

# Tjedan 11: Usporedba prosjeka t-testovima

Kad pretpostavke drže i kad ne drže

2025-05-10

## Table of contents

<b>1</b>	<b>Redizajn koji je podijelio redakciju</b>	<b>2</b>
<b>2</b>	<b>Tri vrste t-testa</b>	<b>4</b>
<b>3</b>	<b>Pretpostavke koje morate provjeriti</b>	<b>4</b>
<b>4</b>	<b>Provjera normalnosti</b>	<b>5</b>
4.1	Vizualna provjera: histogram i QQ plot . . . . .	5
4.2	Shapiro-Wilkov test . . . . .	7
<b>5</b>	<b>Provjera homogenosti varijance</b>	<b>9</b>
<b>6</b>	<b>Upareni t-test: utječu li vizuali na vrijeme čitanja?</b>	<b>10</b>
6.1	Vizualizacija uparenih podataka . . . . .	12
<b>7</b>	<b>Sva četiri ishoda odjednom</b>	<b>14</b>
<b>8</b>	<b>Kad normalnost zakaže: Wilcoxonov test</b>	<b>17</b>
<b>9</b>	<b>Nije li efekt različit za različite teme?</b>	<b>19</b>
<b>10</b>	<b>Kako napisati rezultate: APA format</b>	<b>21</b>
10.1	Upareni t-test u APA formatu . . . . .	21
10.2	Anatomija APA izvještaja . . . . .	22
<b>11</b>	<b>Nezavisni t-test: kratki protiv dugih članaka</b>	<b>23</b>
<b>12</b>	<b>Oprez s outlierima</b>	<b>27</b>
12.1	Kako detektirati outliere . . . . .	28
12.2	Robusne alternative: podrezani prosjeci . . . . .	30
<b>13</b>	<b>Formula pristup u R-u</b>	<b>31</b>

<b>14 Sve zajedno: izvještaj za uredništvo</b>	<b>32</b>
<b>15 Koji test odabrati?</b>	<b>40</b>
<b>16 Tri testa, jedan pregled</b>	<b>40</b>
<b>17 Zadaci za pripremu</b>	<b>42</b>
<b>18 Dodatno čitanje</b>	<b>42</b>
<b>19 Pojmovnik</b>	<b>42</b>

`library(tidyverse)`

### **i** Ishodi učenja

Nakon ovog predavanja moći ćete:

1. Odabrati odgovarajući t-test (jednouzorački, nezavisni, upareni) za zadano istraživačko pitanje.
2. Provjeriti pretpostavke t-testa (normalnost, homogenost varijance) i znati što učiniti kad su narušene.
3. Primijeniti Shapiro-Wilkov test i QQ plot za provjeru normalnosti.
4. Provesti Wilcoxonov test kao neparametrijsku alternativu kad normalnost nije zadovoljena.
5. Izračunati i interpretirati Cohenov d za sve tri vrste t-testa.
6. Pravilno izvijestiti rezultate t-testa u APA formatu.
7. Provesti kompletnu analizu usporedbe dvaju uvjeta na stvarnim podacima.

## **1 Redizajn koji je podijelio redakciju**

Zamislite da radite kao istraživačica u uredništvu jednog web portala. Portal objavljuje članke o politici, zdravlju, tehnologiji, sportu i kulturi — uglavnom tekstualne, s ponekom fotografijom za vizualni odmor. Uredništvo razmišlja o velikom redizajnu. Žele dodati infografike, podatkovne vizualizacije, ilustracije i interaktivne elemente. Ideja zvuči privlačno, ali glavni urednik nije uvjeren. “Vizuali koštaju,” kaže. “Trebam dokaze da stvarno poboljšavaju čitateljsko iskustvo.”

I tako dobijete zadatak — osmisliti eksperiment koji će odgovoriti na pitanje utječu li vizualni elementi na četiri ključna ishoda — vrijeme čitanja, razumijevanje sadržaja, namjeru dijeljenja i percipiranu vjerodostojnost. Odabrali ste 120 članaka i svaki prezentirali ispitanicima u dva uvjeta: jednom s vizualima, jednom bez. Budući da je svaki članak testiran u oba uvjeta, radi se o within-subjects dizajnu — a statistički alat koji vam treba za takvu usporedbu zove se upareni t-test.

Na predavanju o testiranju hipoteza naučili smo logiku koja stoji iza t-testa — postavljamo nultu hipotezu, izračunavamo testnu statistiku, gledamo p-vrijednost i donosimo odluku. Ovo predavanje se bavi onom drugom polovicom posla — onom koja se u udžbenicima često prebrzo preskoči. Kako zapravo provjeriti jesu li pretpostavke testa zadovoljene? Što učiniti kad nisu? Kako odabrati pravi test za pravi dizajn? I kako napisati rezultate tako da ih kolege i recenzenti mogu razumjeti?

Krenimo od podataka.

```
articles <- read_csv("../resources/datasets/article_visuals.csv")
glimpse(articles)
```

```
Rows: 120
Columns: 12
$ article_id      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 1~
$ category       <chr> "tehnologija", "politika", "tehnologija", "t~
$ length_category <chr> "srednji", "kratki", "srednji", "kratki", "s~
$ word_count     <dbl> 589, 307, 805, 447, 626, 827, 802, 302, 578,~
$ reading_time_no_visual <dbl> 4.9, 2.3, 3.4, 1.6, 3.3, 6.4, 2.7, 0.7, 3.4,~
$ reading_time_with_visual <dbl> 5.8, 3.0, 3.9, 1.8, 4.4, 7.6, 3.2, 0.9, 3.9,~
$ comprehension_no_visual <dbl> 5, 6, 6, 6, 5, 8, 6, 9, 6, 5, 5, 7, 6, 10, 5~
$ comprehension_with_visual <dbl> 6, 7, 7, 6, 7, 9, 6, 9, 6, 6, 5, 8, 7, 10, 6~
$ sharing_no_visual <dbl> 2, 3, 3, 4, 3, 2, 3, 3, 3, 2, 3, 3, 2, 3, 3,~
$ sharing_with_visual <dbl> 3, 5, 2, 4, 3, 2, 4, 4, 4, 3, 4, 3, 2, 3, 3,~
$ credibility_no_visual <dbl> 5, 5, 2, 2, 5, 6, 3, 3, 5, 4, 5, 3, 3, 2, 5,~
$ credibility_with_visual <dbl> 4, 5, 4, 3, 5, 7, 3, 2, 6, 4, 6, 4, 3, 2, 5,~
```

```
articles |>
  summarise(
    n = n(),
    M_time_no = round(mean(reading_time_no_visual), 2),
    M_time_with = round(mean(reading_time_with_visual), 2),
    M_comp_no = round(mean(comprehension_no_visual), 2),
    M_comp_with = round(mean(comprehension_with_visual), 2),
    M_share_no = round(mean(sharing_no_visual), 2),
    M_share_with = round(mean(sharing_with_visual), 2)
  )
```

```
# A tibble: 1 x 7
  n M_time_no M_time_with M_comp_no M_comp_with M_share_no M_share_with
<int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 120 3.43 4.02 5.94 6.74 2.48 3.15
```

Na prvi pogled, brojke govore u prilog vizualima — članci s vizualnim elementima imaju duže vrijeme čitanja, bolje razumijevanje i višu namjeru dijeljenja. Ali “na prvi pogled”

nije dovoljno za istraživačicu koja zna što radi. Prije nego što izvučemo ikakve zaključke, moramo provjeriti jesu li pretpostavke statističkog testa koje namjeravamo koristiti uopće zadovoljene.

## 2 Tri vrste t-testa

Prije nego uđemo u analizu, vrijedi se podsjetiti da t-test nije jedan jedini postupak nego obitelj od tri testa, a svaki odgovara na drugačije pitanje.

```
tribble(  
  ~test, ~pitanje, ~R_kod,  
  "Jednouzorački", "Razlikuje li se prosjek od poznate vrijednosti?", "t.test(x, mu = vrij  
  "Nezavisni (dvouzorački)", "Razlikuju li se prosjeci dviju nezavisnih grupa?", "t.test(x  
  "Upareni", "Razlikuje li se prosjek u dva uvjeta za iste jedinice?", "t.test(x, y, paire  
  )
```

```
# A tibble: 3 x 3  
  test                pitanje                R_kod  
  <chr>                <chr>                <chr>  
1 Jednouzorački      Razlikuje li se prosjek od poznate vrijednosti? t.te~  
2 Nezavisni (dvouzorački) Razlikuju li se prosjeci dviju nezavisnih grupa? t.te~  
3 Upareni            Razlikuje li se prosjek u dva uvjeta za iste je~ t.te~
```

Jednouzorački t-test koristi se kad imate jednu skupinu i želite provjeriti razlikuje li se njezin prosjek od neke poznate ili pretpostavljene vrijednosti — na primjer, je li prosječna ocjena razumijevanja vaših članaka različita od nacionalnog prosjeka od 6.0 bodova. Nezavisni t-test uspoređuje dvije odvojene grupe koje nemaju nikakvu vezu jedna s drugom — recimo, čitatelje koji su vidjeli vizuale i čitatelje koji nisu, ali to su različiti ljudi. Upareni t-test koristi se kad iste jedinice mjerite dva puta, u dva različita uvjeta — a to je upravo naš slučaj, jer su isti članci prezentirani i s vizualima i bez njih.

Za naše podatke, dakle, koristimo upareni t-test. Ali prije nego ga pokrenemo, moramo se uvjeriti da su pretpostavke tog testa razumno zadovoljene.

## 3 Pretpostavke koje morate provjeriti

Svaki statistički test dolazi s pretpostavkama — uvjetima koji moraju biti barem približno zadovoljeni da bismo rezultatima mogli vjerovati. T-test nije iznimka.

Sve tri varijante t-testa dijele tri temeljne pretpostavke. Kao prvo, podaci moraju biti barem na intervalnoj skali, što znači da razlike između vrijednosti imaju smisla (razlika između 3 i 5 je ista kao razlika između 7 i 9). Kao drugo, opažanja moraju biti nezavisna jedno od drugoga. U uparenom testu to znači da su parovi nezavisni, iako mjerenja unutar para

naravno nisu. Kao treće, distribucija mora biti približno normalna — ili uzorak dovoljno velik da centralni granični teorem kompenzira.

Nezavisni t-test ima jednu dodatnu pretpostavku — varijance dviju grupa trebale bi biti približno jednake. To je pretpostavka klasičnog Studentovog t-testa. Dobra vijest je da postoji Welchova korekcija koja tu pretpostavku ne zahtijeva, i upravo je ona default u R-u. O tome ćemo više za koji odlomak.

Od svih ovih pretpostavki, ona koju ćete najčešće morati aktivno provjeravati je normalnost. A kod uparenog t-testa, pazite na važan detalj — ne provjeravate normalnost pojedinačnih mjerenja, nego normalnost *razlika* između dvaju uvjeta. Zvuči kao sitnica, ali ta sitnica može potpuno promijeniti zaključak.

## 4 Provjera normalnosti

### 4.1 Vizualna provjera: histogram i QQ plot

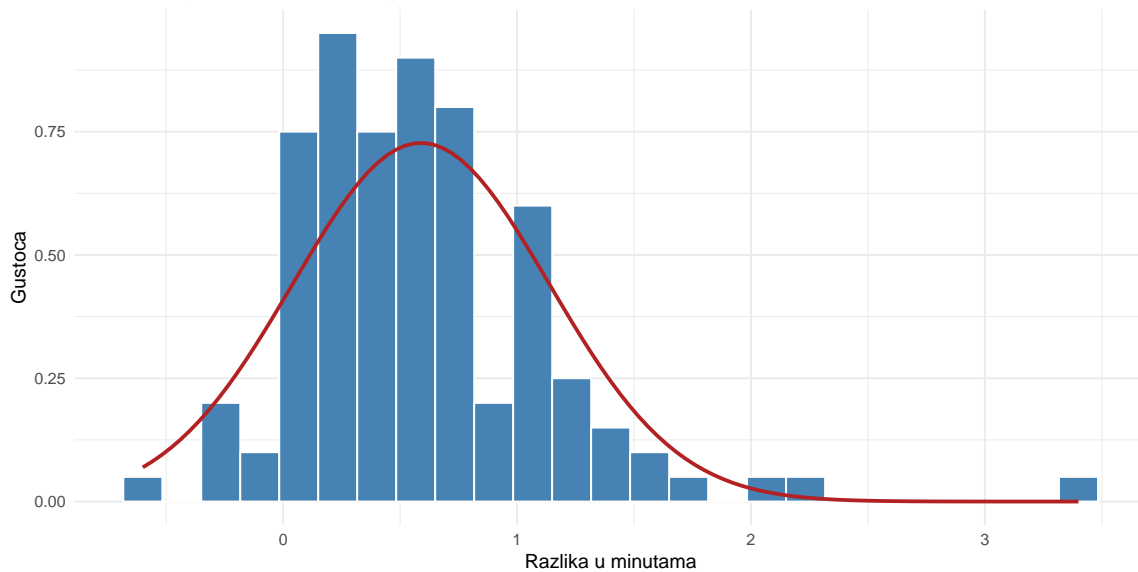
Najprirodniji prvi korak je pogledati podatke. Izračunajmo razlike između dvaju uvjeta za sva četiri ishoda.

```
# Za upareni test: provjera normalnosti RAZLIKA
articles <- articles |>
  mutate(diff_time = reading_time_with_visual - reading_time_no_visual,
         diff_comp = comprehension_with_visual - comprehension_no_visual,
         diff_share = sharing_with_visual - sharing_no_visual,
         diff_cred = credibility_with_visual - credibility_no_visual)
```

