományos közlemény.....	96

Előszó

Ez a jegyzet egy gyakorlati összefoglaló, melynek célja áttekinteni egy rövid kérdőíves vizsgálat teljes folyamatát a megtervezésétől a kivitelezésén át az elemzésig. Az összefoglalás kiterjed az epidemiológiára, egészségügyi informatikára és biostatistikára. Az egyes területek részletes elméleti háttérével külön tantárgyak foglalkoznak, ám a három érintett terület szervesen kapcsolódik egymáshoz. A gyakorlati megvalósítás segítheti az elmélet könnyebb elsajátítását, valamint a megtanult elméletet hivatott kiteljesíteni. Törekszünk arra, hogy a legfontosabb fogalmakat a gyakorlati alkalmazás szempontjai szerint határozzuk meg. Az egyes fejezetek logikus sorrendben követik egymást, de önmagukban is hivatottak egy-egy terület lefedésére. A könnyebb elsajátítás érdekében a kulcsinformációk vázlatos formában kerülnek feltüntetésre, valamint képernyőképek (screenshot) könnyítik a megértést/követést, legyenek ezek a használt programok telepítésével (install) kapcsolatosak, vagy szemléltessenek egy-egy eredményterméket (output). Végig haladunk a programok installálásától a megfelelő módszer kiválasztásán és kivitelezésén át, a kapott eredmények értelmezéséig. Az angol szaknyelv magas fokú elterjedtsége okán a gyakran használt idegennyelvű szakszavak is feltüntetésre kerülnek. Hasznos linkekkel és a felhasznált irodalommal zárjuk az anyagot.

Dr. Nagy Attila Csaba

„Non scholae sed vitae discimus.”

Epidemiológiai kitekintés

Az epidemiológiát gyakran a járványtannal azonosítják, ám ennél szélesebb körű, általánosabb tudomány. Gyakorlatilag kiterjed a fertőző és nemfertőző megbetegedések jellemzőinek leírására, kialakulásukat és lefolyásukat befolyásoló tényezők vizsgálatára. A különböző összefüggéseket és jellemzőket úgynevezett epidemiológiai mutatók írják le. Az epidemiológia a megelőzés teljes spektrumának szerves része a primer prevenciótól a terciér prevencióig.

A megelőzés vagy prevenció három fő részre osztható:

- Primer: egészségmegőrzés (a meglévő egészség fenntartása a cél), betegségmegelőzés, rizikótényezők (pl. dohányzás) kiküszöbölése például egészségnevelés vagy védőoltások révén
- Szekunder: szűrés, a már kialakult, de még tünetmentes betegség felismerése a cél, a mielőbbi terápia megkezdése érdekében
- Tercier: megfelelő gondozás/ellátás, lehetőség szerint teljes rehabilitáció

Gyakorisági mutatók

Az abszolút számokból gyakorisági következtetések nem vonhatók le, mivel a nevezők nem állnak rendelkezésre. Ha A megyében 10-en haltak meg, B megyében pedig 20-an, nem lehet a halálozás szintjét összehasonlítani a teljes megyei lakosságok ismerete nélkül. A gyakorisági mutatók viszont már alkalmasak következtetések levonására. Ilyenkor van egy viszonyítási alapunk, amely kiküszöböli az eltérő nevezők hatását.

- Prevalencia: $P=n/N$, vagyis az eseteket (lehet betegség, vagy állapot) elosztjuk a teljes populációval. Szinonimái: gyakoriság, előfordulás, állandó esetek. A 2-es típusú diabétesz prevalencia Magyarországon 8% körül van (800 000/1 000 000).
- Incidencia: új esetek előfordulását méri, két típusa van:
 - o Kumulatív incidencia: $KI=n/N$, a fő különbség a prevalenciához képest, hogy egy követési időtartamra vonatkozik, tehát az adott időtartam alatt jelentkező új esetek számát elosztjuk a kezdetben kockázatnak kitett populációval. Például, ha 100 emberből 10 betegszik meg egy év alatt, akkor 10% az abszolút kockázat (kumulatív incidencia).
 - o Incidencia sűrűség: $IS=n/SZI$, inkább dinamikus vizsgálati populációra vonatkozik, új esetek száma egy adott időtartam alatt, elosztva az összes kockázatnak kitett személyidővel (ameddig valaki él, a vizsgálatban van, képes megbetegedni, de nem beteg). Például, 5 új eset lesz 10 személyév alatt, akkor $5/10=0,5$ személyév⁻¹ az abszolút veszély (incidencia sűrűség).

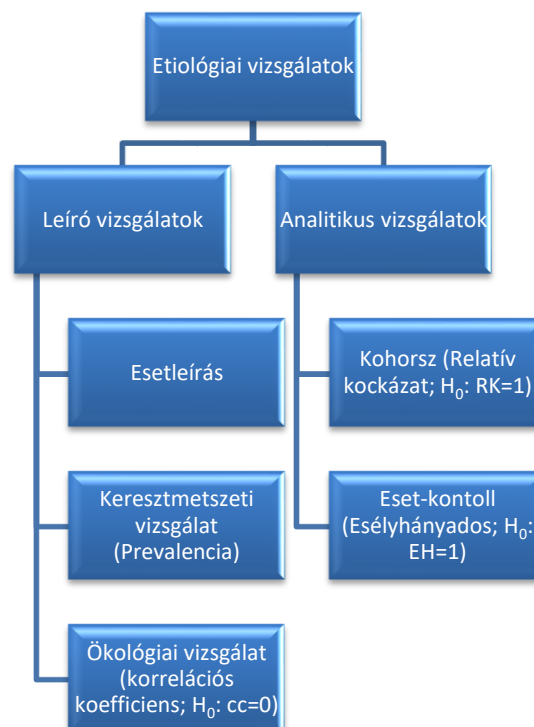
Kapcsolati mutatók

A gyakorisági mutatók ismeretében kapcsolati mutatókat tudunk számítani. Kapcsolati mutatók kiszámíthatóak az exponált (rizikó/protektív tényezőnek kitett) és nem exponált egyének csoportjára.

- Különbség mutatók:
 - Járulékos Kockázat: $JK=KI_1-KI_0$ A járulékos (többlet) kockázat kifejezi az expozíciónak köszönhető nettó rizikótöbbletet. Másnéven Kumulatív Incidencia Különbség.
 - Incidencia Sűrűség Különbség: $ISK= IS_1-IS_0$
- Hányados mutatók:
 - Relatív Kockázat: $RK=KI_1/KI_0$ A relatív kockázat kifejezi, hogy az exponált csoport hányszor nagyobb kockázatnak van kitéve a nem exponált csoport tagjaihoz képest (gyakorlatilag a két (exponált/nem exponált) csoport abszolút kockázatainak hányadosa). A semleges hatást az 1-es érték jelzi: 1-szer nagyobb, vagyis ugyanakkora a kockázat a két csoportban, nem valódi befolyásoló tényező.
 - Relatív Veszély: $RV= IS_1/IS_0$

Epidemiológiai vizsgálatok

A különféle epidemiológiai vizsgálatok (diagnosztikus, prognosztikus, etiológiai) közül az etiológiai vizsgálatokat használjuk a leggyakrabban. Ezeket másnéven kóroktani vizsgálatoknak nevezzük, hiszen a kimenetelt (betegség, halál stb.) potenciálisan befolyó tényezőket (rizikó vagy protektív) azonosítjuk a segítségükkel. A csoportosításuk:



1. ábra: Etiológiai vizsgálatok csoportosítása

- Deskriptív (leíró) vizsgálatok: nevükből fakadóan jelenségek leírása, hipotézisek generálása lehetséges a segítségükkel.
 - Esetleírás: egyedi/ritka eseteket ír le kórtörténettel
 - Keresztmetszeti: Prevalenciák mérhetőek segítségükkel, ide tartoznak például az egészségfelmérések révén mért betegséggyakoriságok
 - Ökológiai: a vizsgálat egységei a csoportok, korrelációs koefficiensek számíthatóak a segítségükkel. Azonban, ha a csoport szinten tapasztalható megfigyeléseket egyén szintjén próbáljuk meg értelmezni (helytelenül), akkor ökológiai hibát követünk el.
- Analitikus (számszerűsítő) vizsgálatok: a leíró vizsgálatok által generált hipotéziseket tudjuk tesztelni, valamint a kapcsolat erősségét számszerűsíteni.
 - Kohorsz: zárt csoport (a szó római eredetének megfelelően, ami zárt katonai csoportot jelentett), a vizsgálat elején mindenki egészséges, majd a nyomonkövetés során betegség alakul ki az exponált és a nem exponált csoportban is. A számítható kapcsolati mutató a relatív kockázat. Ezen vizsgálat típusok idő- és erőforrásigényesek, ezért nem alkalmasak ritka betegségek vizsgálatára. Általában prospektív (előre tekintő) vizsgálatokról beszélünk, de csökkenthető a költség a retrospektív (visszatekintő) vizsgálatok révén. Csoportosítható még a minta jellege alapján, lehet dinamikusan (nyílt kohorsz) változó (elköltözés/új belépés), vagy nem változó (zárt kohorsz), illetve (gyakran az utóbbi alá sorolják) fix kohorsz (expozíciós kategória sem változik).
 - Eset-kontroll: röviden a kohorsz ellentéte, tehát gyors, olcsó, alkalmas ritka betegség vizsgálatára, de nem alkalmas ritka expozíció vizsgálatára. A már beteg eset csoporthoz választunk véletlenszerűen kontrollokat (azonos forráspopulációból, hogy azokat reprezentálják, függetlenül az expozíciós státuszuktól, hiszen pont erről szeretnénk képet kapni), akik elkaphatják a betegséget, de nem betegek. A számítható kapcsolati mutató az Esélyhányados.

Hitelesség

A hitelesség (validitás) egy fontos kívánalom a vizsgálati eredményekkel kapcsolatban. Azt jelenti, hogy az összegyűjtött információ milyen mértékben válaszol pontosan a kutatási kérdésre. Két csoportra bontható:

- **Belső:** azt jelenti, hogy a vizsgálati eredmény érvényes a vizsgálatban résztvevőkre, feltétele a szisztematikus hibáktól mentesség.
- **Külső:** azt jelenti, hogy a vizsgálati eredmény a teljes forráspopulációra is érvényes, feltétele a belső hitelesség és a reprezentativitás.

A szisztematikus hibák 3 csoportra bonthatóak:

- **Megfigyelési hiba:** a megfigyelő (adatrögzítő/kitöltő) téves információt rögzít, akár visszaemlékezési torzítás miatt, ezáltal torzul például az exponált vagy a nem exponált csoportba kerülés.
- **Kiválasztási hiba:** hibás a csoportkiválasztás, például nem véletlenszerű a kontroll választás, hanem inkább nem dohányzókat teszünk a kontroll csoportba, amely így nem lesz reprezentatív a forráspopulációra (akire a következtetést le szeretnénk vonni a véletlen mintámon végzett vizsgálat alapján).
- **Zavaró tényező:** három kritérium teljesülése esetén beszélünk zavaró tényezőről:
 - o Független kockázati tényezője a kimeneti változónak
 - o Kapcsolatban van a feltételezett expozícióval (amit vizsgálunk)
 - o Nincs rajta az oksági láncon

Fiktív példa a kávéfogyasztás (expozíció) és hasnyálmirigyrák potenciális kapcsolatát elemző vizsgálat, melynél a dohányzás (a mindhárom kritériumnak eleget tevő) zavaró tényező. Ez utóbbi felelős a valós hatásért. (Általános zavaró tényező a nem és az életkor.)

Lehetőség szerint megelőzni kell az előfordulását:

- o Randomizálás (véletlenszerűen kiválasztott minta, ez a legjobb megoldás)
- o Korlátozás (kizárom az egyik alcsoportot, de ezzel információt veszítek)
- o Illesztés (zavarótényező mentén illesztem a csoportokat)

Amennyiben nem sikerült megelőzni, akkor korrigálni kell a hatásukra:

- o Rétegzett elemzés (zavaró tényező mentén rétegzünk)
- o Regresszió (gyorsabb és jobb megoldás, főleg sok zavaró tényező/réteg esetén)

Szűrési teszjt jellemzők

A másodlagos (szekunder) prevenció egy szóval: szűrés. A szűrési teszteknek négy fő jellemzőjük van:

- Szenzitivitás: érzékenység, a valódi pozitív arányt mutatja meg, vagyis mennyire jó a tesztünk a betegek megtalálásában ($a/a+c$)
- Specificitás vagy fajlagosság: a valódi negatív arányt mutatja meg, vagyis mennyire jó a tesztünk betegségtől mentesek azonosításában ($d/b+d$)
- Pozitív prediktív érték (PPÉ): megmutatja, hogy a szűrési teszt alapján pozitívként azonosított egyén, milyen valószínűséggel beteg ($a/a+b$)
- Negatív prediktív érték (NPÉ): megmutatja, hogy a szűrési teszt alapján negatívnak talált egyén milyen valószínűséggel mentes a betegségtől. ($d/c+d$)
- Prevalencia: betegség előfordulási gyakorisága $(a+c)/(a+b+c+d)$

		Diagnózis		
		beteg	nem beteg	
Teszt	pozitív	a	b	a+b
	negatív	c	d	c+d
		a+c	b+d	a+b+c+d

1. táblázat: szűrési teszteredmény viszonya a valós diagnózishoz

Az egyes cellákban lévő jelölések:

- a: valódi pozitív, beteg és tudja
- b: álpozitív, nem betegek, de hamis betegségtudata lehet
- c: álnegatív, betegek, de nem tud róla, nem kap időben kezelést
- d: valódi negatív, nem beteg és tudja

Adatgyűjtés

Kérdőívekről

A kérdőív az adatgyűjtés jól strukturált módszere, amely kérdések sorozatából áll. A kérdőíves felmérések általánosan elterjedt formái az információgyűjtésnek, melyeket az egészségfelmérésektől a célzott vizsgálatokon át a közvéleménykutatásokig széleskörűen alkalmaznak.

Kérdőívek tervezése

A kérdőíves felméréseknél használhatunk már létező, validált kérdőívet (természetesen a megfelelő citáció mellett), vagy létrehozhatunk új kérdőívet. Az újonnan létrehozott validálandó kérdőíveknél lényegesek a szabad szöveggel befejezhető nyitott kérdések. A végleges kérdőíveknél a zárt kérdéseket tartalmazó kérdőívek az elterjedtebbek az elemzés könnyebbége okán is. Fontos a kérdések sorrendje, megfelelő kérdéscsoportok kapcsán, valamint azokon belül is. Visszaulva a fiktív példánkra: a rizikótényezők feltárása során érdemes a kávéval kezdeni, majd a dohányzással folytatni, végül feltenni az alkoholfogyasztással kapcsolatos kérdéseket. Ezen fontos egészségmagatartással kapcsolatos tényezők felmérésénél a kitöltő könnyebben válaszol egy kevésbé érzékeny kérdésre (kávé), majd miután oldódik, jöhet a következő kérdés (alkohol). A kérdések mellett a válaszoknak is logikus sorrendben érdemes követniük egymást. Ellentmondásos a kérdésszám helyzete, mivel a több kérdés több információt szolgáltat, viszont a kitöltő hajlamos idővel elfáradni és kevésbé pontos válaszokat adni. Régebben az optimális kérdésszámot 20-30 körülire, az elvi maximumot pedig 70 körüli kérdésre tették. A modern online kérdőíveknél érdekesebb inkább kitöltési időben gondolkozni, az optimális hossz 10-20 perc.

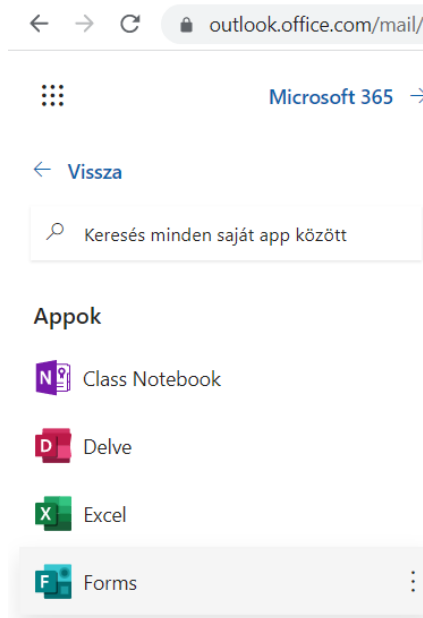
Kérdőívek szerkesztése

Korábban papír alapú kérdőíveket nyomtattak, manapság ezeket felváltja az online, mely lehet saját honlapon elhelyezett, valamely kérdőívekkel foglalkozó cégtől vett, vagy pedig ingyenes. Utóbbiak közül a Google Forms népszerű, mégis érdekesebb a Microsoft Formsot használni a szigorúbb adatkezelés és adatbiztonság miatt.

A Microsoft Forms használata

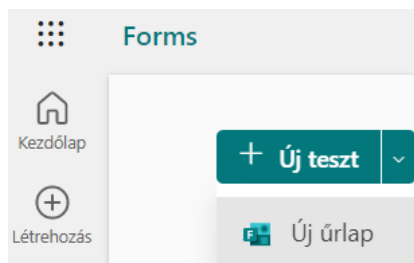
A következőkben képernyőképeken (screenshot) keresztül kerül bemutatásra a Microsoft Forms elérése, valamint használata.

A Microsoft Forms a Microsoft Office/365 programcsomag része.



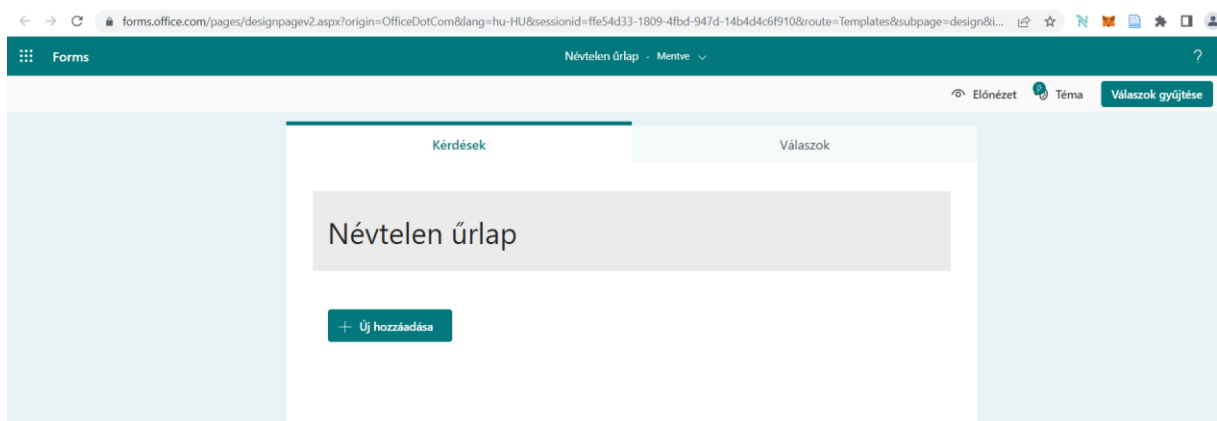
2. ábra: Microsoft Forms ikon helye

Új tesztek, vagy új űrlapok választhatók indítás után:



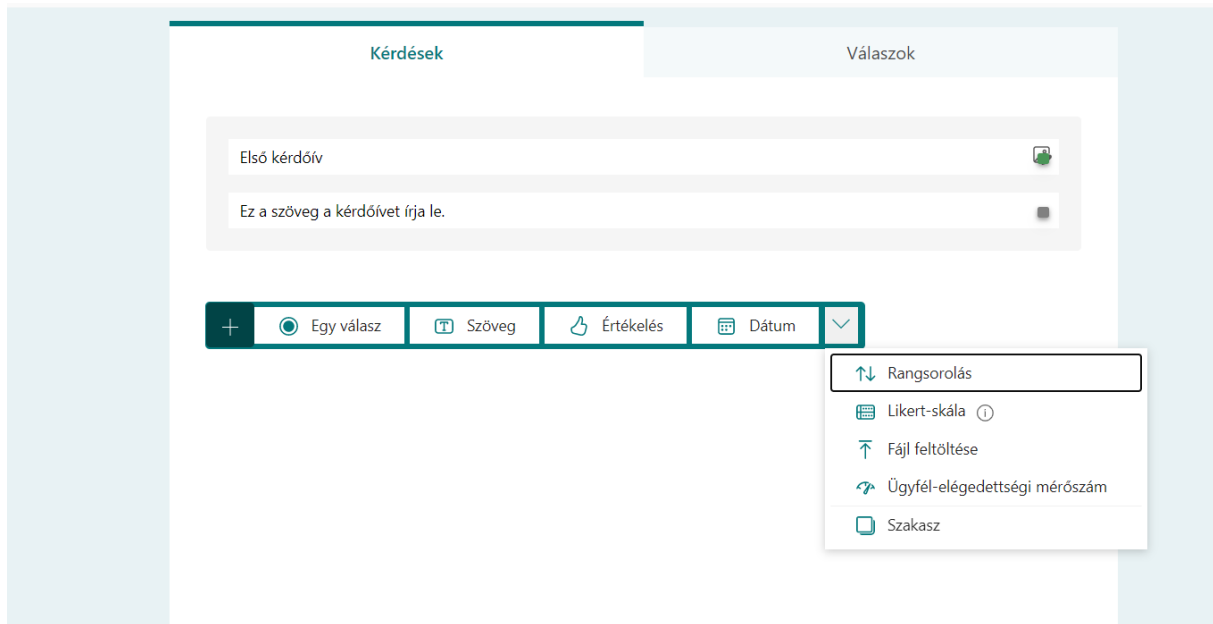
3. ábra: Új teszt létrehozása

Első lépésként érdemes elnevezni az új űrlapunkat/kérdőívünket:



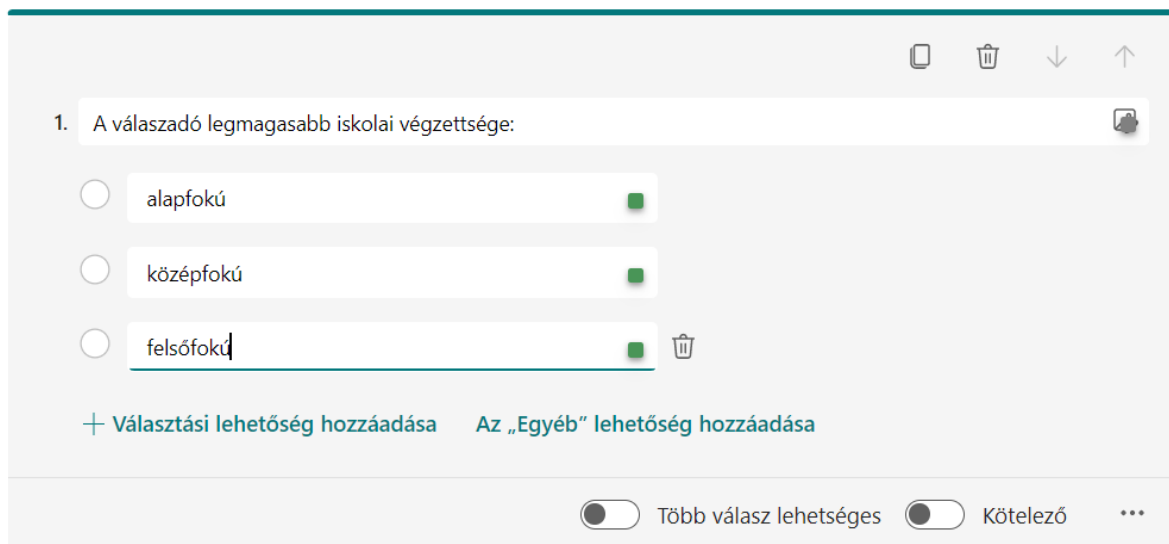
4. ábra: Űrlap elnevezése

A kérdőív elnevezése után lehetőségünk van egy rövid tájékoztató megadására is a felméréssel kapcsolatosan, utána pedig különféle kérdéstípusokból választhatunk:



5. ábra: Kérdéstípusok

Az egy választ váró kérdéseknél az „Egy válasz” opcióra kattintsunk. A kérdésünk lehet „Kötelező”, valamint „Több válasz lehetséges” opció is használható, amennyiben több válasz is megadható a kérdésre. A kérdések szerkeszthetőek, törölhetőek, illetve a sorrend is változtatható, mind a kérdések, mind a válaszok tekintetében:



6. ábra: Választási lehetőségek hozzáadása

1. A válaszadó legmagasabb iskolai végzettsége:

- alapfokú
- középfokú
- felsőfokú

2. Szabad szöveget írjon:

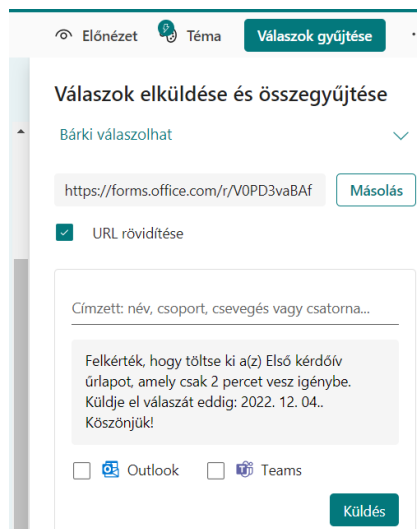
Írja be a választ

3. Többet is megjelölhet:

- első opció
- második opció

7. ábra: Fő kérdéstípusok

A kérdőív elkészülte után használható/terjeszthető a „Válaszok gyűjtése” opció révén.

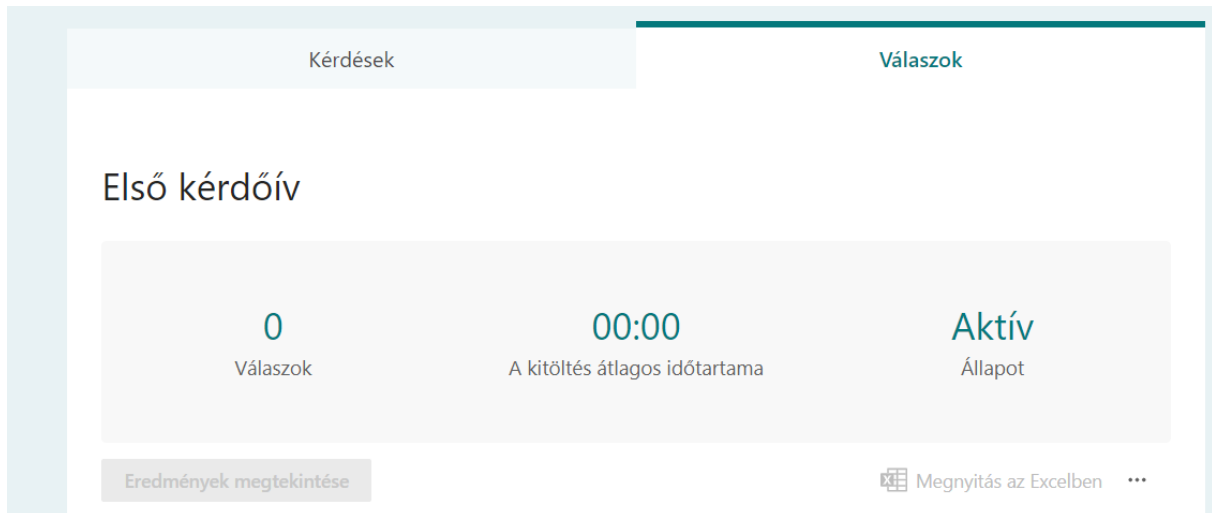


8. ábra: Kérdőív terjesztése

Érdemes beállítani a „Bárki válaszolhat” és az „URL rövidítése” opciót a kérdőív könnyebb terjesztés érdekében.

A „Válaszok” menüpontban részletes információ áll rendelkezésre a kitöltőkkel kapcsolatban:

- kitöltők száma
- kitöltés átlagos ideje
- eredmények diagramokkal
- letölthető adatbázis MS-Excel formátumban



9. ábra: Válaszok jellemzői

Adatok felvétele/adatbevitel

Fontos elv a „garbage in, garbage out”, vagyis az „értéktelen” adatokból „értéktelen” elemzéseket lehet végezni. Az elv nem pusztán az adatok felvételére vonatkozik, hanem az adatok megfelelő rögzítésére is. A kész (rögzített) adatbázist tisztítani kell. A kérdések számáról, hosszáról már ejtettünk szót. A kérdéseket érdemes csoportosítani/tagolni. Hivatkozható mintául szolgálhat az Európai Lakossági Egészségfelmérés (ELEF) kérdőíve (https://www.ksh.hu/elef/elef2019_kerdoiv.pdf).

A papír alapú kérdőívek külön adatrögzítést igényelnek. A rögzítés során a papír alapú információt digitális karakterekké (számokká és betűkké) alakítjuk. Az adatbevitel során a hiányzó értékeket „9”-cel kódoljuk, hogy el lehessen különíteni a hiányzó érték okát (kitöltő nem írta oda, vagy a rögzítő nem rögzítette). Annyi „9” szükséges, ami nem reális érték már az adott kérdésre (pl. életkornál „999”). Az elemzéshez úgynevezett kódszótárt kell készítenünk, mely részletesen leírja a kérdéseket, lehetséges válaszopciókat és az adatbázisban használt változónevet és kódolást. A statisztikai programok az ékezet nélküli, rövid változónevekkel boldogulnak könnyen. A kódszótárakban feltüntetjük a változó nevét, a hozzá tartozó kérdést és a potenciális válaszokat a megfelelő kódokkal.

	A	B	C	D	
1	Változó	Kérdés	Válasz kódja	Válasz opció	
2	nem	1. Mi az Ön neme?	1	férfi	
3			2	nő	
4	iskola	2. Mi a legmagasab	1	alapfokú	
5			2	középfokú	
6			3	felsőfokú	
7					
8					
9					
10					

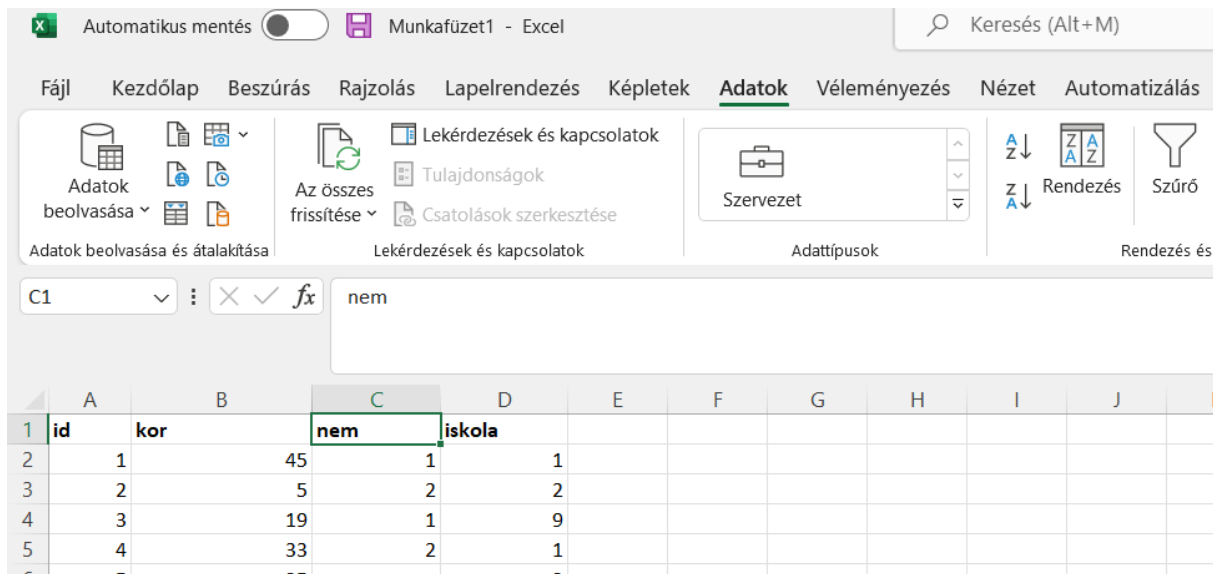
10. ábra: Kódszótár

Kettős adatbevitel

Amennyiben papír alapú kérdőívet szeretnénk rögzíteni, érdemes megfontolni a kettős adatbevitelt. Ez egyszerűen azt jelenti, hogy kétszer kerül rögzítésre minden kérdőív. Ennek a hátránya a többlet idő és a többlet költség. A nagy előnye a rögzítési hibáktól mentes adatbázis, mivel a dupla adatbázis összevethető, az eltérések pedig visszakereshetők.

