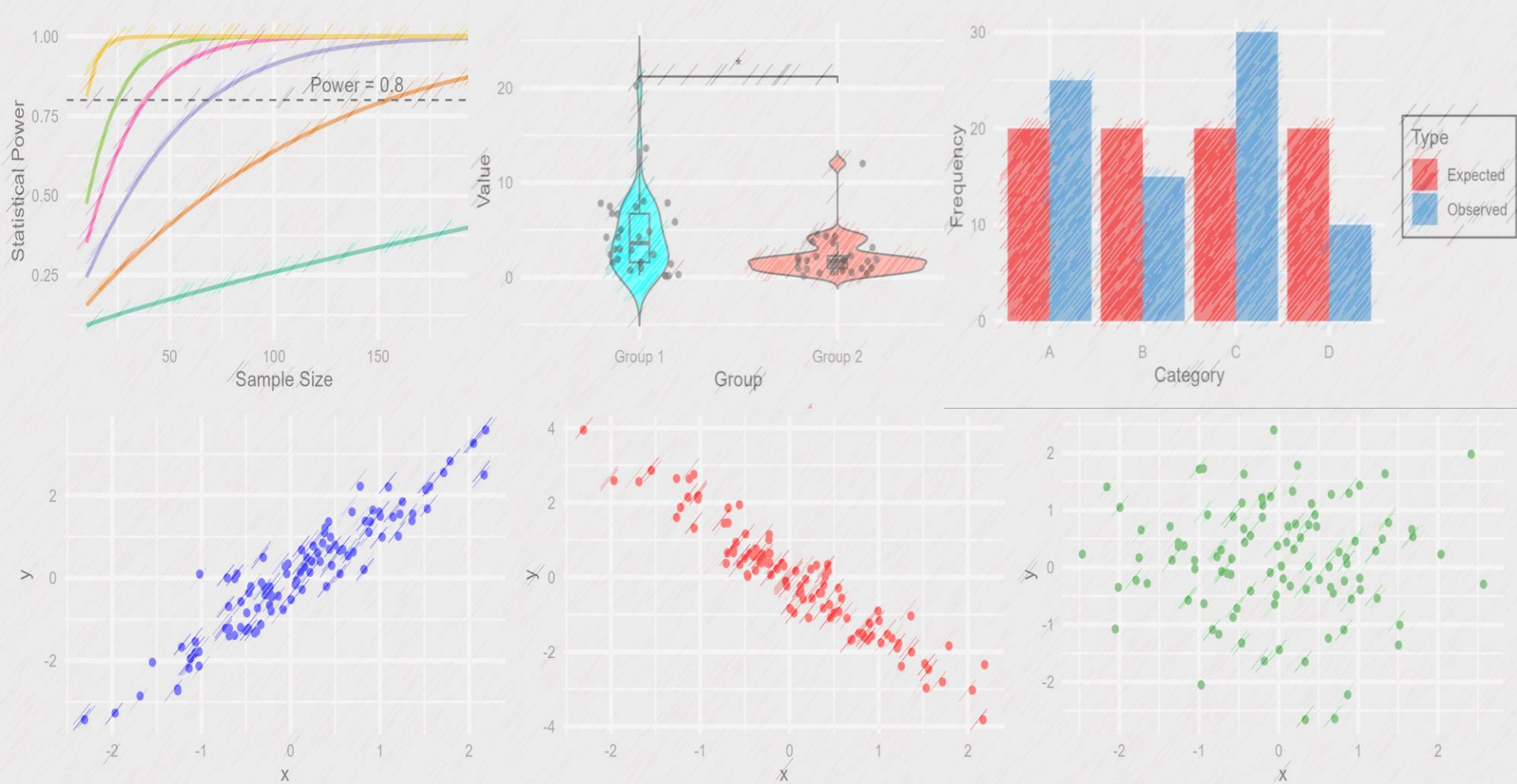




Seminar 7: kategorički podaci; snaga testa; korelacija i regresija

1. Kategorički podaci

- Škola je provela anketu među 60 učenika kako bi utvrdila njihove sklonosti prema tri različite metode poučavanja: Metoda A, Metoda B i Metoda C. Odgovori su dalje kategorizirani prema spolu (muški i ženski). Rezultati ankete sažeti su u tablici u nastavku:

	Metoda A	Metoda B	Metoda C	Ukupno
Muški	10	15	5	30
Ženski	5	7	18	30
Ukupno	15	22	23	60

- Koji statistički test biste koristili da testirate postoji li preferencija za određenu metodu poučavanja s obzirom na spol? Zašto? Koja je nulta hipoteza?