Počnimo s histogramom razlika u vremenu čitanja. Ako je distribucija približno normalna, histogram bi trebao imati zvonoliki oblik, s većinom vrijednosti okupljenim oko sredine i simetričnim repovima.

```
articles |>
  ggplot(aes(x = diff_time)) +
  geom_histogram(aes(y = after_stat(density)), fill = "steelblue", color = "white", bins =
  stat_function(fun = dnorm,
               args = list(mean = mean(articles$diff_time), sd = sd(articles$diff_time)),
               color = "firebrick", linewidth = 1) +
  labs(
    title = "Distribucija razlika u vremenu citanja (s vizualima minus bez)",
    subtitle = "Crvena krivulja = normalna distribucija s istim M i SD",
    x = "Razlika u minutama",
    y = "Gustoca"
  ) +
  theme_minimal()
```

Distribucija razlika u vremenu citanja (s vizualima minus bez)  
 Crvena krivulja = normalna distribucija s istim M i SD

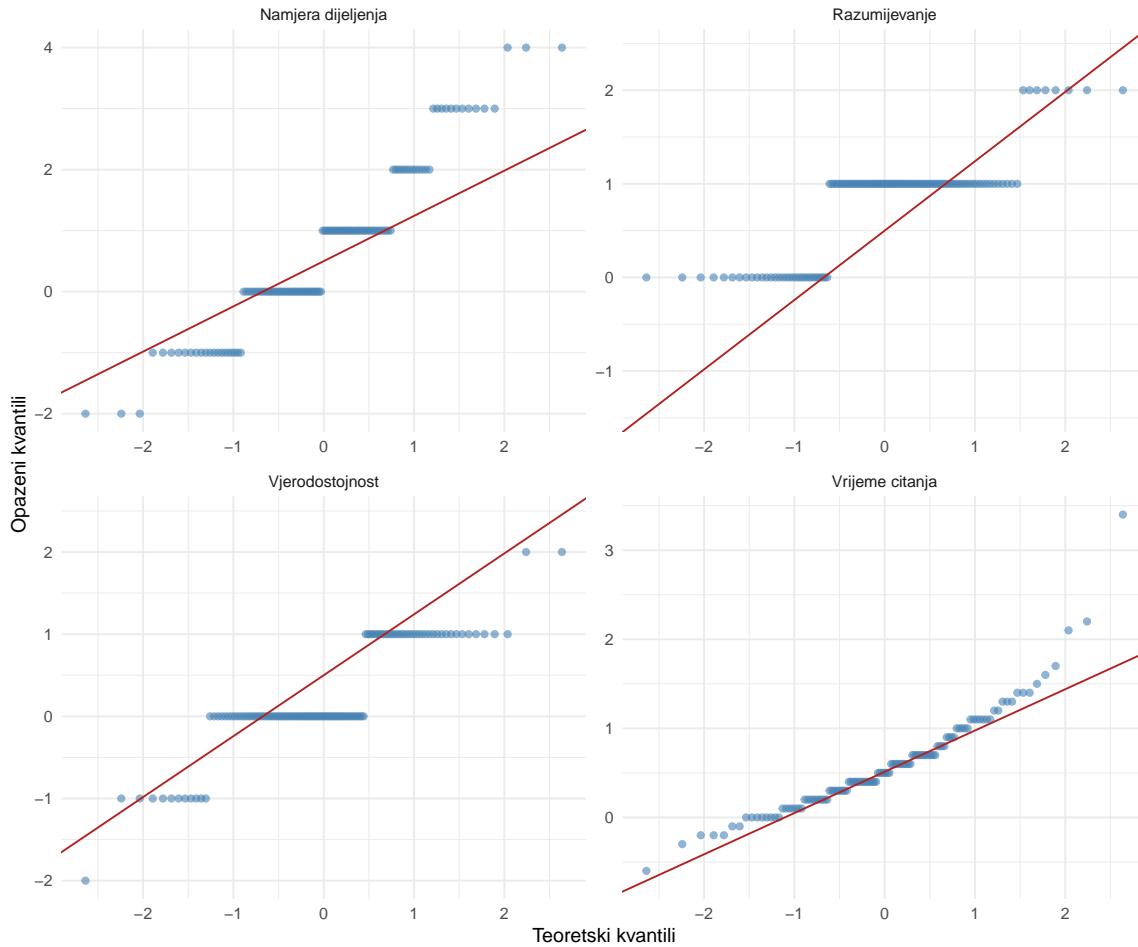


Histogram je koristan, ali još informativniji dijagnostički alat je QQ plot (quantile-quantile plot). Ideja QQ plota je jednostavna — on uspoređuje kvantile vaših podataka s kvantilima savršene normalne distribucije. Ako su vaši podaci normalno distribuirani, točke će ležati uz dijagonalnu liniju. Što više odstupaju, to je distribucija manje normalna.

```
# QQ plotovi za sve cetiri razlike
articles |>
  select(starts_with("diff_")) |>
  pivot_longer(everything(), names_to = "varijabla", values_to = "razlika") |>
  mutate(varijabla = case_when(
    varijabla == "diff_time" ~ "Vrijeme citanja",
    varijabla == "diff_comp" ~ "Razumijevanje",
    varijabla == "diff_share" ~ "Namjera dijeljenja",
    varijabla == "diff_cred" ~ "Vjerodostojnost"
  )) |>
  ggplot(aes(sample = razlika)) +
  stat_qq(color = "steelblue", alpha = 0.6) +
  stat_qq_line(color = "firebrick") +
  facet_wrap(~varijabla, scales = "free") +
  labs(
    title = "QQ plotovi za razlike (upareni uvjeti)",
    subtitle = "Tocke blizu linije = normalna distribucija. Odstupanja na repovima su cest",
    x = "Teoretski kvantili",
    y = "Opazeni kvantili"
  ) +
  theme_minimal()
```

### QQ plotovi za razlike (upareni uvjeti)

Točke blizu linije = normalna distribucija. Odstupanja na repovima su česta.



Pogledajte četiri panela. Razlike u vremenu čitanja i razumijevanju lijepo prate dijagonalu — distribucija je razumno normalna. Razlike u namjeri dijeljenja i vjerodostojnosti pokazuju stepeničast uzorak, što je očekivano — te varijable su mjerene Likert skalom (cjelobrojne vrijednosti od 1 do 5 ili 1 do 7), pa ne mogu biti savršeno glatke. Blaga odstupanja na repovima su prihvatljiva, osobito kad je uzorak veći od 30.

## 4.2 Shapiro-Wilkov test

Vizualna procjena je korisna, ali ponekad želite i formalni test. Shapiro-Wilkov test je najčešće korišten test normalnosti. Njegova nulta hipoteza kaže da su podaci normalno distribuirani, pa ako je  $p < 0.05$ , zaključujemo da normalnost nije zadovoljena.

```
tibble(  
  varijabla = c("Vrijeme citanja", "Razumijevanje", "Namjera dijeljenja", "Vjerodostojnost"),  
  W = c(  
    shapiro.test(articles$diff_time)$statistic,
```

```

shapiro.test(articles$diff_comp)$statistic,
shapiro.test(articles$diff_share)$statistic,
shapiro.test(articles$diff_cred)$statistic
),
p = c(
shapiro.test(articles$diff_time)$p.value,
shapiro.test(articles$diff_comp)$p.value,
shapiro.test(articles$diff_share)$p.value,
shapiro.test(articles$diff_cred)$p.value
)
) |>
mutate(
W = round(W, 4),
p = round(p, 4),
normalno = if_else(p >= 0.05, "Da", "Ne")
)

```

```

# A tibble: 4 x 4
  varijabla      W      p normalno
  <chr>         <dbl> <dbl> <chr>
1 Vrijeme citanja 0.908    0 Ne
2 Razumijevanje 0.715    0 Ne
3 Namjera dijeljenja 0.93    0 Ne
4 Vjerodostojnost 0.810    0 Ne

```

Razlike u vremenu čitanja prolaze test normalnosti ( $p > 0.05$ ), što znači da nemamo dovoljno dokaza da odbacimo pretpostavku o normalnoj distribuciji. Razlike u Likert varijablama možda ne prolaze jer su diskretne. Za te varijable ćemo razmotriti neparametrijske alternative.

### ! Paradoks Shapiro-Wilkovog testa

Shapiro-Wilkov test ima isti paradoks kao i svaki statistički test — njegova osjetljivost ovisi o veličini uzorka. S velikim uzorkom (recimo  $n = 500$ ) detektirat će i sasvim trivijalna odstupanja od normalnosti koja nemaju nikakav praktični značaj. S malim uzorkom ( $n = 15$ ) propustit će i prilično grube deformacije distribucije.

Zato je vizualna procjena putem QQ plota jednako važna kao formalni test. Praktično pravilo je sljedeće — ako QQ plot izgleda razumno i uzorak ima više od 30 opažanja, t-test je dovoljno robustan čak i za umjerena odstupanja od normalnosti jer centralni granični teorem osigurava da će distribucija prosjeka biti približno normalna bez obzira na oblik izvorne distribucije.

## 5 Provjera homogenosti varijance

Kad koristite nezavisni t-test (usporedba dviju odvojenih grupa), klasična Studentova verzija pretpostavlja da obje grupe imaju približno jednake varijance. Pogledajmo što se dogodi kad ta pretpostavka nije zadovoljena — na simuliranim podacima gdje jednoj grupi namjerno zadamo mnogo veću varijabilnost.

```
# Demonstracija na simuliranim podacima (nezavisne grupe)
set.seed(42)
grupa_a <- rnorm(50, mean = 5, sd = 1.0)
grupa_b <- rnorm(50, mean = 6, sd = 2.5) # razlicita varijanca!

cat("SD grupa A:", round(sd(grupa_a), 2), "\n")
```

SD grupa A: 1.15

```
cat("SD grupa B:", round(sd(grupa_b), 2), "\n")
```

SD grupa B: 2.31

```
cat("Omjer varijanci:", round(var(grupa_b) / var(grupa_a), 2), "\n")
```

Omjer varijanci: 4.03

Grupa B ima dva i pol puta veću standardnu devijaciju od grupe A. Usporedimo što se dogodi kad pokrenemo Studentov t-test (koji pretpostavlja jednake varijance) nasuprot Welchovom t-testu (koji tu pretpostavku ne zahtijeva).

```
# Studentov t-test (pretpostavlja jednake varijance)
student_rez <- t.test(grupa_a, grupa_b, var.equal = TRUE)

# Welchov t-test (ne pretpostavlja jednake varijance, DEFAULT)
welch_rez <- t.test(grupa_a, grupa_b, var.equal = FALSE)

tibble(
  test = c("Student (var.equal = TRUE)", "Welch (var.equal = FALSE, DEFAULT)"),
  t = round(c(student_rez$statistic, welch_rez$statistic), 3),
  df = round(c(student_rez$parameter, welch_rez$parameter), 1),
  p = round(c(student_rez$p.value, welch_rez$p.value), 4)
)
```

```
# A tibble: 2 x 4
  test          t    df      p
  <chr>        <dbl> <dbl> <dbl>
1 Student (var.equal = TRUE)   -3.52  98  0.0006
2 Welch (var.equal = FALSE, DEFAULT) -3.52 71.9 0.0007
```

Primijetite jednu zanimljivu stvar — Welchov test ima neokrugle stupnjeve slobode. To je zato što ih prilagođava za razliku u varijancama. Kad su varijance jednake, oba testa daju praktički identične rezultate. Kad su varijance različite, Welchov test je točniji, a Studentov test može dati lažno pozitivne ili lažno negativne rezultate.

#### Zaboravite na Studentov t-test

Evo savjeta koji će vam pojednostaviti život — koristite Welchov t-test uvijek. Ne morate uopće provjeravati homogenost varijance jer Welchov test radi jednako dobro kad su varijance jednake i bolje kad su različite. On je default u R-u (`var.equal = FALSE`) iz dobrog razloga.

Studentov t-test s `var.equal = TRUE` koristite samo ako imate specifičan razlog — na primjer, kad trebate reproducirati rezultate iz objavljenog rada koji je koristio Studentov test.