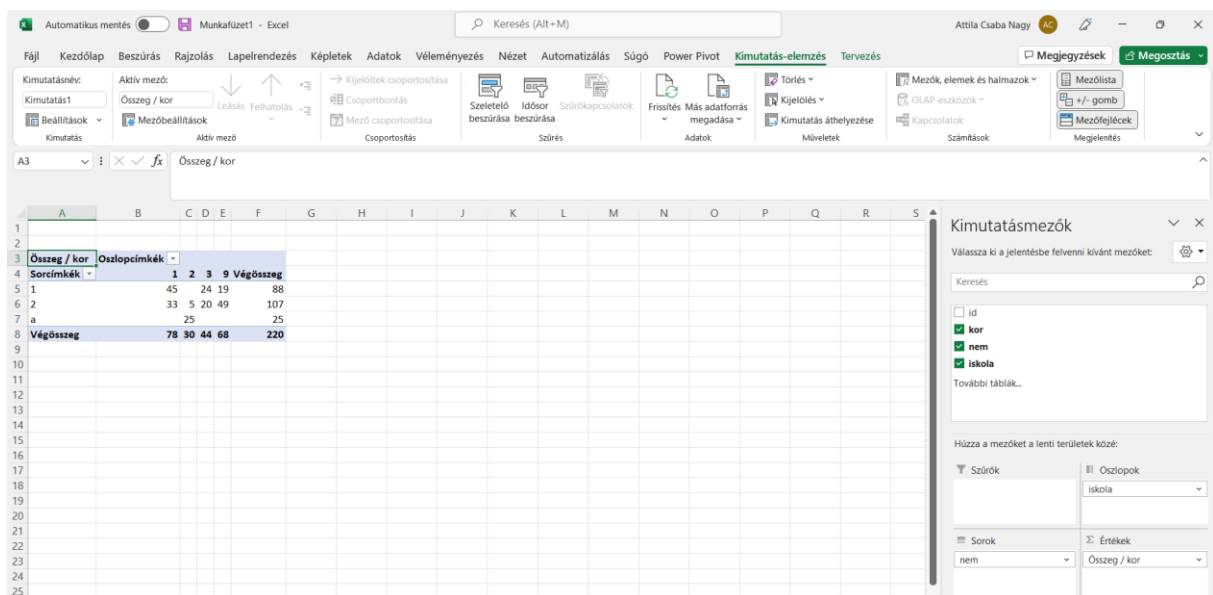
Adattisztítás

Elemzés előtt a nyers (rögzített vagy kapott) adatbázist tisztítani szükséges. Ez a folyamat akár MS-Excelben is kivitelezhető. Nagyon hasznos a „Szűrő” funkció.



11. ábra: Szűrő megjelenítése

A szűrő megmutatja a potenciális értékeket az adott oszlopban, így egyből láthatóvá válik az oda nem illő (hibásan kitöltött vagy bevitt) érték. A hibás, irreális és 9-cel kódolt értékek hiányzóra cserélhetők. Egyéb funkciók mellett (pl. aggregálás) az adattisztításban is segít a „Kimutatás”, mely a „Beszúrás” menüpont alatt található.



12. ábra: Kimutatás létrehozása

Egyszerű „fogd és vidd” (drag and drop) módszerrel az egyes változók a sor vagy oszlop részre húzhatók. Akár több réteg/alcsoport is rendezhető egy tengely mentén. Itt is található egy szűrő, ami segít leszűkíteni az adatbázist. Az „Értékek” révén pedig nem pusztán az összegek számíthatók ki, de akár az elemszámok is (darab).

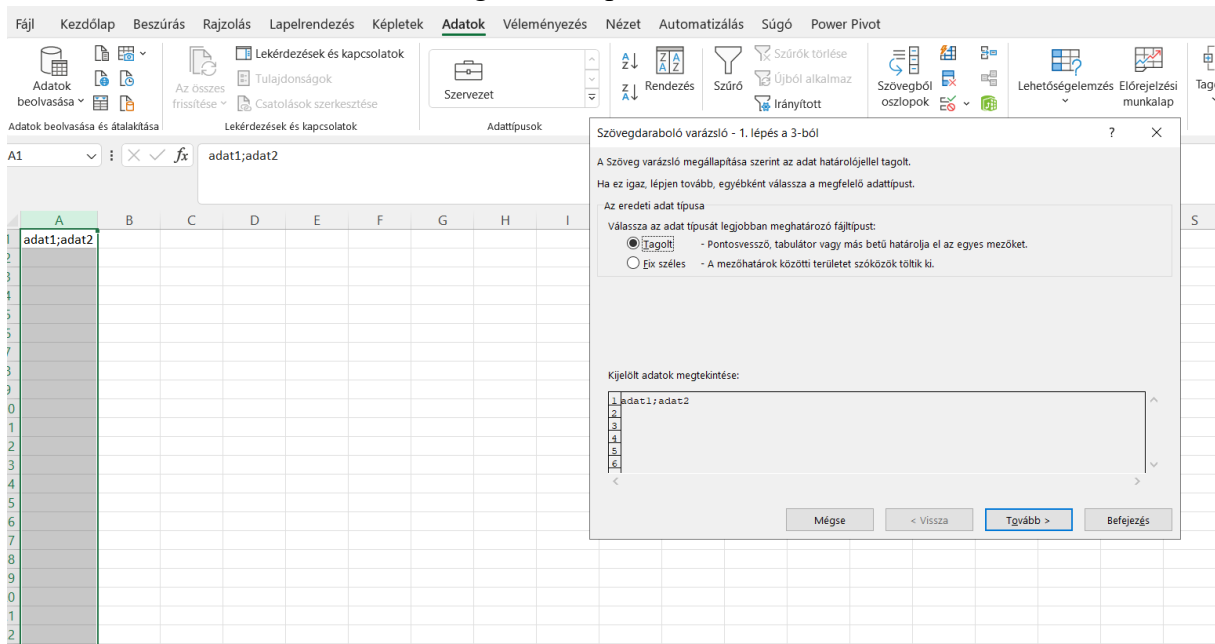
Szöveg az MS-Excelben

Az egyik legáltalánosabb adatbázis formátum a szövegfile. Leggyakrabban az értékek vesszővel vannak elválasztva (.csv), tehát vesszővel elválasztott értékek (comma-separated

values) szerepelnek a fájlban. Általánosabb a bármilyen egyéb karakterrel elválasztott (szóköz, tabulátor stb.) adatfájl (.txt). Az angol és magyar adatkezelésben a tizedesjel a legfőbb eltérés, ami forrása lehet további adathibáknak (pl. dátummá konvertálódó számok). A magyar .csv kissé megtévesztő, hiszen itt nem vessző, hanem pontosvessző az elválasztó karakter.

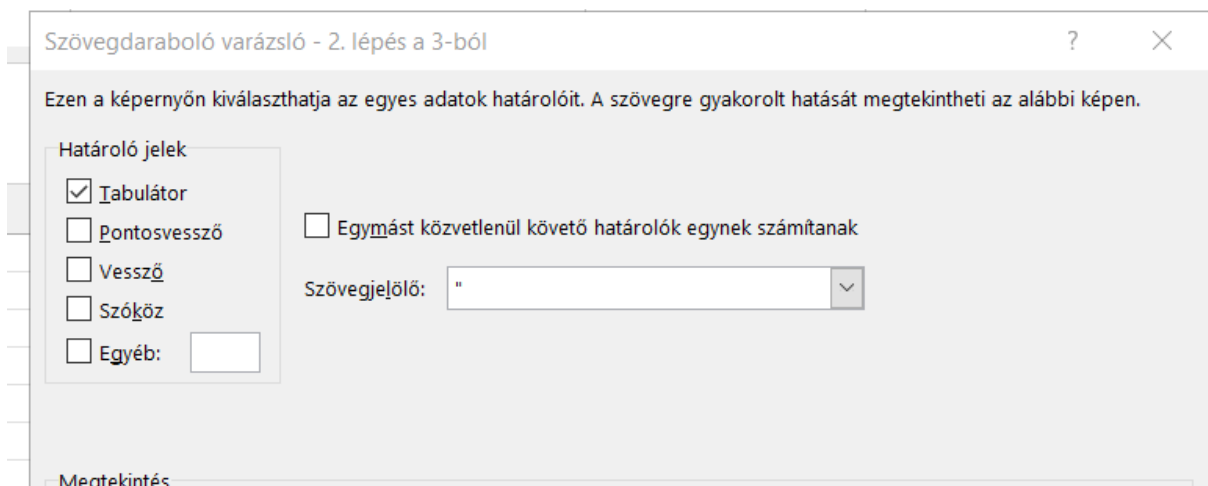
A „Szövegdaraboló varázsló” két módon hozható elő:

- szöveges adat Excel-be másolás
- „Adatok” fülön belül „Szövegből oszlopok” kiválasztása



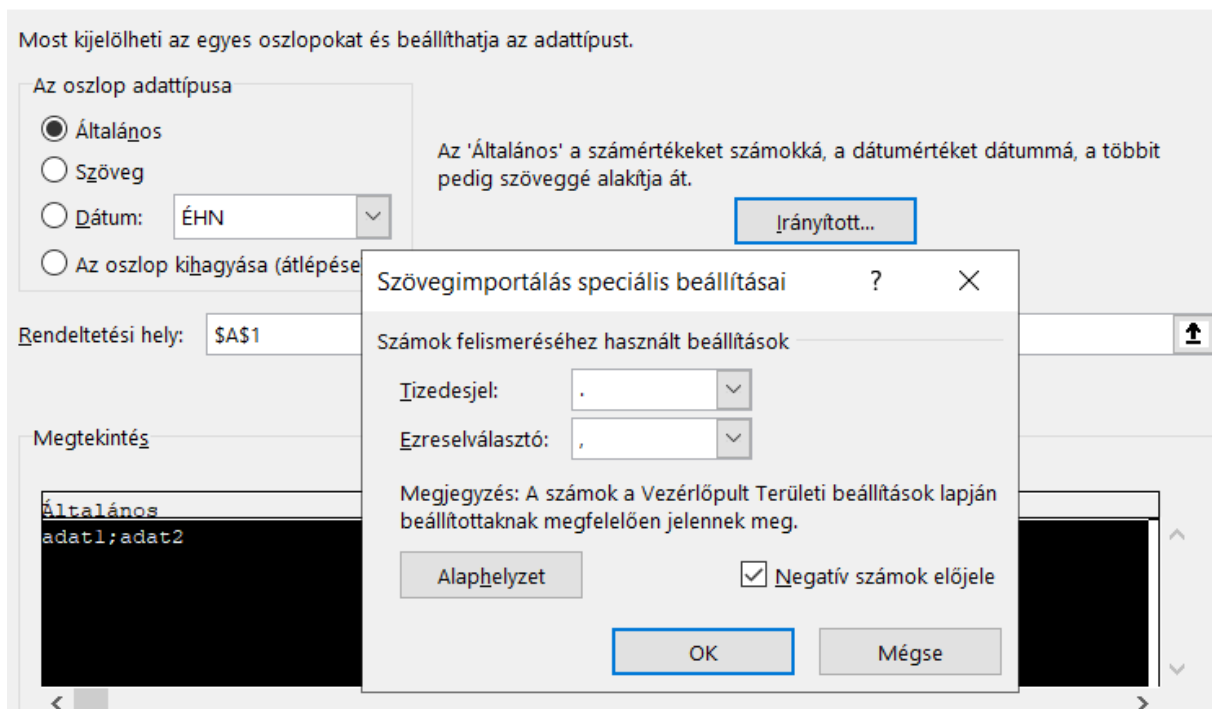
13. ábra: Szövegdaraboló varázsló

Mindkettőnél fontos az elválasztó (field separator) kiválasztása, amely a felsoroltak mellett bármilyen tetszőleges karakter is lehet.



14. ábra: Mező elválasztó

Amennyiben angol adatbázist illesztünk be, mindenképpen fontos a következő oldalon található tizedesjel és ezreselválasztó.



15. ábra: Tizedesjel beállítása

Az automatikus konverzió elkerülése érdekében a tizedesjel pont legyen, az ezreselválasztó pedig vessző.

Adatok összekötése / FKERES függvény

Az adatbázisok összekötésének két fő típusa van.

- *hozzáfűzés (append)*: azonos struktúrájú adatbázisok összefűzése történhet akkor, ha több ember visz be kérdőíveket azonos rendszerben és a végén egy nagy adatbázist szeretnének létrehozni. Egyszerű egymás alá másolásról van szó, a biztonság kedvéért a címsor alulra másolható, visszaellenőrzés után pedig törölhető a másolat és az új címsor.

	A	B	C	D
id	kor	nem	iskola	
	1	34	1	1
	2	12	2	2
	3	25	1	2
	4	33	1	2
id	kor	nem	iskola	
id	kor	nem	iskola	
	5	15	2	2
	6	30	2	3

16. ábra: Adatbázisok hozzáfűzése

- *összefűzés (merge)*: gyakrabban használatos, ilyenkor a meglévő adatbázist egy másikkal fűzzük össze. Leegyszerűsítve értékekkel egészítjük ki az adatbázist egyértelműen azonosító változó mentén.

Az FKERES függvény két fő funkciója a pontos egyezés keresés és a kategorizálás.

A pontos egyezésnél egyértelmű azonosító mentén keressük a vágópontot egy másik adatbázisból. Jelen példánknál az iskolai végzettséget hozzuk át egy másik adatbázisból:

id	kor	nem	iskola	id	iskola
1	34	1	s	1	34
2	12	2		2	12
3	25	1		3	25
4	33	1		4	33

17. ábra: FKERES függvény beállítása pontos egyezés esetén

Fontos a „Táblázat” értékénél abszolút hivatkozást használni a másodlagos tartomány kijelölésénél (\$ jel), így a képlet másolásánál nem fog elcsúszni a másodlagos adattáblát jelölő tartomány.

A kategorizálás során folytonos változót alakítunk át kategorikussá (pl. pontszám-jegy; életkor-korcsoport):

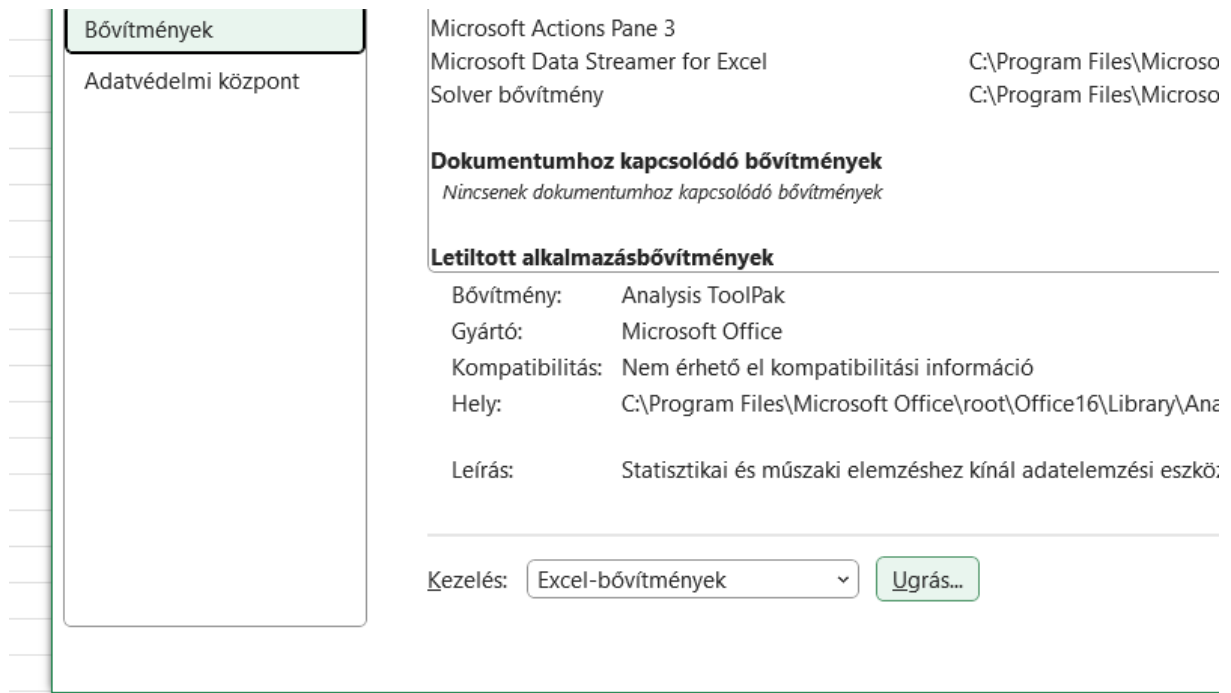
id	pont	jegy	pont	jegy
1	55	IGAZ	0	1
2	85		60	2
3	95		70	3
4	33		80	4
			90	5

18. ábra: FKERES függvény beállítása közelítőlegesen egyezés esetén

Ilyenkor is abszolút hivatkozást érdemes használni, viszont a „Tartományi_ keresés” „IGAZ”, vagyis az előző példával ellentétben (ahol „HAMIS” volt) nem a pontos egyezés esetén ad vissza értéket, hanem értéktartományok mentén dolgozik.

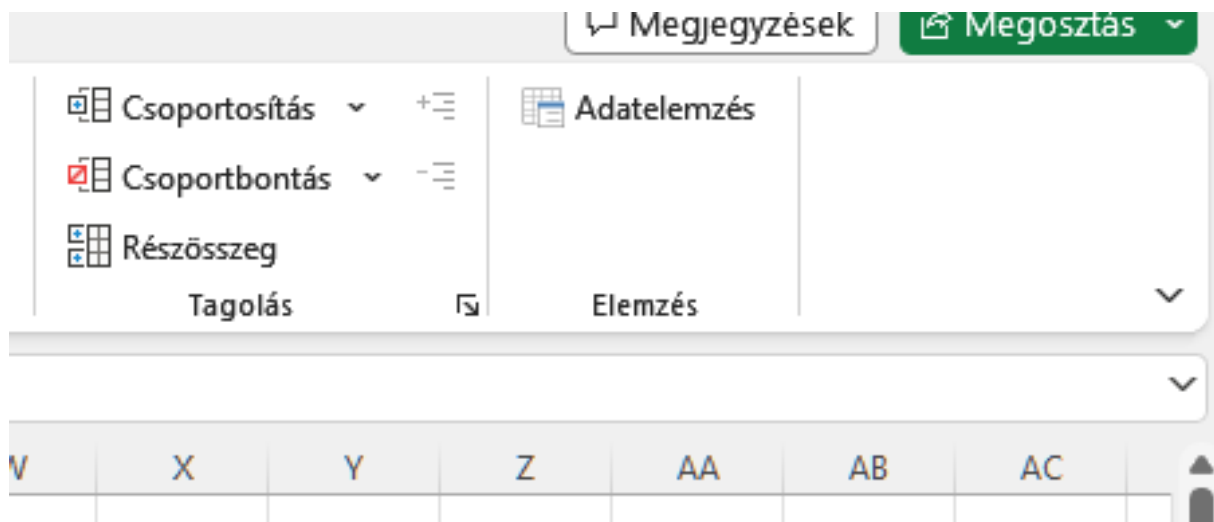
Adatelemzés az MS-Excelben

Az MS-Excel meglepően sokoldalúan képes az adatok megjelenítésére és kezelésére. A beépített függvények révén sokféle statisztikai módszer elérhető. Vannak ingyenesen és pénzért letölthető további bővítmények is. Van viszont egy beépített bővítmény, ami nem érhető el alapértelmezetten, ez az „Adatelemzés”. Előhozásához a „Fájl”, „Beállítások”, „Bővítmények” kattintások után az „Excel-bővítmények” lehetőségre kell „Ugrani”:

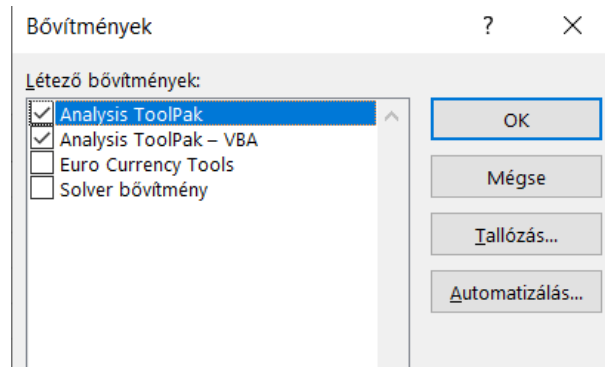


19. ábra: Excel-bővítmények elérése

Ezután bepipálva az „Analysis ToolPak” opciókat, elérhetővé válik az „Adatelemzés” az „Adatok” menüpont végén.

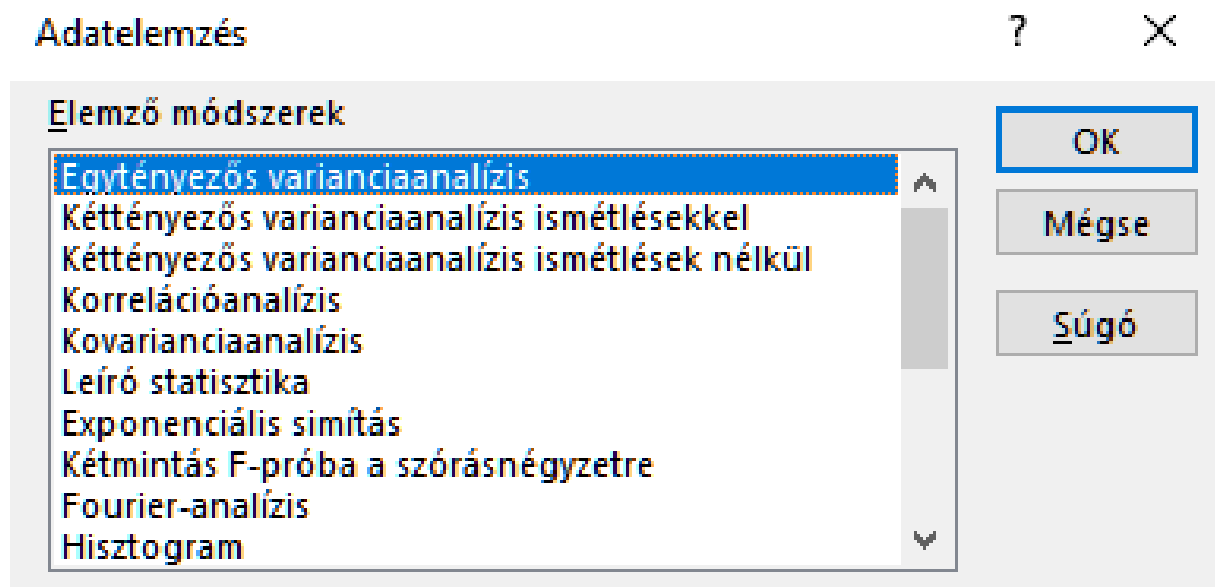


20. ábra: Adatelemzés gomb elhelyezkedése



21. ábra: Adatelemzés megjelenítéséhez szükséges bővítmények

Itt már a leíró statisztika mellett az egytényezős varianciaanalízisen (ANOVA) túl, többszörös lineáris regresszió is kivitelezhető (lásd később).



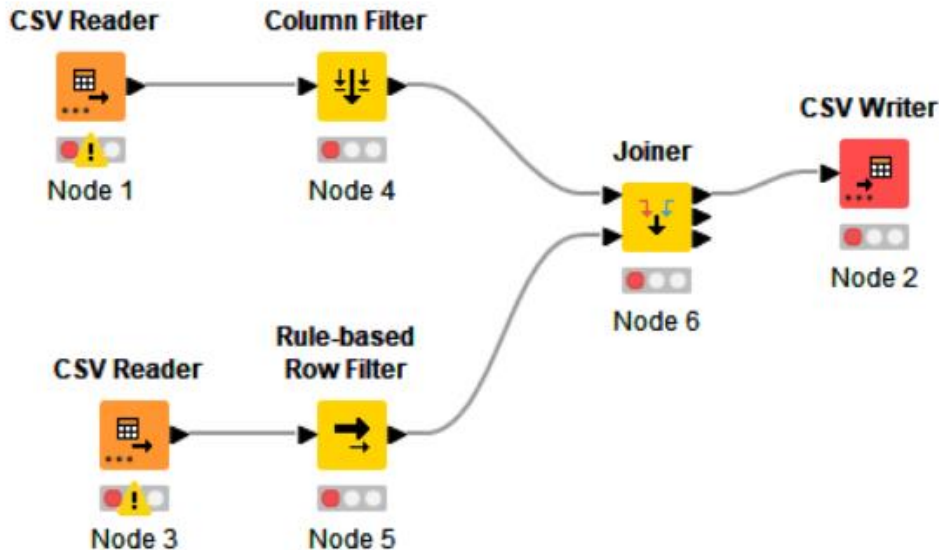
22. ábra: Adatelemzésben elérhető elemző módszerek

Statisztikai programokról

A statisztikai elemzéseket lehetővé tevő programok széles palettán mozognak. Vannak, amiknél másodlagos funkció az elemzés (MS-Excel) és csak korlátozott módszerek állnak rendelkezésre. A kimondottan statisztikai elemzésre kifejlesztett szoftverek között találhatunk ingyenes (pl.: R, PSPP), fizetős (pl.: Stata, SPSS, SAS) változatokat, valamint programnyelveket (pl.: Python) statisztikai modulokkal.

Nagy adatbázisok kezelése (KNIME)

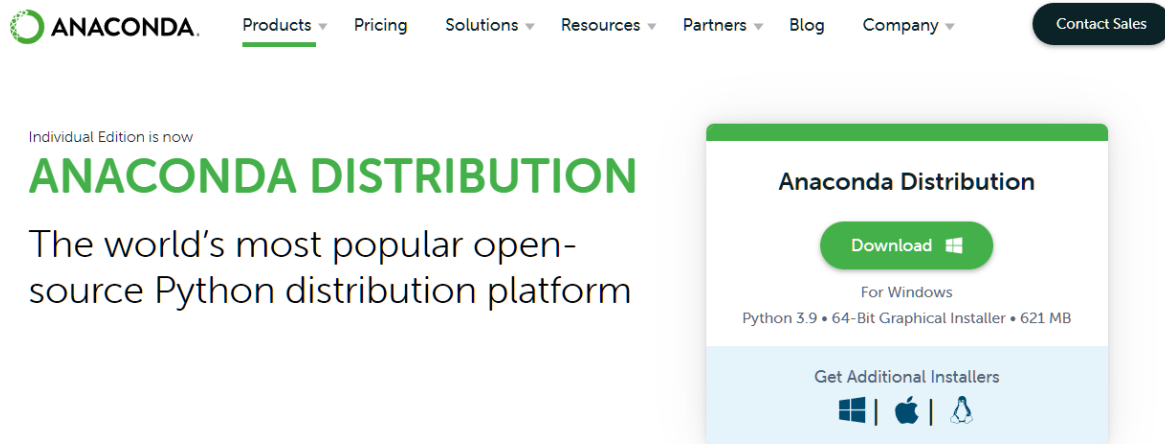
Manapság egyre több és nagyobb adatbázis keletkezik, úgynevezett „Big Data” méretben. Itt a hatalmas méret (volume) mellett (pl.: gigabyte, terabyte, petabyte) a dinamikus változás is jellemző. Az MS-Excel korlátja 1 048 576 sor és 16 384 oszlop. Az Excel valójában nem adatbáziskezelő, hanem táblázatkezelő. Az ennél nagyobb adatbázist érdemes speciális adatbáziskezelő szoftverben vagy statisztikai programban megnyitni. Viszonylag „kisebb” adatbázisokat is érdemes speciális programok segítségével átalakítani vagy akár megnyitni. Érdemes pár szót ejtenünk a szintén ingyenes Knime-ről. A statisztikai programokat is megelőzi a fájlműveletek gyorsaságának tekintetében. A program úgynevezett csomópontokkal (node) operál. Beolvashatók fájlok (CSV Reader), szűrhetők (Row/Column Filter) az adatok, egyesíthetők (Joiner) adatbázisok, valamint kiírhatók (CSV Writer) állományba.



23. ábra: KNIME csomópontok

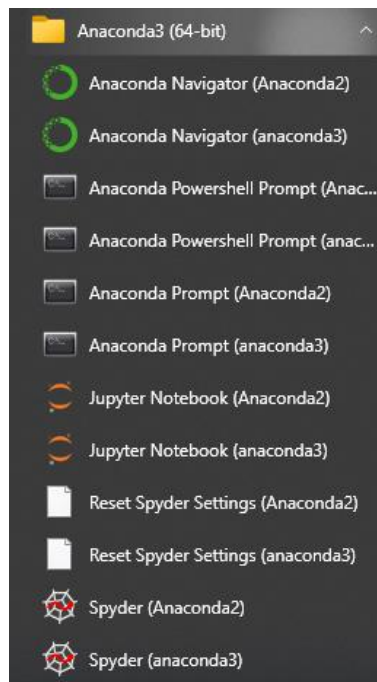
Python

A Python az egyik legelterjedtebb programnyelv. Mivel egyre több az adattudománnyal kapcsolatos csomagja, így egyre nagyobb népszerűségnek örvend az adattudósok (data scientists) körében is. Gyakorlatilag minden statisztikai módszert ismer alapértelmezetten, vagy ingyenes kiegészítő csomagok révén. Nagy előnye az ingyenesség mellett, a gyorsasága és széleskörű alkalmazhatósága. Minden operációs rendszerre letölthető, az egyik legnépszerűbb disztribúciója az Anaconda (<https://www.anaconda.com/products/distribution>).



24. ábra: Anaconda Python disztribúció honlapja

Letöltés és telepítés után a „Jupyter Notebook (anaconda3)” ikont érdemes futtatni.



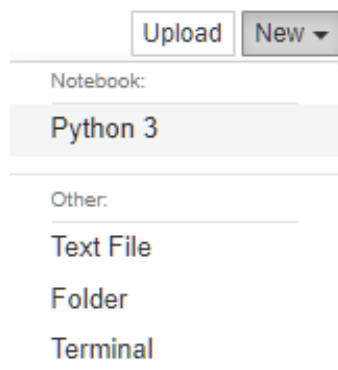
25. ábra: Anaconda3 a Start menüben

Futtatás után az alapértelmezett webböngészőben nyílik meg a felülete.



26. ábra Anaconda felülete az alapértelmezett böngészőben (<http://localhost:8888/tree>)

Új állomány létrehozásához a „New” majd „Python 3” gombra kell kattintani.



27. ábra: Új Python állomány létrehozása

Az alapértelmezett állomány, amely a kódokat és az eredményeket tartalmazza .ipynb kiterjesztésű és a felhasználó könyvtárában helyezkedik el (C:\Users\felhasználónév\).

A Python programozásához számos segédlet érhető el az interneten, jelen jegyzet keretei között csupán néhány alapvető paranccsal ismerkedünk meg.

A 28. ábrán található parancsok, és futtatásuk esetén kapott eredmények a következők:

In[1]: első parancs (2+3)

Out[1]: a parancs eredménye (5)

A beírt parancsot a Shift+Enter billentyűkombinációval futtathatjuk le.

A „pip install pandas” parancs installálja az egyik leggyakrabban használt csomagot (pandas). Ez alapértelmezetten a rendszer része, ezt jelzi a figyelmeztetés (warning/notice).

Az „import panda as pd” parancs betölti a „pandas” csomagot és „pd”-ként lehet majd hivatkozni rá a jövőben.

A „df = pd.read_csv(r'c:/minta/minta.csv')” parancs egy df nevű dataframe változóba tölti be a minta.csv fájlunkat. Amennyiben az elválasztó (separator) nem az alapértelmezett vessző, akkor ezt külön meg kell adni, vagy előtte át kell cserélni a csv állományban.

A „df.head()” parancs kilistázza a df változó tartalmát.

A „df.age.mean()” parancs pedig a kor változó átlagát adja vissza.

