

Navodila za urejanje podatkov

Prosimo vas, da podatke zbrane v Excelu pred statistično analizo uredite na naslednji način:

- Če je le možno naj bodo vsi podatki za analizo zbrani v enem zavihku (*sheet*). To naj bo prvi zavihek *.xlsx* dokumenta. V nadaljevanju bomo ta zavihek imenovali **osnovni**.
- Osnovni zavihek naj vsebuje zgolj in samo podatke – nobenih pojasnil, dopisov ob ali pod tabelo. To tabelo bomo v nadaljevanju imenovali **osnovna tabela**.
- Osnovna tabela naj bo formatirana kot tabela. V ta namen si lahko pomagate s funkcijo **Format as table oz. Oblikuj kot tabelo** (izberete lahko katerokoli ponujeno barvno shemo). Ta funkcija tabelo uredi tako, da ji doda *header*-je (imena stolpcev). Vsak stolpec pa do neke mere tudi uredi. Če so npr. vsa polja razen enega v nekem stolpcu poravnana desno, je zelo verjetno, da je v tistem polju nekaj čudnega (npr. decimalna pika namesto vejica ali obratno). Poleg tega nudi še druge koristne funkcije – ob kliku na puščico lahko v odprtem okenčku izberete razvrščanje (*sort*), prikazane so vrednosti tega stolpca, glede na katere lahko tabelo tudi filtrirate.
- Osnovna tabela naj bo ustrezno formatirana. Pazite na **decimalne pike, vejice** ipd. – uporabite zgolj in samo to kar ustreza nastavitvam vašega sistema (to preverite tako, da v prazen zavihek v polji *A1* in *B1* vnesete dve decimalni števili ter ju poskusite sešteti v polju *C1* s formulo $=A1 + B1$).
- Polja v osnovni tabeli, za katere nimate na voljo dejanske meritve naj bodo prazna. Povedano drugače, **manjkajoče vrednosti** v osnovni tabeli naj bodo označene s praznimi polji.
- **Imena stolpcev** v osnovni tabeli naj bodo:
 - brez presledkov, posebnih znakov (npr. % ali &) in šumnikov. V primeru, da imena stolpcev sestavljate iz več besed, predlagamo, da uporabite velike črke za začetke novih besed (npr. *ImeStolpca1*), lahko tudi *ime_stolpca1* ali pa *ime.stolpca1*.
 - čim krajša in brez nepotrebnih dodatnih informacij (npr. merske enote – le-te zapišite v opis podatkov, ki naj bo v drugem zavihku (*sheet*) Excel datoteke, več o tem kasneje).
 - jasno in kratko opisujejo vrednost, ki je v nekem stolpcu vnesena. Če je le možno naj bodo imena taka, da jih bomo lahko brez večjih težav povezali s količinami, omenjenimi v opisu oz. protokolu vaše študije.
- V primeru, da osnovna tabela vsebuje **stolpec z več kategorijami** (npr. *DomacaZival: pes, macka, riba*) pazite na naslednje:
 - Ime ene kategorije naj bo vedno enako, tj. *moski, moški* in *Moski* so različne kategorije.
 - Imena kategorij naj ne vsebujejo šumnikov. Lahko vsebujejo presledke, pazite pa naj ne bodo predolga.
 - Namesto besednih imen lahko uporabite številke (tj. kodiranje), pri tem pojasnite kodiranje v drugem zavihku Excel datoteke, več o tem kasneje. Prav tako prosim pojasnite imena kategorij, če le-ta niso za laika očitna.
- Podatke **prečistite tako, da za vsak stolpec preverite**:
 - Za opisne spremenljivke: Ali so v njem natanko določene kategorije (npr. *zenski* in *moski*, ne pa tudi *Moski*) ter prazne celice (za manjkajoče vrednosti)? Kliknite na puščico v stolpcu, v odprtem okenčku se vam bodo prikazale vse vrednosti v tem stolpcu.