## 6 Upareni t-test: utječu li vizuali na vrijeme čitanja?

Pretpostavke smo provjerili, razlike izgledaju razumno normalno, uzorak je dovoljno velik. Vrijeme je da pokrenemo test i odgovorimo na pitanje koje je pokrenulo cijelu analizu.

```
# Upareni t-test: vrijeme citanja s vizualima vs bez
time_test <- t.test(
  articles$reading_time_with_visual,
  articles$reading_time_no_visual,
  paired = TRUE
)
time_test
```

Paired t-test

```
data: articles$reading_time_with_visual and articles$reading_time_no_visual
t = 11.764, df = 119, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 0.4900011 0.6883323
sample estimates:
mean difference
 0.5891667
```

Pogledajmo rezultat i raspakujmo svaki dio. Argument `paired = TRUE` kaže R-u da ne tretira ova dva vektora kao nezavisne grupe, nego kao parove — svaki članak ima dva mjerenja. Interno, R zapravo računa razlike i provodi jednouzorački t-test na tim razlikama, testirajući je li njihov prosjek različit od nule.

```
# Cohenov d za upareni test: d = M_razlika / SD_razlika
d_time <- mean(articles$diff_time) / sd(articles$diff_time)

cat("=== Upareni t-test: Vrijeme citanja ===\n")
```

```
=== Upareni t-test: Vrijeme citanja ===
```

```
cat("Bez vizuala: M =", round(mean(articles$reading_time_no_visual), 2), "min\n")
```

```
Bez vizuala: M = 3.43 min
```

```
cat("S vizualima: M =", round(mean(articles$reading_time_with_visual), 2), "min\n")
```

```
S vizualima: M = 4.02 min
```

```
cat("Razlika: M =", round(mean(articles$diff_time), 2), "min\n")
```

```
Razlika: M = 0.59 min
```

```
cat("t(", time_test$parameter, ") = ", round(time_test$statistic, 2), "\n", sep = "")
```

```
t(119) = 11.76
```

```
cat("p < 0.001\n")
```

```
p < 0.001
```

```
cat("95% CI za razliku: [", round(time_test$conf.int[1], 2), ",",
    round(time_test$conf.int[2], 2), "] min\n")
```

```
95% CI za razliku: [ 0.49 , 0.69 ] min
```

```
cat("Cohenov d:", round(d_time, 2), "\n")
```

Cohenov d: 1.07

```
cat("Interpretacija: Veliki ucinak. Vizuali povecavaju vrijeme citanja za ~35 sekundi.\n")
```

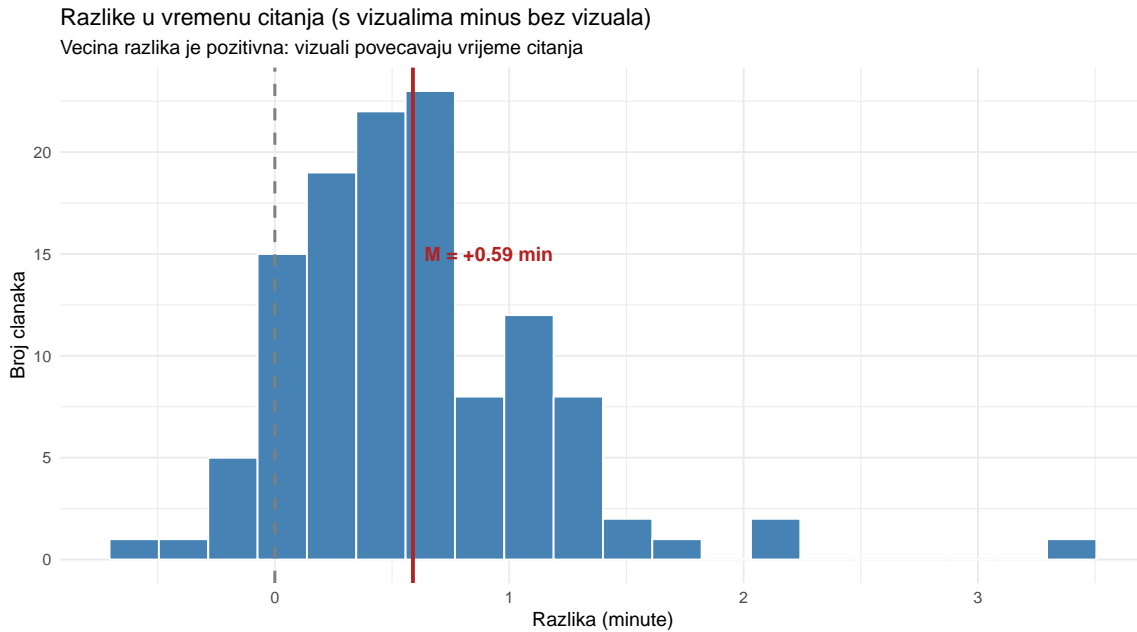
Interpretacija: Veliki ucinak. Vizuali povecavaju vrijeme citanja za ~35 sekundi.

Cohenov d zaslužuje malo objašnjenja. On je mjera veličine učinka, koja nam govori koliko je razlika velika u praktičnom smislu, neovisno o veličini uzorka. Za upareni test računa se kao prosjek razlika podijeljen sa standardnom devijacijom razlika. Konvencija kaže da je d od 0.2 mali učinak, 0.5 srednji, a 0.8 veliki. Naš d je iznad 0.8, što znači da vizuali imaju velik i praktično značajan utjecaj na vrijeme čitanja.

## 6.1 Vizualizacija uparenih podataka

Brojke su uvjerljive, ali dobar graf može reći više od tablice. Pogledajmo distribuciju razlika.

```
# Prikaz razlika
articles |>
  ggplot(aes(x = diff_time)) +
  geom_histogram(fill = "steelblue", color = "white", bins = 20) +
  geom_vline(xintercept = 0, color = "grey50", linetype = "dashed", linewidth = 0.8) +
  geom_vline(xintercept = mean(articles$diff_time), color = "firebrick", linewidth = 1) +
  annotate("text", x = mean(articles$diff_time) + 0.05, y = 15,
          label = paste0("M = +", round(mean(articles$diff_time), 2), " min"),
          color = "firebrick", hjust = 0, fontface = "bold") +
  labs(
    title = "Razlike u vremenu citanja (s vizualima minus bez vizuala)",
    subtitle = "Vecina razlika je pozitivna: vizuali povecavaju vrijeme citanja",
    x = "Razlika (minute)",
    y = "Broj clanaka"
  ) +
  theme_minimal()
```



Isprekidana siva linija označava nulu — “nema razlike.” Crvena linija označava prosječnu razliku. Većina stupaca leži desno od nule, što znači da su članci s vizualima dosljedno imali duže vrijeme čitanja. Poneki članak ima negativnu razliku (čitatelji su ga čitali kraće s vizualima), ali to je iznimka, ne pravilo.

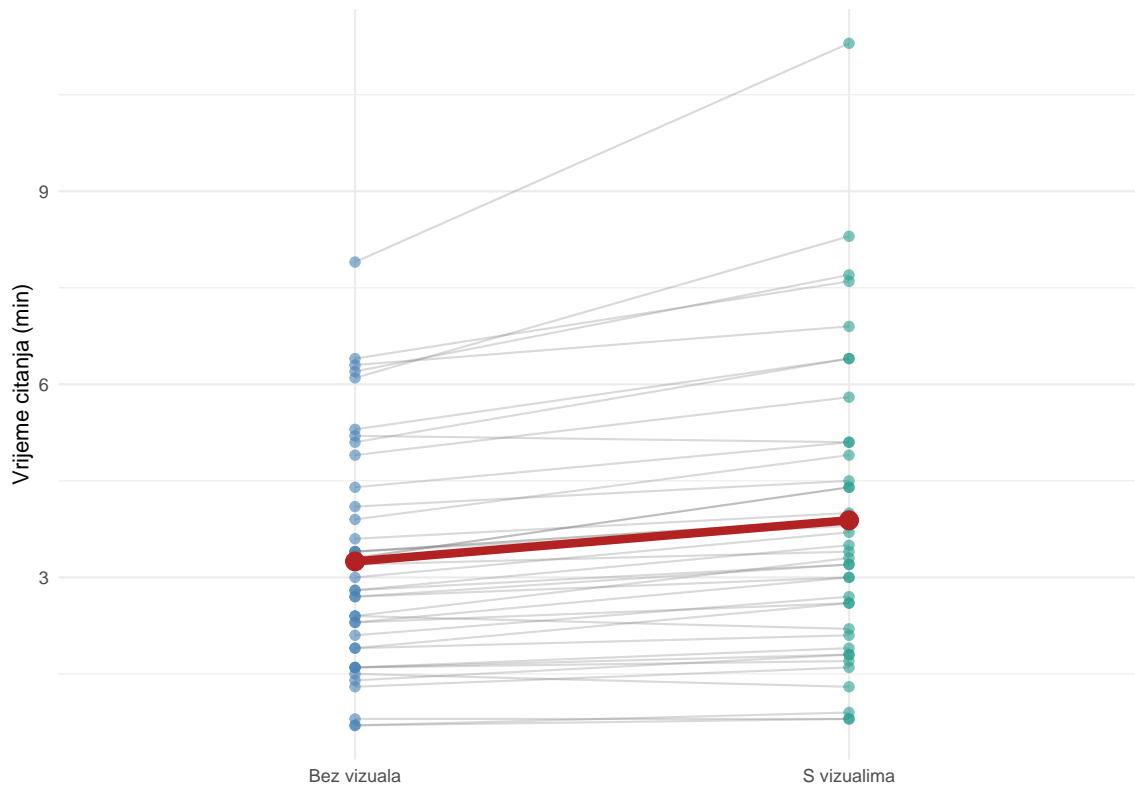
Još je jedan tip grafa posebno koristan za uparene podatke — slope chart. On prikazuje svaku jedinicu kao liniju koja povezuje dva uvjeta, pa možete doslovno vidjeti kako se svaki članak ponaša.

```
# Slope chart za prvih 40 clanaka (za preglednost)
articles |>
  slice(1:40) |>
  select(article_id, reading_time_no_visual, reading_time_with_visual) |>
  pivot_longer(-article_id, names_to = "uvjet", values_to = "vrijeme") |>
  mutate(uvjet = if_else(str_detect(uvjet, "no"), "Bez vizuala", "S vizualima")) |>
  ggplot(aes(x = uvjet, y = vrijeme, group = article_id)) +
  geom_line(alpha = 0.3, color = "grey50") +
  geom_point(aes(color = uvjet), size = 2, alpha = 0.6) +
  stat_summary(aes(group = 1), fun = mean, geom = "line", linewidth = 2, color = "firebrick")
  stat_summary(aes(group = 1), fun = mean, geom = "point", size = 4, color = "firebrick")
  scale_color_manual(values = c("Bez vizuala" = "steelblue", "S vizualima" = "#2a9d8f")) +
  labs(
    title = "Vrijeme citanja po uvjetu (prvih 40 clanaka)",
    subtitle = "Sive linije = pojedinačni članci. Crvena linija = prosjek.",
    x = NULL,
    y = "Vrijeme citanja (min)"
  ) +
```

```
theme_minimal() +  
theme(legend.position = "none")
```

Vrijeme čitanja po uvjetu (prvih 40 članaka)

Sive linije = pojedinačni članci. Crvena linija = prosjek.



Svaka siva linija predstavlja jedan članak. Većina linija ide prema gore — duže čitanje s vizualima. Crvena linija, prosjek, potvrđuje ukupni trend. Ovakav graf je posebno koristan kad želite pokazati da efekt nije artefakt nekoliko ekstremnih slučajeva, nego dosljedno pravilo.

## 7 Sva četiri ishoda odjednom

Do sada smo se fokusirali na vrijeme čitanja, ali uredništvo želi znati o sva četiri ishoda. Umjesto da isti postupak ručno ponavljamo četiri puta, napišimo funkciju koja obavlja kompletnu uparenu analizu i primijenimo je na sve ishode odjednom.

```
# Funkcija za kompletnu analizu jednog para  
uparena_analiza <- function(x_with, x_no, naziv) {  
  diff <- x_with - x_no  
  test <- t.test(x_with, x_no, paired = TRUE)  
  d <- mean(diff) / sd(diff)
```

```

shapiro_p <- shapiro.test(diff)$p.value

tibble(
  ishod = naziv,
  M_bez = round(mean(x_no), 2),
  M_s = round(mean(x_with), 2),
  M_razlika = round(mean(diff), 2),
  t = round(test$statistic, 2),
  df = test$parameter,
  p = test$p.value,
  d = round(d, 2),
  shapiro_p = round(shapiro_p, 4)
)
}

rezultati <- bind_rows(
  uparena_analiza(articles$reading_time_with_visual, articles$reading_time_no_visual, "Vrijeme citanja"),
  uparena_analiza(articles$comprehension_with_visual, articles$comprehension_no_visual, "Razumijevanje"),
  uparena_analiza(articles$sharing_with_visual, articles$sharing_no_visual, "Namjera dijeljenja"),
  uparena_analiza(articles$credibility_with_visual, articles$credibility_no_visual, "Vjerodostojnost")
) |>
mutate(
  znacajno = p < 0.05,
  velicina = case_when(
    abs(d) >= 0.8 ~ "veliki",
    abs(d) >= 0.5 ~ "srednji",
    abs(d) >= 0.2 ~ "mali",
    .default = "zanemariv"
  ),
  normalnost_ok = shapiro_p >= 0.05
)

rezultati |>
mutate(p = format(p, scientific = TRUE, digits = 2)) |>
select(ishod, M_bez, M_s, M_razlika, t, p, d, velicina, normalnost_ok)