Egy adott sorra kattintva kijelölhető a sor, jelenleg az In[10] van kijelölve. Az „x” billentyű kitörli (kivágja) az adott sort, a „b” billentyű pedig új parancssort szúr be.

```

In [1]: 2+3
Out[1]: 5

In [2]: pip install pandas
Requirement already satisfied: pandas in c:\users\a\anaconda3\lib\site-packages (1.1.2)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\a\anaconda3\lib\site-packages (from pandas) (2.8.1)
Requirement already satisfied: pytz>=2017.2 in c:\users\a\anaconda3\lib\site-packages (from pandas) (2019.3)
Requirement already satisfied: numpy>=1.15.4 in c:\users\a\anaconda3\lib\site-packages (from pandas) (1.21.6)
Requirement already satisfied: six>=1.5 in c:\users\a\anaconda3\lib\site-packages (from python-dateutil>=2.7.3->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip available: 22.3.1 -> 23.0
[notice] To update, run: python.exe -m pip install --upgrade pip

In [3]: import pandas as pd

In [7]: df = pd.read_csv(r'c:/minta/minta.csv')

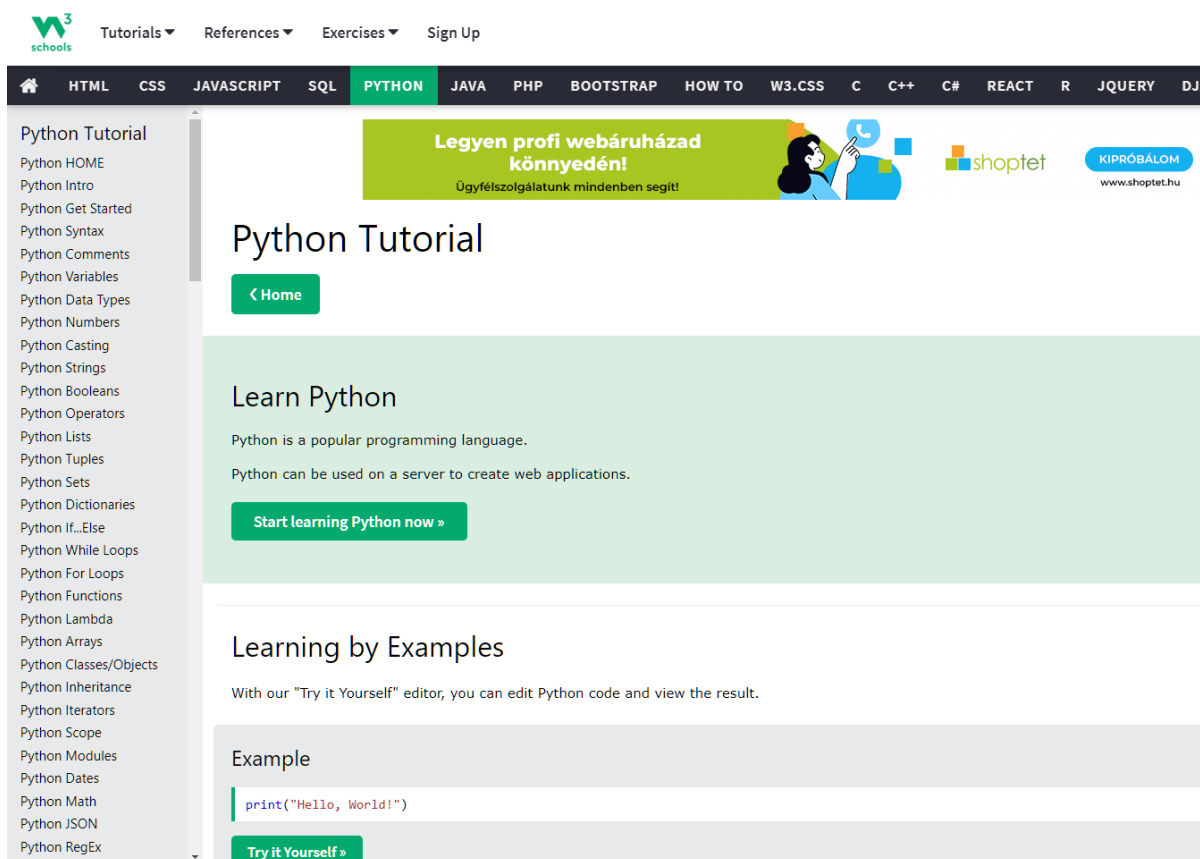
In [10]: df.head()
Out[10]:
   id  gender  age  edu  lab  bin_out  cont_out
0  1     1     53   1   95         1         49
1  2     2     88   3   80         0         54
2  3     1     7    1   96         1         35
3  4     2     5    3   16         0         55
4  5     1    19    1   50         1         55

In [9]: df.age.mean()
Out[9]: 54.56410256410256

```

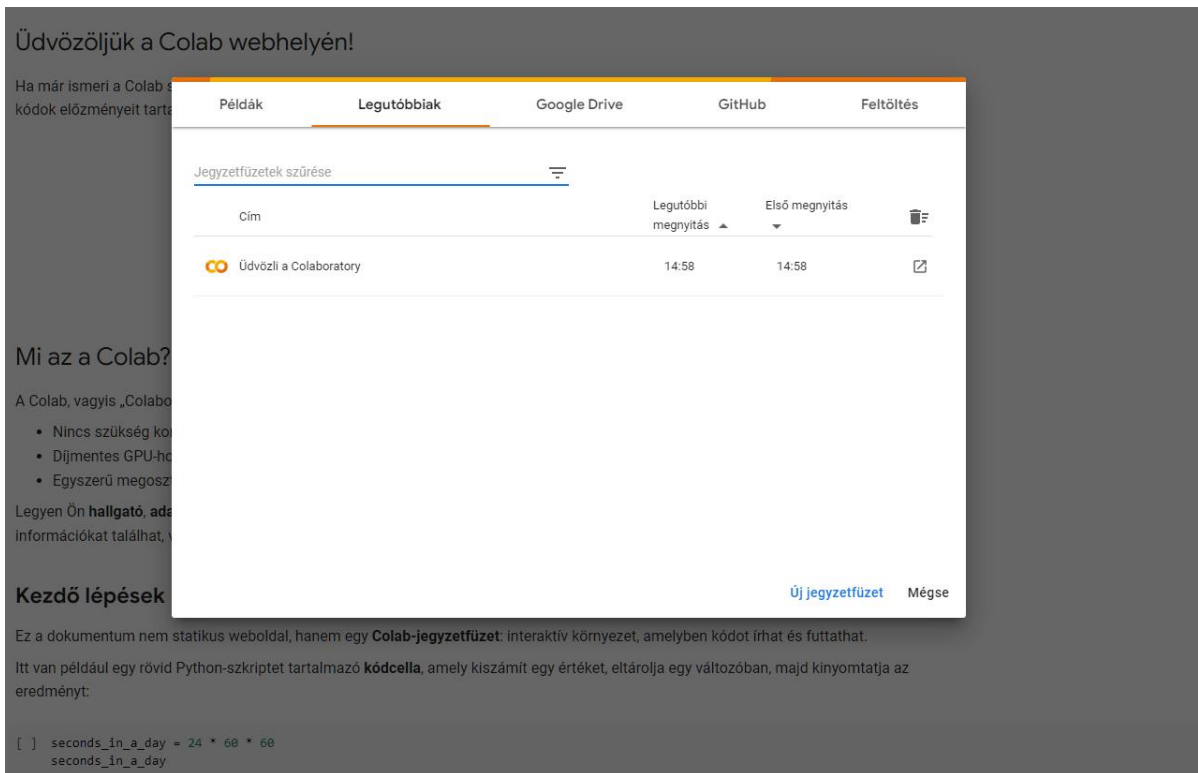
28. ábra: Python parancsok és eredmények

A W3 Schools oldala (<https://www.w3schools.com/python/default.asp>) az alapoktól a haladó szintig segít szemléletes példákkal a programnyelv elsajátításában.



29. ábra: W3Schools honlap Python oktatóanyaga

Aki nem szeretné letölteni az Anaconda programot, annak lehetősége nyílik online felület használatára is, amennyiben rendelkezik Google fiókkal.



Üdvözöljük a Colab webhelyén!

Ha már ismeri a Colab s
kódok előzményeit tart

Példák Legutóbbiak Google Drive GitHub Feltöltés

Jegyzetfüzetek szűrése

Cím	Legutóbbi megnyitás	Első megnyitás	
Üdvözlő a Colaboratory	14:58	14:58	

Mi az a Colab?

A Colab, vagyis „Colabo

- Nincs szükség ko
- Díjmentes GPU-ho
- Egyszerű megosz

Legyen Ön **hallgató**, ad

információkat találhat,

Kezdő lépések

Ez a dokumentum nem statikus weboldal, hanem egy **Colab-jegyzetfüzet**: interaktív környezet, amelyben kódot írhat és futtathat.

Itt van például egy rövid Python-szkriptet tartalmazó **kódcella**, amely kiszámít egy értéket, eltárolja egy változóban, majd kinyomtatja az eredményt:

```
[ ] seconds_in_a_day = 24 * 60 * 60  
seconds_in_a_day
```

Új jegyzetfüzet Mégse

30. ábra: Google Colaboratory (<https://colab.research.google.com/>)

Adatelemzés

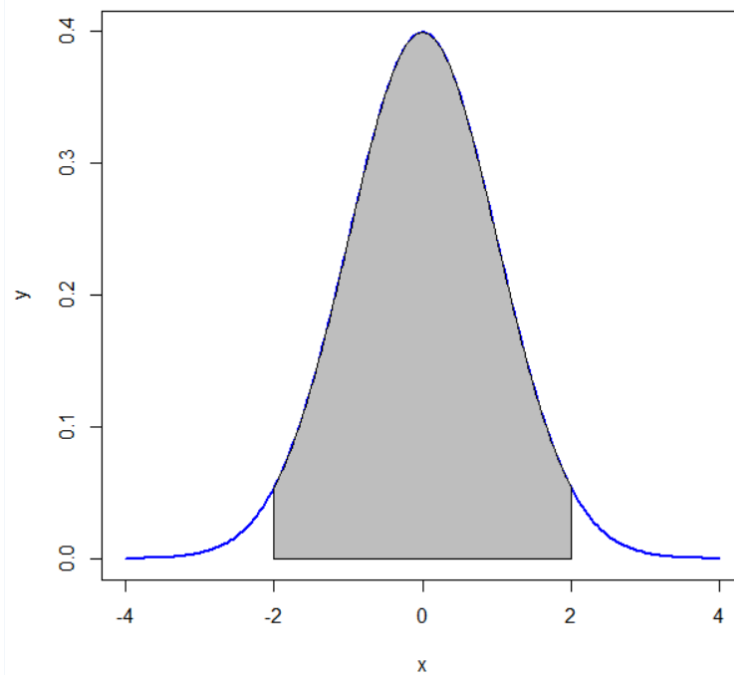
Az adatok tisztítása után az adatelemzéshez érkezünk. Az adatelemzés három fő részre osztható:

- leíró elemzés, mely segítségével képet kapunk a vizsgálati populációról, különös tekintettel a főbb jellemzőkre (beleértve a vizsgált kimenetelt)
- egyszerű (simple) elemzések révén kapunk képet a potenciális feltevésekről (hipotézis) (pl. t-próba, khi-négyzet próba)
- többszörös (multiple) elemzések révén pedig a tiszta hatást látjuk (zavaró tényezőkre korrigálva)

A részletek tárgyalása előtt érdemes röviden feleleveníteni néhány fogalmat. A p-érték az angol probability szóból ered. Legáltalánosabban 5%-ot hagyunk a véletlen szerepének ($p=0,05$) az elemzések során. Amennyiben 5%-nál nagyobb a véletlen szerepe, nem beszélünk tényleges összefüggésről, nem vetjük el az úgynevezett nullhipotézist (H_0 , mely a különbség/összefüggés hiányára utal). A kutatási hipotézis (H_1 , vagyis a nullhipotézis ellentétje), a feltevés az összefüggésre a potenciális befolyásoló tényező és a kimenetel között. Ezt a leíró etiológiai vizsgálatok révén generáljuk, illetve az elemzés kapcsán a leíró elemzések adják a képet, amit az egyszerű elemzés kristályosít, az összetett (többszörös) pedig véglegesít. A leggyakoribb p-értékhez (0,05) 95%-os megbízhatósági tartomány tartozik. Ez utóbbi a becslés bizonytalanságát szemlélteti, minél szűkebb, annál pontosabb a becslés, annál kisebb különbségek mutathatók ki. Azt jelenti, ha 100-szor megismételjük (különböző véletlen mintákon) a vizsgálatot, az eredmények 95-ször ezen tartományon belül lesznek. A másik jelentés, hogy a keresett populációs paraméter (amire kíváncsiak vagyunk, és ami miatt végezzük a vizsgálatot a reprezentatív mintán) 95%-os valószínűséggel van a tartományban. Az általános képlete a 95%-os megbízhatósági tartománynak:

$95\%MT = \text{pontbecslés} \pm 1,96 \text{ standard hiba (standard error)}$

A standard hiba számított mutatóként (pontbecslésenként) változik. Az 1,96 a standard normális eloszlásból származik (aminek várható értéke 0, szórása pedig 1). A legtöbb populációs paraméter (testsúly, vércukor stb.) normális eloszlást követ.



31. ábra: Standard normális eloszlás

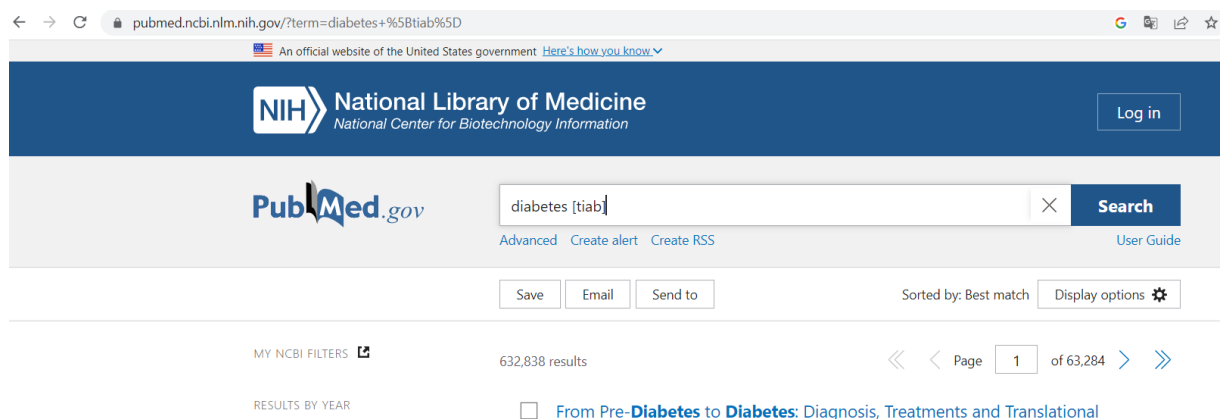
A teljes göbre alatti terület a 100%, azaz 1-es valószínűség. A pontbecsléshez tartozó 95%-os megbízhatósági tartomány kiszámítása során ezt a görbét adaptáljuk a mi helyzetünkhöz (saját szórás, saját mintaelemszám, saját várható érték (átlag)).

Szakirodalom kutatása

A teljes populáción nem tudunk méréseket végezni a korlátozott erőforrások miatt (idő és pénz). Ezért veszünk egy reprezentatív (minél több jellemzőben a teljes populációhoz (forrás populáció) hasonlós) mintát (legtöbbször véletlen mintavétellel) és körükben végezzük el a vizsgálatot. A reprezentativitás mellett fontos a megfelelő mintaelemszám is. A kívánt hatás eléréséhez szükséges minimális mintaelemszámot meg tudják becsülni a statisztikai programok.

Ehhez (és a további elemzésekhez) a könnyen elérhető és ingyenes R programot fogjuk használni.

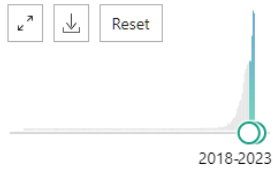
Az első lépés a releváns szakirodalom kutatása. A PubMed-en TIAB TAG a leggyakrabban használt, mely a találatokat leszűkíti a címben és absztraktban előfordulásokra, pl.:



32. ábra: PubMed keresés [tiab] TAG használatával

A cikkeket tovább szűkíthetjük megjelenés dátuma, nyelv stb. alapján. A releváns cikkekből kiszedhetjük a számunkra fontos eredményeket, amelyek segítségével elvégezhető lesz a mintaelemszám becslés. Fontosabb tématerületi áttekintés terén érdemes a „Review” vagy a „Systematic Review” szűrőt bekapcsolni, valamint az elmúlt 5 év releváns eredményeit nézni. A 33. ábrán látható feltételeknél még az ingyenes elérhető eredményekre szűrtünk rá („Free full text”).

RESULTS BY YEAR



TEXT AVAILABILITY

- Abstract
- Free full text
- Full text

ARTICLE ATTRIBUTE

- Associated data

ARTICLE TYPE

- Books and Documents
- Clinical Trial
- Meta-Analysis
- Randomized Controlled Trial
- Review
- Systematic Review

PUBLICATION DATE

- 1 year
- 5 years
- 10 years
- Custom Range

Additional filters

Filters applied: Free full text, Review, in the last 5 years. [Clear all](#)

- Genetics of **diabetes** mellitus and **diabetes** complications.**
- 1 Cole JB, Florez JC.
Cite Nat Rev Nephrol. 2020 Jul;16(7):377-390. doi: 10.1038/s41581-020-0278-5. Epub 2020 May 12. PMID: 32398868 [Free PMC article](#). [Review](#).
- Share **Diabetes** is one of the fastest growing diseases worldwide, projected to affect 693 million adults by 2045. ...The explosion of new genomic datasets, both in terms of biobanks and aggregation of worldwide cohorts, has more than doubled the number of genetic discoveries for ...
- Emerging Targets in Type 2 **Diabetes** and Diabetic Complications.**
- 2 Demir S, Nawroth PP, Herzig S, Ekim Üstünel B.
Cite Adv Sci (Weinh). 2021 Sep;8(18):e2100275. doi: 10.1002/advs.202100275. Epub 2021 Jul 28. PMID: 34319011 [Free PMC article](#). [Review](#).
- Share Type 2 **diabetes** is a metabolic, chronic disorder characterized by insulin resistance and elevated blood glucose levels. ...Overall, the molecular mechanisms of how type 2 **diabetes** develops and leads to irreparable organ damage remain elusive. ...
- From Pre-**Diabetes** to **Diabetes**: Diagnosis, Treatments and Translational Research.**
- 3 Khan RMM, Chua ZJY, Tan JC, Yang Y, Liao Z, Zhao Y.
Cite Medicina (Kaunas). 2019 Aug 29;55(9):546. doi: 10.3390/medicina55090546. PMID: 31470636 [Free PMC article](#). [Review](#).
- Share This unawareness and ignorance lead to further complications. Pre-**diabetes** is the preceding condition of **diabetes**, and in most of the cases, this ultimately leads to the development of **diabetes**. **Diabetes** can be classified into three types, namely type ...
- Type 2 **diabetes**: a multifaceted disease.**
- 4 Pearson ER.
Cite Diabetologia. 2019 Jul;62(7):1107-1112. doi: 10.1007/s00125-019-4909-y. Epub 2019 Jun 3. PMID: 31161345 [Free PMC article](#). [Review](#).
- Share Type 2 **diabetes** is a complex disease usually diagnosed with little regard to aetiology. ...Beyond this, however, type 2 **diabetes** is a highly heterogeneous polygenic disease. This review outlines the recent developments that recognise this heterogeneity by deconvolut ...
- Type 2 **Diabetes** Mellitus: A Review of Multi-Target Drugs.**
- 5 Artasensi A, Pedretti A, Vistoli G, Fumagalli L.
Cite Molecules. 2020 Apr 23;25(8):1987. doi: 10.3390/molecules25081987.

33. ábra: PubMed szűrési feltételek beállítása

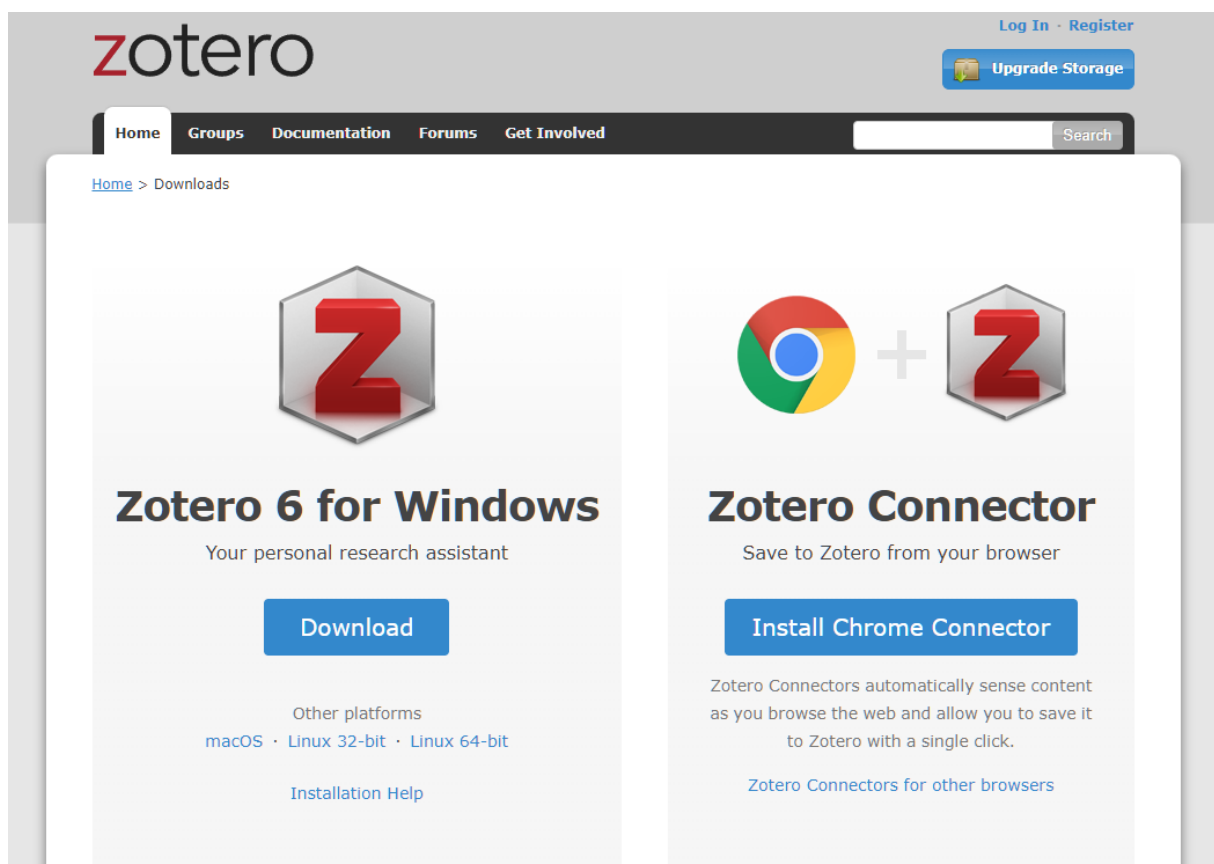
Referenciaszerkesztő

A megfelelő formátumú irodalmi hivatkozás elengedhetetlen egy projektmunkától a szakdolgozatokon át a cikkekig.

A legtöbb folyóiratnak saját előírása van az elvárt irodalmi hivatkozásokra, mégis a két legáltalánosabban elterjedt forma a Harvard és a Vancouver típus. A Vancouver típusnál számozzuk az citációkat és a végén számsorrendben tüntetjük fel őket. A Harvard stílusnál a szerző vagy szerzők nevei és a publikáció éve kerül feltüntetésre, majd a munka végén ABC sorrendben szerepelnek az idézett művek.

Eltérő a könyvek, a cikkek és a weboldalak hivatkozási formája tekintettel a szerzők számára, kiadóra, hozzáférés dátumára.

Az egyik leggyakrabban használt ingyenes referenciaszerkesztő a Zotero.



34. ábra: Zotero letöltése (<https://www.zotero.org/download/>)

Érdeemes a fő program mellett a Chrome Connector-t is telepíteni. A program integrálódik a böngészőbe és a Word-be is.

Használat lépései:

- Zotero elindítása a Start menüből
- Kitűzés a böngésző bővítményeknél

Hasznos cikk/oldal esetén le tudjuk menteni a hivatkozást a bővítményre kattintva, ami a 35. ábrán látható.



35. ábra: Hivatkozás mentése Zotero-ba

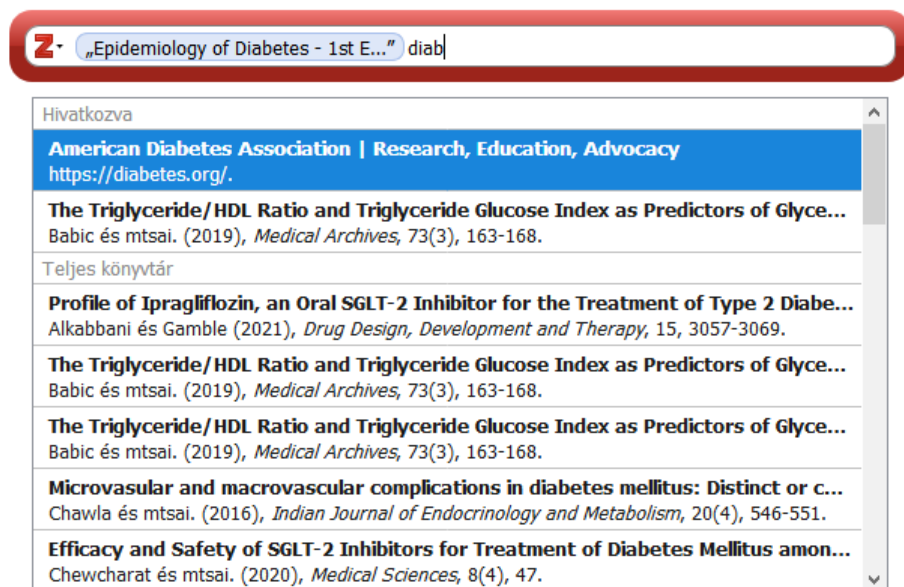
A megfelelő hivatkozások összegyűjtése után a Word-ben tudjuk beszűrni őket (Add Citation), valamint elkészíteni az Irodalomjegyzéket (Add Bibliography).



36. ábra: Zotero a Word-ben

Új hivatkozást tudunk létrehozni, valamint a meglévőt szerkeszteni az „Add/Edit Citation” gombra kattintva, ahogyan a 37. ábra mutatja.

Első hivatkozás [1], második hivatkozás helye [2], harmadik hivatkozás helye [3].



37. ábra: Hivatkozás szerkesztése Zotero-val

Az Irodalomjegyzék elkészítése az „Add/Edit Bibliography” gomb segítségével történik.

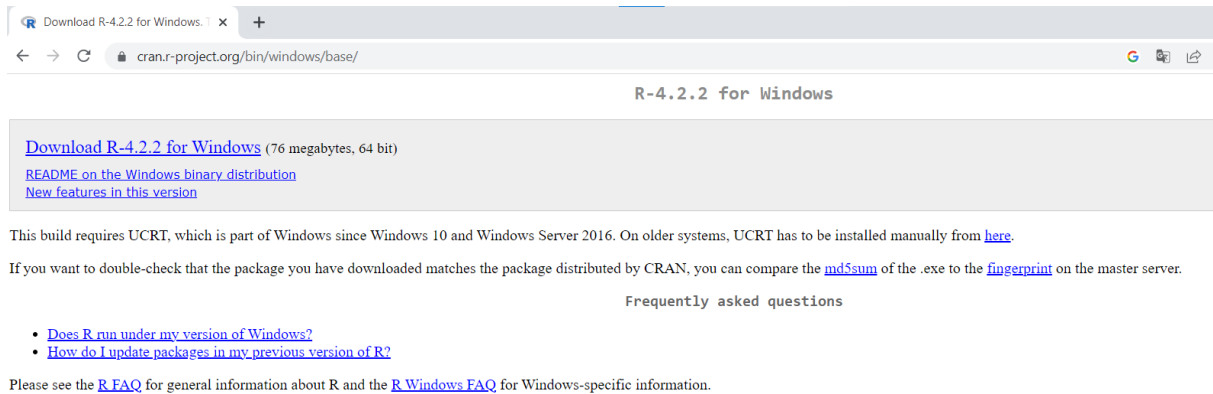
Első hivatkozás [1], második hivatkozás helye [2], harmadik hivatkozás helye [3].

- [1] „Epidemiology of Diabetes - 1st Edition”. <https://www.elsevier.com/books/epidemiology-of-diabetes/moini/978-0-12-816864-6> (elérés 2023. február 4.).
- [2] „American Diabetes Association | Research, Education, Advocacy”. <https://diabetes.org/> (elérés 2023. február 4.).
- [3] N. Babic, A. Valjevac, A. Zaciragic, N. Avdagic, S. Zukic, és S. Hasic, „The Triglyceride/HDL Ratio and Triglyceride Glucose Index as Predictors of Glycemic Control in Patients with Diabetes Mellitus Type 2”, *Med Arch*, köt. 73, sz. 3, o. 163–168, jún. 2019, doi: 10.5455/medarh.2019.73.163-168.

38. ábra: Zotero által készített Irodalomjegyzék

R program

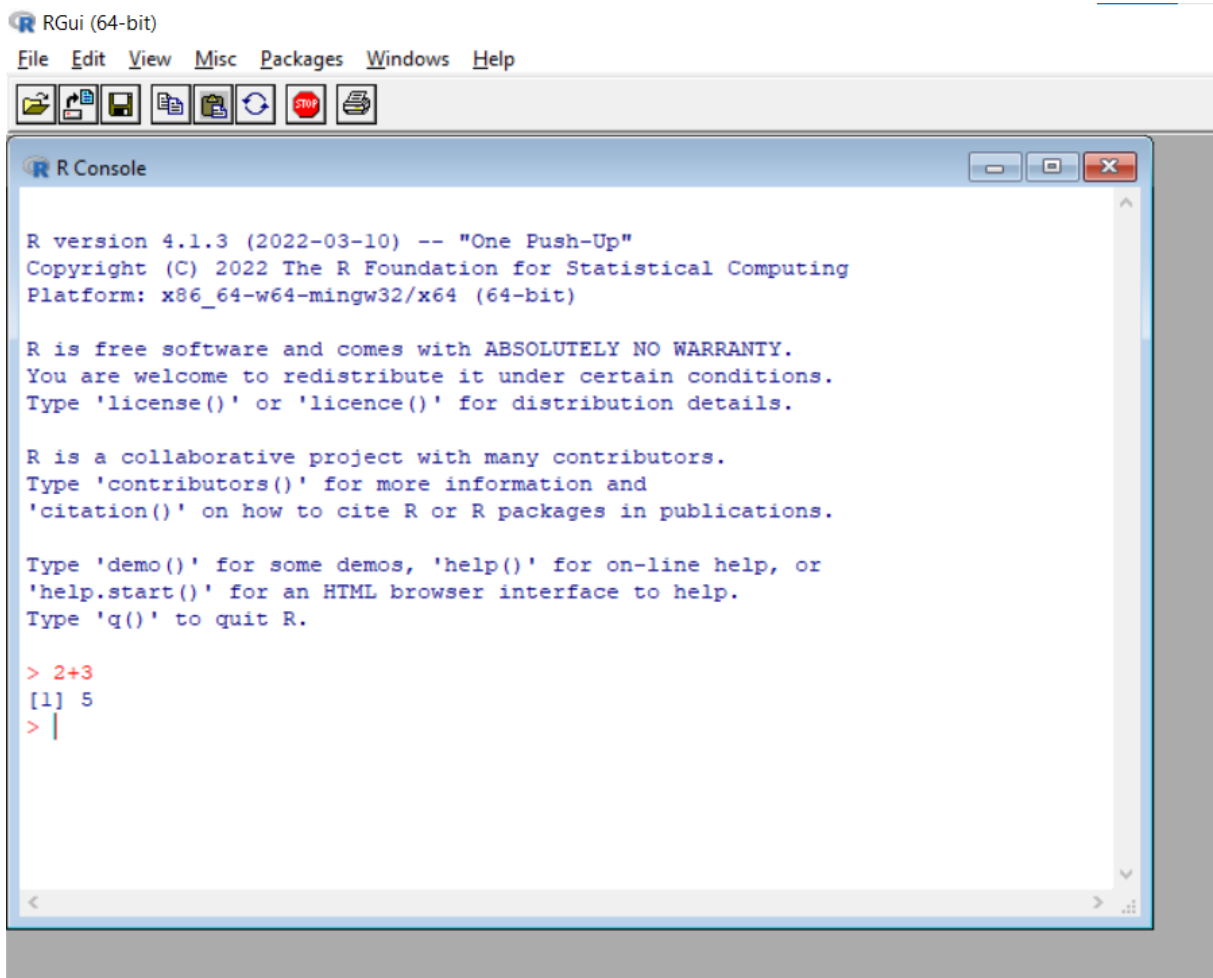
Az R program az R-project honlapjáról tölthető le:



39. ábra: R program letöltése

Érdemes a 64-bites verziót használni (x64) a jobb memóriakezelés miatt.