- Za številke spremenljivke: Ali so vse vrednosti znotraj pričakovanega razpona (npr. starost glede na vključitveni kriterij, teža v pričakovanem razponu, odgovori na vprašanja glede na Likartovo lestvico vprašalnika, ipd.)? Kliknite na puščico v stolpcu, v odprtem okenčku lahko izberete naraščajočo ali padajočo razvrstitev (*sort*).
- Glede na naravo podatkov preverite še, ali so stolpci »smiselni« en na drugega. Primer: če imamo stolpec *jemljeZdravila* z vrednostmi *da/ne* in stolpec *vrstaZdravil*, potem moramo imeti pri preiskovancih z vrednostjo *ne*, prazno celico v stolpcu *vrstaZdravil*. Kliknite na puščico v stolpcu *jemljeZdravila*, v odprtem okenčku in odkljukajte le tiste z vrednostjo *ne* – prikazali se vam bodo le takšni preiskovanci. Nato kliknite na puščico v stolpcu *vrstaZdravil* in v okenčku boste lahko videli vse vrednosti. Preverite tudi, ali si datumi sledijo v pravilnem časovnem sosledju – če boste analizirali čas do dogodka in uporabljali analizo preživetja (ang. *survival analysis*, *event history analysis*), glejte navodila na koncu.
- Pri vseh zgornjih ukazih (razvrščanje in filtriranje) bodite pozorni, da imate znotraj tabele vse podatke. V nasprotnem primeru bo del podatkov ostal v originalni razvrstitvi, te vrednosti torej ne bodo več pripadale pravim preiskovancem!
- **Prosimo, da ste pri čiščenju podatkov dosledni in natančni**, saj je od tega odvisna pravilnost analize.
- V Excelovo datoteko vključite še 2. zavihek, ki bo predstavljal **šifrant**. Po vrsticah zapišite vse spremenljivke s takšnimi imeni kot v prvem zavihku. Ob tem lahko dodajate daljši opis spremenljivke, merske enote za številke spremenljivke (če je to potrebno), kodiranje za opisne spremenljivke (če je to potrebno). Poleg tega vas prosimo, da ob vsaki spremenljivki zabeležite še, ali je številka (ang. *numerical*), opisna imenska (ang. *nominal*) ali opisna urejenostna (ang. *ordinal*).
- V osnovni tabeli **lahko zase izboljšate preglednost** tako, da si obarvate spremenljivke iz »enake skupine«, npr. z eno barvo si obarvate rezultate hemograma, z drugo barve pa rezultate vprašalnika, ipd. Pri tem obarvajte celoten stolpec in ne le eno celico, tj. opozorila, ki so bila namenjena le vam, razrešite in nato umaknite iz končne verzije podatkov (npr. z rumeno so označeni vsi, ki jih moramo še dodatno poklicati). V primeru, da je za vas informacija, ki ste jo označevali z neko barvo pomembna (npr. se je nehal držati protokola), potem to vključite v podatke kot dodatno spremenljivko, tj. dodate stolpec, ki to informacijo vključuje (npr. 1 se je držal protokola in 0 sicer). **Upoštevajte torej, da si z barvami lahko izboljšate preglednost, ne bodo pa upoštevane v analizi.**
- Pazite na **GDPR**. V poslani datoteki tako naj ne bo imen, naslovov, telefonskih števil, mailov itd. V primeru, da ste do sedaj paciente beležili npr. z imeni, jih v svoji verziji osnovne tabele ohranite, poleg stolpca z imeni pa dodate še stolpec s šiframi (imenovan npr. *ID*). Na koncu urejanja podatkov naredite verzijo datoteke, v kateri izbrišete vse stolpce z osebnimi podatki, ohranite stolpec *ID* in nam pošljite takšno datoteko.
- Osnovna tabela naj ne vsebuje podatkov, ki v analizo zagotovo ne bodo vključeni (npr. posamezni odgovori anketnega vprašalnika, ki ga ne bomo analizirali).

Dodatna navodila, če boste analizirali čas do dogodka in uporabljali analizo preživetja (ang. survival analysis, event history analysis):

- Preverite, ali je sosledje vseh datumov ustrezno, npr. po vrsti si sledijo datumi diagnoze, začetka zdravljenja, konca zdravljenja, relapsa, smrti.
- Za vsak dogodek, ki ga preučujete (npr. smrt ali relaps), naj bo v podatkih:
 - Stolpec, ki označuje, ali je tekom spremljanja posameznik dogodek imel (tipično oznaka 1) ali ne (tipično oznaka 0) – za posameznike brez dogodka (oznaka 0) pravimo, da so krnjeni (ang. censored).
 - Stolpec z datumom dogodka oz. krnjenja, natančneje:
 - Če je posameznik imel dogodek, potem zapišite datum tega dogodka (npr. datum smrti ali relapsa).
 - Če posameznik dogodka ni imel, zapišite zadnji čas, za katerega veste, da posameznik dogodka ni imel – to je lahko datum konca raziskave ali pa nek bolj zgoden datum, ko smo posameznika izgubili izpod nadzora raziskave (ang. lost to follow-up). Za primer smrti je to datum ob koncu raziskave, ko ste v registru preverili, da je posameznik še živ. Če preučujete relaps ali progres bolezni, pa mora biti to zadnji datum, ob katerem je bil posameznik zagotovo brez relapsa ali progressa, tj. datum zadnje kontrole (in tipično ne sovpada z datumom iz registra smrti).