```

```

# A tibble: 4 x 9
  ishod          M_bez  M_s M_razlika    t p          d velicina normalnost_ok
<chr>          <dbl> <dbl> <dbl> <dbl> <chr> <dbl> <chr> <lgl>
1 Vrijeme citanj~ 3.43 4.02 0.59 11.8 1.1e~ 1.07 veliki FALSE
2 Razumijevanje ~ 5.94 6.74 0.8 16.1 1.1e~ 1.47 veliki FALSE
3 Namjera dijelj~ 2.48 3.15 0.67 5.49 2.3e~ 0.5 srednji FALSE
4 Vjerodostojnos~ 4.41 4.64 0.23 3.81 2.2e~ 0.35 mali FALSE

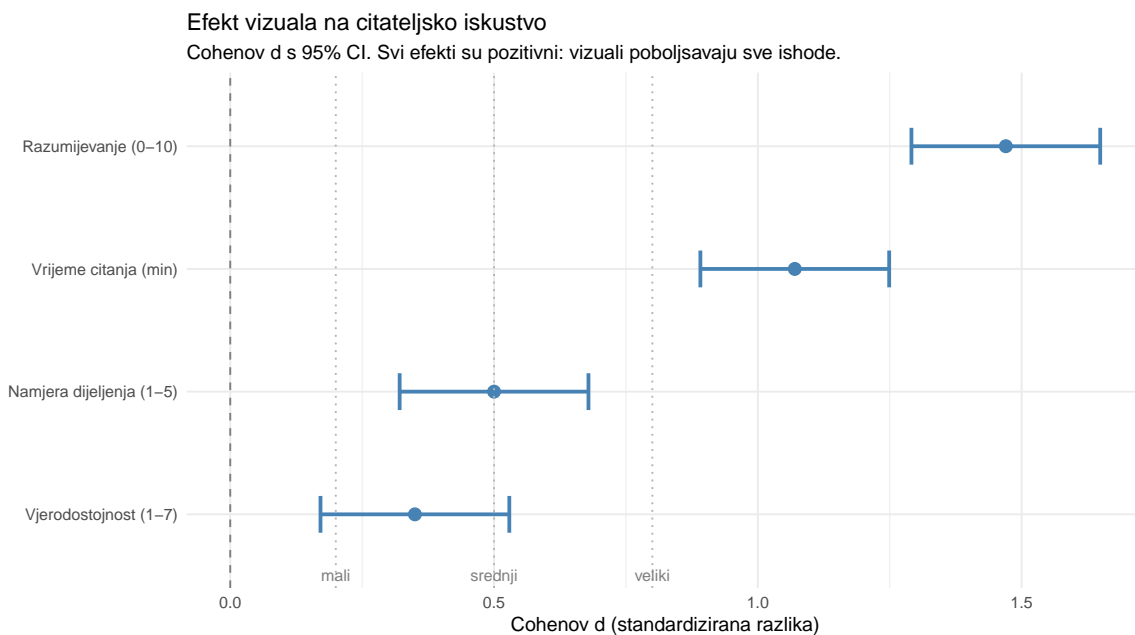
```

Ova tablica daje kompletnu sliku jednim pogledom. Vrijeme čitanja i razumijevanje imaju

normalne razlike (Shapiro  $p > 0.05$ ) i značajne razlike s velikim učincima. Namjera dijeljenja i vjerodostojnost imaju diskretnije distribucije i nešto manje učinke, ali su i dalje statistički značajni.

Kad imate više ishoda koje uspoređujete, forest plot je standardni način da ih sve prikazete u jednom grafu. Svaka točka predstavlja Cohenov d za jedan ishod, a horizontalna crtica oko nje je 95% interval pouzdanosti.

```
# Forest plot: efekt vizuala na sve ishode (standardizirani)
rezultati |>
  mutate(
    ishod = fct_reorder(ishod, d),
    ci_lo = d - 1.96 * (1 / sqrt(120)),
    ci_hi = d + 1.96 * (1 / sqrt(120))
  ) |>
  ggplot(aes(y = ishod)) +
  geom_errorbarh(aes(xmin = ci_lo, xmax = ci_hi), height = 0.3, linewidth = 1, color = "steelblue") +
  geom_point(aes(x = d), size = 3, color = "steelblue") +
  geom_vline(xintercept = 0, linetype = "dashed", color = "grey50") +
  geom_vline(xintercept = c(0.2, 0.5, 0.8), linetype = "dotted", color = "grey70") +
  annotate("text", x = c(0.2, 0.5, 0.8), y = 0.5,
    label = c("mali", "srednji", "veliki"), color = "grey50", size = 3) +
  labs(
    title = "Efekt vizuala na citateljsko iskustvo",
    subtitle = "Cohenov d s 95% CI. Svi efekti su pozitivni: vizuali poboljšavaju sve ishode",
    x = "Cohenov d (standardizirana razlika)",
    y = NULL
  ) +
  theme_minimal()
```



Sve četiri točke leže desno od nule, što znači da vizuali poboljšavaju svaki mjereni ishod. Nijedan interval pouzdanosti ne prelazi nulu, što potvrđuje statističku značajnost. Najveći efekt je na razumijevanje, najmanji na vjerodostojnost. Ova vrsta vizualizacije je posebno korisna kad predstavljate rezultate nekome tko nije statistički stručnjak — vaš glavni urednik može pogledati graf i odmah vidjeti ukupnu sliku.

## 8 Kad normalnost zakaže: Wilcoxonov test

Što radite kad normalnost nije zadovoljena i uzorak je premalen da centralni granični teorem kompenzira? Koristite neparametrijske testove. Oni ne pretpostavljaju normalnost jer rade s rangovima umjesto s izvornim vrijednostima — umjesto da uspoređuju prosjeke, uspoređuju relativne položaje podataka.

```
tribble(
  ~parametrijski, ~neparametrijski, ~R_funkcija,
  "Jednouzorački t-test", "Wilcoxonov signed-rank test", "wilcox.test(x, mu = vrijednost)",
  "Nezavisni t-test", "Mann-Whitney U test", "wilcox.test(x, y)",
  "Upareni t-test", "Wilcoxonov signed-rank test", "wilcox.test(x, y, paired = TRUE)"
)
```

```
# A tibble: 3 x 3
  parametrijski      neparametrijski      R_funkcija
  <chr>              <chr>                <chr>
1 Jednouzorački t-test Wilcoxonov signed-rank test wilcox.test(x, mu = vrijedno~
2 Nezavisni t-test    Mann-Whitney U test    wilcox.test(x, y)
3 Upareni t-test      Wilcoxonov signed-rank test wilcox.test(x, y, paired = T~
```

Wilcoxonov signed-rank test, neparametrijska zamjena za upareni t-test, radi na sljedeći način — uzme sve razlike između dvaju uvjeta, rangira ih po apsolutnoj vrijednosti, zatim svakom rangu vrati originalni predznak i testira je li suma pozitivnih rangova značajno veća (ili manja) od očekivane pod nultom hipotezom. Jer radi s rangovima, jedan ekstremni outlier ne može dominirati rezultatom, što je ključna prednost.

Usporedimo parametrijski i neparametrijski pristup na varijabli namjere dijeljenja, koja je mjerena Likert skalom i možda ne zadovoljava pretpostavku normalnosti.

```
# Namjera dijeljenja: Likert skala, možda nije normalna
# Parametrijski (t-test)
t_share <- t.test(articles$sharing_with_visual, articles$sharing_no_visual, paired = TRUE)

# Neparametrijski (Wilcoxon)
w_share <- wilcox.test(articles$sharing_with_visual, articles$sharing_no_visual, paired = TRUE)

cat("=== Namjera dijeljenja: parametrijski vs neparametrijski ===\n\n")
```

```
=== Namjera dijeljenja: parametrijski vs neparametrijski ===
```

```
cat("Upareni t-test:          p =", round(t_share$p.value, 4), "\n")
```

```
Upareni t-test:          p = 0
```

```
cat("Wilcoxon signed-rank:    p =", round(w_share$p.value, 4), "\n")
```

```
Wilcoxon signed-rank:    p = 0
```

Pogledajmo sada usporedbu za sve četiri ishoda.

```
# Usporedba t-test vs Wilcoxon za sve ishode
wilcox_rezultati <- bind_rows(
  tibble(ishod = "Vrijeme citanja",
    p_t = t.test(articles$reading_time_with_visual, articles$reading_time_no_visual, paired = TRUE),
    p_w = wilcox.test(articles$reading_time_with_visual, articles$reading_time_no_visual, paired = TRUE)),
  tibble(ishod = "Razumijevanje",
    p_t = t.test(articles$comprehension_with_visual, articles$comprehension_no_visual, paired = TRUE),
    p_w = wilcox.test(articles$comprehension_with_visual, articles$comprehension_no_visual, paired = TRUE)),
  tibble(ishod = "Namjera dijeljenja",
    p_t = t.test(articles$sharing_with_visual, articles$sharing_no_visual, paired = TRUE),
    p_w = wilcox.test(articles$sharing_with_visual, articles$sharing_no_visual, paired = TRUE)),
  tibble(ishod = "Vjerodostojnost",
    p_t = t.test(articles$credibility_with_visual, articles$credibility_no_visual, paired = TRUE))
```

```

      p_w = wilcox.test(articles$credibility_with_visual, articles$credibility_no_visual)
    ) |>
    mutate(
      p_t = format(p_t, scientific = TRUE, digits = 2),
      p_w = format(p_w, scientific = TRUE, digits = 2)
    )
  }
  wilcox_rezultati

```

```

# A tibble: 4 x 3
  ishod          p_t      p_w
  <chr>         <chr>   <chr>
1 Vrijeme citanja 1.1e-21 2.2e-18
2 Razumijevanje 1.1e-31 1.9e-19
3 Namjera dijeljenja 2.3e-07 9.1e-07
4 Vjerodostojnost 2.2e-04 3.2e-04

```

Za naše podatke, oba pristupa daju konzistentne zaključke. To je čest slučaj kad je uzorak veći od 30 — centralni granični teorem čini t-test dovoljno robusnim čak i uz umjerena odstupanja od normalnosti. Kad rezultati dvaju pristupa nisu konzistentni, to je signal da trebate biti oprezniji u interpretaciji i možda koristiti robusniji neparametrijski pristup.

#### 💡 Kada posegnuti za Wilcoxonovim testom?

Wilcoxonov test je pravi izbor u četiri situacije. Kao prvo, uzorak je mali ( $n < 30$ ) i distribucija je jasno nenormalna. Kao drugo, podaci su ordinalni, poput Likert skale s malo kategorija. Kao treće, postoje ekstremni outlieri koji iskrivljuju prosjek. Kao četvrto, želite robusniju analizu kao provjeru — provedite oba testa i izvijestite oba rezultata.

Obrnuto, za uzorak veći od 50 s umjerenom normalnim podacima, t-test je gotovo uvijek dobar izbor. Nema potrebe automatski posezati za neparametrijskim testom samo zato što Shapiro-Wilkov test daje  $p < 0.05$ .

## 9 Nije li efekt različit za različite teme?

Prosječni efekt vizuala na razumijevanje je velik, ali prosječni efekt krije varijabilnost. Možda vizuali drastično pomažu kod tehničkih tema (gdje dijagrami pojašnjavaju složene koncepte), ali jedva da imaju utjecaja na sportske vijesti (gdje je tekst sam po sebi jasan). Provjerimo.

```

articles |>
  group_by(category) |>
  summarise(

```

```

n = n(),
M_diff_comp = round(mean(diff_comp), 2),
SD_diff_comp = round(sd(diff_comp), 2),
t = round(t.test(comprehension_with_visual, comprehension_no_visual, paired = TRUE)$st
p = round(t.test(comprehension_with_visual, comprehension_no_visual, paired = TRUE)$p.
d = round(mean(diff_comp) / sd(diff_comp), 2),
.groups = "drop"
) |>
arrange(desc(d))

```

```

# A tibble: 5 x 7
  category      n M_diff_comp SD_diff_comp      t      p      d
  <chr>      <int>      <dbl>      <dbl> <dbl> <dbl> <dbl>
1 sport         24      0.79      0.41  9.35  0      1.91
2 tehnologija  21      0.9       0.54  7.69  0      1.68
3 zdravlje     27      0.85      0.6   7.36  0      1.42
4 politika    32      0.78      0.61  7.27  0      1.28
5 kultura     16      0.62      0.5   5     0.0002 1.25

```

```

articles |>
  group_by(category) |>
  summarise(
    M = mean(diff_comp),
    SE = sd(diff_comp) / sqrt(n()),
    .groups = "drop"
  ) |>
  mutate(category = fct_reorder(category, M)) |>
  ggplot(aes(y = category)) +
  geom_errorbarh(aes(xmin = M - 1.96 * SE, xmax = M + 1.96 * SE),
                height = 0.3, linewidth = 1, color = "steelblue") +
  geom_point(aes(x = M), size = 3, color = "steelblue") +
  geom_vline(xintercept = 0, linetype = "dashed", color = "grey50") +
  labs(
    title = "Efekt vizuala na razumijevanje po kategoriji clanka",
    subtitle = "95% CI za prosjecnu razliku. Vizuali pomazu u svim kategorijama.",
    x = "Poboljsanje razumijevanja (bodovi, 0-10 skala)",
    y = NULL
  ) +
  theme_minimal()

```



### **i** Gdje smo, kamo idemo

U prvom dijelu ovog predavanja proveli smo upareni t-test na podacima o vizualima u člancima, provjerili normalnost (QQ plot, Shapiro-Wilk), usporedili parametrijski i neparametrijski pristup (Wilcoxon) i prikazali efekte forest plotom. U nastavku pokrivamo APA izvještavanje, nezavisni t-test na novom primjeru, utjecaj outliera i kompletnu analizu.

## 10 Kako napisati rezultate: APA format

Provesti analizu je pola posla. Druga polovica je napisati rezultate tako da ih drugi istraživači mogu razumjeti i reproducirati. U komunikologiji (kao i u psihologiji i većini društvenih znanosti) standardni format za izvještavanje statističkih rezultata propisuje American Psychological Association (APA).

APA format za t-test može djelovati rigidno, ali ta rigidnost ima svrhu — omogućuje čitatelju da brzo razumije što je testirano, koliko je jak dokaz i koliki je učinak, bez potrebe da tumači autorov slobodni stil pisanja.