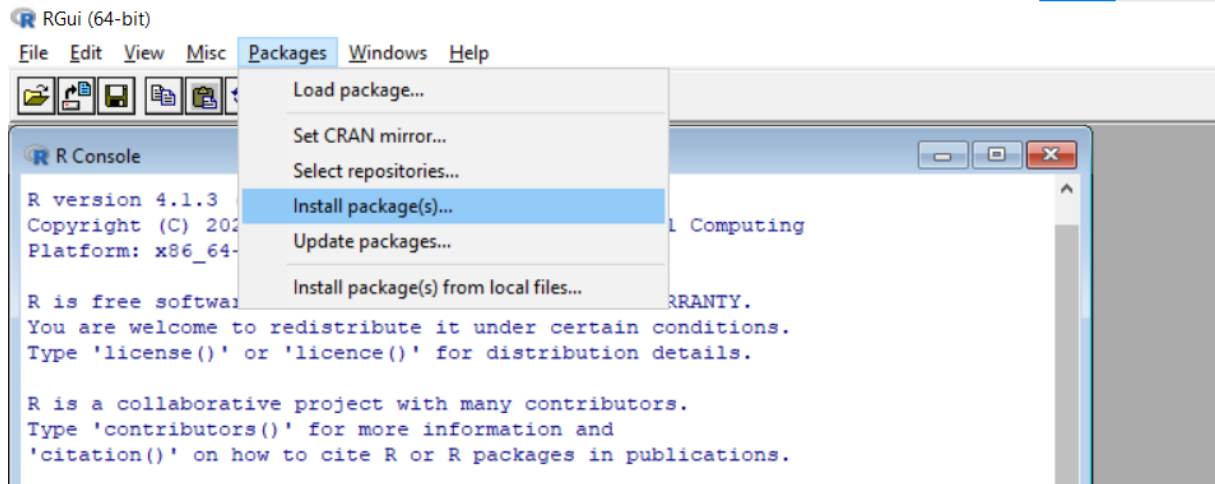
Telepítés után (érdemes az alapértelmezett módon) indítható a program. Parancsok beírása után az eredmény egyből láthatóvá válik.



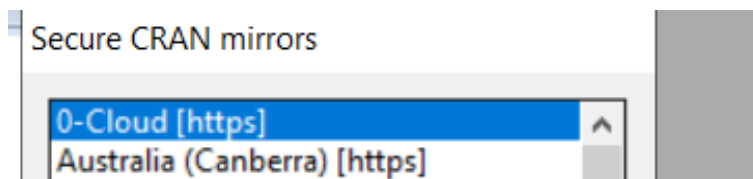
40. ábra: R program felülete

Az egyes csomagokat (packages) telepíteni (install.package) kell, ezután pedig be kell tölteni (library), hogy elérhetőek legyenek. A csomag installálás történhet parancssoros módon, illetve menüből kattintva. A parancs alatt szerepeltetjük a menüs elérést:

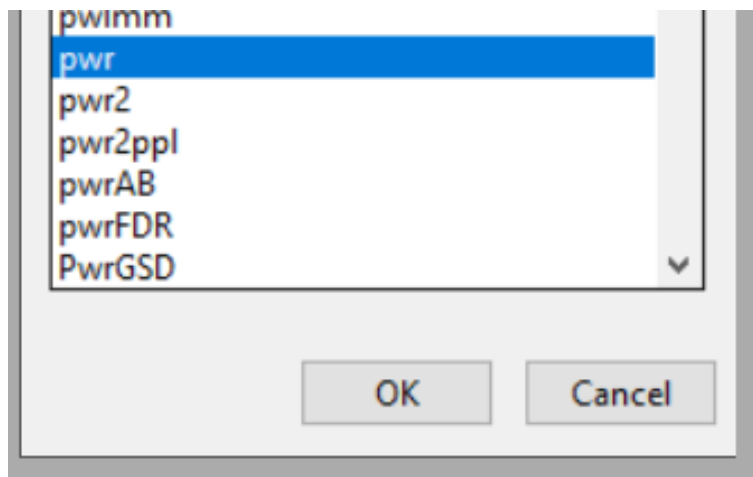
```
install.packages("pwr")
```



41. ábra: R programcsomag telepítése

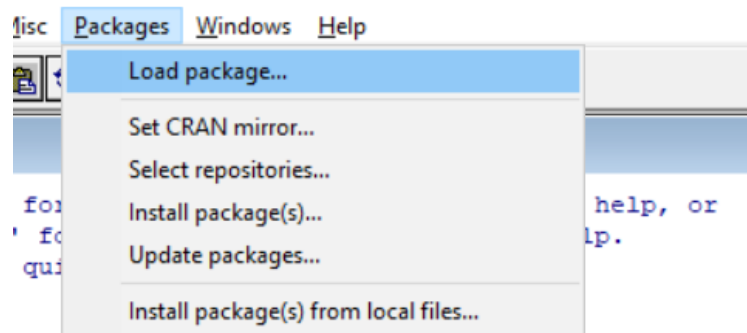


42. ábra: R tükörszerver kiválasztása a telepítéshez

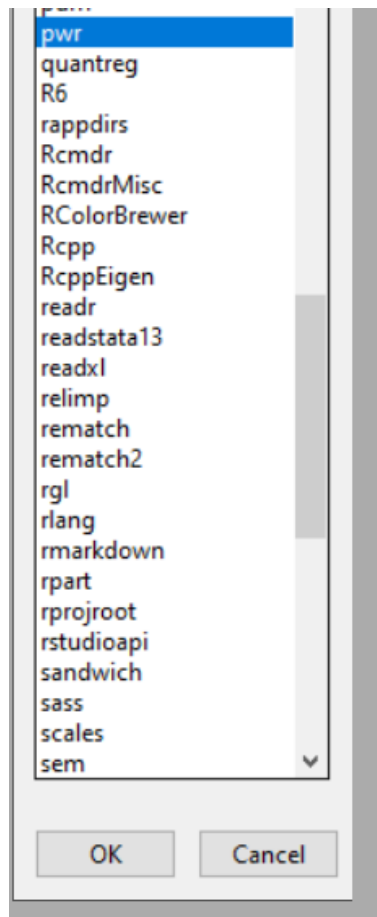


43. ábra: „pwr” csomag telepítése

A telepített csomag betöltése:
library(pwr)



44. ábra: R programcsomag betöltése



45. ábra: „pwr” csomag betöltése

Az `install.packages` parancsnál szükséges az idézőjel, a `library` parancsnál viszont tilos.

Mintaelemszám becslés

A csomag telepítése és betöltése után meg tudjuk nézni pl. egy 0,5-ös különbség detektálásához szükséges mintaelemszámot 0,05-ös p-érték és 80%-os statisztikai erő (a másodfajú hiba komplementere, minél nagyobb annál jobb, legalább 80% legyen) esetén kétmintás t-próbánál.

```
pwr.t.test(d=0.5, sig.level=0.05, power=0.80, type="two.sample", alternative="greater")
```

Two-sample t test power calculation

```
      n = 50.1508
      d = 0.5
sig.level = 0.05
  power = 0.8
alternative = greater
```

NOTE: n is number in *each* group

46. ábra: Mintaelemszám becslés

A becslés alapján csoportonként legalább 50 fő szükséges. Saját függvények is létrehozhatóak. Az alábbi kód begépelése után egy új parancs válik elérhetővé, ami pontosabb mintaelemszám becslést tesz lehetővé az átlagok és szórások felhasználása révén (amiket pl. PubMed-ről tudunk).

```
sampsi.means<-function(m1, m2, sd1, sd2=NA, ratio=1, power=.90, alpha=.05,
two.sided=TRUE, one.sample=FALSE){
effect.size<-abs(m2-m1)
sd2<-ifelse(!is.na(sd2), sd2, sd1)
z.pow<-qt(1-power, df=Inf, lower.tail=FALSE)
z.alph<-ifelse(two.sided==TRUE, qt(alpha/2, df=Inf, lower.tail=FALSE), qt(alpha, df=Inf,
lower.tail=FALSE))
ct<-(z.pow+z.alph)
n1<-(sd1^2+(sd2^2)/ratio)*(ct)^2/(effect.size^2)
n<-(ct*sd1/effect.size)^2
if(one.sample==FALSE){
col1<-c("alpha", "power", "m1", "m2", "sd1", "sd2", "effect size", "n2/n1", "n1", "n2")
col2<-c(alpha, power, m1, m2, sd1, sd2, effect.size, ratio, ceiling(n1), ceiling(n1*ratio))
}
else{
col1<-c("alpha", "power", "null", "alternative", "n")
col2<-c(alpha, power, m1, m2, ceiling(n))
}
ret<-as.data.frame(cbind(col1, col2))
ret$col2<-as.numeric(as.character(ret$col2))
colnames(ret)<-c("Assumptions", "Value")
description<-paste(ifelse(one.sample==FALSE, "Two-sample", "One-sample"),
ifelse(two.sided==TRUE, "two-sided", "one-sided"), "test of means")
retlist<-list(description, ret)
return(retlist)
}
```

Ezután az alábbi parancs begépelhető:

```
sampsi.means(10, 30, sd1=15, sd2=20, alpha=.05, ratio=1)
```

```
> sampsi.means(10, 30, sd1=15, sd2=20, alpha=.05, ratio=1)
[[1]]
[1] "Two-sample two-sided test of means"

[[2]]
  Assumptions Value
1      alpha 0.05
2     power 0.90
3        m1 10.00
4        m2 30.00
5        sd1 15.00
6        sd2 20.00
7 effect size 20.00
8      n2/n1 1.00
9         n1 17.00
10        n2 17.00
```

47. ábra: A mintaelemszám becslés eredménye

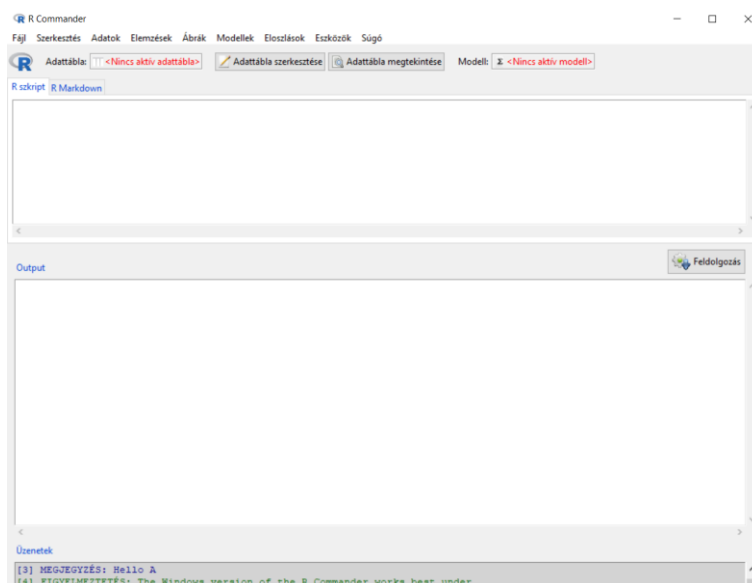
Vagyis 10 ± 15 (átlag \pm szórás) és 30 ± 20 csoportok esetén a különbség szignifikáns lesz ($p=0,05$), ha csoportonként minimum 17 fővel dolgozunk.

Leíró statisztika

A kategorikus változók (pl. nemek vagy iskolai végzettség) jellemzése százalékos megoszlással történik. A folytonos változók jellemzése (normalitás függvényében) átlag \pm szórás, vagy medián és interkvartilis tartomány (nem normális eloszlás esetén) segítségével történik.

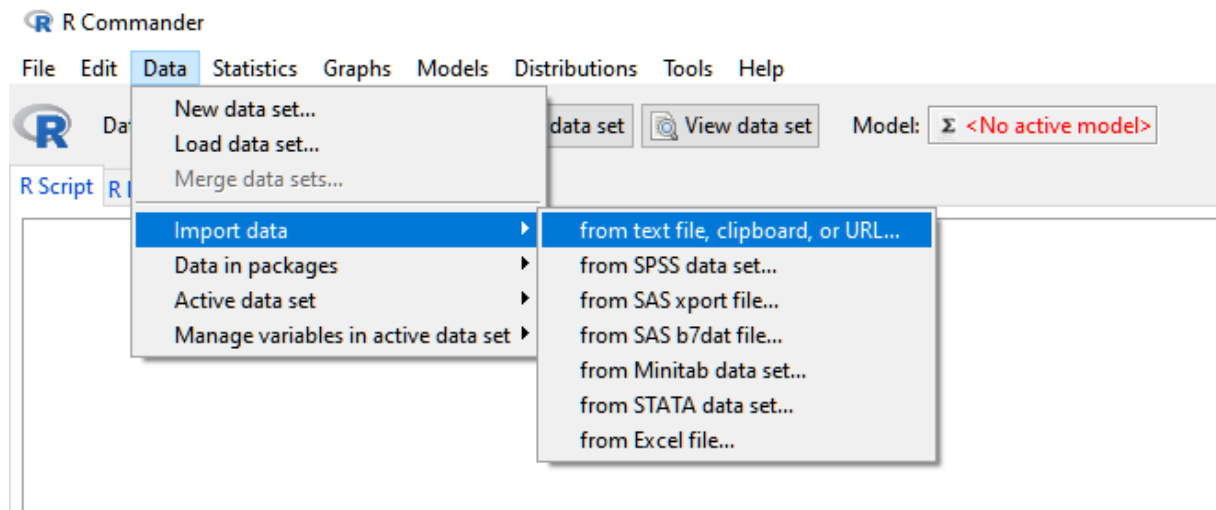
Leíró statisztika készítéséhez első lépésben telepítsük fel és töltsük be a grafikus felülettel rendelkező Rcommander (Rcmdr) csomagot.

```
install.packages("Rcmdr")
library(Rcmdr)
```



48. ábra: R commander

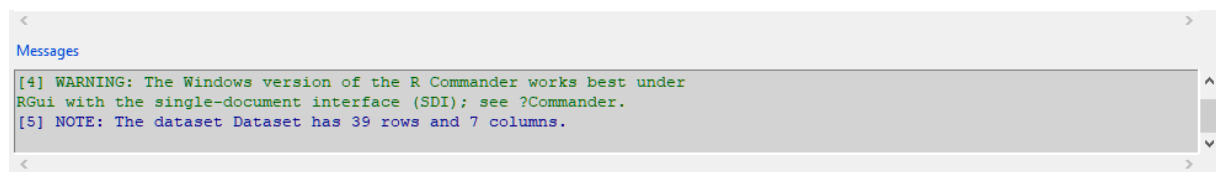
A megnyitás után lehetőségünk van az adatbázis beolvasására. A jegyzetünk végén megtalálható (és kimásolható) a felhasznált adatbázis, amit egy *minta.csv*-ből fogunk beolvasni. Magyar nyelv is letölthető, de a szakszavak használata miatt érdemes az angol verziót használni.



49. ábra: Fájl beolvasás R commander segítségével

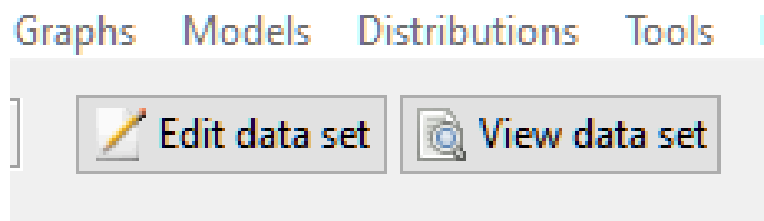
Hagyhatunk alapértelmezetten mindent, főleg az elválasztóra (field separator) kell figyelniük, ami jelen esetünkben a pontosvessző (semicolon).

Sikeres betöltés esetén kiírja az adatbázis (dataset) sor (row) és oszlopszámát (column).



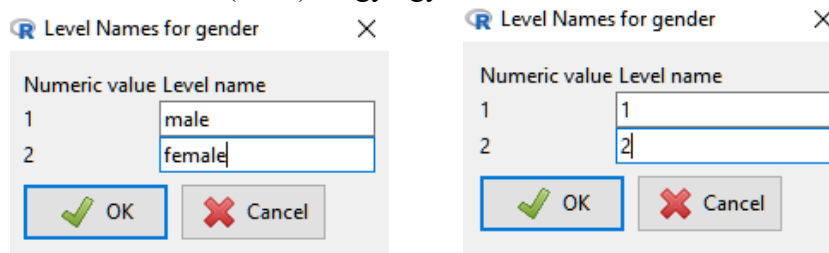
50. ábra: Sikeres adatbeolvasás

Lehetőségünk van az adatbázis szerkesztésére vagy pusztán megtekintésére.



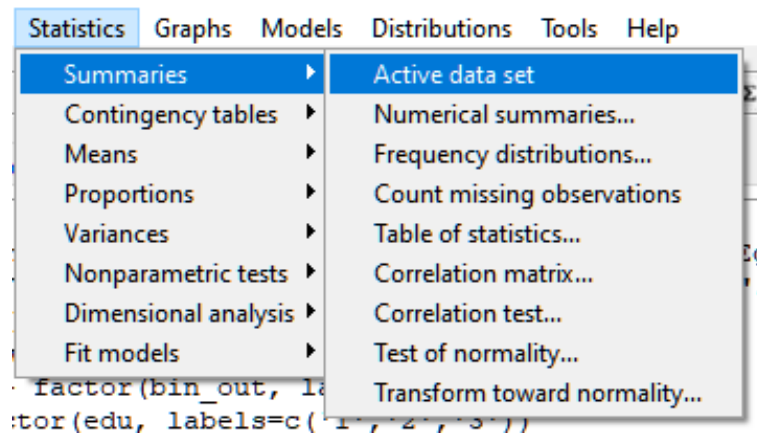
51. ábra: Adatbázis megtekintése/szerkesztése

A meglévő változók felülírhatók (overwrite). Az „OK” gomb megnyomása után be kell írunk az értékekhez tartozó címkeket (label), vagy egyszerűen az értéket, átcímkezés nélkül.



54. ábra: Változók átcímkezése

Kérhetünk egy általános leíró jellemzést az összes változóról.



55. ábra: Leíró statisztika lekérése

Ilyenkor folytonos változóként kezel mindent a rendszer és megadja a változók következő paramétereit:

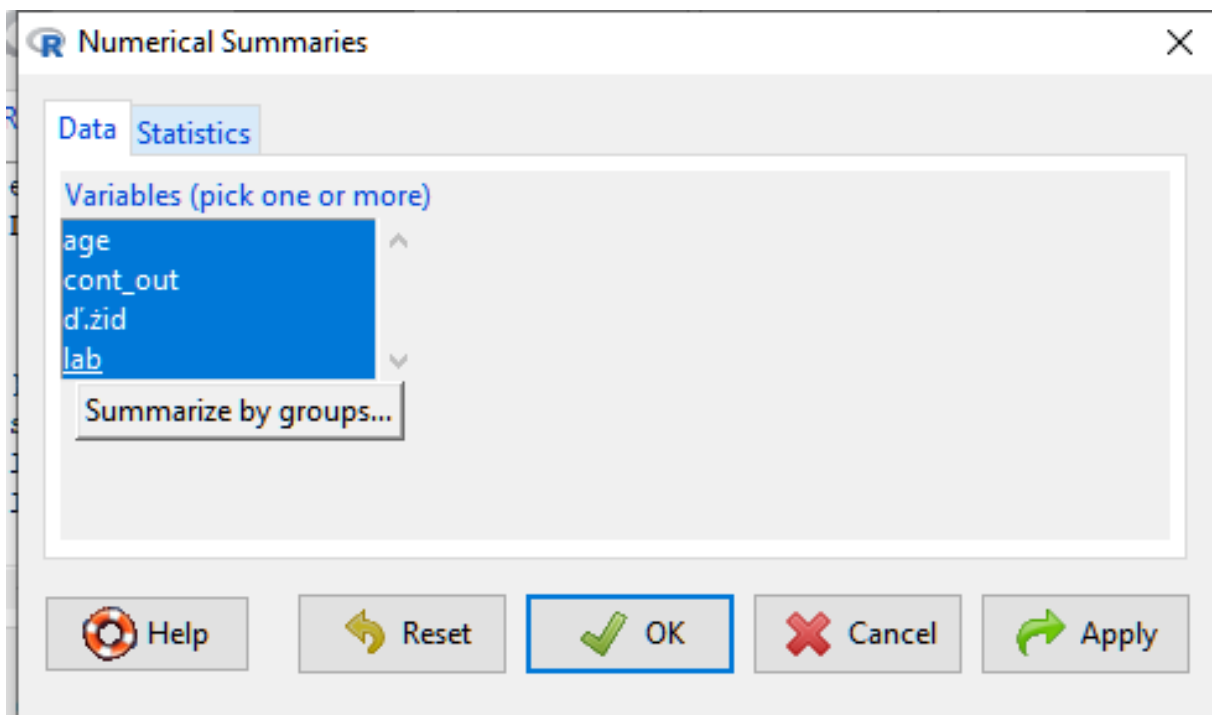
- minimum
- első kvartilis (Q1), ami a 25%-os percentilis érték
- medián (Q2), vagy 50%-os percentilis érték
- harmadik kvartilis (Q3), vagy 75%-os percentilis érték
- maximum

```
> summary(Dataset)
   d.żid   gender   age   edu   lab   bin_out   cont_out
Min.   : 1.0   1:22  Min.   : 3.00  1:12  Min.   :12.00  0:19  Min.   :11.00
1st Qu.:10.5  2:17  1st Qu.:29.50  2:16  1st Qu.:25.50  1:20  1st Qu.:35.00
Median :20.0  Median :53.00  3:11  Median :60.00  Median :49.00
Mean   :20.0  Mean   :54.56  Mean   :57.08  Mean   :46.23
3rd Qu.:29.5  3rd Qu.:82.50  3rd Qu.:85.00  3rd Qu.:58.00
Max.   :39.0  Max.   :97.00  Max.   :99.00  Max.   :78.00
```

56. ábra: Leíró statisztika

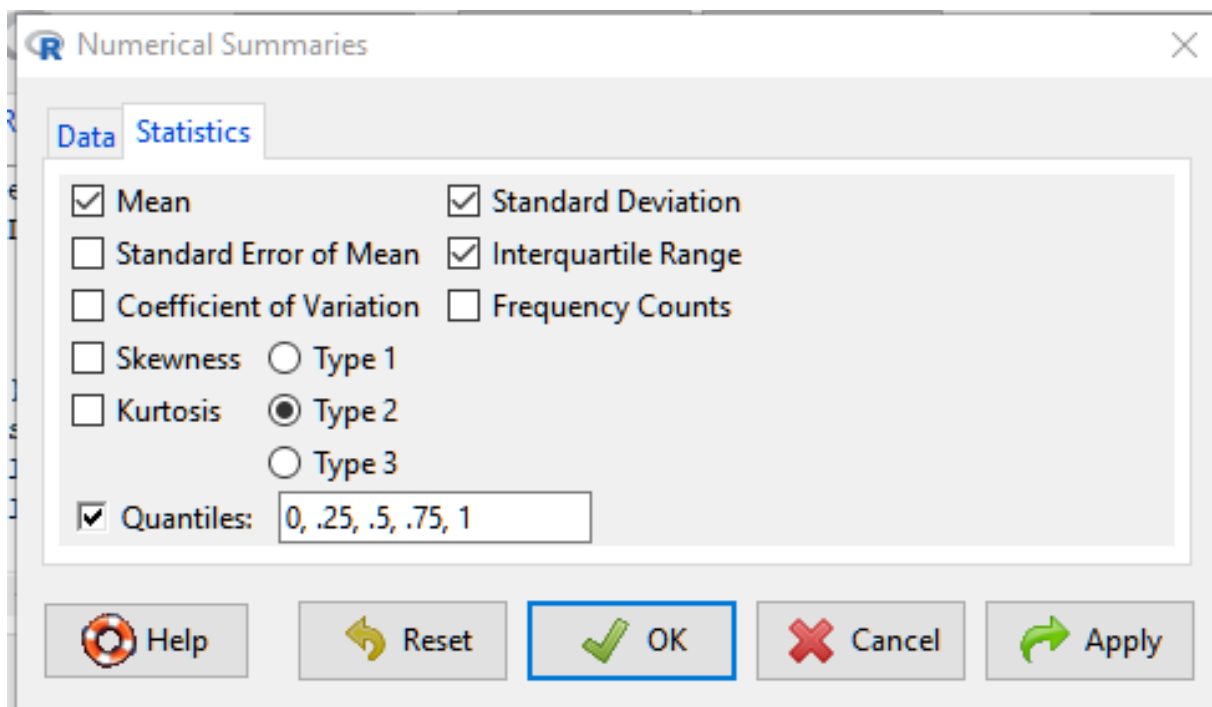
A program kiírja a parancsokat is (*summary*), ami segítségével a grafikus felület (GUI), vagyis az Rcommander nélkül is végrehajtható az utasítás.

A folytonos változók külön a „Numerical Summaries” révén jellemezhetők:



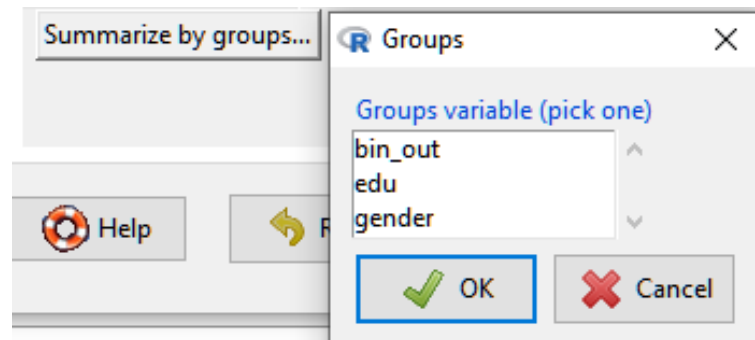
57. ábra: Folytonos változók kijelölése jellemzés céljából

Szintén alkalmazható többször kijelölés, valamint kiválasztható a számunkra érdekes statisztika.



58. ábra: Statisztikák beállítása

Lehetőségünk van rétegenkénti megjelenítésre is, (ahogyan az 57. ábra is látszik) egy adott csoportosító változó mentén (Summarize by groups).



59. ábra: Rétegzett elemzés csoportonként

A mutatott jellemzők a következők:

- átlag (mean)
- szórás (sd)
- interkvartilis tartomány (IQR)
- percentilisek
 - o 0% (minimum)
 - o 25% (Q1)
 - o 50% (Q2), vagy medián
 - o 75% (Q3)
 - o 100% (maximum)
 - o mintaelemszám (N)

A kategorikus változók (faktorral alakítás után) százalékos megoszlással jellemezhetők:

```
counts:
gender
  1  2
22 17

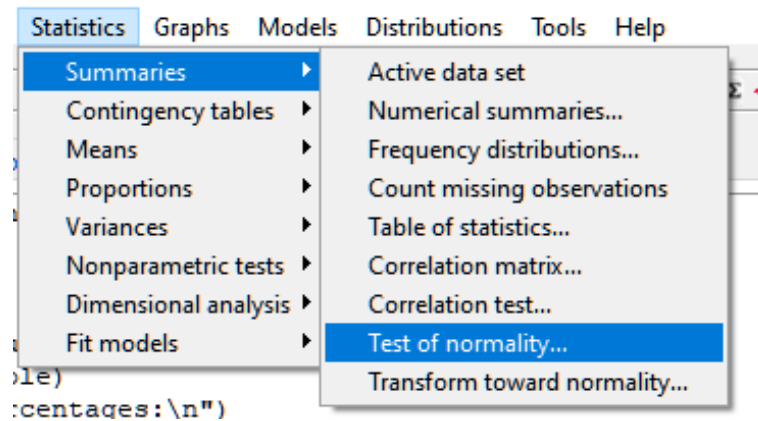
percentages:
gender
   1    2
56.41 43.59
```

60. ábra: Kategorikus változók százalékos megoszlása

Az eredmény (output) tábla tartalmazza az abszolút számokat (count) és a százalékos megoszlásokat (percentages).

A folytonos változók esetén a normális eloszlástól való eltérés (normalitás) vizsgálata szintén fontos. Az egyik fő ok a leírása a változónak. Ha az eloszlás normális (vagy attól nem tér el szignifikánsan, $p > 0,05$) akkor használjuk az átlag \pm szórást. Általában két tizedesjegyre kerekítjük a számokat. A p-értéket három tizedesjegyre kerekítjük. Nem normális eloszlás esetén medián és IQR jellemzést használunk.

A normalitás vizsgálatához az egyik leggyakrabban használt módszer a Shapiro-Wilk próba.



61. ábra: Normalitás tesztelése

```
> normalityTest(~cont_out, test="shapiro.test", data=Dataset)

Shapiro-Wilk normality test

data:  cont_out
W = 0.95184, p-value = 0.09477
```

62. ábra: Normalitás tesztelésének eredménye

Eredményünk alapján elmondható, hogy a folytonos kimenetel (cont_out) normális eloszlást követ, attól nem tér el szignifikánsan ($p=0,095$).

Egyszerű elemzés

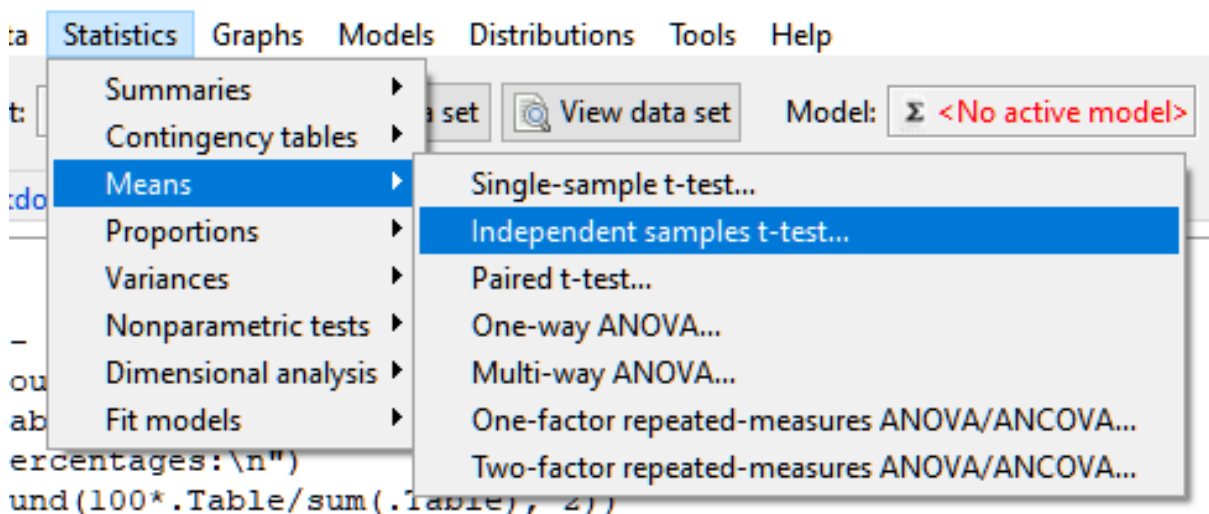
A leíró elemzést az egyszerű (simple) elemzés követi. Néha ezeket egyváltozós elemzésnek is hívják, mivel egy kimenetel (outcome, vagy függő változó) és egy potenciális befolyásoló tényező (magyarázó, vagy független változó) kapcsolatát vizsgáljuk. A nyers elemzések segítségével hipotézis generálható, melyet a korrigált (többszörös) elemzésekkel fogunk majd értékelni.

Paraméteres próbák

A paraméteres próbáknak nagyobb a statisztikai erejük, ám bizonyos feltételek megléte szükséges az alkalmazásukhoz. A leggyakoribb feltétel a folytonos változó normális (vagy normális közeli) megoszlása. A másik leggyakoribb feltétel a kellően nagy mintaelemszám.