2. Snaga testa

- Istraživač provodi studiju kako bi ispitao smanjuje li novi lijek sistolički krvni tlak učinkovitije od placeba. Planira koristiti t-test s dva uzorka s razinom značajnosti (α) od 0,05. Međutim, njegov preliminarni izračun snage pokazuje da studija ima samo 60% snage da otkrije srednju razliku od 5 mmHg. Koja bi od sljedećih promjena povećala snagu statističkog testa? (Odaberite sve primjenjivo.)
- A) Povećajte veličinu uzorka u obje skupine.
- B) Povećajte razinu značajnosti (α) od 0,05 do 0,10.
- C) Smanjite varijabilnost (standardnu devijaciju) u izmjerenim vrijednostima krvnog tlaka, ako je moguće.
- D) Smanjite veličinu učinka kako biste otkrili manje razlike u krvnom tlaku.
- E) Koristite jednostrani test umjesto dvostranog testa, pod pretpostavkom da je smjer učinka poznat.

2. Tumačenje i značajnost koeficijenta korelacije

- **Pitanje:** U istraživanju povezanosti raspoloženja i količine tekućine unesene pijenjem tijekom dana dobivena je povezanost $r = 0.12$, $P = 0.003$. Je li ispravno zaključiti kako postoji značajna povezanost raspoloženja i količine popijene tekućine?

Udovičić, Martina ; Baždarić Ksenija ; Bilić-Zulle Lidija ; Petrovečki Mladen Što treba znati kada izračunavamo koeficijent korelacije? // Biochemia medica, 17 (2007), 1; 10-15

2. Tumačenje i značajnost koeficijenta korelacije

- **Pitanje:** U istraživanju povezanosti raspoloženja i količine tekućine unesene pijenjem tijekom dana dobivena je povezanost $r = 0.12$, $P = 0.003$. Je li ispravno zaključiti kako postoji značajna povezanost raspoloženja i količine popijene tekućine?
- **Odgovor:** Nije ispravno. P-vrijednost manja od 0.05 znači da je koeficijent korelacije značajan i da se smije tumačiti. Budući da je vrijednost $r = 0.12$, možemo zaključiti da nema povezanosti između ispitivanih varijabli, i to možemo tvrditi jer je koeficijent korelacije značajan ($P < 0.05$)

Udovičić, Martina ; Baždarić Ksenija ; Bilić-Zulle Lidija ; Petrovečki Mladen Što treba znati kada izračunavamo koeficijent korelacije? // Biochemia medica, 17 (2007), 1; 10-15

3. Tumačenje i značajnost koeficijenta korelacije

- **Pitanje:** U istraživanju povezanosti koncentracije alkohola u krvi i prometnih nesreća utvrđeni su $r = 0.78$ i $P = 0.002$. Možemo li zaključiti kako uzimanje alkohola uzrokuje prometne nesreće, tj. Promatrane prometne nesreće su posljedica uzimanja alkohola?

Udovičić, Martina ; Baždarić Ksenija ; Bilić-Zulle Lidija ; Petrovečki Mladen Što treba znati kada izračunavamo koeficijent korelacije? // Biochemia medica, 17 (2007), 1; 10-15

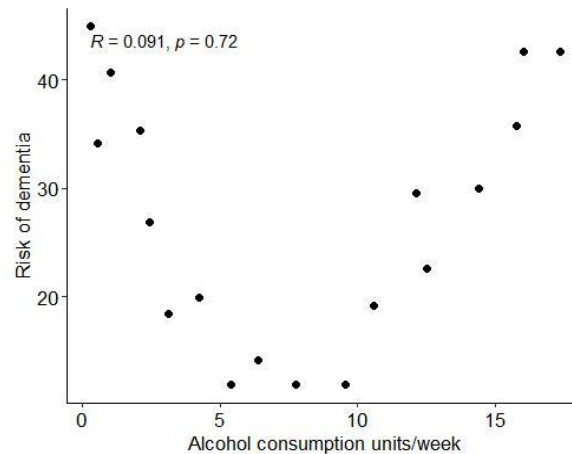
3. Tumačenje i značajnost koeficijenta korelacije

- **Pitanje:** U istraživanju povezanosti koncentracije alkohola u krvi i prometnih nesreća utvrđeni su $r = 0.78$ i $P = 0.002$. Možemo li zaključiti kako uzimanje alkohola uzrokuje prometne nesreće, tj. Promatrane prometne nesreće su posljedica uzimanja alkohola?
- **Odgovor:** Ne, ne možemo. Korelacija govori o povezanosti, a ne o uzročno posljedičnim vezama. Moguće je da postoje i drugi neistraženi čimbenici povezani s učestalošću nesreća koji utječu na učestalnost prometnih nesreća (ispravnost vozila, uvjeti na cesti, bolesti, druge farmakološki aktivne tvari, ...)