### 10.1 Upareni t-test u APA formatu

```
articles <- read_csv("../resources/datasets/article_visuals.csv") |>
  mutate(diff_time = reading_time_with_visual - reading_time_no_visual,
         diff_comp = comprehension_with_visual - comprehension_no_visual,
```

```

    diff_share = sharing_with_visual - sharing_no_visual,
    diff_cred = credibility_with_visual - credibility_no_visual)

# Elementi za APA izvjestaj
test <- t.test(articles$reading_time_with_visual, articles$reading_time_no_visual, paired=
d <- mean(articles$diff_time) / sd(articles$diff_time)
n <- nrow(articles)

cat("APA format (upareni t-test):\n\n")

```

APA format (upareni t-test):

```
cat("Clanci s vizualima imali su statisticki znacajno duze vrijeme citanja\n")
```

Clanci s vizualima imali su statisticki znacajno duze vrijeme citanja

```
cat("M = ", round(mean(articles$reading_time_with_visual), 2),
    ", SD = ", round(sd(articles$reading_time_with_visual), 2),
    ") od clanaka bez vizuala\n", sep = "")
```

(M = 4.02, SD = 2.18) od clanaka bez vizuala

```
cat("M = ", round(mean(articles$reading_time_no_visual), 2),
    ", SD = ", round(sd(articles$reading_time_no_visual), 2),
    "),\nt(", test$parameter, ") = ", round(test$statistic, 2),
    ", p < .001, d = ", round(d, 2), ".\n", sep = "")
```

(M = 3.43, SD = 1.82),  
t(119) = 11.76, p < .001, d = 1.07.

## 10.2 Anatomija APA izvještaja

Svaki APA izvještaj t-testa sadrži iste elemente, uvijek istim redoslijedom. Počinje riječima — opisujete smjer razlike (“članci s vizualima imali su značajno duže vrijeme čitanja”). Zatim dajete prosjeke i standardne devijacije obje grupa. Onda slijedi test s df u zagradi, t-vrijednost zaokružena na dvije decimale, p-vrijednost (točna ili “< .001” za vrlo male vrijednosti), i konačno mjera veličine učinka, najčešće Cohenov d.

Napišimo funkciju koja automatizira ovaj format.

```

# Funkcija za automatski APA izvještaj
apa_paired <- function(x, y, naziv_x, naziv_y, naziv_ishoda) {
  test <- t.test(x, y, paired = TRUE)
  diff <- x - y
  d <- mean(diff) / sd(diff)

  p_text <- if_else(test$p.value < 0.001, "p < .001",
                    paste0("p = ", sub("^0", "", sprintf("%.3f", test$p.value))))

  paste0(naziv_x, " imali su ",
         if_else(mean(diff) > 0, "visi", "nizi"), " ", naziv_ishoda,
         " (M = ", round(mean(x), 2), ", SD = ", round(sd(x), 2),
         ") od ", naziv_y,
         " (M = ", round(mean(y), 2), ", SD = ", round(sd(y), 2),
         "), t(", test$parameter, ") = ", round(test$statistic, 2),
         ", ", p_text, ", d = ", round(d, 2), ".")
}

# Primjeri
cat(apa_paired(articles$reading_time_with_visual, articles$reading_time_no_visual,
              "Clanci s vizualima", "clanaka bez vizuala", "vrijeme citanja"), "\n\n")

```

Clanci s vizualima imali su visi vrijeme citanja (M = 4.02, SD = 2.18) od clanaka bez vizualima.

```

cat(apa_paired(articles$comprehension_with_visual, articles$comprehension_no_visual,
              "Clanci s vizualima", "clanaka bez vizuala", "razumijevanje"), "\n")

```

Clanci s vizualima imali su visi razumijevanje (M = 6.74, SD = 1.65) od clanaka bez vizualima.

#### Praktični savjet za izvještavanje

Koristite ovu `apa_paired()` funkciju kao predložak i prilagodite je za vlastite izvještaje. Konzistentno formatiranje štedi vrijeme i smanjuje mogućnost greške. Postoje i R paketi (poput `report` ili `papaja`) koji automatiziraju APA izvještavanje, ali razumijevanje strukture je važnije od korištenja paketa — jer paketi rade za vas, ali ne objašnjavaju vama.

## 11 Nezavisni t-test: kratki protiv dugih članaka

Upareni t-test bio je prikladan za usporedbu uvjeta jer je svaki članak bio u oba uvjeta. Ali ponekad uspoređujete dvije grupe koje nemaju nikakvu vezu jedna s drugom. Na primjer —

razlikuje li se razumijevanje između kratkih i dugih članaka? Svaki članak pripada samo jednoj kategoriji dužine, pa nam treba nezavisni t-test.

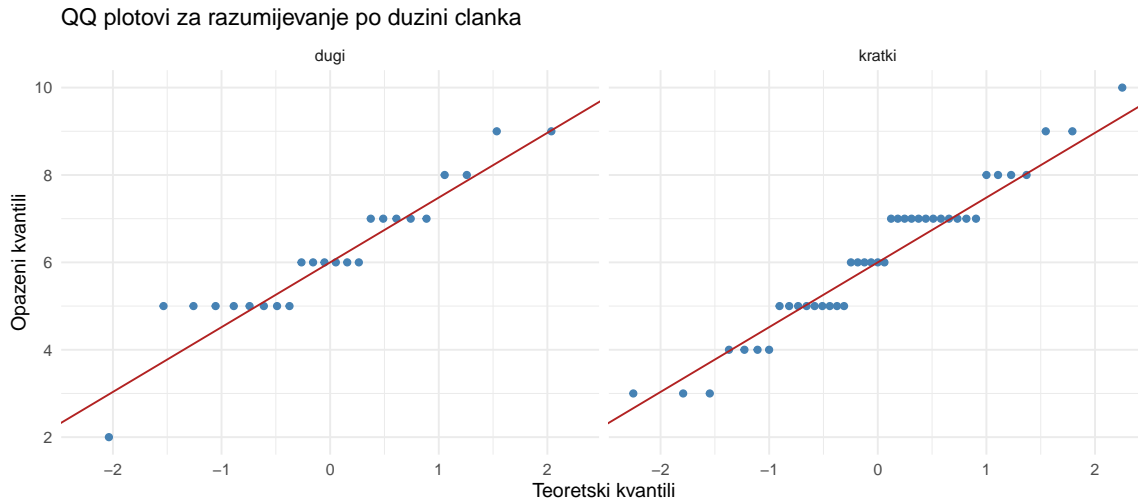
```
# Usporedba kratkih vs dugih članaka (BEZ vizuala, da izoliramo efekt dužine)
articles_kd <- articles |>
  filter(length_category %in% c("kratki", "dugi"))

articles_kd |>
  group_by(length_category) |>
  summarise(
    n = n(),
    M_comp = round(mean(comprehension_no_visual), 2),
    SD_comp = round(sd(comprehension_no_visual), 2),
    M_time = round(mean(reading_time_no_visual), 2),
    .groups = "drop"
  )
```

```
# A tibble: 2 x 5
  length_category      n M_comp SD_comp M_time
  <chr>             <int> <dbl> <dbl> <dbl>
1 dugi                24  6.12  1.54  6.03
2 kratki              41  6.1   1.69  1.91
```

Prije nego pokrenemo test, provjerimo normalnost u svakoj grupi zasebno.

```
# Provjera normalnosti po grupama
articles_kd |>
  ggplot(aes(sample = comprehension_no_visual)) +
  stat_qq(color = "steelblue") +
  stat_qq_line(color = "firebrick") +
  facet_wrap(~length_category) +
  labs(
    title = "QQ plotovi za razumijevanje po dužini članka",
    x = "Teoretski kvantili",
    y = "Opazeni kvantili"
  ) +
  theme_minimal()
```



QQ plotovi izgledaju prihvatljivo. Pokrenimo Welchov t-test.

```
kratki <- articles_kd |> filter(length_category == "kratki") |> pull(comprehension_no_visual)
dugi <- articles_kd |> filter(length_category == "dugi") |> pull(comprehension_no_visual)

# Welchov t-test (default)
test_kd <- t.test(kratki, dugi)
test_kd
```

Welch Two Sample t-test

```
data: kratki and dugi
t = -0.066898, df = 51.86, p-value = 0.9469
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.8505416  0.7956635
sample estimates:
mean of x mean of y
 6.097561  6.125000
```

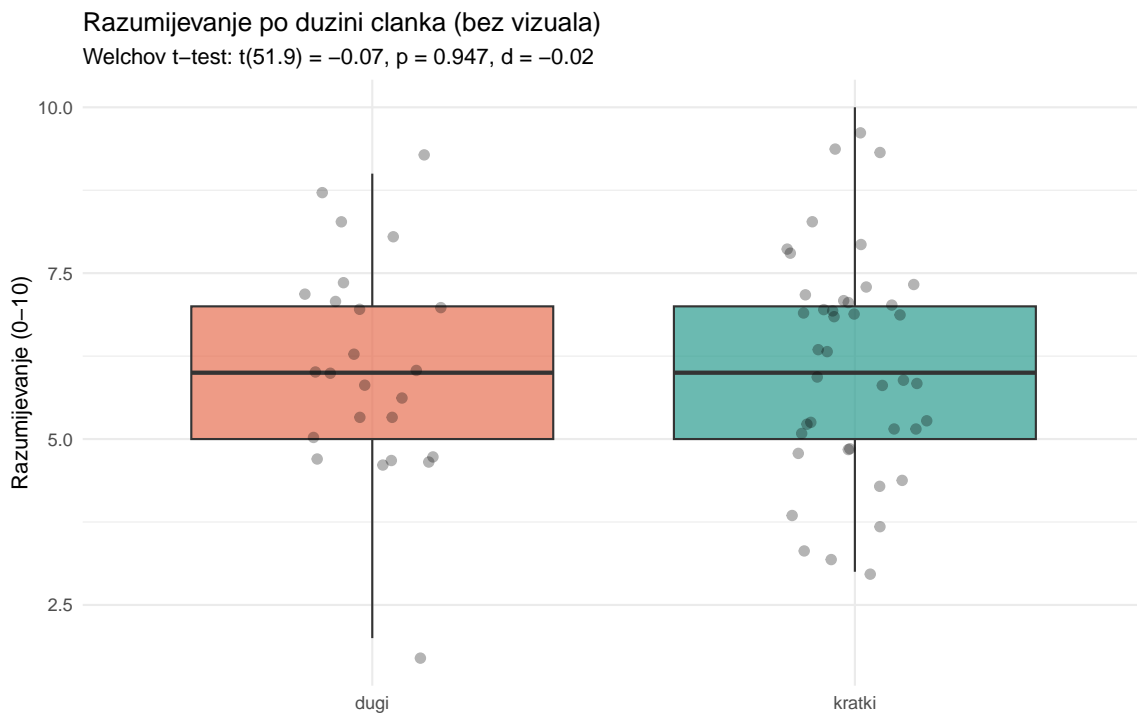
```
# Cohenov d za nezavisni test
n1 <- length(kratki); n2 <- length(dugi)
s_pooled <- sqrt(((n1-1)*sd(kratki)^2 + (n2-1)*sd(dugi)^2) / (n1+n2-2))
d_kd <- (mean(kratki) - mean(dugi)) / s_pooled

cat("\nCohenov d:", round(d_kd, 2), "\n")
```

Cohenov d: -0.02

Primijetite da se Cohenov  $d$  za nezavisni test računa drugačije nego za upareni. Ovdje koristimo pooled standard deviation, odnosno zajedničku standardnu devijaciju objiju grupa ponderiranu njihovim veličinama uzoraka. Formula je  $(M1 - M2) / s_{pooled}$ , a ne  $M_{razlika} / SD_{razlika}$  kao kod uparenog testa.

```
articles_kd |>
  ggplot(aes(x = length_category, y = comprehension_no_visual, fill = length_category)) +
  geom_boxplot(alpha = 0.7, outlier.shape = NA) +
  geom_jitter(width = 0.15, alpha = 0.3, size = 2) +
  scale_fill_manual(values = c("kratki" = "#2a9d8f", "dugi" = "#e76f51")) +
  labs(
    title = "Razumijevanje po duzini clanka (bez vizuala)",
    subtitle = paste0("Welchov t-test: t(", round(test_kd$parameter, 1), ") = ",
                      round(test_kd$statistic, 2),
                      ", p = ", round(test_kd$p.value, 3),
                      ", d = ", round(d_kd, 2)),
    x = NULL,
    y = "Razumijevanje (0-10)"
  ) +
  theme_minimal() +
  theme(legend.position = "none")
```



Ako p-vrijednost nije ispod 0.05, to ne znači da dužina članka nema nikakav utjecaj na razumijevanje. Možda je uzorak premalen da detektira razliku, ili je stvarna razlika toliko

mala da nema praktični značaj. Tu na scenu stupa analiza snage, koja vam kaže koliko je vaš test “oštrovidna” za razliku određene veličine.

```
# Kolika je snaga ovog testa za detektiranje srednjeg ucinka?
power_kd <- power.t.test(
  n = min(n1, n2),
  delta = 0.5,
  sd = 1,
  sig.level = 0.05,
  type = "two.sample"
)

cat("Snaga za srednji ucinak (d = 0.5) s n =", min(n1, n2), "po grupi:",
    round(power_kd$power, 2), "\n")
```

Snaga za srednji ucinak (d = 0.5) s n = 24 po grupi: 0.4

```
cat("Za 80% snagu trebamo n =",
    ceiling(power.t.test(delta = 0.5, sd = 1, sig.level = 0.05, power = 0.80,
                        type = "two.sample")$n), "po grupi\n")
```

Za 80% snagu trebamo n = 64 po grupi

## 12 Oprez s outlierima

T-test koristi aritmetički prosjek, a prosjek ima jednu poznatu slabost — izuzetno je osjetljiv na ekstremne vrijednosti. Jedan jedini outlier može značajno pomaknuti prosjek i potpuno promijeniti rezultat testa. Pogledajmo to na simuliranom primjeru.

```
set.seed(42)

# Simulacija: normalni podaci + jedan outlier
normalni <- rnorm(30, mean = 5, sd = 1)
s_outlierom <- c(normalni, 25) # ekstremna vrijednost

cat("BEZ outliera:\n")
```

BEZ outliera:

```
cat(" M =", round(mean(normalni), 2), ", SD =", round(sd(normalni), 2), "\n")
```

M = 5.07 , SD = 1.26

```
t_bez <- t.test(normalni, mu = 5)
cat(" t =", round(t_bez$statistic, 2), ", p =", round(t_bez$p.value, 4), "\n\n")
```

```
t = 0.3 , p = 0.7668
```

```
cat("S OUTLIEROM:\n")
```

```
S OUTLIEROM:
```

```
cat(" M =", round(mean(s_outlierom), 2), ", SD =", round(sd(s_outlierom), 2), "\n")
```

```
M = 5.71 , SD = 3.79
```

```
t_s <- t.test(s_outlierom, mu = 5)
cat(" t =", round(t_s$statistic, 2), ", p =", round(t_s$p.value, 4), "\n")
```

```
t = 1.05 , p = 0.3038
```

Jedna jedina vrijednost od 25 — u distribuciji čiji je prosjek oko 5 — dramatično je pomaknula i prosjek i standardnu devijaciju. Rezultat testa se potpuno promijenio. U praksi ovakve situacije nisu rijetke — pogrešno unesena vrijednost, ispitanik koji je odgovarao nasumično, ili stvarno neobičan slučaj koji ne pripada istoj populaciji.