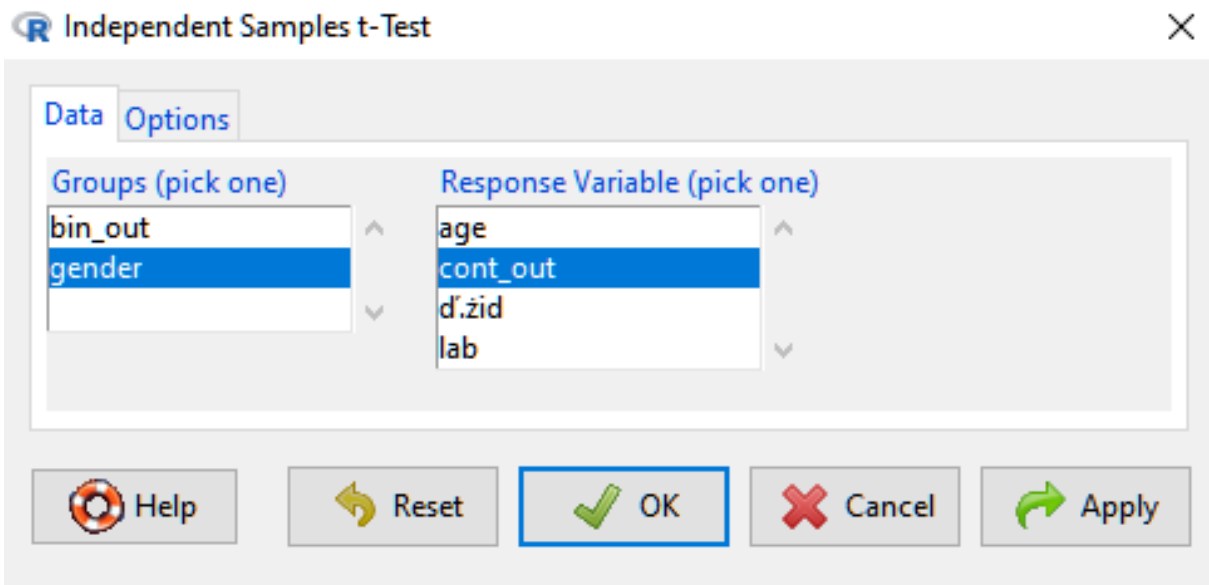
T-próba

A Student-féle t-próba csoportátlagok összehasonlítására alkalmas. Lehet a csoportátlagot egy konstans számhoz (állandóhoz) hasonlítani (egymintás t- próba; single-sample t-test), valamint egy csoport két átlagértéke is összehasonlítható (előtte-utána; páros t-próba, paired t-test), míg két független csoport (nem feltétlenül megegyező mintaelemszámokkal) átlaga is összevethető (kétmintás t-próba; independent samples t-test).



63. ábra: Kétmintás t-próba

Összevethetjük például, hogy van-e eltérés a folytonos kimenetelben (*cont_out*) nemenként (*gender*).



64. ábra: Kétmintás t-próba beállítása

Az eredmény alapján elmondhatjuk, hogy nincs szignifikáns különbség a két csoportátlag (48,32 vs. 43,53) között ($p=0,474$) (marad érvényben a nullhipotézis, mely szerint a csoportátlagok között nincs eltérés).

```
> t.test(cont_out~gender, alternative='two.sided', conf.level=.95, var.equal=FALSE, data=Dataset)

Welch Two Sample t-test

data: cont_out by gender
t = 0.72469, df = 31.449, p-value = 0.474
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-8.680565 18.258105
sample estimates:
mean in group 1 mean in group 2
 48.31818      43.52941
```

65. ábra: Kétmintás t-próba eredménye

ANOVA

Több, mint két csoportátlag összehasonlítása esetén használatos az ANOVA (Analysis of variance). A név kissé megtévesztő, de a varianciákkal (szórásnégyzetekkel) csak a háttérben számol a próba, valójában átlagok összehasonlítására alkalmas.

The screenshot shows the RStudio interface. The 'Statistics' menu is open, and 'One-way ANOVA...' is selected. The console below shows the following R code and output:

```

> AnovaModel.2 <- aov(cont_out ~ edu, data=Dataset)

> summary(AnovaModel.2)
          Df Sum Sq Mean Sq F value Pr(>F)
edu         2    523   261.3    0.646  0.53
Residuals  36 14556   404.3

```

```

> with(Dataset, numSummary(cont_out, groups=edu, statistics=c("mean", "sd")))
      mean      sd data:n
1 42.58333 13.80684     12
2 45.12500 21.46586     16
3 51.81818 23.55342     11

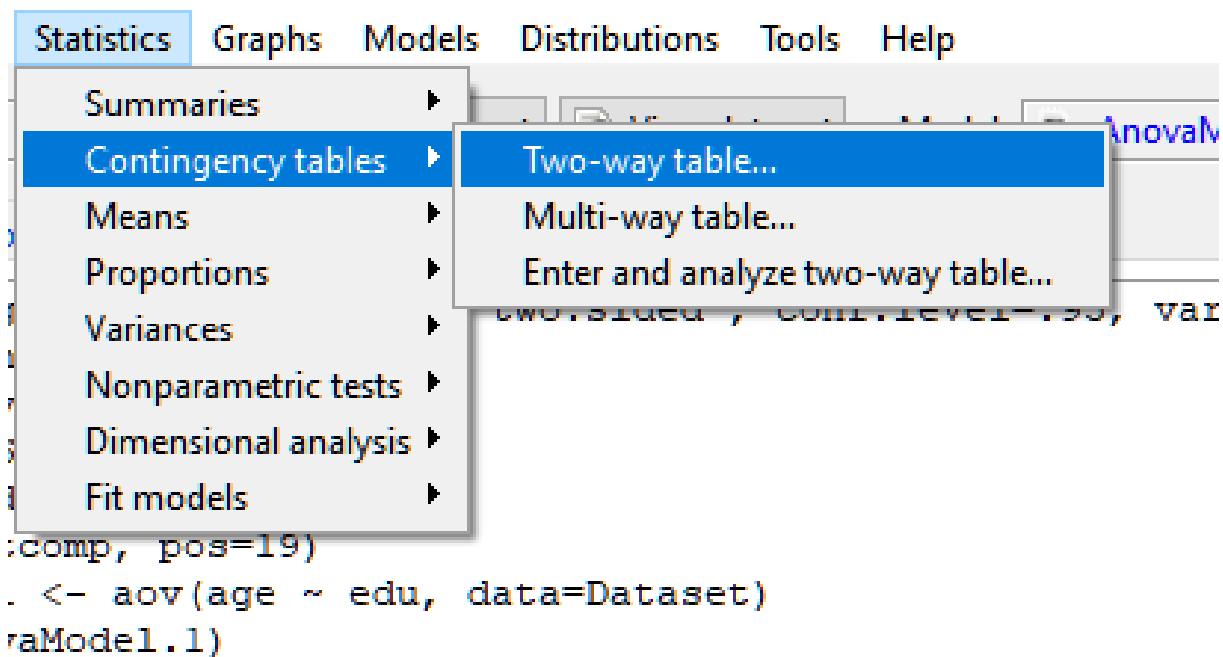
```

66. ábra: ANOVA beállítása

Az eredmény alapján elmondható, hogy iskolai végzettség tekintetében nincs különbség a folytonos kimenetel átlagértékei között ($p=0,530$).

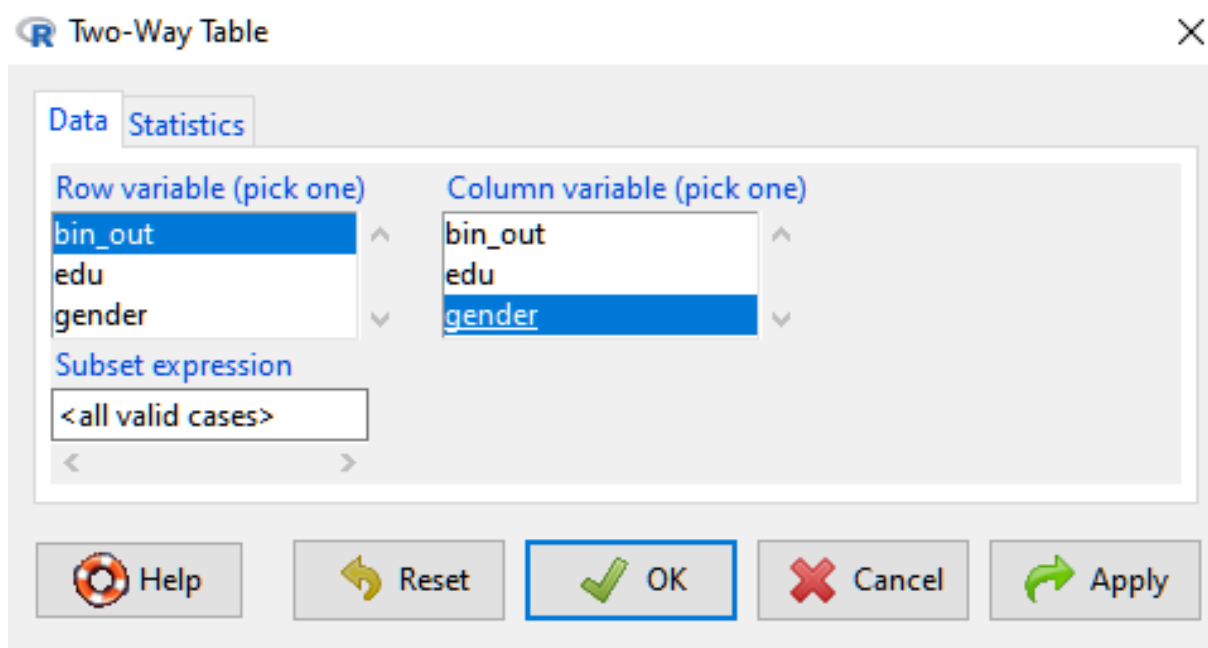
Khi-négyzet próba

Kategorikus változók közötti összefüggéseket a Pearson-féle khi-négyzet próba tud detektálni. Szintén a mintaelemszám egy fontos kritérium az alkalmazhatóságnál. Ha nincs legalább 5 megfigyelés a cellák legalább 80%-ában, akkor nem alkalmazható (helyette a Fisher-féle egzakt próba alkalmazandó).



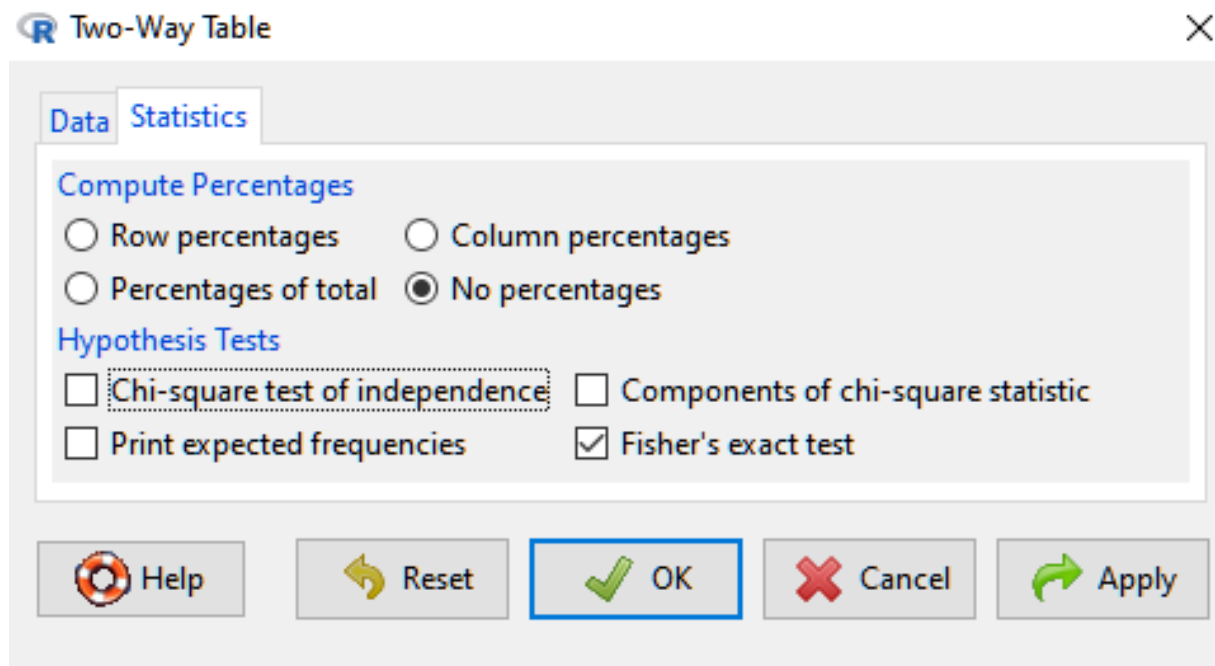
67. ábra: Khi-négyzet próba kiválasztása

Vizsgáljuk meg a bináris (két kategória, igen-nem) kimenetelünk (*bin_out*) összefügg-e a nemmel (*gender*).



68. ábra: Khi-négyzet próba beállítása

A Statistics fülön kérhetjük a Fisher egzakt tesztet is (most maradjunk az alapértelmezett khi-négyzetnél), illetve pl. sorszázalékot is kérhetünk (row percentages).



69. ábra: Fisher-féle egzakt próba kiválasztása

Eredményünk alapján elmondható, hogy szignifikáns az összefüggés ($p=0,016$) a nem és a bináris kimenetünk között. A sorszázalékokon látjuk, hogy a kimenettel rendelkezők körében több a férfi, míg a kimenetlől mentesek körében a nők vannak többen.

```

Frequency table:
  gender
bin_out  1  2
  0  7 12
  1 15  5

Row percentages:
  gender
bin_out   1   2 Total Count
  0 36.8 63.2   100    19
  1 75.0 25.0   100    20

      Pearson's Chi-squared test

data:  .Table
X-squared = 5.7696, df = 1, p-value = 0.01631

```

70. ábra: Khi-négyzet próba eredménye

Nemparaméteres próbák

A feltételek (assumptions) teljesülésének hiányában a statisztikailag gyengébb nemparaméteres (non parametric) tesztekkel kell használnunk.

Wilcoxon próba

A t-próbák alternatívája. Átlag helyett mediánokat hasonlít össze (pontosabban eloszlásokat, azonos mediánok esetén is lehet akár szignifikáns eltérés).

- Egymintás t-próba alternatíva
 - o Single sample Wilcoxon test
- Páros t-próba alternatíva
 - o Wilcoxon féle előjeles rangpróba (Wilcoxon signed-rank test/ Paired samples Wilcoxon test)
- Kétmintás t-próba alternatíva
 - o Wilcoxon-féle előjeles rangösszeg próba/ Mann–Whitney U próba/ Wilcoxon-féle rangösszegteszt (Two-sample Wilcoxon test/ Wilcoxon rank sum test)

A mintánkban az életkor változó (*age*) nem követ normálist eloszlást ($p=0,025$), amely nemparaméteres teszt alkalmazását teszi indokolttá.

```
> normalityTest(~age, test="shapiro.test", data=Dataset)

      Shapiro-Wilk normality test

data:  age
W = 0.93457, p-value = 0.02525
```

71. ábra: Normalitás tesztelése

Nézzük meg, hogy a mediánok (eloszlások) között van-e különbség a nemek tekintetében.

```
> wilcox.test(age ~ gender, alternative="two.sided", data=Dataset)

      Wilcoxon rank sum test with continuity correction

data:  age by gender
W = 137, p-value = 0.1608
alternative hypothesis: true location shift is not equal to 0
```

72. ábra: Mann-Whitney-Wilcoxon teszt

A teszt eredménye alapján nincs különbség ($p=0,161$) a két medián (eloszlás) tekintetében.

Fisher-féle egzakt próba

A kategorikus változók összevetésére alkalmas, kisebb mintaelemszám esetén. Az előbbi példánál maradva (sor és oszlop felcserélhető, a p-érték marad), a mintaelemszám kellően nagy volt a khi-négyzet próbához, minden cellában volt legalább 5 megfigyelés. Ha legalább az egyik

cellában pl. 4 fő lett volna, akkor az összes cella 25%-ban nem teljesült volna a feltétel, tehát nem lett volna meg az 5 fő minimum a cellák 80%-ban (csak 75%-ban).

```

Frequency table:
      bin_out
gender 0  1
      1  7 15
      2 12  5

      Pearson's Chi-squared test

data: .Table
X-squared = 5.7696, df = 1, p-value = 0.01631

      Fisher's Exact Test for Count Data

data: .Table
p-value = 0.02484
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.03881034 0.91928487
sample estimates:
odds ratio
 0.2037389

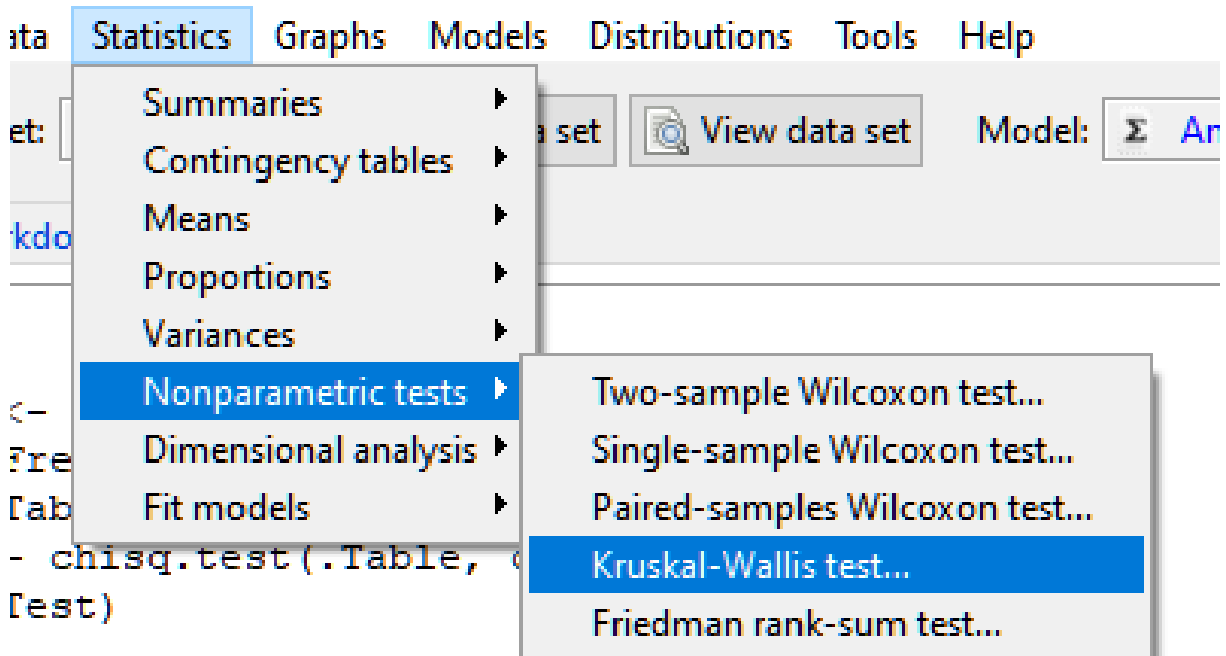
```

73. ábra: Fisher-féle egzakt próba eredménye

Kiszámítva a Fisher próba p-értékét, hasonló értéket kapunk (0,016 vs 0,025). Azonban a borderline szignifikáns (0,05 közeli) p-érték átbillenthet, ezért figyeljük a feltétel teljesülését!

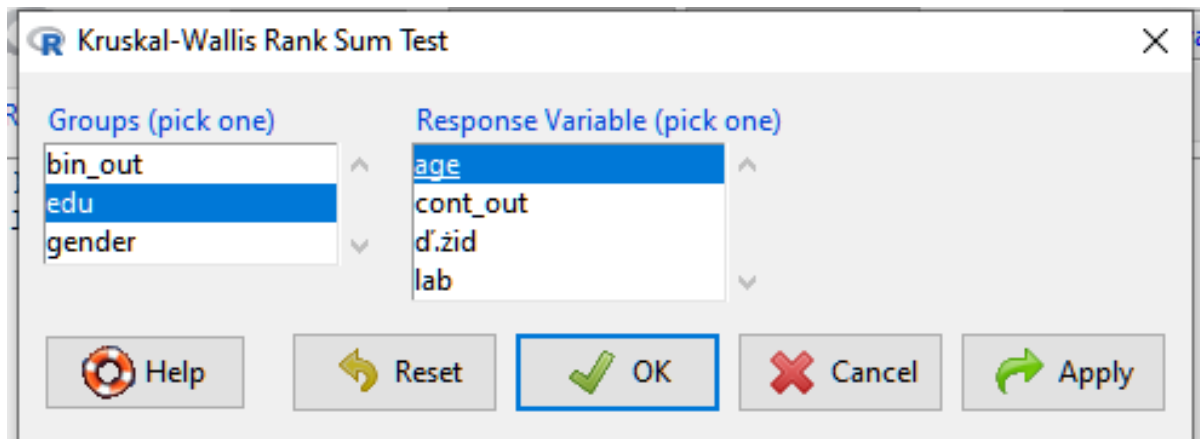
Kruskal-Wallis ANOVA

Az ANOVA nemparaméteres párja.



74. ábra: Kruskal-Wallis ANOVA kiválasztása

Vizsgáljuk meg, hogy van-e különbség medián életkorok tekintetében a képzettségi szintek mentén.



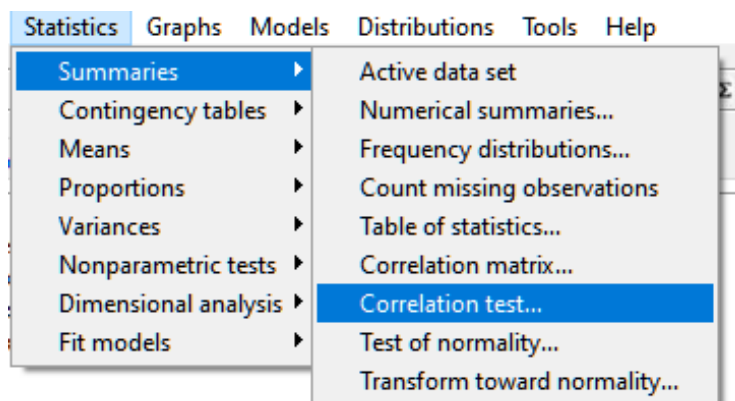
75. ábra: Kruskal-Wallis ANOVA beállítása

A teszt eredménye alapján van különbség (legalább 2 csoport között), mivel $p=0,042$. Hogy pontosan mely két csoport között van különbség, azt az úgynevezett post hoc tesztek mutatják meg. (Ilyen post hoc tesztre példa ANOVA esetén a Bonferroni, KW ANOVA esetén pedig a Dunn teszt.)

Korreláció

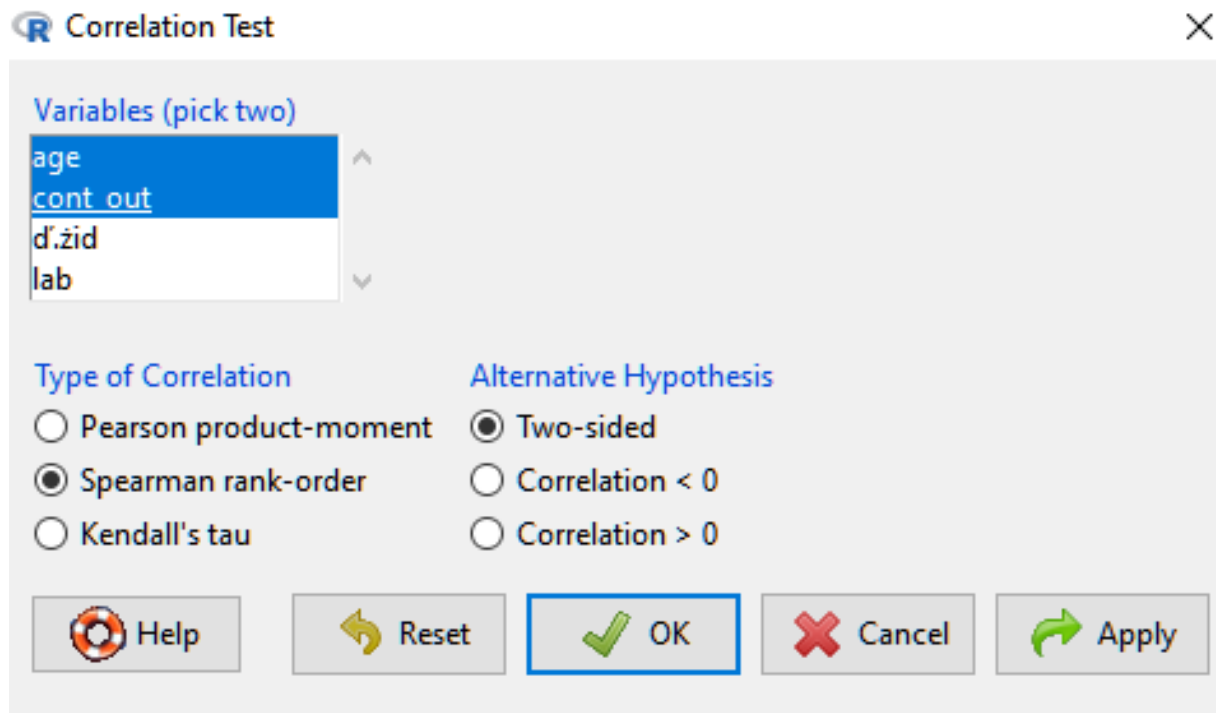
Két folytonos változó kapcsolatának elemzésére a korrelációs elemzés használható. A pontdiagram két tengelye mentén (x és y) a két változó van feltüntetve (tengely felcserélés csupán az értelmezésen változtathat, de a p -értéken és korrelációs koefficiensen nem). A statisztikai módszerek átfedésére jó példa a korreláció. Egyszerű lineáris regresszió esetén a regressziós koefficiens p -értéke tökéletesen megegyezik a korrelációs koefficiens p -értékével. A különbség csak annyi, hogy a korrelációs koefficiens a két folytonos változó százalékos együtt mozgását mutatja meg a -1 és $+1$ közötti tartományon belül, míg a regressziós koefficiens egy egységnyi változás (a független változóban) esetén mutatja meg a kimenetel (függő változó) változását egy $-\infty$ és $+\infty$ skálán.

Normális eloszlás esetén a Pearson-féle korreláció használandó, mely nemparaméteres párja a Spearman-féle korreláció.



76. ábra: Korreláció kiválasztása

Megvizsgálva az életkor (*age*, mely a korábbi normalitás teszt eredménye alapján nem mutat normális eloszlást) és a folytonos kimenetelünk együtt mozgását (korrelációját):



77. ábra: Korreláció beállítása

Elmondható, hogy Spearman-féle korreláció eredménye alapján nagyon gyenge (5,67%) együtt mozgás figyelhető meg, ami nem is szignifikáns ($p=0,732$).

```
> with(Dataset, cor.test(age, cont_out, alternative="two.sided", method="spearman"))  
  
Spearman's rank correlation rho  
  
data: age and cont_out  
S = 9319.6, p-value = 0.7316  
alternative hypothesis: true rho is not equal to 0  
sample estimates:  
rho  
0.05672035
```

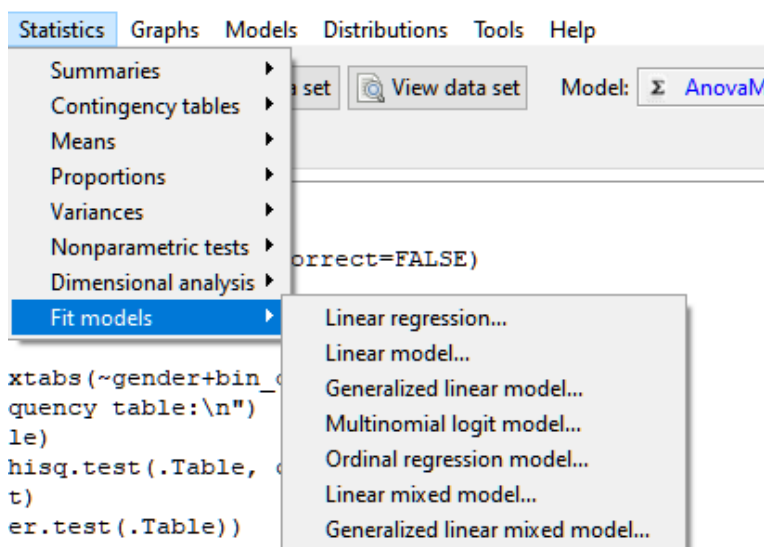
78. ábra: Spearman korreláció eredménye

Többszörös elemzés

A leíró statisztikák alapján képet kaptunk az adatbázisról. Az egyszerű elemzések kapcsán megnéztük, hogy mely változók befolyásolják szignifikánsan a kimenetelünket. A többszörös elemzések kapcsán pedig korrigált mutatókat kapunk, melyek a potenciális zavaró hatásokra korrigálva vannak. Sokszor keverik a többszörös (multiple) és többváltozós (multivariate) fogalmakat. A többváltozós azt jelenti, hogy nem egy vizsgálat kimenetelünk van, hanem több (pl. a diabétesz és a hipertónia kapcsán is szeretnénk tesztelni a magyarázó változók hatását). A többszörös modell azt jelenti, hogy nem csak egy magyarázó (befolyásoló) tényezőnk van (mint az egyszerű elemzéseknél), hanem több.

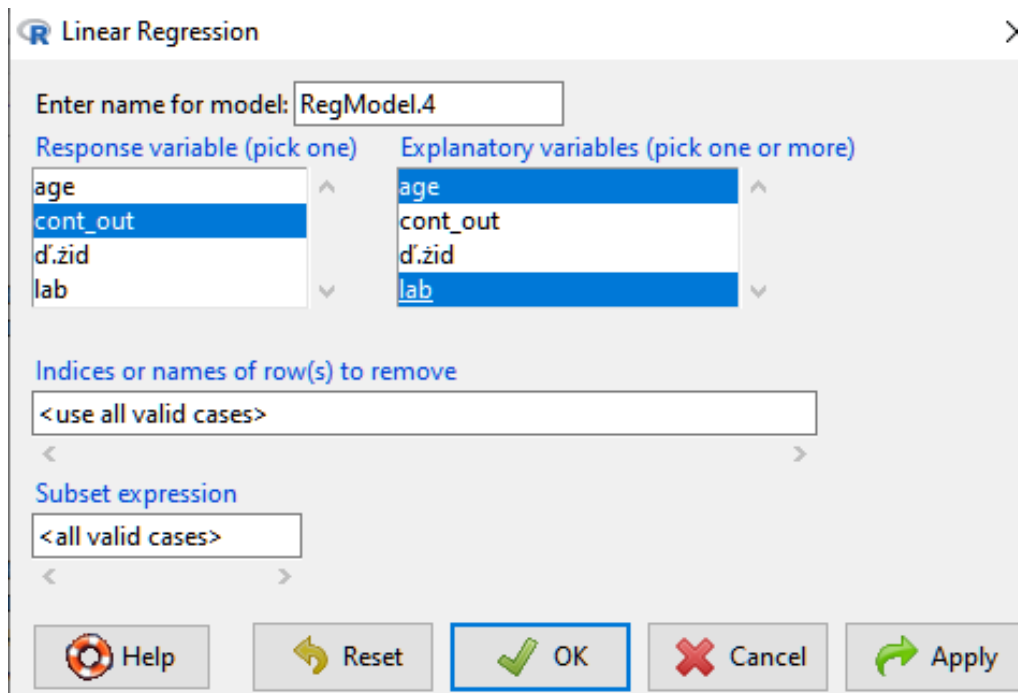
Lineáris regresszió

A többszörös modellek közül akkor választjuk a lineáris regressziót, ha a kimenetelünk folytonos változó. Itt gyakran megvizsgálják a kimenetel eloszlását (valójában az OLS regressziónál a pontdiagrammra illesztett egyenes esetén a maradékok eloszlását kellene vizsgálni), amennyiben nem közel normális (tehát nem kell 0,05 felett lenni a p-értéknek, elég, ha közel van hozzá, akár, ha alatta is van kicsit) az eloszlás, akkor vagy át kell alakítani a változónkat (transzformálni), vagy statisztikailag kicsit gyengébb módszert választani (pl. robusztus regresszió).



79. ábra: Lineáris regresszió kiválasztása

Nézzük meg, hogy a folytonos kimenetelünket befolyásolja-e a kor és/vagy a laborérték.



