Centralni granični teorem i primjena parametrijske i neparametrijske statistike: normalnost podataka i veliki uzorci

1. Uvod

Centralni granični teorem (CGT) jedan je od temeljnih koncepata inferencijalne statistike i predstavlja ključni argument za korištenje parametrijskih testova čak i kada distribucija podataka nije savršeno normalna. Kako se u stvarnim istraživanjima rijetko susreće idealna normalnost, CGT omogućuje znanstvenicima i analitičarima pouzdanu primjenu parametrijske statistike u širokom rasponu situacija, posebno kod velikih uzoraka. U slučajevima kada su uzorci mali ili kada je distribucija izrazito odstupajuća, oslanjamo se na neparametrijsku statistiku.

2. Što je centralni granični teorem?

Centralni granični teorem navodi da će, kako veličina uzorka raste, **uzorkovna distribucija sredine** biti približno normalna, bez obzira na to kakva je stvarna populacijska distribucija (Field, 2013; Howell, 2012). Drugim riječima, pojedinačni podaci **ne moraju** biti normalno raspoređeni da bi se prosjeci tih podataka ponašali kao da dolaze iz normalne distribucije.

To je posebno važno jer parametrijski testovi (npr. t-test, ANOVA, Pearsonova korelacija) pretpostavljaju normalnost uzorkovne distribucije statistika — ne nužno normalnost samih sirovih podataka.

3. Koliko velik uzorak se smatra „dovoljno velikim“?

Prema relevantnoj literaturi:

- **n > 30** — uzorkovna distribucija sredine već poprima približno normalan oblik (Tabachnick & Fidell, 2013).

- **n ≥ 100** — odstupanja od normalnosti postaju manje statistički problematična (Howitt & Cramer, 2017).
- **n ≥ 200** — parametrijski testovi vrlo su stabilni i pri značajnim odstupanjima od normalnosti (Kim & Feldt, 2008).

Zaključak:
U praksi, **uzorci veći od n = 200** omogućuju vrlo sigurnu primjenu parametrijskih testova zahvaljujući stabilnom djelovanju centralnog graničnog teorema.

4. Kakvi podaci mogu biti uključeni u parametrijsku statistiku?

1) Intervalne i omjerne skale — idealne za parametrijske testove

Primjeri: dob, visina, težina, vrijeme, ukupni skorovi skala.

2) Likert skale (5 ili više kategorija)

Likert skale su tehnički ordinalne, ali u velikim uzorcima ponašaju se kao kvazi-intervalne. **Zbog CGT-a**, njihova srednja vrijednost je normalno raspodijeljena → mogu se koristiti parametrijski testovi.

3) Dihotomni podatci (0/1, da/ne)

Mogu se koristiti u regresijskim modelima ili kao prediktori/kontrolne varijable. Kod velikih uzoraka parametrijski modeli izvrsno funkcioniraju.

4) Nominalne kategorije

Mogu se uključiti u parametrijske analize preko:

- dummy kodiranja,
- ANOVA modela (ako su skupine dovoljno velike),
- linearne i logističke regresije.

5) Kada koristiti neparametrijske testove?

Neparametrijska statistika (Mann–Whitney U, Kruskal–Wallis, Spearman) preporučuje se kada:

- $n < 30$,
- distribucija je izrazito asimetrična,
- postoje teški outlieri,
- skala ima vrlo malo kategorija (npr. 3 ili manje),
- varijance između skupina su nejednake.

5. Primjer primjene centralnog graničnog teorema

Procjena prosječnog vremena odziva korisničke podrške

Pretpostavimo da tvrtka želi procijeniti prosječno vrijeme odziva svoje korisničke službe. Vrijeme odziva je poznato po tome da ima **desno-asimetričnu distribuciju** (neki korisnici čekaju vrlo dugo, što povlači rep distribucije).

Tvrtka dnevno ima oko 5.000 upita.

Koraci:

1. Svaki dan slučajno odaberu uzorak od **n = 250 korisničkih upita**.
2. Izračunaju prosječno vrijeme odziva tog uzorka.
3. Ponove postupak svakog dana kroz dva mjeseca (oko 60 uzoraka).
4. Iako su pojedinačna vremena odziva izrazito nenormalna, **prosječne vrijednosti uzoraka od n = 250** slijede približno normalnu distribuciju.
5. Tvrtka sada može koristiti parametrijske metode (t-test, ANOVA, interval pouzdanosti) za procjenu prosječnog vremena odziva i usporedbu razdoblja.

Zaključak primjera:

Parametrijska statistika pouzdano se koristi jer je uzorak dovoljno velik i CGT osigurava normalnost uzorkovne distribucije sredina.

6. Zaključak

Centralni granični teorem omogućuje široku i pouzdanu primjenu parametrijskih testova čak i u slučajevima kada sami podaci nisu normalno raspodijeljeni. Veći uzorci (posebno $n \geq 200$) osiguravaju stabilnost procjena, robusnost testova i smanjuju važnost odstupanja od normalnosti.

Podaci mogu biti:

- intervalni,
- omjerni,
- Likert skale,
- dihotomni,
- nominalni (uz odgovarajuće kodiranje).

Neparametrijska statistika koristi se kada su uzorci mali ili kada distribucija značajno odstupa od uvjeta koje parametrijski testovi mogu tolerirati.

CGT ostaje jedan od najvažnijih teorema u statistici jer omogućuje primjenu naprednih metoda i smisleno interpretiranje rezultata čak i u realnim, nesavršenim uvjetima.

Link pdf:
https://www.StatistischeBeratung.de/images/services/www_Statistika_hr_Centralni_granicni_teorem_i_primjena_parametrijske_i_neparametrijske_statistike.pdf

Cite this article in your research paper:

APA

Statistische Beratung Leonardo Miljko (*01.12.2025.*) **Centralni granični teorem i primjena parametrijske i neparametrijske statistike: normalnost podataka i veliki uzorci**. Retrieved from: https://www.StatistischeBeratung.de/images/services/www_Statistika_hr_Centralni_granicni_teorem_i_primjena_parametrijske_i_neparametrijske_statistike.pdf.

Literatura

Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). Sage.

Gravetter, F. J., & Wallnau, L. B. (2014). *Essentials of statistics for the behavioral sciences* (8th ed.). Cengage Learning.

Howell, D. C. (2012). *Statistical methods for psychology* (8th ed.). Cengage Learning.

Howitt, D., & Cramer, D. (2017). *Introduction to statistics in psychology* (7th ed.). Pearson.

Khamis, H., & El-Sayed, A. (1999). On large sample tests for the normality of data. *Journal of Modern Applied Statistical Methods, 4*(1), 48–57.

Kim, J. O., & Feldt, L. S. (2008). Statistical procedures for classification and grading of items in educational measurement and evaluation. *Educational and Psychological Measurement, 68*(5), 829–844. https://doi.org/10.1177/0013164407313751

Opić, S. (2011). Centralni granični teorem i njegova uloga u inferencijalnoj statistici. *Školski vjesnik, 60*(2), 181–198.

Stevens, J. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). Routledge.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Wilcox, R. R. (2017). *Introduction to robust estimation and hypothesis testing* (4th ed.). Academic Press.