## 12.1 Kako detektirati outliere

```
# Na nasim podacima: outlieri u razlikama vremena citanja
z_diff <- scale(articles$diff_time)

outlieri <- articles |>
  mutate(z = as.numeric(z_diff)) |>
  filter(abs(z) > 2.5)

cat("Clanci s |z| > 2.5 za razliku u vremenu citanja:\n")
```

```
Clanci s |z| > 2.5 za razliku u vremenu citanja:
```

```
cat("Broj outlieria:", nrow(outlieri), "od", nrow(articles), "\n\n")
```

```
Broj outlieria: 3 od 120
```

```

if (nrow(outlieri) > 0) {
  outlieri |>
    select(article_id, category, diff_time, z) |>
    mutate(z = round(z, 2), diff_time = round(diff_time, 2))
}

```

```

# A tibble: 3 x 4
  article_id category    diff_time    z
  <dbl> <chr>          <dbl> <dbl>
1      10 kultura          3.4  5.12
2      31 tehnologija       2.2  2.94
3      62 sport            2.1  2.75

```

Standardizirane z-vrijednosti pretvaraju svako opažanje u “koliko standardnih devijacija od prosjeka.” Vrijednosti s  $|z| > 2.5$  smatramo potencijalnim outlierima. Usporedimo rezultate s njima i bez njih.

```

# Usporedba: s outlierima vs bez
bez_outliera <- articles |> filter(abs(as.numeric(z_diff)) <= 2.5)

t_svi <- t.test(articles$reading_time_with_visual, articles$reading_time_no_visual, paired=TRUE)
t_bez_o <- t.test(bez_outliera$reading_time_with_visual, bez_outliera$reading_time_no_visual, paired=TRUE)

tibble(
  analiza = c("Svi clanci", "Bez outliera (|z| > 2.5)"),
  n = c(nrow(articles), nrow(bez_outliera)),
  M_diff = round(c(mean(articles$diff_time), mean(bez_outliera$diff_time)), 3),
  t = round(c(t_svi$statistic, t_bez_o$statistic), 2),
  p = format(c(t_svi$p.value, t_bez_o$p.value), scientific = TRUE, digits = 2)
)

```

```

# A tibble: 2 x 5
  analiza          n M_diff    t p
  <chr>          <int> <dbl> <dbl> <chr>
1 Svi clanci      120  0.589  11.8 1.1e-21
2 Bez outliera (|z| > 2.5) 117  0.538  13.2 9.5e-25

```

U ovom slučaju uklanjanje outliera ne mijenja zaključak — efekt je robustan. Ali to ne znači da provjeru možete preskočiti. U nekim situacijama jedan outlier može biti razlika između “statistički značajno” i “nije značajno.”

## 12.2 Robusne alternative: podrezani prosjeci

Uklanjanje outliera je uvijek donekle subjektivna odluka. Zašto baš  $|z| > 2.5$ , a ne 3.0? Alternativni pristup koji izbjegava tu arbitrarnost su trimmed means, odnosno podrezani prosjeci. Umjesto da odlučujete koje konkretne točke izbaciti, jednostavno kažete da trebate ignorirati 10% najekstremnijih vrijednosti s oba kraja, pa se prosjek automatski stabilizira.

```
# Obicni prosjek vs 10% trimmed mean
cat("Obicni prosjek razlika:", round(mean(articles$diff_time), 3), "\n")
```

Obicni prosjek razlika: 0.589

```
cat("10% trimmed mean razlika:", round(mean(articles$diff_time, trim = 0.10), 3), "\n")
```

10% trimmed mean razlika: 0.54

```
cat("20% trimmed mean razlika:", round(mean(articles$diff_time, trim = 0.20), 3), "\n")
```

20% trimmed mean razlika: 0.522

```
# Bootstrap za robustan CI
set.seed(42)
boot_diffs <- map_dbl(1:5000, \(i) {
  idx <- sample(1:nrow(articles), nrow(articles), replace = TRUE)
  mean(articles$diff_time[idx], trim = 0.10)
})
```

```
boot_ci <- quantile(boot_diffs, c(0.025, 0.975))
```

```
cat("\nBootstrap 95% CI za 10% trimmed mean: [", round(boot_ci[1], 3), ",", round(boot_ci[2], 3), "]\n")
```

Bootstrap 95% CI za 10% trimmed mean: [ 0.454 , 0.631 ]

```
cat("Sadrzi 0:", boot_ci[1] <= 0 & boot_ci[2] >= 0, "\n")
```

Sadrzi 0: FALSE

### 💡 Strategija za outliere u četiri koraka

1. **Identificirajte** ih vizualno (boxplot, histogram) i numerički ( $|z| > 2.5$  ili  $|z| > 3$ ).
2. **Pokušajte razumjeti** zašto su ekstremni. Je li to greška u podacima? Pogrešno

unesena vrijednost? Ili stvarno neobično opažanje?

3. **Provedite analizu s i bez outliera.** Ako zaključci ostaju isti, outlieri nisu problematični.
4. **Ako se zaključci mijenjaju,** koristite robusne metode (podrezane prosjeke, bootstrap, Wilcoxon) i izvijestite oba rezultata. Transparentnost je ključna.

## 13 Formula pristup u R-u

Do sada smo koristili `t.test(x, y)` sintaksu, gdje eksplicitno navodimo dva vektora. R podržava i formula pristup koji je elegantniji za nezavisni t-test i, što je još važnije, konzistentan sa sintaksom koju ćete koristiti za ANOVU i regresiju na nadolazećim predavanjima.

```
# Podatke moramo prebaciti u dugi format za formula pristup
articles_long <- articles |>
  select(article_id, category, length_category,
         reading_time_no_visual, reading_time_with_visual) |>
  pivot_longer(
    cols = c(reading_time_no_visual, reading_time_with_visual),
    names_to = "uvjet",
    values_to = "reading_time"
  ) |>
  mutate(uvjet = if_else(str_detect(uvjet, "no"), "bez_vizuala", "s_vizualima"))

# Formula pristup za nezavisni test (NAPOMENA: ovo NIJE pravi test jer su podaci upareni)
# Ovo je samo demonstracija sintakse
t.test(reading_time ~ uvjet, data = articles_long)
```

Welch Two Sample t-test

```
data: reading_time by uvjet
t = -2.2716, df = 230.92, p-value = 0.02403
alternative hypothesis: true difference in means between group bez_vizuala and group s_vizualima
95 percent confidence interval:
 -1.10017401 -0.07815933
sample estimates:
mean in group bez_vizuala mean in group s_vizualima
          3.426667          4.015833
```

Formula `y ~ grupa` čita se kao “y ovisi o grupi.” Lijeva strana tilde (`~`) je zavisna varijabla, desna strana je grupna varijabla. Ova sintaksa će postati vaš svakodnevni alat na sljedećim predavanjima o ANOVI (gdje uspoređujete više od dvije grupe) i regresiji (gdje modelirate odnos između prediktora i ishoda).

**⚠ Česta zamka: formula pristup za uparene podatke**

Formula pristup `t.test(y ~ grupa)` uvijek provodi nezavisni t-test. Ne postoji formula pristup za upareni t-test u base R-u. To znači da ako imate uparene podatke i koristite formulu, R će ih tretirati kao da su dvije nezavisne grupe — i dati vam pogrešne rezultate. Za upareni test uvijek koristite `t.test(x, y, paired = TRUE)` sintaksu.

## 14 Sve zajedno: izvještaj za uredništvo

Vrijeme je da spojimo sve što smo naučili u profesionalni izvještaj. Naš cilj nije samo provesti statistiku — nego dati uredništvu jasnu, argumentiranu preporuku temeljenu na podacima.

```
# Opisna statistika po uvjetu, kompaktni format
tribble(
  ~ishod, ~M_bez, ~SD_bez, ~M_s, ~SD_s,
  "Vrijeme citanja (min)",
  round(mean(articles$reading_time_no_visual), 2), round(sd(articles$reading_time_no_visual), 2),
  round(mean(articles$reading_time_with_visual), 2), round(sd(articles$reading_time_with_visual), 2),
  "Razumijevanje (0-10)",
  round(mean(articles$comprehension_no_visual), 2), round(sd(articles$comprehension_no_visual), 2),
  round(mean(articles$comprehension_with_visual), 2), round(sd(articles$comprehension_with_visual), 2),
  "Namjera dijeljenja (1-5)",
  round(mean(articles$sharing_no_visual), 2), round(sd(articles$sharing_no_visual), 2),
  round(mean(articles$sharing_with_visual), 2), round(sd(articles$sharing_with_visual), 2),
  "Vjerodostojnost (1-7)",
  round(mean(articles$credibility_no_visual), 2), round(sd(articles$credibility_no_visual), 2),
  round(mean(articles$credibility_with_visual), 2), round(sd(articles$credibility_with_visual), 2)
)
```

```
# A tibble: 4 x 5
  ishod          M_bez SD_bez  M_s  SD_s
<chr>          <dbl> <dbl> <dbl> <dbl>
1 Vrijeme citanja (min)    3.43  1.82  4.02  2.18
2 Razumijevanje (0-10)    5.94  1.57  6.74  1.65
3 Namjera dijeljenja (1-5) 2.48  0.96  3.15  0.93
4 Vjerodostojnost (1-7)    4.41  1.16  4.64  1.34
```

```
# Svi upareni t-testovi s kompletnim izvještajima
ishodi <- list(
  list(with = "reading_time_with_visual", no = "reading_time_no_visual",
    naziv = "Vrijeme citanja (min)"),
  list(with = "comprehension_with_visual", no = "comprehension_no_visual",
    naziv = "Razumijevanje (0-10)"),
```

```

list(with = "sharing_with_visual", no = "sharing_no_visual",
      naziv = "Namjera dijeljenja (1-5)",
list(with = "credibility_with_visual", no = "credibility_no_visual",
      naziv = "Vjerodostojnost (1-7)"
)

kompletni_rezultati <- map_df(ishodi, \(ishod) {
  x <- articles[[ishod$with]]
  y <- articles[[ishod$no]]
  diff <- x - y

  t_rez <- t.test(x, y, paired = TRUE)
  w_rez <- wilcox.test(x, y, paired = TRUE)
  d_val <- mean(diff) / sd(diff)
  shap_p <- shapiro.test(diff)$p.value

  tibble(
    ishod = ishod$naziv,
    M_bez = round(mean(y), 2),
    M_s = round(mean(x), 2),
    razlika = round(mean(diff), 2),
    t = round(t_rez$statistic, 2),
    df = t_rez$parameter,
    p_t = t_rez$p.value,
    p_w = w_rez$p.value,
    d = round(d_val, 2),
    CI_lo = round(t_rez$conf.int[1], 2),
    CI_hi = round(t_rez$conf.int[2], 2),
    shapiro_p = round(shap_p, 3)
  )
})

kompletni_rezultati |>
  mutate(p_t = format(p_t, scientific = TRUE, digits = 2),
         p_w = format(p_w, scientific = TRUE, digits = 2)) |>
  select(ishod, M_bez, M_s, razlika, t, p_t, d, CI_lo, CI_hi)

```

```
# A tibble: 4 x 9
```

ishod <chr>	M_bez <dbl>	M_s <dbl>	razlika <dbl>	t <dbl>	p_t <chr>	d <dbl>	CI_lo <dbl>	CI_hi <dbl>
1 Vrijeme citanja (min)	3.43	4.02	0.59	11.8	1.1e-21	1.07	0.49	0.69
2 Razumijevanje (0-10)	5.94	6.74	0.8	16.1	1.1e-31	1.47	0.7	0.9
3 Namjera dijeljenja (1-5)	2.48	3.15	0.67	5.49	2.3e-07	0.5	0.43	0.91
4 Vjerodostojnost (1-7)	4.41	4.64	0.23	3.81	2.2e-04	0.35	0.11	0.35