80. ábra: Lineáris regresszió beállítása

```
> RegModel.1 <- lm(cont_out~age+lab, data=Dataset)
> summary(RegModel.1)

Call:
lm(formula = cont_out ~ age + lab, data = Dataset)

Residuals:
    Min       1Q   Median       3Q      Max
-34.050 -11.403   1.799  12.665  32.897

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  42.54656    9.01198   4.721 0.0000351 ***
age           0.03908    0.11241   0.348   0.730
lab           0.02719    0.11024   0.247   0.807
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20.41 on 36 degrees of freedom
Multiple R-squared:  0.00547, Adjusted R-squared:  -0.04978
F-statistic: 0.09901 on 2 and 36 DF, p-value: 0.906
```

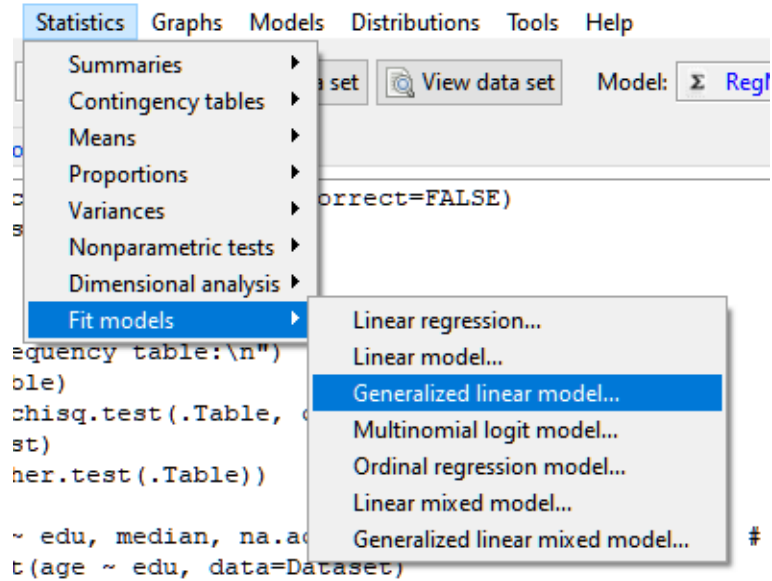
81. ábra: Lineáris regresszió eredménye

Úgy tűnik, egyik tényező (age, lab) sem befolyásolja szignifikánsan a kimenetelt ($p=0,348$; $p=0,807$).

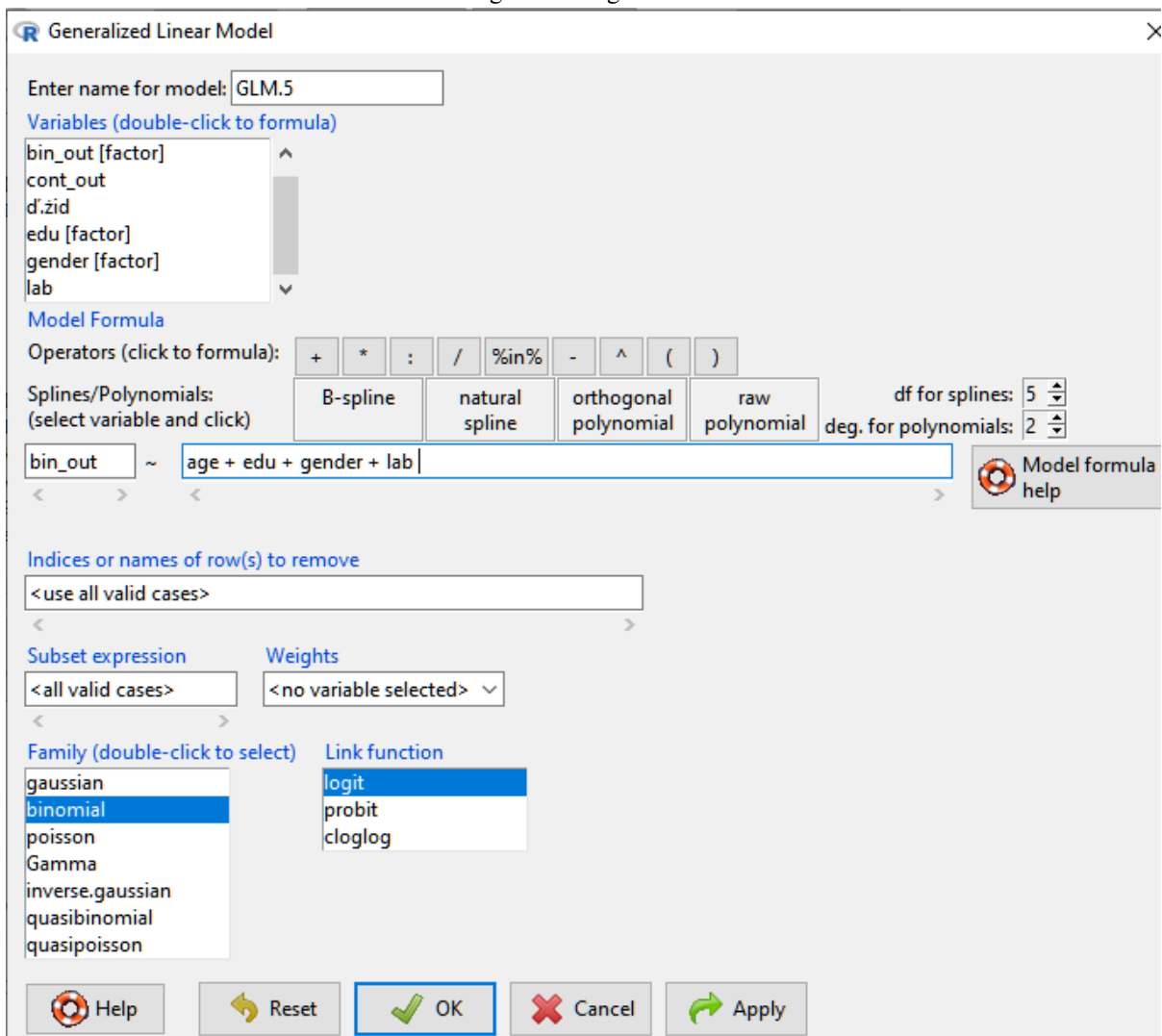
Logisztikus regresszió

Logisztikus regressziónál bináris a kimenetelünk (igen-nem; 0-1). Bár a modell ugyanúgy megadja a koefficienseket, gyakoribb az esélyhányados (Odds Ratio) értelmezése.

Vizsgáljuk meg a bináris kimenetelünk potenciális befolyásoló tényezőit.



82. ábra: Logisztikus regresszió kiválasztása



83. ábra: Logisztikus regresszió beállítása

```

> GLM.5 <- glm(bin_out ~ age + edu + gender + lab, family=binomial(logit), data=Dataset)
> summary(GLM.5)

Call:
glm(formula = bin_out ~ age + edu + gender + lab, family = binomial(logit),
    data = Dataset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6336  -0.7798   0.2215   0.7345   1.6044

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.41024    1.31429  -1.073  0.28327
age          -0.03434    0.02124  -1.617  0.10585
edu[T.2]      1.55378    1.27606   1.218  0.22336
edu[T.3]      1.18113    1.50671   0.784  0.43309
gender[T.2]  -1.92077    0.93891  -2.046  0.04078 *
lab           0.05569    0.02142   2.600  0.00933 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 54.040  on 38  degrees of freedom

```

84. ábra: Logisztikus regresszió eredménye

Úgy tűnik a nem (*gender*; $p=0,041$) és a laborérték (*lab*; $p=0,009$) szignifikáns befolyásoló tényezők.

Legörgetve megkapjuk az esélyhányadosokat is.

```

Residual deviance: 35.793  on 33  degrees of freedom
AIC: 47.793

Number of Fisher Scoring iterations: 5

> exp(coef(GLM.5)) # Exponentiated coefficients ("odds ratios")
(Intercept)      age      edu[T.2]      edu[T.3] gender[T.2]      lab
  0.2440845  0.9662393  4.7293281  3.2580580  0.1464938  1.0572679

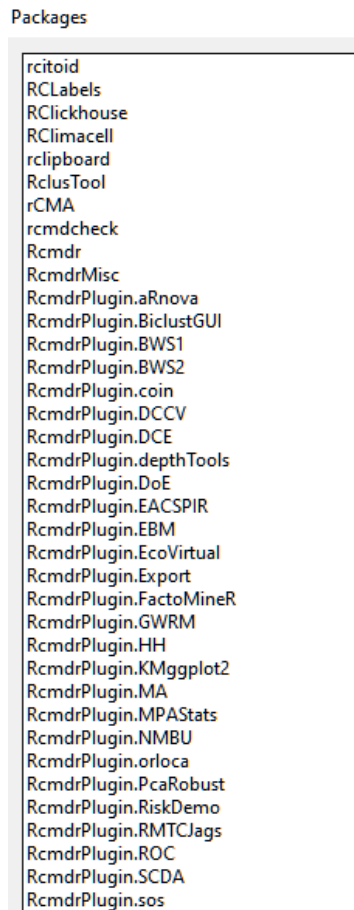
```

85. ábra: Logisztikus regresszió esélyhányadosai

A *nem* protektív tényező, előremutató értelmezésként 0,15-ször kisebb eséllyel rendelkezik majd a nő a kimenetellel a férfيهz képest. A laborérték 1,05-szörös rizikótényezőnek tűnik.

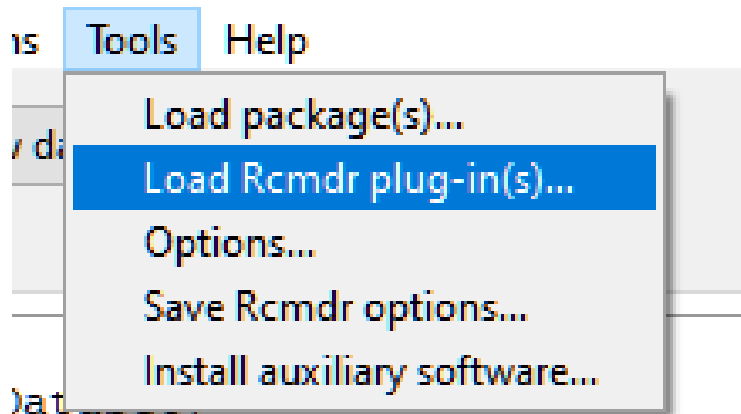
Cox-regresszió

A túlélés elemzésre használt Cox-regressziónál szintén bináris a kimenetel (él, nem él). Annyival több egy logisztikus regressziónál, hogy a nyomkövetési idő is számításba van véve. Első lépésként az R packages-ből installáljuk az „*RcmdrPlugin.survival*” plugint:

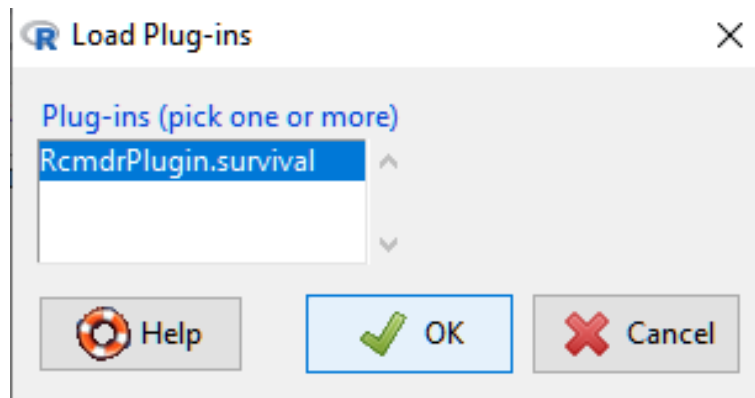


86. ábra: Cox-regresszióhoz szükséges csomag telepítése

Ezt követően az Rcommander-ből ezt be tudjuk tölteni.

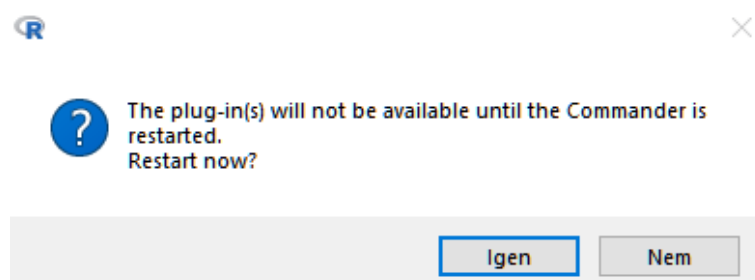


87. ábra: R csomag betöltése



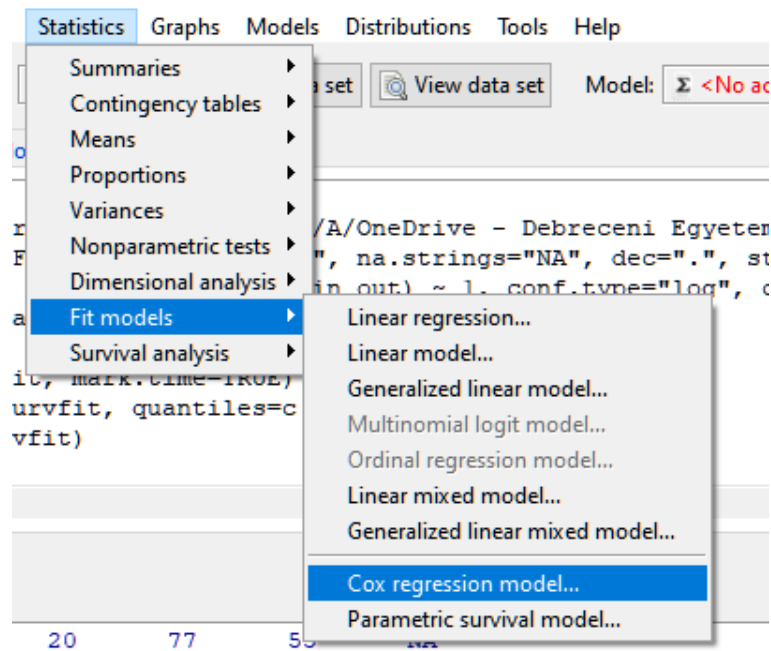
88. ábra: Túlélés elemzéshez használandó csomag kiválasztása

A betöltés az R commander újraindítását igényli (a meglévő adatbázis törlődik a memóriából).



89. ábra: R commander újraindítása

Megkaptuk az új menüpontot.



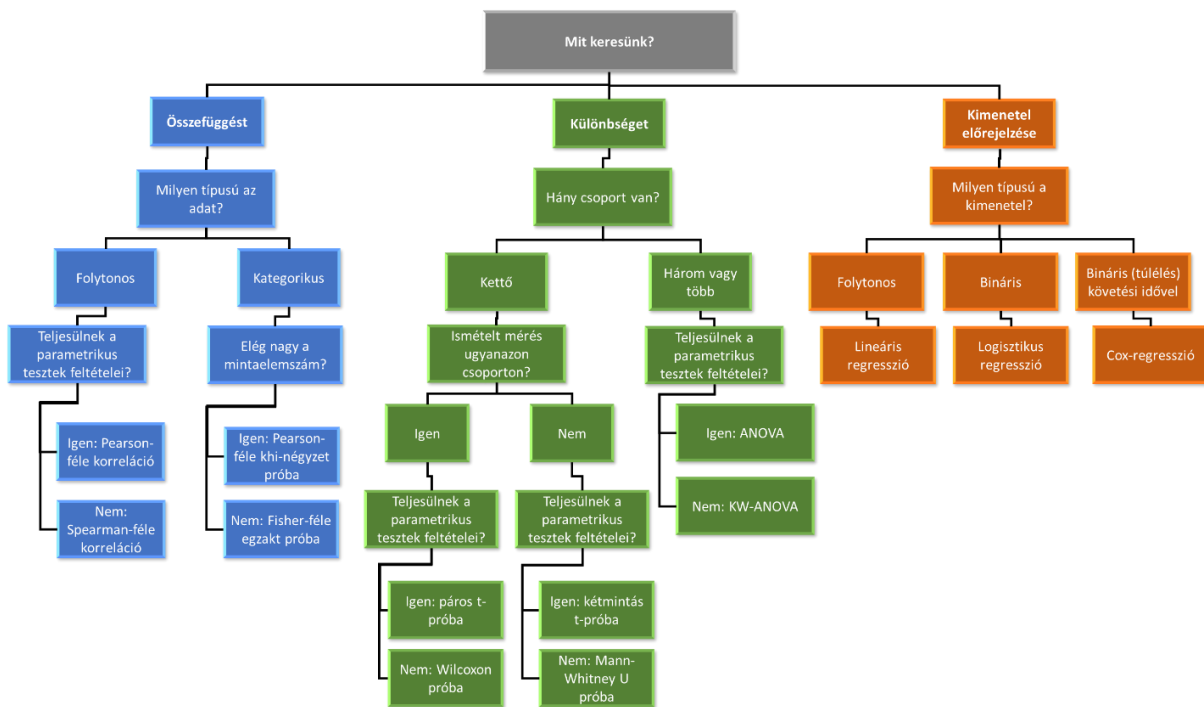
90. ábra: Cox-regresszió kiválasztása

A Cox regressziónál be tudjuk állítani a követési időt (time), a kimenetelt (event), melynél általában 1-es a halál bekövetkeztének kódja. Veszélyhányadosokat ad vissza (Hazard Ratio) melyek értelmezése az esélyhányadoshoz hasonlatos.

Összefoglaló az elemzési módszerekről

A statisztikai módszerek alkalmazásáról látható egy összefoglaló ábra a kutatási kérdések és a rendelkezésre álló adatok függvényében. A leggyakrabban vizsgálat feltétel a normalitás megléte, illetve a megfelelő mintaelemszám.

Az alábbi összefoglaló ábra a kutatási kérdéstől indul, majd az adatok alapján haladunk a végső adekvát módszerig a feltételek teljesülése alapján.



91. ábra: Összefoglaló az elemzési módszerekről

Eredmények ábrázolása

A megfelelő vizualizáció elengedhetetlen fontosságú, az eredményeink bemutatása kapcsán. A különféle adatok a természetükből fakadóan könnyebben áttekinthetők a megfelelő diagramtípus kiválasztását követően. A teljesség igénye nélkül a főbb dolgokat megemlítve:

- időtrendeket vonaldiagramon érdemes ábrázolni (line chart) (pl. prevalencia alakulása hosszabb időtávon)
- százalékos megoszlásokat kördiagramon (pie chart) szemléltetünk általában (pl. nemek megoszlása)
- folytonos változót hisztogramon szemléltetünk (histogram)
- nemnormális folytonos változót doboz ábrán mutatunk be (box plot)
- folytonos változók kapcsolatát a pontdiagram írja le (scattergram)
- kategorikus változók (pl.: iskolai végzettség) esetén oszlopdiagram is használható (bar chart)

Utószó

A Felhasznált irodalom rendhagyó módon hasznos linkeket is tartalmaz rövid leírással/magyarázattal. A felhasznált adatbázis kódszótárral áll rendelkezésre és teszi lehetővé a bemutatott módszerekkel elért eredmények reprodukcióját. A statisztikára fokozottan igaz a „Gyakorlat teszi a mestert” mondás. Amennyiben egy választott módszer alkalmazási feltételeiben vagy adott probléma esetén használhatóságában vagyunk bizonytalanok, akkor keressünk utána. A PubMed segít a kutatási kérdésnek megfelelő módszerek megtalálásában/visszaellenőrzésében. A YouTube sokszor részletesen végig visz minket a kivitelezésen, akár még az értelmezésen is. A Google a barátunk (Google Is Your Friend), megfelelő kulcsszavak használatával értékes oldalakra találhatunk. Érdeemes a kulcsfontosságú szóösszetételeket idézőjelbe tennünk, illetve, ha konkrét fájl típusra keresünk, használjuk a megfelelő opciót (pl.: filetype:pdf). Gyakoroljunk minél többet!

Felhasznált irodalom

1. A **Debreceni Egyetem, Egészségtudományi Kar** releváns előadásanyagai részletes kiegészítéssel bírnak a megemlített összefoglaló gyakorlati aspektusokhoz.
2. Részletes elméleti kiegészítő Dr. Takács Péter: **Probléma alapú kutatómódszertani és biostatistikai ismeretek** című jegyzet.
3. A kérdőíveknél használható validált mintával szolgálhat az **Európai Lakossági Egészségfelmérés** (https://www.ksh.hu/elef/elef2019_kerdoiv.pdf)
4. A **Microsoft Forms** (<https://www.office.com/launch/forms?auth=2>) megbízható beviteli felület szerkesztő.
5. Hasznos fordító a Google Translate mellett a **DeepL** (<https://www.deepl.com/translator>)
6. Ingyenes és könnyen használható referenciaszerkesztő a **Zotero** (<https://www.zotero.org/>), mely integrálódik a böngészőbe és a Microsoft Word-be is.
7. Számos online **adatbázis** elérhető nemzetközi összehasonlításra alkalmas adatokkal, pl. HFA explorer (<https://gateway.euro.who.int/en/hfa-explorer/>), Global Burden of Disease (GBD) (<https://vizhub.healthdata.org/gbd-compare/>)
8. A **Laerd Statistics** (<https://statistics.laerd.com/spss-tutorials/linear-regression-using-spss-statistics.php>) részletesen leírja az alkalmazási feltételeket (assumptions) és a lebonnyítás részleteit is bemutatja SPSS-ben.
9. Szintén részletes bemutató található értelmezéssel az UCLA: **Statistical Consulting Group** oldalán (<https://stats.oarc.ucla.edu/other/dae/>) Stata, SAS, SPSS, Mplus és R programokra.

Felhasznált adatbázis (pontosvesszővel elválasztott értékek)

A példákhoz felhasznált adatbázis (minta.csv) tartalma magyar elválasztóval (; separator).

id;gender;age;edu;lab;bin_out;cont_out

1;1;53;1;95;1;49
2;2;88;3;80;0;54
3;1;7;1;96;1;35
4;2;5;3;16;0;55
5;1;19;1;50;1;55
6;2;92;3;12;0;58
7;2;97;3;63;0;14
8;1;50;2;20;1;53
9;1;3;1;90;1;40
10;1;29;1;88;1;38
11;2;94;3;39;0;75
12;2;82;2;74;0;14
13;1;77;3;89;1;78
14;1;21;1;46;0;65
15;1;29;1;21;0;57
16;2;37;2;42;0;36
17;1;86;2;60;0;15
18;2;15;2;26;0;38
19;1;54;3;99;1;52
20;1;28;3;17;0;77
21;1;77;2;95;1;48
22;2;70;1;77;0;35
23;2;45;3;22;0;11
24;2;44;2;59;1;73
25;1;43;2;39;1;50
26;1;72;1;12;0;18
27;1;45;2;41;0;58
28;1;30;1;99;1;23
29;2;62;2;88;0;72
30;1;65;2;25;1;39
31;2;9;2;23;1;13
32;2;97;1;16;0;47
33;1;49;2;70;1;26
34;1;89;2;35;1;39
35;1;84;2;77;1;75
36;2;90;3;97;1;64
37;1;78;2;64;0;73
38;2;30;1;82;1;49
39;2;83;3;82;1;32

Felhasznált adatbázis (vesszővel elválasztott értékek)

A példákhoz felhasznált adatbázis (minta.csv) vesszővel elválasztott formában (ami az eredeti angol comma-separated values).

id,gender,age,edu,lab,bin_out,cont_out

1,1,53,1,95,1,49
2,2,88,3,80,0,54
3,1,7,1,96,1,35
4,2,5,3,16,0,55
5,1,19,1,50,1,55
6,2,92,3,12,0,58
7,2,97,3,63,0,14
8,1,50,2,20,1,53
9,1,3,1,90,1,40
10,1,29,1,88,1,38
11,2,94,3,39,0,75
12,2,82,2,74,0,14
13,1,77,3,89,1,78
14,1,21,1,46,0,65
15,1,29,1,21,0,57
16,2,37,2,42,0,36
17,1,86,2,60,0,15
18,2,15,2,26,0,38
19,1,54,3,99,1,52
20,1,28,3,17,0,77
21,1,77,2,95,1,48
22,2,70,1,77,0,35
23,2,45,3,22,0,11
24,2,44,2,59,1,73
25,1,43,2,39,1,50
26,1,72,1,12,0,18
27,1,45,2,41,0,58
28,1,30,1,99,1,23
29,2,62,2,88,0,72
30,1,65,2,25,1,39
31,2,9,2,23,1,13
32,2,97,1,16,0,47
33,1,49,2,70,1,26
34,1,89,2,35,1,39
35,1,84,2,77,1,75
36,2,90,3,97,1,64
37,1,78,2,64,0,73
38,2,30,1,82,1,49
39,2,83,3,82,1,32

Felhasznált adatbázis (QR kód)

Az alábbi QR kód tartalmazza az adatbázist. A QR kód generálás az ingyenes QR Code Generátorral történt. (<https://goqr.me/>)



A fenti kép kamerán kívül beolvasható online QR kód olvasókkal. (<https://blog.qr4.nl/Online-QR-Code-Decoder.aspx>)

Felhasznált adatbázis kódszótára

id – egyedi azonosító

gender – nem

- 1 – férfi
- 2 – nő

age – életkor (év)

edu – iskolai végzettség

- 1 – alapkú
- 2 – középkú
- 3 – felsőpkú

lab – vizsgálat laboratóriumi paraméter (folytonos változó)

bin_out – bináris kimenetel

- 0 – nem
- 1 – igen

cont_out – folytonos kimenetel

Feladatok

1. Számolja ki Excelben a minta.csv-ben található kor (age) változó átlagát és ezen pontbecslés 95% megbízhatósági tartományát, valamint ez alapján nyilatkozzon, hogy eltér-e szignifikánsan a minta átlagéletkora a 35-től! Ábrázolja az átlagot és a megbízhatósági tartományt!

2. Adott 50 normál BMI-vel rendelkező férfi, 80 elhízott férfi, 90 normál BMI-vel rendelkező nő és 20 elhízott nő. Van -e összefüggés a nem és az elhízás között?

3. Nyomonkövetéssel vizsgálat során száz embert követünk egy évig (2022.01.01-2022.12.31). Tíz embernél történik valamilyen esemény (betegség/halál/költözés), a fennmaradó kilencven embernél ezen események egyike sem következik be. A sárga cella a betegséget jelöli, a zöld az elköltözést, a fekete a halált. A vizsgálat betegség hasonlít az influenzára (ismételten fertőzhető), de nem korlátozódik a téli hónapokra. A színes celláknál az adott esemény az adott hónap elsején alakul ki és az utolsó színes hónap utolsó napjáig tart. (a) Mennyi a prevalencia július elsején? (b) Mennyi az egy éves kumulatív incidenciája és (c) mennyi az incidenciasűrűség?

		2022											
Személy	2021. december	2022. január	2022. február	2022. március	2022. április	2022. május	2022. június	2022. július	2022. augusztus	2022. szeptember	2022. október	2022. november	2022. december
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													

4. Az alábbi 2x2-es kontingenciatáblázatok esetén számolja ki a kapcsolati mutatókat az alábbi (a) kohorsz és (b) eset-kontroll vizsgálatokra!

	A	B	C
1	kohorsz	beteg	nem beteg
2	exponált	40	70
3	nem exponált	35	90

	A	B	C
1	eset-kontroll	eset	kontroll
2	exponált	40	50
3	nem exponált	45	90

5. Számolja ki a hiányzó értékeket, ezt követően pedig prevalenciát és a tesztjellemzőket!

	A	B	C	D
1		beteg	nem beteg	total
2	teszt+	40		50
3	teszt-		90	
4	total	45		

Megoldások

1. Számolja ki Excelben a minta.csv-ben található kor (age) változó átlagát és ezen pontbecslés 95% megbízhatósági tartományát, valamint ez alapján nyilatkozzon, hogy eltér-e szignifikánsan a minta átlagéletkora a 35-től! Ábrázolja az átlagot és a megbízhatósági tartományt!

Az átlag [95% megbízhatósági tartomány]: 54,56 [45,28 - 63,84]

A mintám pontbecslése (54,56) szignifikánsan magasabb a hipotetikus 35-től, a megbízhatósági tartomány alapján, mivel a nullhipotézis szerinti érték (35) nincs benne a 95% megbízhatósági tartományban.

Az eredmény egy cellában előállítható a következő függvényekkel:

=ÖSSZEFŰZ(KEREKÍTÉS(D3;2);" [";KEREKÍTÉS(D3;2)-
KEREKÍTÉS(1,96*SZÓRÁS(A2:A40)/GYÖK(DARAB(A2:A40)));2);" -
";KEREKÍTÉS(D3;2)+KEREKÍTÉS(1,96*SZÓRÁS(A2:A40)/GYÖK(DARAB(A2:A40)));2);"]")

Felhasznált függvények:

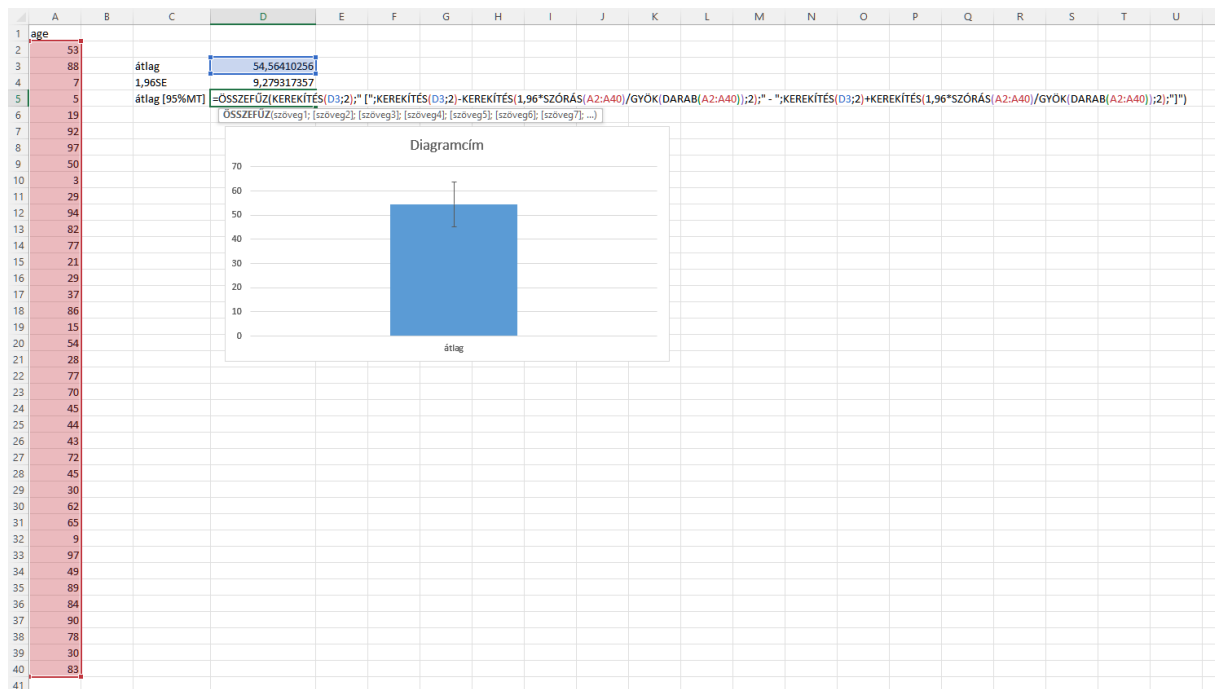
ÖSSZEFŰZÉS: összefűzi a megadott számokat/karaktereket

KEREKÍTÉS: adott tizedesjegyig kerekíti a számot

SZÓRÁS: kijelölt tartomány szórását számolja ki

GYÖK: négyzetgyököt von

DARAB: összeszámolja a cellákat az adott tartományban, jelen esetben a mintaelemszámmal egyezik

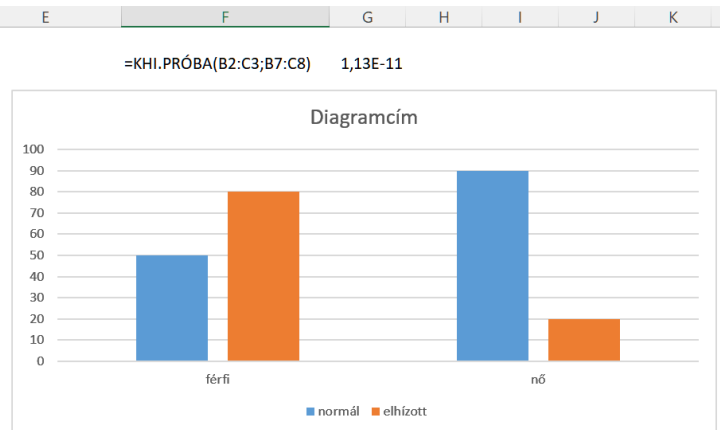


2. Adott 50 normál BMI-vel rendelkező férfi, 80 elhízott férfi, 90 normál BMI-vel rendelkező nő és 20 elhízott nő. Van -e összefüggés a nem és az elhízás között?
A khi-négyzet próba alkalmas a kérdés megválaszolására:

- Kontingencia táblázatban tudjuk feltüntetni az aggregált esetszámokat.
- Ezt követően ki kell számolnunk a sorok és oszlopok összegét.
- Ezután az oszlopokban előforduló gyakoriságok meghatározása következik.
- A várható esetszámok az előbbi számítások eredményeinek felhasználásával számíthatók ki.

Az összefüggés szignifikáns, $p < 0,001$, az ábra alapján a férfiak között több az elhízott.