Udovičić, Martina ; Baždarić Ksenija ; Bilić-Zulle Lidija ; Petrovečki Mladen Što treba znati kada izračunavamo koeficijent korelacije? // Biochemia medica, 17 (2007), 1; 10-15

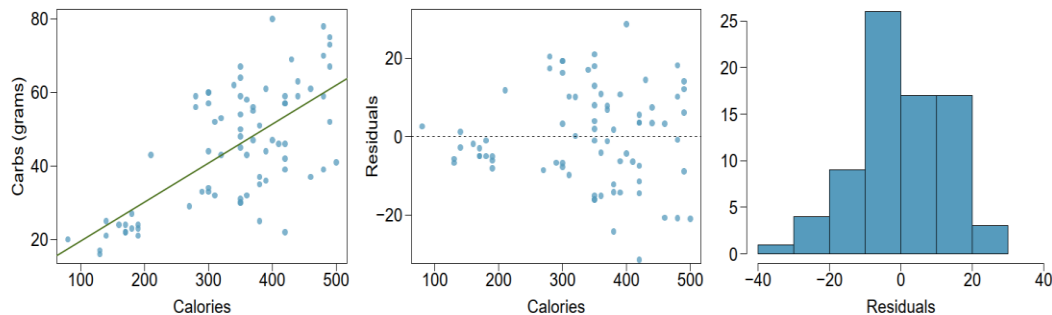
4. Povezanost dviju varijabli

- Provodite istraživanje utjecaja alkohola na rizik razvoja demencije. Što biste na temelju sljedećeg grafičkog prikaza zaključili o povezanosti konzumacije alkohola i rizika za razvoj demencije?



5. Linearna regresija

- Dijagram raspršenosti u nastavku prikazuje odnos između broja kalorija i količina ugljikohidrata (u gramima) različitih proizvoda. Zanima nas predviđanje količine ugljikohidrata na temelju sadržaja kalorija korištenjem linearne regresije.



- Opišite odnos između broja kalorija i količine ugljikohidrata u gramima.
- Navedite nultu i alternativnu hipotezu regresijske analize.
- Na temelju sva tri grafička prikaza, procijenite jesu li ovi podaci prikladni za linearnu regresiju i objasnite svoj odgovor.

6. Linearna regresija

- Proveli smo linearnu regresiju na sljedećim podacima:
 - Nezavisna varijabla (predictor): Količina određene hranjive tvari u biološkom uzorku (npr. koncentracija vitamina C u uzorku krvi).
 - Zavisna varijabla (response): Stopa rasta stanica izloženih hranjivoj tvari (npr. stopa rasta stanica kao postotak maksimalnog rasta).

Te smo dobili sljedeće rezultate:

$$\beta_0 = 10.10$$

$$\beta_1 = 1.98$$

- Napišite regresijsku jednadžbu te interpretirajte koeficijente.

7. Logistička regresija

Istraživači žele znati predviđaju li **PD-L1 ekspresija** i **tumorski mutacijski teret (TMB)** odgovor na imunoterapiju kod raka pluća.

Model je:

Odgovor na imunoterapiju (1 = responder, 0 = non-responder) ~ PDL1 + TMB_High

PD-L1 = kontinuirana mjera (viši = više PD-L1 ekspresije)

TMB_High = 1 ako je visok TMB, 0 ako je nizak

Predictor	Estimate	Std. Error	z value	p-value	Odds Ratio (exp(Estimate))
(Intercept)	-2.30	0.55	-4.18	0.00003	0.10
PDL1	0.60	0.20	3.00	0.003	1.82
TMB_High	0.90	0.35	2.57	0.010	2.46

Koji prediktori su statistički značajni?

- A) Samo PD-L1
- B) Samo TMB_High
- C) Ni jedan
- D) Oba (PD-L1 i TMB_High)

Koji prediktor ima jaču povezanost s odgovorom?

- A) PD-L1 (OR = 1.82)
- B) TMB_High (OR = 2.46)
- C) Jednako su jaki
- D) Ne može se reći bez dodatnih podataka