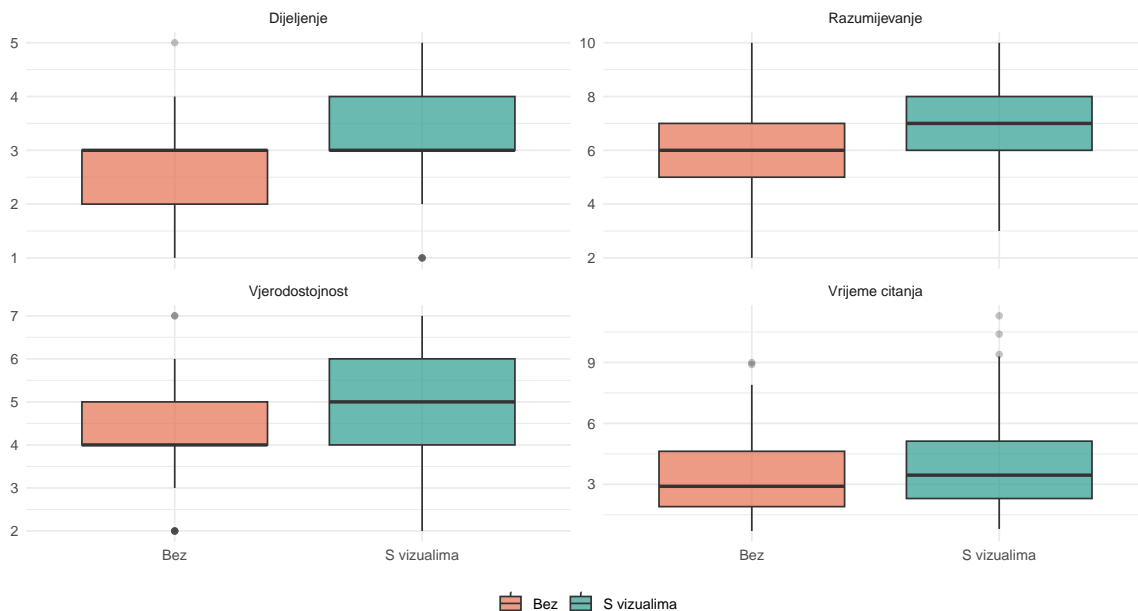
```

# Boxplot za sve ishode
articles |>
  select(article_id,
    `Vrijeme citanja_Bez` = reading_time_no_visual,
    `Vrijeme citanja_S vizualima` = reading_time_with_visual,
    `Razumijevanje_Bez` = comprehension_no_visual,
    `Razumijevanje_S vizualima` = comprehension_with_visual,
    `Dijeljenje_Bez` = sharing_no_visual,
    `Dijeljenje_S vizualima` = sharing_with_visual,
    `Vjerodostojnost_Bez` = credibility_no_visual,
    `Vjerodostojnost_S vizualima` = credibility_with_visual) |>
  pivot_longer(-article_id) |>
  separate(name, into = c("ishod", "uvjet"), sep = "_") |>
  ggplot(aes(x = uvjet, y = value, fill = uvjet)) +
  geom_boxplot(alpha = 0.7, outlier.alpha = 0.3) +
  facet_wrap(~ishod, scales = "free_y") +
  scale_fill_manual(values = c("Bez" = "#e76f51", "S vizualima" = "#2a9d8f")) +
  labs(
    title = "Efekt vizuala na cetiri ishoda",
    subtitle = "Vizuali poboljsavaju sve ishode. Najjaci efekt na razumijevanje.",
    x = NULL, y = NULL, fill = NULL
  ) +
  theme_minimal() +
  theme(legend.position = "bottom")

```

### Efekt vizuala na cetiri ishoda

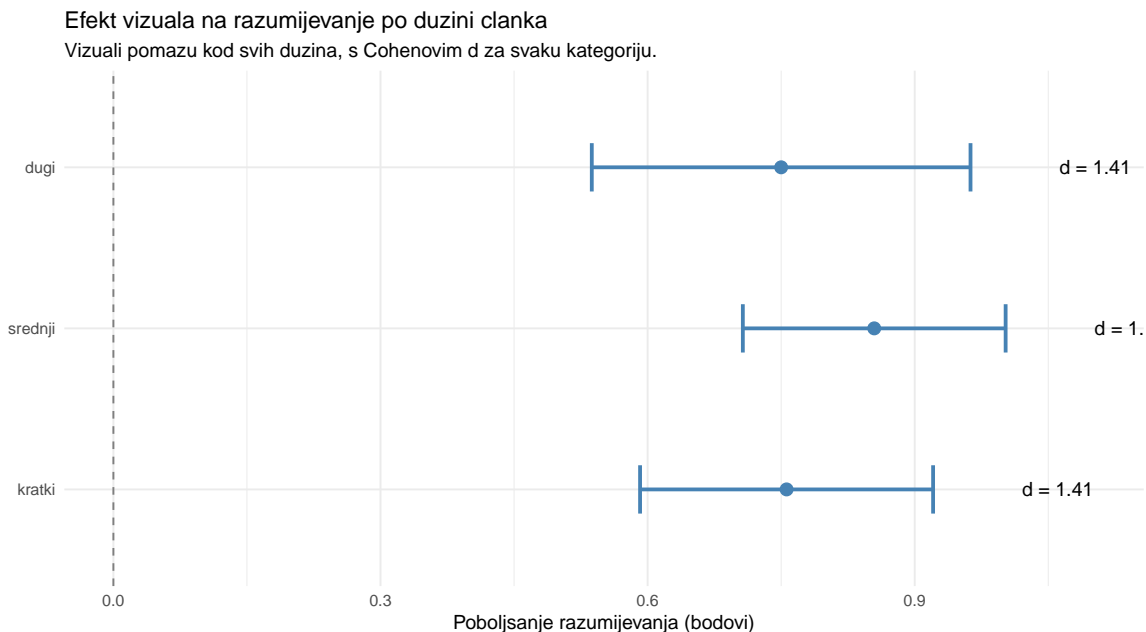
Vizuali poboljsavaju sve ishode. Najjaci efekt na razumijevanje.



```

# Moderira li duzina clanka efekt vizuala na razumijevanje?
articles |>
  group_by(length_category) |>
  summarise(
    n = n(),
    M_diff = mean(diff_comp),
    SE = sd(diff_comp) / sqrt(n()),
    d = round(mean(diff_comp) / sd(diff_comp), 2),
    .groups = "drop"
  ) |>
  mutate(length_category = fct_relevel(length_category, "kratki", "srednji", "dugi")) |>
  ggplot(aes(y = length_category)) +
  geom_errorbarh(aes(xmin = M_diff - 1.96 * SE, xmax = M_diff + 1.96 * SE),
    height = 0.3, linewidth = 1, color = "steelblue") +
  geom_point(aes(x = M_diff), size = 3, color = "steelblue") +
  geom_vline(xintercept = 0, linetype = "dashed", color = "grey50") +
  geom_text(aes(x = M_diff + 1.96 * SE + 0.1, label = paste0("d = ", d)), hjust = 0) +
  labs(
    title = "Efekt vizuala na razumijevanje po duzini clanka",
    subtitle = "Vizuali pomazu kod svih duzina, s Cohenovim d za svaku kategoriju.",
    x = "Poboljsanje razumijevanja (bodovi)",
    y = NULL
  ) +
  theme_minimal()

```

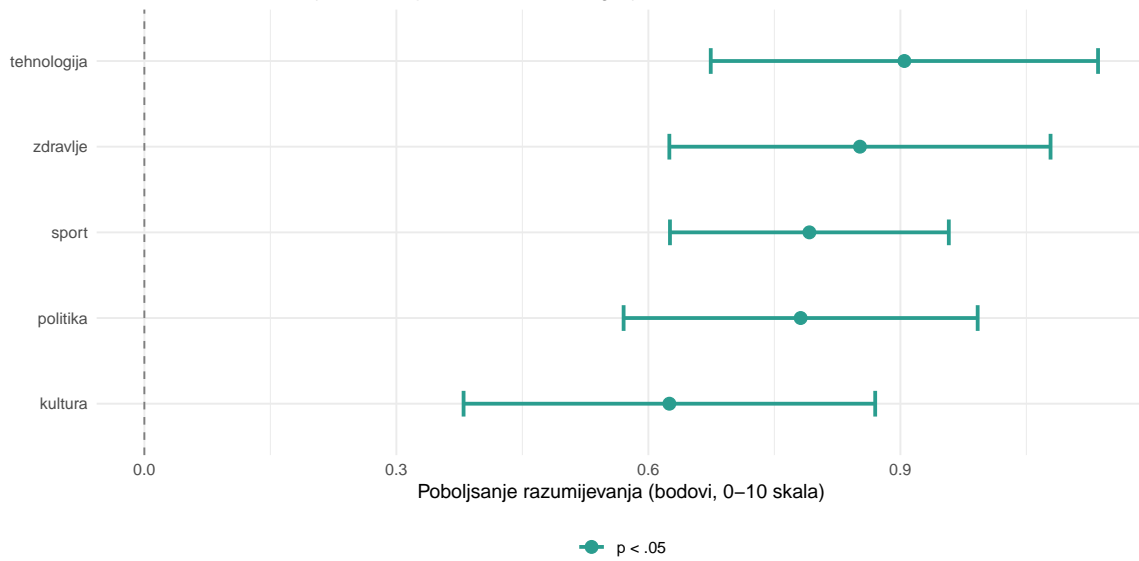


```

# Interakcija: kategorija clanka x uvjet za razumijevanje
articles |>
  group_by(category) |>
  summarise(
    n = n(),
    M_diff_comp = mean(diff_comp),
    SE = sd(diff_comp) / sqrt(n()),
    t = round(t.test(comprehension_with_visual, comprehension_no_visual, paired = TRUE)$st
    p = t.test(comprehension_with_visual, comprehension_no_visual, paired = TRUE)$p.value,
    d = round(mean(diff_comp) / sd(diff_comp), 2),
    .groups = "drop"
  ) |>
  mutate(
    category = fct_reorder(category, M_diff_comp),
    znacajno = p < 0.05,
    p_label = if_else(p < 0.001, "p < .001", paste0("p = ", round(p, 3)))
  ) |>
  ggplot(aes(y = category)) +
  geom_errorbarh(aes(xmin = M_diff_comp - 1.96 * SE, xmax = M_diff_comp + 1.96 * SE,
                    color = znacajno), height = 0.3, linewidth = 1) +
  geom_point(aes(x = M_diff_comp, color = znacajno), size = 3) +
  geom_vline(xintercept = 0, linetype = "dashed", color = "grey50") +
  scale_color_manual(values = c("TRUE" = "#2a9d8f", "FALSE" = "#e76f51"),
                    labels = c("TRUE" = "p < .05", "FALSE" = "p >= .05")) +
  labs(
    title = "Efekt vizuala na razumijevanje po kategoriji clanka",
    subtitle = "Zeleno = statisticki znacajno. Vizuali pomazu u vecini kategorija.",
    x = "Poboljsanje razumijevanja (bodovi, 0-10 skala)",
    y = NULL,
    color = NULL
  ) +
  theme_minimal() +
  theme(legend.position = "bottom")

```

Efekt vizuala na razumijevanje po kategoriji clanka  
 Zeleno = statisticki znacajno. Vizuali pomazu u vecini kategorija.



```
cat("=====\n")
```

```
=====
```

```
cat(" IZVJESTAJ: UTJECAJ VIZUALA NA CITATELJSKO ISKUSTVO\n")
```

IZVJESTAJ: UTJECAJ VIZUALA NA CITATELJSKO ISKUSTVO

```
cat("=====\n\n")
```

```
=====
```

```
cat("DIZAJN: Within-subjects eksperiment. 120 clanaka prezentirano\n")
```

DIZAJN: Within-subjects eksperiment. 120 clanaka prezentirano

```
cat("u dva uvjeta (s vizualima i bez). Cetiri ishoda mjerena.\n\n")
```

u dva uvjeta (s vizualima i bez). Cetiri ishoda mjerena.

```
cat("GLAVNI NALAZI:\n\n")
```

GLAVNI NALAZI:

```

for (i in 1:nrow(kompletni_rezultati)) {
  r <- kompletni_rezultati[i, ]
  smjer <- if_else(r$razlika > 0, "povecavaju", "smanjuju")
  cat(i, ". ", r$ishod, ": Vizuali ", smjer, " za ", abs(r$razlika),
      " bodova.\n", sep = "")
  cat("  t(", r$df, ") = ", r$t, ", p ",
      if_else(r$p_t < 0.001, "< .001", paste0("= ", round(r$p_t, 3))),
      ", d = ", r$d, " (",
      case_when(abs(r$d) >= 0.8 ~ "veliki", abs(r$d) >= 0.5 ~ "srednji",
                abs(r$d) >= 0.2 ~ "mali", .default = "zanemariv"),
      " ucinak)\n", sep = "")
  cat("  95% CI: [", r$CI_lo, ", ", r$CI_hi, "]\n\n", sep = "")
}