	A	B	C	D
1	tényleges	normál	elhízott	total
2	férfi	50	80	=SZUM(B2:C2)
3	nő	90	20	=SZUM(B3:C3)
4	total	=SZUM(B2:B3)	=SZUM(C2:C3)	=SZUM(D2:D3)
5	gyakoriság	=B4/D4	=C4/D4	
6	várható	normál	elhízott	
7	férfi	=B\$5*\$D2	=C\$5*\$D2	
8	nő	=B\$5*\$D3	=C\$5*\$D3	
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				



3. Nyomonkövetéssel vizsgált során száz embert követünk egy évig (2022.01.01-2022.12.31). Tíz embernél történik valamilyen esemény (betegség/halál/költözés), a fennmaradó kilencven embernél ezen események egyike sem következik be. A sárga cella a betegséget jelöli, a zöld az elköltözést, a fekete a halált. A vizsgálat betegség hasonlít az influenzára (ismételten fertőzhető), de nem korlátozódik a téli hónapokra. A színes celláknál az adott esemény az adott hónap elsején alakul ki és az utolsó színes hónap utolsó napjáig tart. (a) Mennyi a prevalencia július elsején? (b) Mennyi az egy éves kumulatív incidenciája és (c) mennyi az incidenciasűrűség?

	2022												
Személy	2021. december	2022. január	2022. február	2022. március	2022. április	2022. május	2022. június	2022. július	2022. augusztus	2022. szeptember	2022. október	2022. november	2022. december
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													

(a) $P = n/N = 3/96 = 0,0312$

A prevalencia 3,12%, mivel július elsején három beteg van, a teljes populáció pedig 100-4 (három személy elköltözött, egy pedig elhunyt.)

(b) $KI = n/N = 5/95 = 0,0526$

Az 1-éves kumulatív incidenciája 5,26%, mivel 5 új esemény (néhány személynél többször előfordult) következett be, kezdetben pedig 100-5 fő (5 fő már eleve beteg) volt kockázatnak kitéve.

(c) $IS = n/\text{összes személyidő} = 5/1115 = 0,0045/\text{személyhó} = 0,0538/\text{személyév}$

Összesen öt új esemény következett be. Kilencven személynél nem fordult elő esemény, a személyidejük $90 \cdot 12 = 1080$ hónap, ameddig kockázatnak (nem beteg és a vizsgálatban van) voltak kitéve. A fennmaradó 10 fő személyideje $7 + 0 + 4 + 11 + 0 + 1 + 1 + 3 + 1 + 7 = 35$ hónap, összesen tehát $1080 + 35 = 1115$ hónap. Ez a személyhónap $1115/12 = 92,92$ személyévnek felel meg.

4. Az alábbi 2x2-es kontingenciatáblázatok esetén számolja ki a kapcsolati mutatókat az alábbi kohorsz és eset-kontroll vizsgálatokra!
(a)

	A	B	C
1	kohorsz	beteg	nem beteg
2	exponált	40	70
3	nem exponált	35	90

Kohorsz vizsgálat esetén a kapcsolati mutató a relatív kockázat, amikor a két abszolút kockázatot (kumulatív incidenciát) osztjuk el egymással.

	A	B	C	D
1	kohorsz	beteg	nem beteg	
2	exponált	40	70	=B2/(B2+C2)
3	nem exponált	35	90	=B3/(B3+C3)
4				
5			RK	=D2/D3
6				1,30

A relatív kockázat alapján az exponáltaknak 1,3x nagyobb kockázata van a megbetegedésre, a nem-exponáltakhoz képest.

(b)

	A	B	C
1	eset-kontroll	eset	kontroll
2	exponált	40	50
3	nem exponált	45	90

	A	B	C
1	eset-kontroll	eset	kontroll
2	exponált	40	50
3	nem exponált	45	90
4		=B2/B3	=C2/C3
5			
6	EH	=B4/C4	
7		1,6	

Az expozíciós esélyek hányadosa az esélyhányados. Jelen esetben 1,6x nagyobb az expozíció esélye az esetek körében.

5. Számolja ki a hiányzó értékeket, ezt követően pedig prevalenciát és a tesztjellemzőket!

	A	B	C	D
1		beteg	nem beteg	total
2	teszt+	40		50
3	teszt-		90	
4	total	45		

	A	B	C	D
1		beteg	nem beteg	total
2	teszt+	40	=D2-B2	50
3	teszt-	=B4-B2	90	=B3+C3
4	total	45	=C3+C2	=B4+C4
5				
6		P	31%	=B4/D4
7		Szenzitivitás	89%	=B2/B4
8		Specificitás	90%	=C3/C4
9		PPV	80%	=B2/D2
10		NPV	95%	=C3/D3

A betegség gyakorisága (prevalencia) 31%.

A teszt a betegek 89%-át azonosítja helyesen. (szenzitivitás)

A teszt az egészségesek 90%-át azonosítja helyesen. (specificitás)

A pozitívnak talált egyének 80%-os valószínűséggel betegek. (PPV)

A negatívnak talált egyének 95%-os valószínűséggel egészségesek. (NPV)

Melléklet

Az alábbiakban néhány epidemiológiai szempontból fontos kérdőívet vagy kérdőív részletet mellékelünk forrásmegjelöléssel.

Cukorbetegség kockázat mérő FINDRISC kérdőív ¹

A Findrisk vizsgálat adatlapja (magyar változat)

A közelmúltban igen egyszerű, ugyanakkor jól használható kérdőívek váltak népszerűvé, amelyek a 2-es típusú cukorbetegség korai felismerését segítik. Kérjük, szánjon néhány percet a kitöltésre, majd adja át az asszisztensnőnek, aki segít Önnek az értékelésben. A válaszadás egyszerű, javasoljuk, kezdje el Ön, Kedves Olvasó, de ne feledkezzék meg családtagjairól sem! Kérjen számukra is kérdőívet, amelyet kitöltés után hasonló módon juttasson vissza.

Adatlap 2-es típusú cukorbetegség kockázatának felmérésére

Az üres kockába írja be az Önre vonatkozó pontértéket, majd adja össze azokat a kérdőív végén!

1. Életkor

- 0. pont 45 év alatt
- 2. pont 45-54 év között
- 3. pont 55-64 év között
- 4. pont 64 év felett

2. Testtömeg index (BMI)

A testtömeg index kiszámolása: Testsúly kg-ban, osztva a méterben mért testmagasság négyzetével. A számításban segítenek Önnek. Pl: ha az Ön magassága 165 cm, súlya 70 kg, a számítás: $70/(1,65 \times 1,65) = 25,7$

- 0. pont alacsonyabb, mint 25 kg/m²
- 1. pont 25-30 kg/m²
- 3. pont nagyobb, mint 30 kg/m²

3. Haskőrfogat

(az alsó bordák és a csípőtővis közötti távolság felénél mérve - gyakorlatilag a köldök vonalában, közepes belégzés után)

Pont	Férfiak	Pont	Nők
0	Kevesebb, mint 94 cm	0	Kevesebb, mint 80 cm
3	94-102 cm	3	80-88 cm
4	Több, mint 102 cm	4	Több, mint 88 cm

4. Végez e legalább 30 perces fizikai tevékenységet munkaköréből adódóan/vagy szabadidejében?

- 0. pont Igen
- 2. pont Nem

5. Milyen gyakran fogyaszt zöldséget, vagy gyümölcsöt?

- 0. pont Minden nap
- 1. pont Nem minden nap

¹ Lindström J, Tuomilehto J. The Diabetes Risk Score: A practical tool to predict type 2 diabetes risk. Diabetes Care. 2003 Mar 1;26(3):725-31.
Magyar változat: <https://www.reg-med.hu/findrisk.pdf>

6. Szed e rendszeresen vérnyomáscsökkentő gyógyszereket?

0. pont Nem

2. pont Igen

7. Mértek e Önnél valaha magasabb vércukor értéket (orvosi vizsgálatkor, betegség, terhesség esetén)

0. pont Nem

5. pont Igen

8. Van e családtagjai között, vagy közeli rokonságában 1-es (ifjúkori típusú), vagy 2-es (időskori típusú) cukorbeteg?

0. pont Nem

3. pont Igen: nagyszülő, nagynéni, nagybácsi,
vagy elsőfokú unokatestvér

5. pont Igen: szülő, testvér, vagy saját gyermek

Értékelés: összesen pont

Kockázat a cukorbetegség 10 éven belüli kialakulására:

Kevesebb, mint 7 pont	alacsony. Becslések szerint 100 főből 1 esetben lesz valaki cukorbeteg,
7-11	enyhén fokozott kockázat. Becslések szerint 25 főből 1 esetben lesz valaki cukorbeteg,
12-14	fokozott kockázat. Becslések szerint 6 főből 1 esetben lesz valaki cukorbeteg,
15-20	magas. Becslések szerint 3 főből 1 esetben lesz valaki cukorbeteg,
Több mint 20 pont	igen magas. Becslések szerint 2 főből 1 cukorbeteg lesz 10 éven belül.

Alkoholfogyasztási rendellenességek azonosítási tesztje²

ALKOHOLFOGYASZTÁSI RENDELLENESÉGEK AZONOSÍTÁSI TESZTJE (AUDIT-10) kérdőív

- 1. Milyen gyakran iszik alkohol tartalmú italt?**
Soha = 0 pont
Havonta vagy kevesebbszer = 1 pont
Kétszer vagy négyszer egy hónapban = 2 pont
Kétszer vagy háromszor egy héten = 3
Négyszer vagy többször egy héten = 4
- 2. Mennyi alkoholtartalmú italt fogyaszt egy olyan napon, amikor iszik?**
1 vagy 2 = 0 pont
3 vagy 4 = 1 pont
5 vagy 6 = 2 pont
7 vagy 9 = 3 pont
10 vagy több = 4 pont
- 3. Milyen gyakran iszik hat vagy annál több alkoholtartalmú italt egy alkalommal?**
Soha = 0 pont
Kevesebbszer, mint havonta = 1 pont
Havonta = 2 pont
Hetente = 3 pont
Naponta vagy szinte soha = 4 pont
- 4. Az elmúlt évben milyen gyakran érezte, hogy nem bírja abbahagyni az ivást, ha már elkezdte?**
Soha = 0 pont
Kevesebbszer, mint havonta = 1 pont
Havonta = 2 pont
Hetente = 3 pont
Naponta vagy szinte soha = 4 pont
- 5. Az elmúlt évben milyen gyakran érezte, hogy nem bírta teljesíteni azt, amit Öntől elvártak, mert ivott?**
Soha = 0 pont
Kevesebbszer, mint havonta = 1 pont
Havonta = 2 pont
Hetente = 3 pont
Naponta vagy szinte soha = 4 pont
- 6. Az elmúlt évben milyen gyakran volt szüksége egy első italra reggel, hogy elkezdje a napot egy erős italozás után?**
Soha = 0 pont
Kevesebbszer, mint havonta = 1 pont
Havonta = 2 pont
Hetente = 3 pont
Naponta vagy szinte soha = 4 pont

² AUDIT : the Alcohol Use Disorders Identification Test : guidelines for use in primary health care [Internet]. [cited 2023 Feb 16]. Available from: <https://www.who.int/publications-detail-redirect/WHO-MSD-MSB-01.6a>
Magyar változat: <http://efi.szentendre.hu/wp-content/uploads/2020/07/AlkoholF%C3%BCgg%C5%91s%C3%A9g-teszt.pdf>

7. Az elmúlt évben milyen gyakran érezte rosszul vagy volt büntudata ivás után?
 Soha = 0 pont
 Kevesebbszer, mint havonta = 1 pont
 Havonta = 2 pont
 Hetente = 3 pont
 Naponta vagy szinte soha = 4 pont
8. Az elmúlt évben milyen gyakran fordult elő, hogy nem emlékezett arra, mi történt az előző este az italozás miatt?
 Soha = 0 pont
 Kevesebbszer, mint havonta = 1 pont
 Havonta = 2 pont
 Hetente = 3 pont
 Naponta vagy szinte soha = 4 pont
9. Megsérült-e Ön vagy valaki más italozása miatt?
 Nem = 0 pont
 Igen, de nem az elmúlt évben = 2 pont
 Igen a múlt évben = 4 pont
10. Aggódott-e már rokon, barát, orvos vagy egy egészségügyi dolgozó az ivása miatt és javasolta a csökkentést?
 Nem = 0 pont
 Igen, de nem az elmúlt évben = 2 pont
 Igen a múlt évben = 4 pont

Összpontszám:pont

Kiértékelés: Javasolt beavatkozás:

- 0-7 pont = Alacsony kockázatú ivás Felvilágosítás;
- 8-15 pont = Magas kockázatú ivás Rövid intervenció, újra értékelés (kétévente);
- 16-19 pont = Ártalmas ivás Rövid intervenció, rendszeres monitorozás;
- 20- pont = Alkoholfüggőség valószínű Szakellátásba irányítás diagnosztikai és terápiás céllal

Fagerström-féle nikotinfüggőségi teszt³

FAGERSTRÖM-FÉLE NIKOTINFÜGGŐSÉGI TESZT, MINIMÁLINTERVENCIÓN

Tabulátor - balra

Név:
TAJ:
Születési idő:
Életkor:
Vizsgálatot végző neve:
Vizsgálat dátuma:

Válaszoljon az alábbi kérdésekre!

- 1. Ébredés után mennyi idővel gyújt rá az első cigarettára?**

60 perccel később	0 pont
31-60 perc között	1 pont
6-30 perc között	2 pont
5 percen belül	3 pont
- 2. Problémát okoz-e Önnek, ha olyan helyen van, ahol nem lehet rágyújtani?**

Igen	1 pont
Nem	0 pont
- 3. A nap folyamán melyik cigarettát élvezzi a legjobban?**

A reggeli első rágyújtást	1 pont
Bármilyen más alkalommal	0 pont
- 4. Naponta hány cigarettát szokott elszívni?**

10 szál vagy kevesebb	0 pont
11-20 szál között	1 pont
21-30 szál között	2 pont
31 szál vagy annál több	3 pont
- 5. Többet szív reggel, mint a nap hátralévő részében?**

Igen	1 pont
Nem	0 pont
- 6. Rá szokott-e gyújtani olyankor is, amikor olyan beteg, hogy fel sem tud kelni az ágyból?**

Igen	1 pont
Nem	0 pont

Adja össze a pontszámait:

Az ön függésének szintjét a kérdőívben kapott pontszámával jellemezhetjük:

- | | |
|-------------------|-----------------|
| 0-2 pont: | Nagyon alacsony |
| 3-4 pont: | Alacsony |
| 5 pont: | Közepes |
| 6-7 pont: | Magas |
| 8-10 pont: | Nagyon magas |

³ Fagerstrom Test for Nicotine Dependence - an overview | ScienceDirect Topics [Internet]. [cited 2023 Feb 16]. Available from: <https://www.sciencedirect.com/topics/medicine-and-dentistry/fagerstrom-test-for-nicotine-dependence>
Magyar változat: <https://alapellatas.okfo.gov.hu/wp-content/uploads/2021/11/Fagerstorm-fele-nikotinfuggosegi-teszt-minimalintervencio.pdf>

Nemzetközi Fizikai Aktivitás Kérdőív Rövidített Változata⁴

IPAQ (International Physical Activity Questionnaire Short Form)

IPAQ (International Physical Activity Questionnaire Short Form)

Nemzetközi Fizikai Aktivitás Kérdőív Rövidített Változata

Szeretnénk megismerni az Ön mindennapos, jellemző fizikai aktivitását. A kérdések az utolsó hét napra vonatkoznak, de ha ezek lényegesen különböznenek az év nagyobb részétől, akkor írja be zárójelben az egyes kérdések mögé azt a számot, amit Ön a testmozgását, fizikai aktivitását az egész évben (annak nagyobb részében) jellemezi.

Csak azokat a testmozgásokat tüntesse fel, amelyek legalább 10 percig tartanak, az ennél rövidebbeket ne.

Intenzívnek az olyan testmozgást nevezzük, ami jelentősen fokozza a lihegését, növeli a pulzusszámát, amitől fél-egy óra alatt elfárad (ha nem szokott hozzá az évek során).

1. Az elmúlt 7 nap során hány napon végzett intenzív fizikai mozgást (nagy súlyok emelése, ásás, kocogás-futás, izzasztó kerékpározás)?
.....napon az elmúlt héten
.....percen át naponta
2. Mérsékelt erősségű (kissé lihegtető) testmozgást mennyit végzett az elmúlt 7 nap során legalább 10 percen át?
.....napon az elmúlt héten
.....percen át naponta
3. Mennyi időt töltött gyaloglással az elmúlt 7 nap során? Ide sorolható a háztartásban járkálás, az utcai gyalogos közlekedés, a kirándulás, és minden egyéb gyaloglás, amelyik egyfolytában legalább 10 percig tartott.
.....napon az elmúlt héten
.....percen át naponta
4. Naponta mennyi időn át ült az elmúlt 7 napban?
.....órán át naponta
.....percen át naponta

⁴ Cardol M, de Haan RJ, de Jong BA, van den Bos GA, de Groot IJ. Psychometric properties of the Impact on Participation and Autonomy Questionnaire. Arch Phys Med Rehabil. 2001 Feb;82(2):210–6.

Magyar változat:

http://felnottszivhiba.hu/images/yootheme/FVSZA_forum_10_2018/2018_Sio_Eszter_IPAQ_Nemzetkozi_Fizikai_Aktivitas_Kerdoiv.pdf

Értékelés:

1. Teljes inaktivitás, vagy a csekély aktivitás nem éri el a következő kategóriákat.
2. Minimális aktivitás:
 - a. 3 vagy több napon intenzív testmozgás legalább 20 percig; vagy
 - b. 5 vagy több napon mérsékelt aktivitás vagy gyaloglás legalább 30 percen át; vagy
 - c. 5 vagy több napon kombináltan gyaloglás, mérsékelt és intenzív testmozgás legalább 600 MET/perc/hét.
3. Egészségjavító fizikai aktivitás:
 - a. intenzív testmozgás legalább 3 napon összesen legalább 1500 MET/perc/hét; vagy
 - b. 7 napon többféle mozgás kombinációja (gyaloglás, mérsékelt vagy intenzív testmozgás) összesen legalább 3000 MET/perc/hét.

MET-percek heti összege: intenzitás MET-ben×aktivitás percekben×heti gyakoriság,

ahol:

- a gyaloglás 3,3 MET/perc,
- a mérsékelt/közepes intenzitás 4 MET/perc
- intenzív testmozgás 8 MET/perc

P1.:

3 napon, naponta 60 percen át élénk testmozgás	$3 \times 60 \times 8 =$	1440 MET/perc/hét
6 napon naponta 30 percen át mérsékelt testmozgás	$6 \times 30 \times 4 =$	720 MET/perc/hét
7 napon naponta 40 percen át gyaloglás	$7 \times 40 \times 3,3 =$	<u>924 MET/perc/hét</u>
		Összesen=3084 MET/perc/hét

Ez már éppen egészségjavító fizikai aktivitási szintnek felel meg, hiszen 3000

MET/perc/hétnél több a testmozgás, és a hét minden napján történik valamilyen fizikai aktivitás.

Európai lakossági egészségfelmérés, 2019 – Kérdőív – KSH (releváns részek)⁵

B. EGÉSZSÉGI ÁLLAPOT

Most az egészségére vonatkozóan következnek néhány kérdés.

B1. Milyen az Ön egészsége általában?

- 1 - nagyon jó
- 2 - jó
- 3 - kielégítő
- 4 - rossz
- 5 - nagyon rossz
- 8 - NT
- 9 - NK

HS1

B2. Hogyan változott az Ön egészsége az elmúlt 1 évben?

- 1 - javult
- 2 - nem változott
- 3 - romlott
- 8 - NT
- 9 - NK

HSU_1

B3. Véleménye szerint Ön mennyit tehet az egészségéért?

- 1 - nagyon sokat
- 2 - sokat
- 3 - keveset
- 4 - semmit sem
- 8 - NT
- 9 - NK

HSU_2

B4. Van-e Önnek hosszantartó betegsége vagy egészségi problémája? Hosszantartónak tekintünk egy betegséget, egészségi problémát, ha legalább 6 hónapja tart, vagy várhatóan legalább 6 hónapig fog tartani.

- 1 - van
- 2 - nincs
- 8 - NT
- 9 - NK

HS2

B5. Korlátozza-e Önt valamilyen egészségi probléma a mindennapi tevékenységek elvégzésében?

- 1 - igen, súlyosan korlátozza
- 2 - igen, korlátozza, de nem súlyosan
- 3 - nem korlátozza
- 8 - NT
- 9 - NK

TOVABB A B7. KERDESRE

HS3A

B6. Fennáll-e legalább 6 hónapja ez a korlátozás?

- 1 - igen
- 2 - nem
- 8 - NT
- 9 - NK

HS3B

⁵ European Health Interview Survey - Microdata - Eurostat [Internet]. [cited 2023 Feb 16]. Available from: <https://ec.europa.eu/eurostat/web/microdata/european-health-interview-survey>
Magyar változat: https://www.ksh.hu/elef/elef2019_kerdoiv.pdf

A következő lista krónikus, hosszantartó betegségeket, egészségi problémákat tartalmaz.

B10.1. Az elmúlt 12 HÓNAPBAN volt-e Önnek valamilyen betegsége vagy egészségi problémája az alábbiak közül?

Az 5. VÁLASZLAP segítségével válaszoljon!

☑ **Ha igen:** (CD1=1)

B10.2. A betegséget/egészségi problémát orvos állapította meg?

☑ **Ha igen:** (CDU_1=1)

B10.3. Orvos javaslatára szedett gyógyszert erre a betegségre/egészségi problémára?

Vegyen figyelembe minden olyan gyógyszert és táplálékkiegészítőt (gyógynövénykészítményt, vitamint) is, amit orvosi javaslatra szedett, függetlenül attól, hogy receptre vagy recept nélkül kapható!

B10.4. Az elmúlt 12 HÓNAPBAN milyen rendelésen járt ezzel a betegséggel?

- | |
|---|
| 1 - kizárólag állami finanszírozott rendelésen |
| 2 - kizárólag magánrendelésen |
| 3 - állami finanszírozott és magánrendelésen egyaránt |
| 4 - egyiken sem |
| -8 - NT |
| -9 - NK |

- | |
|----------|
| 1 - igen |
| 2 - nem |
| -8 - NT |
| -9 - NK |

	CD1	CDU_1 CDU_2 CDU_3		
		Ha IGEN:		
	B10.1.	B10.2.	B10.3.	B10.4.
	Az elmúlt 12 hónapban volt-e?	Orvos állapította meg?	Orvos javaslatára szedett rá gyógyszert?	Az elmúlt 12 hónapban milyen rendelésen járt ezzel a betegséggel?
01. Allergia, pl.: szénanátha, ekcéma, ételallergia vagy egyéb allergia (kivéve allergiás asztma)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
02. Asztma (allergiás asztma is)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
03. Krónikus hörghurut, bronchitisz, tüdőtagulás	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
04. Szívinfarktus (szívroham) vagy ennek krónikus következményei	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
05. Szívkoszorúér- (koronária-) megbetegedés, angina	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
06. Szabálytalan szívverés, szívritmuszavar, pitvarfibrilláció	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
07. Bármely egyéb szívbetegség	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
08. Magasvérnyomás-betegség (hipertónia)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
09. Magas koleszterinszint, magas vérzsírszint, magas trigliceridszint	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Agyvérzés (szélütés, gutaütés, stroke, agyérgörcs) vagy ennek krónikus következményei	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Arthrosis, ízületi porc kopása (izületi gyulladás kivételével)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Rheumatoid arthritis, idült ízületi gyulladás	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Derék- vagy hátfájás vagy egyéb krónikus derék vagy hátgerinc probléma	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Nyaki fájdalom vagy egyéb krónikus nyaki gerincprobléma	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Csontritkulás	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. Cukorbetegség	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17. Gyomor- vagy nyombélfekély	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

I. BETEGSÉGEK ELLENI VÉDEKEZÉS, SZŰRÉSEK

A következő kérdések az influenza elleni védőoltásra vonatkoznak.

11. Kapott-e Ön valaha influenza elleni védőoltást?

1 - igen

2 - nem

-8 - NT

-9 - NK

TOVÁBB AZ I3. KÉRDÉSRE

PA1M_1

12. Mikor kapott Ön utoljára influenza elleni védőoltást?

Kérjük, adja meg az évet és a hónapot, amikor utoljára védőoltást kapott!

..... év hónap

TOVÁBB AZ I3. KÉRDÉSRE

7777 - nem emlékszem pontosan

-9 - NK

TOVÁBB AZ I3. KÉRDÉSRE

PA1M_2

Ha nem emlékszik pontosan (PA1M_2=7777), ezt kérdezd:

A tavalyi évnél régebben?

1 - igen

2 - nem

-8 - NT

-9 - NK

PA1M_3

A továbbiakban a vérnyomás, a koleszterin- és a vércukorszint ellenőrzésével kapcsolatos kérdések következnek.

13. Mikor mérte meg a vérnyomását utoljára egészségügyi dolgozó?

1 - az elmúlt 12 hónapban

2 - több mint 1 éve, de 3 éven belül

3 - több mint 3 éve, de 5 éven belül

4 - több mint 5 éve

5 - soha nem mérte még

-8 - NT

-9 - NK

PA2

14. Mikor mérte meg koleszterin-/vérzsírszintjét utoljára egészségügyi dolgozó?

1 - az elmúlt 12 hónapban

2 - több mint 1 éve, de 3 éven belül

3 - több mint 3 éve, de 5 éven belül

4 - több mint 5 éve

5 - soha nem mérte még

-8 - NT

-9 - NK

PA3

K. AZ EGÉSZSÉGÜGYI ELLÁTÁSRA VONATKOZÓ VÉLEMÉNY

Most az egészségügyi ellátásra vonatkozó véleményével kapcsolatos kérdések következnek.

Mennyire elégedett általában a következő egészségügyi ellátásokkal Magyarországon?
A 11. VALASZLAP segítségével válaszoljon!

K1. Először kizárólag a társadalombiztosítás által finanszírozott ellátásokat vegye figyelembe, magán egészségügyi ellátásokat ne!

1 - nagyon elégedett	4 - elégedetlen
2 - elégedett	5 - nagyon elégedetlen
3 - nem elégedett, de nem is elégedetlen	6 - nincs közvetlen tapasztalatom
	-8 - NT
	-9 - NK

SAU_1	KÓDOK:
a. Kórházak (beleértve a sürgősségi ellátást)	<input type="checkbox"/>
b. Fogorvos, fogszabályozó szakorvos, más fogászati szakember	<input type="checkbox"/>
c. Járóbeteg-ellátó intézményben dolgozó szakorvos	<input type="checkbox"/>
d. Házi orvos, házi gyermekorvos	<input type="checkbox"/>
e. Védőnő	<input type="checkbox"/>
f. Otthoni ápolás	<input type="checkbox"/>
g. Sürgősségi mentőszolgálat	<input type="checkbox"/>
h. Betegszállítás	<input type="checkbox"/>

K2. Most a magán egészségügyi ellátásokra gondoljon!
Mennyire elégedett általában a következő magán egészségügyi ellátásokkal Magyarországon? A 11. VALASZLAP segítségével válaszoljon!

1 - nagyon elégedett	4 - elégedetlen
2 - elégedett	5 - nagyon elégedetlen
3 - nem elégedett, de nem is elégedetlen	6 - nincs közvetlen tapasztalatom
	-8 - NT
	-9 - NK

SAU_2	KÓDOK:
a. Kórházak	<input type="checkbox"/>
b. Fogorvos, fogszabályozó szakorvos, más fogászati szakember	<input type="checkbox"/>
c. Járóbeteg-ellátó intézményben dolgozó szakorvos	<input type="checkbox"/>
d. Házi orvos, házi gyermekorvos	<input type="checkbox"/>
f. Otthoni ápolás	<input type="checkbox"/>
h. Betegszállítás	<input type="checkbox"/>

M. MAGASSÁG, TESTSÚLY

A következő kérdések az Ön magasságára és testsúlyára vonatkoznak.

M1. Milyen magas Ön cipő nélkül?

..... cm →

-8 - NT

-9 - NK

TOVÁBB AZ M3. KÉRDÉSRE

BM1A

--	--	--

M2. Ha nem tudja pontosan, becsülje meg, hogy milyen magas Ön cipő nélkül?

..... cm

-8 - NT

-9 - NK

50 éves vagy fiatalabb nők esetén: Ha várandós, várandósságát megelőző testsúlyát adja meg!

BM1B

--	--	--

M3. Mekkora a testsúlya ruha és cipő nélkül?

..... kg →

-8 - NT

-9 - NK

TOVÁBB AZ N1. KÉRDÉSRE

BM12A

--	--	--

M4. Ha nem tudja pontosan, becsülje meg, hogy mekkora a testsúlya ruha és cipő nélkül?

..... kg

-8 - NT

-9 - NK

BM12B

--	--	--

O. TÁPLÁLKOZÁSI SZOKÁSOK

A következő kérdések a táplálkozási szokásaira vonatkoznak.

O1. Milyen gyakran fogyaszt Ön friss, fagyasztott, szárított vagy konzerv gyümölcsöt? Kérjük, gyümölcslevek fogyasztását ne vegye figyelembe!

A 15. VÁLASZLAP segítségével válaszoljon!

- 1 - naponta többször
- 2 - naponta egyszer
- 3 - hetente 4-6 alkalommal
- 4 - hetente 1-3 alkalommal
- 5 - ritkábban mint hetente egyszer
- 6 - soha
- 8 - NT
- 9 - NK

TOVÁBB AZ O3. KÉRDÉSRE

DH1

O2. Hány adag gyümölcsöt eszik naponta? Kérjük, gyümölcslevek fogyasztását ne vegye figyelembe!

A 16. VÁLASZLAP segítségével válaszoljon!

..... adagot

- 8 - NT
- 9 - NK

DH2

O3. Milyen gyakran fogyaszt Ön friss, fagyasztott, szárított vagy konzerv zöldséget vagy salátát (burgonyát nem számítva)? Kérjük, zöldséglevek, hideg vagy meleg levesek fogyasztását ne vegye figyelembe!

A 15. VÁLASZLAP segítségével válaszoljon!

- 1 - naponta többször
- 2 - naponta egyszer
- 3 - hetente 4-6 alkalommal
- 4 - hetente 1-3 alkalommal
- 5 - ritkábban mint hetente egyszer
- 6 - soha
- 8 - NT
- 9 - NK

TOVÁBB AZ O5. KÉRDÉSRE

DH3

O4. Hány adag zöldséget, salátát eszik naponta (burgonyát nem számítva)? Kérjük, zöldséglevek, hideg vagy meleg levesek fogyasztását ne vegye figyelembe!

A 16. VÁLASZLAP segítségével válaszoljon!

..... adagot

- 8 - NT
- 9 - NK

DH4

O20. Követ-e tudatosan valamilyen étrendet? Legfeljebb három étrendet jelölhet meg!
A 23. VÁLASZLAP segítségével válaszoljon!

- 1 - diabetikus (cukorbeteg) étrend
 - 2 - gluténmentes étrend
 - 3 - laktózmentes étrend
 - 4 - tejfehérjementes étrend
 - 5 - energiaszegény étrend
 - 6 - sószegény étrend
 - 7 - vegetáriánus (vegán, lakto-, ovo-, szemi-vegán stb.) étrend
 - 8 - egyéb étrend
 - 9 - semmilyen étrendet sem követ tudatosan
-
- 8 - NT
-9 - NK

DHU_1A

DHU_1B

DHU_1C

Kérjük, most gondoljon vissza azokra a kérdésekre, amiket az elmúlt percekben megválaszolt a táplálkozási szokásokkal kapcsolatban!

O21. Összességében mennyire érzi biztosnak magát a táplálkozási szokásokkal kapcsolatban adott válaszaiban? Kérjük, egy 0-tól 10-ig terjedő skálán értékelje, ahol a 0 azt jelenti, hogy egyáltalán nem érzi biztosnak, a 10 pedig azt, hogy teljesen biztosnak érzi magát a válaszaiban.

A 28. VÁLASZLAP segítségével válaszoljon!

egyáltalán nem érzi biztosnak (0) 0 1 2 3 4 5 6 7 8 9 10 teljesen biztosnak érzi (10)

- 8 - NT
-9 - NK

TESZT_TAPL1

O22. Milyen volt megválaszolni a táplálkozási szokásokkal kapcsolatos kérdéseket? Kérjük, jelölje meg az **összes** válaszlehetőséget, amiket érvényesnek talál!
A 29. VÁLASZLAP segítségével válaszoljon!

- 01 - könnyen megválaszolhatók voltak
 - 02 - nehezen voltak megválaszolhatók
 - 03 - kényelmetlen, érzékeny vagy zavaró témákat is érintettek
 - 04 - jó érzés volt ezekre a kérdésekre válaszolni
 - 05 - nehezen érthető volt a kérdések megfogalmazása
 - 06 - jól, érthetően voltak megfogalmazva a kérdések
 - 07 - nem kellett sokat gondolkodni ahhoz, hogy válaszoljak
 - 08 - túl sokat kellett gondolkodni ahhoz, hogy válaszoljak
 - 09 - általában volt rám érvényes válaszlehetőség
 - 10 - gyakran nem volt rám érvényes válaszlehetőség
 - 11 - túl sok kérdés volt ebben a témában
 - 12 - hiányoltam témaköröket
- (ha a 12-es választ jelöli: Ha kívánja, ossza meg velünk, mikre gondol:)

- 8 - NT
-9 - NK

TESZT_TAPL2_SZ

TESZT_TAPL2

Az eredmények helyes felépítését demonstrálja az alábbi cikk (leíró, egyszerű, többszörös elemzés).⁶

BMJ Open Factors associated with anxiety and depression among type 2 diabetes outpatients in Malaysia: a descriptive cross-sectional single-centre study

Kurubaran Ganasegeran,¹ Pukunan Renganathan,² Rizal Abdul Manaf,³ Sami Abdo Radman Al-Dubai⁴

To cite: Ganasegeran K, Renganathan P, Manaf RA, et al. Factors associated with anxiety and depression among type 2 diabetes outpatients in Malaysia: a descriptive cross-sectional single-centre study. *BMJ Open* 2014;4:e004794. doi:10.1136/bmjopen-2014-004794

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2014-004794>).

Received 3 January 2014
Revised 3 April 2014
Accepted 7 April 2014



For numbered affiliations see end of article.

Correspondence to
Dr Kurubaran Ganasegeran;
medkuru@yahoo.com

ABSTRACT

Objective: To determine the prevalence and factors associated with anxiety and depression among type 2 diabetes outpatients in Malaysia.

Design: Descriptive, cross-sectional single-centre study with universal sampling of all patients with type 2 diabetes.

Setting: Endocrinology clinic of medical outpatient department in a Malaysian public hospital.

Participants: All 169 patients with type 2 diabetes (men, n=99; women, n=70) aged between 18 and 90 years who acquired follow-up treatment from the endocrinology clinic in the month of September 2013.

Main outcome measures: The validated Hospital Anxiety and Depression Scale (HADS), sociodemographic characteristics and clinical health information from patient records.

Results: Of the total 169 patients surveyed, anxiety and depression were found in 53 (31.4%) and 68 (40.3%), respectively. In multivariate analysis, age, ethnicity and ischaemic heart disease were significantly associated with anxiety, while age, ethnicity and monthly household income were significantly associated with depression.

Conclusions: Sociodemographics and clinical health factors were important correlates of anxiety and depression among patients with diabetes. Integrated psychological and medical care to boost self-determination and confidence in the management of diabetes would catalyse optimal health outcomes among patients with diabetes.

Strengths and limitations of this study

- Malaysia suffers the highest rate of diabetes in the Asian region. People with diabetes are twice more likely to develop anxiety and depression, causing poor health outcomes and increased mortality.
- This study aimed to assess the prevalence and factors associated with anxiety and depression among type 2 diabetes outpatients in Malaysia.
- Integrated psychological and medical care to boost self-determination and confidence in the management of diabetes would catalyse optimal health outcomes among patients with diabetes.
- The absence of a control group and a relatively small sample size from one hospital might limit the generalisability of the study findings. The cross-sectional design of the study limits our ability to make causal inferences.

country suffers the highest rate of diabetes in the Asian region, with prevalence rates rising from 14.9% in 2006 to 20.8% in 2011.⁴

The complex mechanism to cope with chronic diseases requires self-determination to overcome the emotional shock of the diagnoses and proper assimilation of information regarding self-care to prevent disease complications.⁵ The collapse of these coping strategies over time due to low psychological, emotional and social support renders significant comorbid anxiety and depression, exacerbating disease complications and poor prognosis.⁵ People with diabetes were twice at risks to suffer from premorbid anxiety and depression as the general population.^{2 6} The coexistence of anxiety and depression in people with diabetes catalyses serious disease comorbidities, complications, poor quality of life and escalated healthcare expenditures.⁷

Anxious and depressed people with diabetes are less likely to comply with diabetes

INTRODUCTION

Type 2 diabetes is a chronic metabolic disorder characterised by hyperglycaemia due to insulin deficiency.¹ The global prevalence of diabetes is currently estimated to be 285 million and projection rates are expected to rise to over 438 million by the year 2030,² with Asia suffering the bulk of the total diabetes epidemic.³ The Malaysian scenario is more debilitating when figures confirmed that the

⁶ Ganasegeran K, Renganathan P, Manaf RA, Al-Dubai SAR. Factors associated with anxiety and depression among type 2 diabetes outpatients in Malaysia: a descriptive cross-sectional single-centre study. *BMJ Open*. 2014 Apr 23;4(4):e004794.



self-care recommendations.⁶ The diagnosis of diabetes is a life-threatening stressor that demands high mental and physical accommodations due to elevated feelings of fear.⁸ Depression among people with diabetes adds an increased burden to patient adherence, compliance and poor prognosis for quality health outcomes.⁹ Depression in the diabetes population has been associated with potential sociodemographic and clinical factors.⁷ Ageing,² ethnicity,⁸ socioeconomic status,¹⁰ education level¹¹ and unemployment¹² were important correlates for depression among people with diabetes.

Common diabetes vascular complications like ischaemic heart disease (IHD), cerebrovascular accidents (CVAs) and diabetic nephropathy had caused significant rates of mortality and poor quality of life.^{2–11} Malaysia topped the world in diabetic nephropathy, with almost 15 000 patients requiring dialysis and 2000 acquiring kidney transplants.¹³ Diabetes-related complications and associated comorbidities have been proven to amplify psychiatric conditions.²

Numerous studies from developed and developing countries assessed factors affecting anxiety and depression among people with diabetes.^{2–6–14} Irish and Mexican studies concluded that the prevalence of anxiety and depression was considerably higher among people with diabetes in comparison to the general population.^{6–9} A Malaysian study recommended that early psychiatric screening was required owing to elevated risks for anxiety and depression among people with diabetes.⁸ This study aimed to determine the prevalence and factors associated with anxiety and depression among outpatients with diabetes in a Malaysian public hospital.

METHODS

Study setting and population

This cross-sectional single-centre study was conducted in the month of September 2013 among all 169 patients with type 2 diabetes aged between 18 and 90 years who acquired follow-up treatment from the Endocrinology Clinic at the Medical Outpatient Department of Tengku Ampuan Rahimah Hospital (ITAR), Selangor, Malaysia. Objectives and benefits of the study were explained in verbal and written form attached to the questionnaires. Patients were assured that their participation was confidential and would not affect their medical treatment outcomes. A written consent was obtained from those who agreed to participate. Patients with type 1 and gestational diabetes were excluded from the study.

Ethical issues

This study complied with the guidelines convened in the Declaration of Helsinki. The study was conducted as part of a larger study that explored anxiety and depression among outpatients in Malaysia.

Study instruments

A self-administered questionnaire consisting of three parts was used in this study:

The first part included items on sociodemographics (gender, age, ethnicity, marital status, education level, residence, monthly household income and employment status).

The second part assessed anxiety and depression among patients with diabetes. Anxiety is defined as subjective experience of fear and its physical manifestations while depression is defined as anhedonia (reduced positive affect).¹⁵ To explore anxiety and depression among patients with diabetes, we used the Hospital Anxiety and Depression Scale (HADS), originally developed by Zigmond and Snaith,¹⁶ and validated among the Malaysian population.¹⁷ This widely used self-assessment tool measures the level of emotional distress (anxiety and depression) in various clinical settings, including diabetes population.^{2–6–18} HADS is comprised of 14 items, 7 of which measures anxiety (HADS-A) and another 7 measures depression (HADS-D). These items are scored on a four-point Likert scale ranging from 0 (not present) to 3 (considerable). Item scores were summed to provide subscaled scores of anxiety and depression, ranged between 0 and 21, and total summed score ranged from 0 to 42. A higher score represents higher anxiety or depression.¹⁸ The scores are categorised as follows: normal (0–7) and caseness which includes mild distress (8–10), moderate distress (11–14) and severe distress (15–21).¹⁸ The questionnaire was administered in English.

The third part included clinical health information of the patients derived from medical records.

Baseline data definitions

Type 2 diabetes

The presence of diabetes diagnosed by a physician with a fasting plasma glucose value of 7 mmol/L (126 mg/dL) or higher,¹⁹ and patients currently being administered with oral hypoglycaemic drugs or insulin therapy as documented in medical records were included in this study.

Diabetes vascular complications

Vascular complications in diabetes were considered when patients were diagnosed with CVA, IHD or nephropathies. Patients diagnosed with either one vascular complication over the past year were included in this study. CVA was defined as hemiparesis cases proven by medical and CT scan records.²⁰ IHD was considered to exist with a history of angina or acute coronary syndromes elicited among patients with diabetes and documented in medical records.² Nephropathy is defined by proteinuria >500 mg in 24 h among patients with diabetes from medical records.¹

Diabetes comorbid conditions

Patients with diabetes were classified as hypertensive if they were previously diagnosed and were currently on antihypertensive medications² as documented in medical records. Dyslipidaemia was defined as high plasma triglyceride concentration, low high-density

lipoprotein cholesterol concentration and increased concentration of low-density lipoprotein cholesterol²¹ with patients currently on statin medications as documented in medical records.

The Malaysian healthcare system

Public healthcare providers across the nation are mainly entrusted by the Ministry of Health Malaysia with the commitment of 'healthcare access to all'.⁴ The public healthcare is highly subsidised through general revenue and taxation collected by the federal government, and with a minimal registration fee of US\$0.33 or MYR1. Malaysians are granted free access to clinical consultations, treatment and medications both as outpatients or inpatients in all public health facilities within the country.⁴ ITTAR is the second busiest public health facility in terms of patient admissions and outpatient services in Malaysia at the time of this study.⁴

Statistical analysis

Analysis was performed using Statistical Package for Social Sciences (SPSS) (V.16.0, IBM, Armonk, New York, USA). Descriptive analysis was performed for all variables in this study. To check for the validity of the HADS among Malaysian population, an exploratory factor analysis was performed using principal component method with varimax rotation and Cronbach's α was used to test the internal consistency of the scale. Anxiety and depression scores were expressed as mean and SDs. Test of normality was performed for total anxiety and depression subscale scores. T test and analysis of variance (ANOVA) test were applied to compare anxiety and depression across sociodemographic and clinical health variables. In case of ANOVA, post hoc test was used to determine where the significant difference was. Multiple linear regression analysis using 'Enter' technique was employed to obtain significant factors associated with anxiety and depression among patients with diabetes. The accepted level of significance was set below 0.05 ($p < 0.05$). Multicollinearity was checked between independent variables.

RESULTS

Sociodemographic characteristics and clinical health information of the respondents

One hundred sixty-nine patients were included in this study. Of the total, 99 (58.6%) were men and 70 (41.4%) were women. The mean (\pm SD) age of the patients was 36.9 (\pm 15.9) years and the majority aged less than 50 years, 137 (81.1%; table 1).

Baseline clinical data of the patients are summarised in table 2. Of the total patients, 53 (31.4%) were diagnosed for diabetes vascular complications. Twelve patients (7.1%) were diagnosed for CVA, 24 (14.2%) were diagnosed for IHD and 17 (10.1%) developed nephropathy. Forty-four (26.0%) patients developed at least one comorbid condition, while 21 (12.4%) had two

Table 1 Sociodemographic characteristics of the respondents (n=169)

Characteristics	N	Percentage
Gender		
Male	99	58.6
Female	70	41.4
Age (years)		
<50	137	81.1
\geq 50	32	18.9
Ethnicity		
Malay	53	31.3
Chinese	88	52.1
Indian	28	16.6
Marital status		
Single	63	37.3
Married	106	62.7
Highest education level		
High school	75	44.4
Tertiary education	94	55.6
Residence		
Urban	132	78.1
Rural	37	21.9
Monthly household income (MYR)		
<3000	56	33.1
\geq 3000	113	66.9
Current employment status		
Employed	119	70.4
Unemployed	50	29.6

MYR1 is equivalent to US\$0.33 at the time of study.

comorbid conditions. Cronbach's α coefficient for HADS-A subscale was 0.83, while Cronbach's α coefficient for HADS-D subscale was 0.71. Mild anxiety and depression were found in 33 (19.5%) and 49 (29.0%) of the patients, respectively. Moderate anxiety and depression were found in 16 (9.5%) patients respectively. Severe anxiety and depressive symptoms were detected in four (2.4%) and three (1.8%) of the patients, respectively.

Association between anxiety and depression and sociodemographics of the respondents

Table 3 shows the association between anxiety and depression with sociodemographic characteristics. Patients aged 50 years or older had higher anxiety score (9.1 ± 4.6) compared with those aged less than 50 years (6.4 ± 2.7 , $p < 0.001$). Significant associations were observed between anxiety and ethnicity ($p < 0.001$); post hoc tests showed that Indians exhibited higher anxiety score (8.4 ± 4.2) in comparison to Chinese (6.6 ± 2.4 , $p = 0.044$). Patients graduated from high school exhibited higher anxiety score (7.5 ± 4.0) in comparison to those with a tertiary degree (6.4 ± 2.5 , $p = 0.037$). In addition, patients aged 50 years or older were more depressed (9.2 ± 4.0) in comparison to those aged less than 50 years (6.3 ± 2.9 , $p < 0.001$). Significant associations were observed between depression and ethnicity ($p < 0.001$);



Table 2 Clinical health information of the respondents (n=169)

Characteristics	N	Percentage
<i>Diabetes vascular complications</i>		
Cerebrovascular accident		
Yes	12	7.1
No	157	92.9
Ischaemic heart disease		
Yes	24	14.2
No	145	85.8
Diabetic nephropathy		
Yes	17	10.1
No	152	89.9
Comorbidities		
Diabetes alone	104	61.6
Diabetes with hypertension or dyslipidaemia	44	26.0
Diabetes with hypertension and dyslipidaemia	21	12.4
Anxiety		
Normal	116	68.6
Mild	33	19.5
Moderate	16	9.5
Severe	4	2.4
Depression		
Normal	101	59.7
Mild	49	29.0
Moderate	16	9.5
Severe	3	1.8

post hoc tests revealed that Indians exhibited higher depression (9.8 ± 3.5) in comparison to Malays (6.9 ± 3.1) and Chinese (5.9 ± 2.9 , $p < 0.001$, respectively). Similarly, patients who graduated from high school exhibited greater depression (7.7 ± 3.7) in comparison to tertiary graduates (6.2 ± 2.9 , $p = 0.006$). Patients with a monthly household income of less than MYR3000 have higher depression score (8.7 ± 3.6) compared to those with higher income (6.0 ± 2.8 , $p < 0.001$). Similarly, unemployed patients portrayed higher depression score (7.9 ± 3.2) in comparison to those employed (6.4 ± 3.3 , $p = 0.007$).

Association between anxiety and depression and clinical health information of the respondents

Patients diagnosed for IHD exhibited higher anxiety score (8.7 ± 4.2) in comparison to those without such complication (6.6 ± 3.1 , $p = 0.004$). In addition, significant associations were observed between depression and disease comorbidities ($p = 0.010$); post hoc tests showed that patients with associated hypertension or dyslipidaemia had higher depression score (7.5 ± 3.2) in comparison to those without comorbid conditions (6.3 ± 3.4 , $p = 0.009$; table 4).

Factors associated with anxiety among patients with diabetes by multiple linear regression

Table 5 exhibits the factors associated with anxiety among patients with diabetes. Patients aged ≥ 50 years had on

Table 3 Association between anxiety and depression with sociodemographic characteristics of the respondents (n=169)

Characteristics	Anxiety		Depression	
	Mean (SD)	p Value	Mean (SD)	p Value
Gender				
Male	7.0 (3.5)		7.1 (3.5)	
Female	6.8 (3.0)	0.737	6.6 (3.1)	0.345
Age (years)				
<50	6.4 (2.7)		6.3 (2.9)	
≥ 50	9.1 (4.6)	<0.001	9.2 (4.0)	<0.001
Ethnicity				
Malay	6.5 (3.8)		6.9 (3.1)	
Chinese	6.6 (2.4)		5.9 (2.9)	
Indian	8.4 (4.2)	0.035	9.8 (3.5)	<0.001
Marital status				
Single	6.7 (2.7)		6.9 (2.9)	
Married	7.0 (3.6)	0.601	6.8 (3.6)	0.894
Highest education level				
High school	7.5 (4.0)		7.7 (3.7)	
Tertiary	6.4 (2.5)	0.037	6.2 (2.9)	0.006
Residence				
Urban	6.8 (2.9)		6.7 (3.1)	
Rural	7.2 (4.5)	0.569	7.6 (3.9)	0.125
Monthly household income (MYR)				
<3000	7.5 (4.4)		8.7 (3.6)	
≥ 3000	6.6 (2.6)	0.090	6.0 (2.8)	<0.001
Current employment status				
Employed	6.6 (3.2)		6.4 (3.3)	
Unemployed	7.6 (3.4)	0.078	7.9 (3.2)	0.007

the average 2.3 (95% CI 0.9 to 3.6) higher anxiety score in comparison to those aged less than 50 years ($p = 0.001$). Indians had on an average 1.7 (95% CI 0.3 to 3.2) higher anxiety score compared with Malays ($p = 0.018$). Patients diagnosed with IHD had on an average 1.5 (95% CI 0.1 to 2.9) higher anxiety score in comparison to those without such a condition ($p = 0.039$).

Factors associated with depression among patients with diabetes by multiple linear regression

Table 6 shows the factors associated with depression among patients with diabetes. Patients aged ≥ 50 years had on the average 1.4 (95% CI 0.2 to 2.7) higher depression score in comparison to those aged less than 50 years ($p = 0.027$). Indians had on the average 2.7 (95% CI 1.4 to 4.0) higher depression score compared with Chinese ($p < 0.001$). Patients with a monthly household income of less than MYR3000 had on an average 1.9 (95% CI 0.8 to 3.0) higher depression score compared to those with a higher income ($p = 0.001$).

DISCUSSION

This study aimed to determine the prevalence and factors associated with anxiety and depression among

Table 4 Association between anxiety and depression with clinical health information of the respondents (n=169)

Characteristics	Anxiety		Depression	
	Mean (SD)	p Value	Mean (SD)	p Value
<i>Diabetes vascular complications</i>				
Cerebrovascular accident				
Yes	6.6 (4.1)		6.7 (4.8)	
No	6.9 (3.3)	0.742	6.9 (3.2)	0.823
Ischaemic heart disease				
Yes	8.7 (4.2)		7.8 (4.1)	
No	6.6 (3.1)	0.004	6.7 (3.2)	0.131
Diabetic nephropathy				
Yes	7.4 (2.2)		6.4 (3.0)	
No	6.8 (3.4)	0.492	6.9 (3.4)	0.548
<i>Comorbidities</i>				
Diabetes alone	6.6 (3.3)		6.3 (3.4)	
Diabetes with hypertension or dyslipidaemia	7.2 (3.4)		7.5 (3.2)	
Diabetes with hypertension and dyslipidaemia	7.7 (3.0)	0.289	8.4 (2.9)	0.010

diabetes outpatients in Malaysia. Of the 169 patients with diabetes surveyed, 31.4% perceived anxious states while 40.3% exhibited depressive symptoms. The estimated rate of anxiety reported in this study was similar to an Irish sample (32%),⁶ but relatively lower than that found in Mexican (52.9%)⁹ and Pakistani (57.9%) participants.² In contrary, self-reported depression rates reported in this study were similar than that found in Mexican (47.7%)⁹ and Pakistani (43.5%) samples,² but comparatively higher than that found in Irish participants (22.4%).⁶ In the final model, age, ethnicity and history of IHD were significantly associated with anxiety, while factors significantly associated with depression were age, ethnicity and monthly household income.

Ageing appears to accelerate diabetes vascular complications and hyperglycaemic crisis, causing poor functional status and high mortality rates.²² Dysregulation of the hypothalamic-pituitary-adrenal axis and overactivation of the sympathetic nervous system due to fear of hypoglycaemia, complications or mortality are immediate physiological processes that prompt higher anxiety states

among older population.⁵ This study found a significantly higher anxious state among older patients compared with younger ones. Collins *et al*⁶ reported otherwise.

The development of vascular complications is a predictive factor for psychological morbidity among people with diabetes.²³ This study found a significantly higher anxiety level among patients with IHD. Khuwaja *et al*² reported similar associations.

The increased susceptibility to various diseases, disabilities and social isolation among older population causes serious psychological repercussions.²⁴ This study found a significantly higher depression score among older patients in comparison to younger ones. Similar findings were found among patients with diabetes in other countries.^{2, 25}

Latest statistics revealed by the Ministry of Health Malaysia reported that the prevalence of diabetes was the highest among Indian ethnic (24.9%), followed by Malay ethnic (17%) and Chinese ethnic (13.9%).⁴ Minority ethnic groups have been known to experience higher anxiety and depression rates.^{26, 27} This study found a

Table 5 Factors associated with anxiety among patients with diabetes by multiple linear regression (n=169)

Predictors	B	SE	β	p Value	95% CI	
					Lower	Upper
<i>Age (years)</i>						
<50	Ref	Ref	Ref	Ref	Ref	Ref
≥50	2.3	0.7	0.3	0.001	0.9	3.6
<i>Ethnicity</i>						
Malay	Ref	Ref	Ref	Ref	Ref	Ref
Chinese	0.7	0.6	0.1	0.194	-0.4	1.8
Indian	1.7	0.7	0.2	0.018	0.3	3.2
<i>Highest education level</i>						
High school	0.1	0.5	0.0	0.871	-1.0	1.1
Tertiary educated	Ref	Ref	Ref	Ref	Ref	Ref
<i>Having ischaemic heart disease</i>						
Yes	1.5	0.7	0.2	0.039	0.1	2.9
No	Ref	Ref	Ref	Ref	Ref	Ref


Table 6 Factors associated with depression among patients with diabetes by multiple linear regression (n=169)

Predictors	B	SE	β	p Value	95% CI	
					Lower	Upper
Age (years)						
<50	Ref	Ref	Ref	Ref	Ref	Ref
≥ 50	1.4	0.6	0.2	0.027	0.2	2.7
Ethnicity						
Malay	0.4	0.5	0.1	0.458	-0.7	1.4
Indian	2.7	0.7	0.3	<0.001	1.4	4.0
Chinese	Ref	Ref	Ref	Ref	Ref	Ref
Highest education level						
High school	-0.3	0.5	-0.1	0.548	-1.4	0.7
Tertiary educated	Ref	Ref	Ref	Ref	Ref	Ref
Monthly household income (MYR)						
<3000	1.9	0.6	0.3	0.001	0.8	3.0
≥ 3000	Ref	Ref	Ref	Ref	Ref	Ref
Current employment status						
Employed	Ref	Ref	Ref	Ref	Ref	Ref
Unemployed	-1.7	1.5	-0.2	0.265	-4.7	1.3
Comorbidities						
Diabetes alone	Ref	Ref	Ref	Ref	Ref	Ref
Diabetes with hypertension or dyslipidaemia	-2.6	1.5	-0.4	0.080	-5.5	0.3
Diabetes with hypertension and dyslipidaemia	-2.3	1.7	-0.2	0.189	-5.7	1.1

significantly higher anxiety and depression level among Indian patients in comparison to other ethnicities. A recent Malaysian study which reported similar associations postulated that minority ethnic Indians experienced extensive psychological comorbidities due to triadic stressors of socioeconomic constraints, poor education level and perceived discrimination.⁸

Higher depression states in unemployment is caused by reduced sociological functions such as status identity, social contacts, participation in collective purposes and regular activities.¹² This study found a significantly higher depression status among unemployed patients in comparison to those being employed. Kaur *et al*⁸ reported similar consistencies. In addition, this study found a significantly higher depression level in lower income patients. Similar findings were reported in a Malaysian study.⁸ Reduced confidence due to economic instability and increased healthcare expenditures for routine diabetes screening complications, comorbid conditions and adherence to treatment pose substantial psychological illness among people with diabetes.⁹

Diabetes comorbid conditions like hypertension and dyslipidaemia has been known to amplify disease complications and poor treatment outcomes.²¹⁻²⁸ Increased rates of depression have been found in diabetes patients with hypertension.²⁸ An exponential rise of mortality rates due to serious cardiovascular disease complications in dyslipidaemia would contribute to high depression rates among patients with diabetes due to reduced quality of life and poor prognosis.⁷⁻²¹ This study found a significantly higher depression score among patients with diabetes and hypertension or dyslipidaemia. Khuwaja *et al*²⁷ found similar findings.

Higher education attainment has been linked to be a protective factor against anxiety and depression among people with diabetes.^{6-11,29} Education drives individuals towards proper understanding of disease mechanisms and complications, prompting increased compliance and adherence towards disease treatment for better health outcomes. Our study found a significantly lower anxiety and depression level among tertiary educated patients in comparison to high school graduates.

LIMITATIONS

The absence of a control group and a small sample size from a single hospital might limit the generalisability of the study findings. In addition, the heterogeneity of the sample in this study caused by the wide range of age affects the prevalence rates and may limit the exploration of anxiety and depression in the youngest age groups. The cross-sectional design of the study limits our ability to make causal inferences. Further research is needed to address these limitations.

CONCLUSION

Sociodemographics and clinical factors were important correlates of anxiety and depression among patients with diabetes. Age, ethnicity and IID were significantly associated with anxiety. Factors significantly associated with depression were age, ethnicity and monthly household income.

RECOMMENDATIONS

Early recognition of vulnerable factors associated with anxiety and depression in people with diabetes is

necessary to promote patient adherence and compliance to diabetes control. Collaborative teamwork between healthcare providers and patients through a compassionate and holistic approach in recognising early neurotic features is essential to prevent disease comorbidities and mortalities. Rejuvenating primary healthcare policies from an essentially 'reactive-based system' (responding only when individuals are sick) to a 'proactive-based system' (focus on overall mental and physical health well-being) needs to be drafted immediately and amalgamated in all public health facilities within Malaysia. Increasing patient awareness to boost self-determination and confidence through integrated psychological and medical care in the management of diabetes would catalyse optimal health outcomes, as mused Osler (1961):

Care more for the individual patient than for the special features of the disease...Put yourself in his place...The kindly word, the cheerful greeting, the sympathetic look—these the patient understands. Sir William Osler (Aphorisms from his bedside teachings and writings, 1961)

Author affiliations

¹International Medical School, Management and Science University (MSU), Shah Alam, Selangor, Malaysia

²Clinical Research Center, Tengku Ampuan Rahimah Hospital (HTAR), Klang, Selangor, Malaysia

³Community Health Department, Faculty of Medicine, Universiti Kebangsaan Malaysia (UKM), Kuala Lumpur, Malaysia

⁴Department of Community Medicine, International Medical University (IMU), Kuala Lumpur, Malaysia

Contributors KG had the original idea, designed the study, drafted the first version of the manuscript and is responsible for the final version. PR and RAM assisted with the study design, data collection, the literature research and editing of the manuscript. KG and SARA-D contributed to the study coordination, data analysis and preparation of the manuscript. SARA-D revised the final draft critically for important intellectual content. All authors have contributed to and approved the final version of the manuscript.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None.

Patient consent Obtained.

Ethics approval Study protocol was approved by the Medical Research Ethics Committee (MREC), Ministry of Health Malaysia (NMRR-13-643-14711).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

- Fowler MJ. Microvascular and macrovascular complications of diabetes. *Clin Diabetes* 2008;26:77–82.
- Khuwaja AK, Lalani S, Dhanani R, et al. Anxiety and depression among outpatients with type 2 diabetes: a multi-centre study of prevalence and associated factors. *Diabetol Metab Syndr* 2010;2:72.
- International Diabetes Federation. *Diabetes Atlas*. Vol 4. 2010. <http://www.worlddiabetesfoundation.org/composite-35.htm> (accessed 23 Dec 2013).
- Statistics Malaysia and health facts. *Ministry of Health Malaysia*. 2014. <http://www.moh.gov.my> (accessed 31 Mar 2014).
- Gonzalez JS, Sabrina A, Havah E, et al. Psychological issues in adults with type 2 diabetes. In: Pagoto S, eds. *Psychological comorbidities of physical illness: a behavioral medicine perspective, Chapter II*. Springer Science Business Media LLC, 2011:73–121. doi:10.1007/978-1-4419-0029-6_2
- Collins MM, Corcoran P, Perry LJ. Anxiety and depression symptoms in patients with diabetes. *Diabet Med* 2009;26:153–61.
- Engum A, Mykletun A, Midtjell K, et al. Depression and diabetes—a large population-based study of sociodemographic, lifestyle, and clinical factors associated with depression in type 1 and type 2 diabetes. *Diabetes Care* 2005;28:1904–9.
- Kaur G, Tee GH, Aniaratnam S, et al. Depression, anxiety and stress symptoms among diabetics in Malaysia: a cross sectional study in an urban primary care setting. *BMC Fam Pract* 2013;14:69.
- Tovilla-Zarate C, Juarez-Rojop I, Jimenez YP, et al. Prevalence of anxiety and depression among outpatients with type 2 diabetes in the Mexican population. *PLoS ONE* 2012;7:e36887.
- Everson SA, Maty SC, Lynch JW, et al. Epidemiologic evidence for the relation between socioeconomic status and depression, obesity, and diabetes. *J Psychosom Res* 2002;53:891–5.
- Peyrot M, Rubin R. Levels and risks of depression and anxiety symptomatology among diabetic adults. *Diabetes Care* 1997;20:585–90.
- Palizgir M, Bakhtari M, Esteghamati A. Association of depression and anxiety with diabetes mellitus type 2 concerning some sociological factors. *Iran Red Crescent Med J* 2013;15:644–8.
- National Renal Registry Malaysia. *Clinical Research Center Ministry of Health Malaysia*. 2006. <http://www.crc.gov.my> (accessed 15 Nov 2013).
- Katon W, Unutzer J, Russo J. Major depression: the importance of clinical characteristics and treatment response to prognosis. *Depress Anxiety* 2010;27:19–26.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)*. Vol 4. American Psychiatric Association, 2000.
- Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361–70.
- Fatt QK, Atiya AS, Heng NCC, et al. Validation of the Hospital Anxiety and Depression Scale and the psychological disorder among premature ejaculation subjects. *Int J Impot Res* 2007;19:321–5.
- Whelan-Goodinson R, Ponsford J. Validity of the Hospital Anxiety and Depression Scale to assess depression and anxiety following traumatic brain injury as compared with the Structured Clinical Interview for DSM-IV. *J Affect Disord* 2009;114:94–102.
- American Diabetes Association. Standards of medical care in diabetes. *Diabetes Care* 2006;29:4–12.
- Vaz NC, Ferreira AM, Kulkarni MS, et al. Prevalence of diabetic complications in rural Goa, India. *Indian J Community Med* 2011;36:283–6.
- Mooradian AD. Dyslipidemia in type 2 diabetes mellitus. *Nat Clin Pract Endocrinol Metab* 2009;5:150–9.
- Morley JE. The elderly type 2 diabetic patient: special considerations. *Diabet Med* 1998;15:41–6.
- Almawi W, Tamim H, Al-Sayed N, et al. Association of comorbid depression, anxiety, and stress disorders with type 2 diabetes in Bahrain, a country with a very high prevalence of type 2 diabetes. *J Endocrinol Invest* 2008;31:1020–4.
- Ganatra HA, Zafar SN, Qidwai W, et al. Prevalence and predictors of depression among an elderly population of Pakistan. *Aging Ment Health* 2008;12:349–56.
- Mosaku K, Kolawole B, Mume C, et al. Depression, anxiety and quality of life among diabetic patients: a comparative study. *J Natl Med Assoc* 2008;100:73–8.
- Dunlop DD, Song J, Lyons JS, et al. Racial or ethnic differences in rates of depression among pre-retirement adults. *Am J Public Health* 2003;93:1945–52.
- Fisher L, Laurencin G, Chesla CA, et al. Depressive affect among four ethnic groups of male patients with type 2 diabetes. *Diabetes Spectr* 2004;17:215–19.
- Thomas J, Jones G, Scarinci I, et al. A descriptive and comparative study of the prevalence of depressive and anxiety disorders in low-income adults with type 2 diabetes and other chronic illnesses. *Diabetes Care* 2003;26:2311–17.
- Bener A, Al-Hamaq AO, Dafeeah EE. High prevalence of depression, anxiety and stress symptoms among diabetes mellitus patients. *Open Psychiatry J* 2011;5:5–12.