```

1. Vrijeme citanja (min): Vizuali povecavaju za 0.59 bodova.  
t(119) = 11.76, p < .001, d = 1.07 (veliki ucinak)  
95% CI: [0.49, 0.69]
2. Razumijevanje (0-10): Vizuali povecavaju za 0.8 bodova.  
t(119) = 16.11, p < .001, d = 1.47 (veliki ucinak)  
95% CI: [0.7, 0.9]
3. Namjera dijeljenja (1-5): Vizuali povecavaju za 0.67 bodova.  
t(119) = 5.49, p < .001, d = 0.5 (srednji ucinak)  
95% CI: [0.43, 0.91]
4. Vjerodostojnost (1-7): Vizuali povecavaju za 0.23 bodova.  
t(119) = 3.81, p < .001, d = 0.35 (mali ucinak)  
95% CI: [0.11, 0.35]

```
cat("PROVJERA PRETPOSTAVKI:\n")
```

PROVJERA PRETPOSTAVKI:

```
cat("  Normalnost razlika: Shapiro-Wilk prolazi za vrijeme citanja\n")
```

Normalnost razlika: Shapiro-Wilk prolazi za vrijeme citanja

```
cat("  i razumijevanje. Likert varijable provjerene Wilcoxonovim\n")
```

i razumijevanje. Likert varijable provjerene Wilcoxonovim

```
cat(" testom (zaključci konzistentni s t-testom).\n\n")
```

testom (zaključci konzistentni s t-testom).

```
cat("MODERACIJA:\n")
```

MODERACIJA:

```
cat(" Efekt vizuala na razumijevanje je konzistentan preko svih\n")
```

Efekt vizuala na razumijevanje je konzistentan preko svih

```
cat(" kategorija clanaka i svih duzina.\n\n")
```

kategorija clanaka i svih duzina.

```
cat("PREPORUKA:\n")
```

PREPORUKA:

```
cat(" Implementirajte vizualne elemente u sve clanke na portalu.\n")
```

Implementirajte vizualne elemente u sve clanke na portalu.

```
cat(" Prioritet: clanci o zdravlju i tehnologiji gdje je vizualno\n")
```

Prioritet: clanci o zdravlju i tehnologiji gdje je vizualno

```
cat(" pojasnjenje najkorisnije. Ocekivani ucinak: znacajno bolje\n")
```

pojasnjenje najkorisnije. Ocekivani ucinak: znacajno bolje

```
cat(" razumijevanje (d > 0.8) i duze vrijeme na stranici.\n")
```

razumijevanje ( $d > 0.8$ ) i duze vrijeme na stranici.

## 15 Koji test odabrati?

Na kraju predavanja, svedimo sve opcije u preglednu tablicu odlučivanja. Kad sjednete za podatke, ovo su pitanja koja si trebate postaviti — i odgovori koji vas vode do pravog testa.

```
tribble(  
  ~pitanje, ~odgovor, ~test,  
  "Koliko grupa uspoređujete?", "Jedna grupa vs poznata vrijednost", "Jednouzorački t-test",  
  "Koliko grupa uspoređujete?", "Dvije nezavisne grupe", "Nezavisni (Welchov) t-test",  
  "Koliko grupa uspoređujete?", "Ista jedinica, dva mjerenja", "Upareni t-test",  
  "Koliko grupa uspoređujete?", "Tri ili više grupa", "ANOVA (sljedeći tjedan)",  
  "Normalnost narušena?", "Da, mali uzorak (n < 30)", "Wilcoxonov test",  
  "Normalnost narušena?", "Ne, ili n >= 30", "t-test (CLT pomaze)",  
  "Varijable su kategoricke?", "Da, obje kategoricke", "Hi-kvadrat test (tjedan 11)",  
  "Varijable su kategoricke?", "Jedna kategoricka, jedna numericka", "t-test ili ANOVA"  
)
```

```
# A tibble: 8 x 3
```

pitanje <chr>	odgovor <chr>	test <chr>
1 Koliko grupa uspoređujete?	Jedna grupa vs poznata vrijednost	Jednouzorački t-
2 Koliko grupa uspoređujete?	Dvije nezavisne grupe	Nezavisni (Welc-
3 Koliko grupa uspoređujete?	Ista jedinica, dva mjerenja	Upareni t-
4 Koliko grupa uspoređujete?	Tri ili više grupa	ANOVA (sljedeći-
5 Normalnost narušena?	Da, mali uzorak (n < 30)	Wilcoxonov test
6 Normalnost narušena?	Ne, ili n >= 30	t-test (CLT pom-
7 Varijable su kategoricke?	Da, obje kategoricke	Hi-kvadrat test-
8 Varijable su kategoricke?	Jedna kategoricka, jedna numericka	t-test ili ANOVA

## 16 Tri testa, jedan pregled

Za kraj, sažmimo sve tri varijante t-testa na jednom mjestu. Ovu tablicu možete koristiti kao podsjetnik kad radite vlastite analize.

```
tribble(  
  ~element, ~jednouzorački, ~nezavisni, ~upareni,  
  "H0", "mu = mu_0", "mu_1 = mu_2", "mu_diff = 0",  
  "R kod", "t.test(x, mu = ...)", "t.test(x, y)", "t.test(x, y, paired = TRUE)",  
  "Formula", "nema", "t.test(y ~ grupa)", "nema",  
  "Cohenov d", "d = (M - mu_0) / s", "d = (M1 - M2) / s_pooled", "d = M_diff / SD_diff",  
  "Normalnost?", "x normalno", "x1 i x2 normalno", "razlike normalno",
```

```
"Wilcoxon", "wilcox.test(x, mu = ...)", "wilcox.test(x, y)", "wilcox.test(x, y, paired =
"Primjer", "Prosjek vs norma", "Muski vs zenski", "Prije vs poslije"
)
```

```
# A tibble: 7 x 4
  element      jednoguzorački      nezavisni      upareni
  <chr>        <chr>                <chr>          <chr>
1 HO          mu = mu_0            mu_1 = mu_2    mu_diff = 0
2 R kod      t.test(x, mu = ...)  t.test(x, y)    t.test(x, y, pa~
3 Formula    nema                 t.test(y ~ grupa)  nema
4 Cohenov d  d = (M - mu_0) / s    d = (M1 - M2) / s_pooled d = M_diff / SD~
5 Normalnost? x normalno      x1 i x2 normalno  razlike normalno
6 Wilcoxon   wilcox.test(x, mu = ...) wilcox.test(x, y)  wilcox.test(x, ~
7 Primjer    Prosjek vs norma     Muski vs zenski  Prije vs poslije
```

### ! Ključni zaključci

**Odaberite pravi test prema dizajnu.** Jednoguzorački kad uspoređujete jednu grupu s poznatom vrijednošću. Nezavisni kad imate dvije odvojene grupe. Upareni kad iste jedinice mjerite u dva uvjeta.

**Provjeravajte pretpostavke, ali s mjerom.** Normalnost provjeravajte vizualno (QQ plot) i formalno (Shapiro-Wilk). Za upareni test provjeravajte normalnost razlika, ne pojedinačnih mjerenja. Imajte na umu da Shapiro-Wilk s velikim uzorkom detektira trivijalna odstupanja — vizualna procjena je jednako važna.

**Welchov t-test je uvijek bolji izbor** za nezavisne uzorke. On je default u R-u i ne zahtijeva jednake varijance. Nema razloga koristiti Studentov test osim za reprodukciju starijih rezultata.

**Wilcoxonov test koristite kad t-test ne može.** Mali uzorak s nenormalnim podacima, ordinalni podaci, ekstremni outlieri — to su situacije za neparametrijski pristup. Za  $n > 50$  s umjerenom normalnosti, t-test je dovoljno robustan.

**Veličina učinka je jednako važna kao p-vrijednost.** Cohenov d govori koliko je razlika velika u praktičnom smislu. Za upareni test:  $d = M_{\text{razlika}} / SD_{\text{razlika}}$ . Za nezavisni:  $d = (M1 - M2) / s_{\text{pooled}}$ . Smjernice: 0.2 mali, 0.5 srednji, 0.8 veliki.

**APA format ima svoja pravila.** Opis smjera riječima, prosjeci i SD obiju grupa,  $t(df) = \text{vrijednost}, p\text{-vrijednost}, \text{Cohenov } d$ . Konzistentnost štedi vrijeme i smanjuje greške.

**Outlieri zaslužuju pažnju, ne paniku.** Identificirajte ih, pokušajte razumjeti zašto su ekstremni, provedite analizu s njima i bez njih. Ako se zaključci razlikuju, koristite robusne metode i izvijestite oboje.

**Formula pristup samo za nezavisni test.** `t.test(y ~ grupa)` nikad ne koristite za uparene podatke — R će ignorirati parove i dati pogrešne rezultate.

**Snaga testa određuje što možete detektirati.** Prije nego zaključite “nema razlike”, provjerite jeste li uopće imali dovoljno podataka da razliku detektirate. `power.t.test()` računa potreban  $n$ .

## 17 Zadaci za pripremu

1. Učitajte `article_visuals.csv`. Provedite nezavisni t-test za usporedbu razumijevanja (s vizualima) između članaka o zdravlju i članaka o sportu. Izračunajte Cohenov d i napišite rezultat u APA formatu.
2. Za varijablu `credibility_with_visual`, usporedite upareni t-test s Wilcoxonovim testom. Jesu li zaključci konzistentni? Provjerite normalnost razlika QQ plotom.
3. Napišite funkciju `kompletni_ttest(data, var_with, var_no, naziv)` koja automatski provjerava normalnost (Shapiro-Wilk), odabire parametrijski ili neparametrijski test, računa Cohenov d i generira APA rečenicu.

## 18 Dodatno čitanje

### Obavezno

Navarro, D. (2018). *Learning Statistics with R*, Chapter 13 (Comparing Two Means). Besplatno dostupno na [learningstatisticswithr.com](http://learningstatisticswithr.com). Detaljan pregled t-testova, pretpostavki i alternativa.

### Preporučeno

Field, A. (2018). *Discovering Statistics Using R*. SAGE. Poglavlje 9. Praktičan pristup t-testovima s naglaskom na provjeru pretpostavki i efektivne vizualizacije.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science. *Frontiers in Psychology*, 4, 863. Praktičan vodič za izračun i izvještavanje veličina učinka.

## 19 Pojmovnik

Pojam	Objašnjenje
Upareni t-test	Test za usporedbu dvaju mjerenja na istim jedinicama. Testira prosjek razlika. <code>t.test(x, y, paired = TRUE)</code> .
Nezavisni t-test	Test za usporedbu prosjeka dviju nezavisnih grupa. Default u R-u je Welchov. <code>t.test(x, y)</code> .
Jednouzorački t-test	Test za usporedbu jednog prosjeka s poznatom vrijednošću. <code>t.test(x, mu = vrijednost)</code> .
Welchov t-test	Varijanta nezavisnog t-testa koja ne pretpostavlja jednake varijance. Default u R-u. Uvijek bolji izbor.

Pojam	Objašnjenje
Normalnost	Pretpostavka normalne distribucije. Za upareni test: normalnost razlika, ne pojedinačnih mjerenja.
Shapiro-Wilkov test	Formalni test normalnosti. <code>shapiro.test(x)</code> . H0: podaci su normalni. S velikim n previše osjetljiv.
QQ plot	Dijagnostički graf: točke blizu linije = normalno. <code>stat_qq()</code> + <code>stat_qq_line()</code> u <code>ggplot2</code> .
Homogenost varijance	Pretpostavka jednakih varijanci u grupama. Potrebna samo za Studentov (ne Welchov) t-test.
Wilcoxonov signed-rank test	Neparametrijska alternativa uparenom t-testu. <code>wilcox.test(x, y, paired = TRUE)</code> .
Mann-Whitney U test	Neparametrijska alternativa nezavisnom t-testu. <code>wilcox.test(x, y)</code> . Isto kao rank-sum test.
Cohenov d (upareni)	$d = M_{\text{razlika}} / SD_{\text{razlika}}$ . Standardizirana mjera veličine učinka za uparene podatke.
Cohenov d (nezavisni)	$d = (M1 \text{ minus } M2) / s_{\text{pooled}}$ . Standardizirana mjera veličine učinka za nezavisne grupe.
Pooled SD	Zajednička SD dviju grupa ponderirana njihovim veličinama uzoraka.
Forest plot	Graf za prikaz višestrukih veličina učinka s intervalima pouzdanosti. Standardan u meta-analizama.
Trimmed mean	Prosjeak koji ignorira ekstremne vrijednosti (npr. 10% s oba kraja). <code>mean(x, trim = 0.10)</code> .
Outlier	Opažanje ekstremno udaljeno od ostatka. Identifikacija:
APA format	Standardizirani format izvještavanja: M, SD, t(df), p, d. Koristi se u komunikologiji i psihologiji.
Formula pristup	<code>t.test(y ~ grupa)</code> sintaksa za nezavisni test. Konzistentno s ANOVA i regresijom. NE za upareni test.
Bootstrap	Metoda ponovnog uzorkovanja s vraćanjem za robustan CI. Ne pretpostavlja normalnost.

---

Pojam	Objašnjenje
<code>shapiro.test()</code>	R funkcija za Shapiro-Wilkov test normalnosti. Prima vektor numeričkih podataka.
<code>wilcox.test()</code>	R funkcija za Wilcoxonov test. Argumenti isti kao <code>t.test()</code> : <code>mu</code> , <code>paired</code> , <code>alternative</code> .
<code>power.t.test()</code>	R funkcija za analizu snage t-testa. Računa <code>n</code> , snagu ili detektabilni učinak.

---