

Tjedan 12: Usporedba više grupa ANOVA-om

Kad t-test nije dovoljan

2025-05-17

Table of contents

1	Motivacija: vjerodostojnost vijesti po izvoru	2
1.1	Problem višestrukih t-testova	2
2	Naši podaci	4
3	Logika ANOVA-e	5
3.1	Dekompozicija varijance: SS	6
4	ANOVA u R-u	9
5	Pretpostavke ANOVA-e	10
5.1	Provjera normalnosti	10
5.2	Provjera homogenosti varijance: Levenov test	12
6	Vizualizacija ANOVA rezultata	14
7	Post-hoc testovi: Tukey HSD	16
7.1	Compact letter display	19
8	Veličina učinka: eta-kvadrat	19
8.1	Interpretacija eta-kvadrata	20
8.2	Omega-kvadrat: manje pristrana mjera	21
9	Planirane usporedbe	22
10	Kruskal-Wallisov test	23
10.1	Post-hoc za Kruskal-Wallis: Dunn test	23
11	Potpuna analiza: izvještaj	24
12	Dijagram odlučivanja: ANOVA ili nešto drugo?	28

13 Zadaci za pripremu	30
14 Dodatno čitanje	30
15 Pojmovnik	30

`library(tidyverse)`

i Ishodi učenja

Nakon ovog predavanja moći ćete

1. Objasniti zašto višestruki t-testovi nisu primjereni za usporedbu više od dviju grupa.
2. Provesti i interpretirati jednosmjernu ANOVA-u u R-u.
3. Objasniti logiku F-statistike kao omjera varijabilnosti između i unutar grupa.
4. Provjeriti pretpostavke ANOVA-e (normalnost, homogenost varijance).
5. Primijeniti post-hoc testove (Tukey HSD) za identifikaciju specifičnih razlika.
6. Izračunati eta-kvadrat kao mjeru veličine učinka za ANOVA-u.
7. Primijeniti Kruskal-Wallisov test kao neparametrijsku alternativu.
8. Provesti kompletnu analizu s izvještajem.

1 Motivacija: vjerodostojnost vijesti po izvoru

Istraživačko pitanje — percipiraju li ljudi istu vijest kao više ili manje vjerodostojnu ovisno o tome iz kojeg izvora dolazi? Konkretno, razlikuje li se percipirana vjerodostojnost vijesti ovisno o tome pripisuje li se izvoru TV, web portal, društvena mreža, tisak ili podcast?

Imamo pet grupa. Prošli tjedan naučili smo t-test za usporedbu dviju grupa. Zašto ne bismo jednostavno proveli t-test za svaki par?

1.1 Problem višestrukih t-testova

S pet grupa imamo 10 mogućih parova ($5 \text{ choose } 2 = 10$). Ako svaki test provodimo na razini $\alpha = 0.05$, vjerojatnost barem jednog lažno pozitivnog rezultata je:

$$P(\text{barem 1 greska}) = 1 - (1 - 0.05)^{10} = 1 - 0.95^{10} \approx 0.40$$

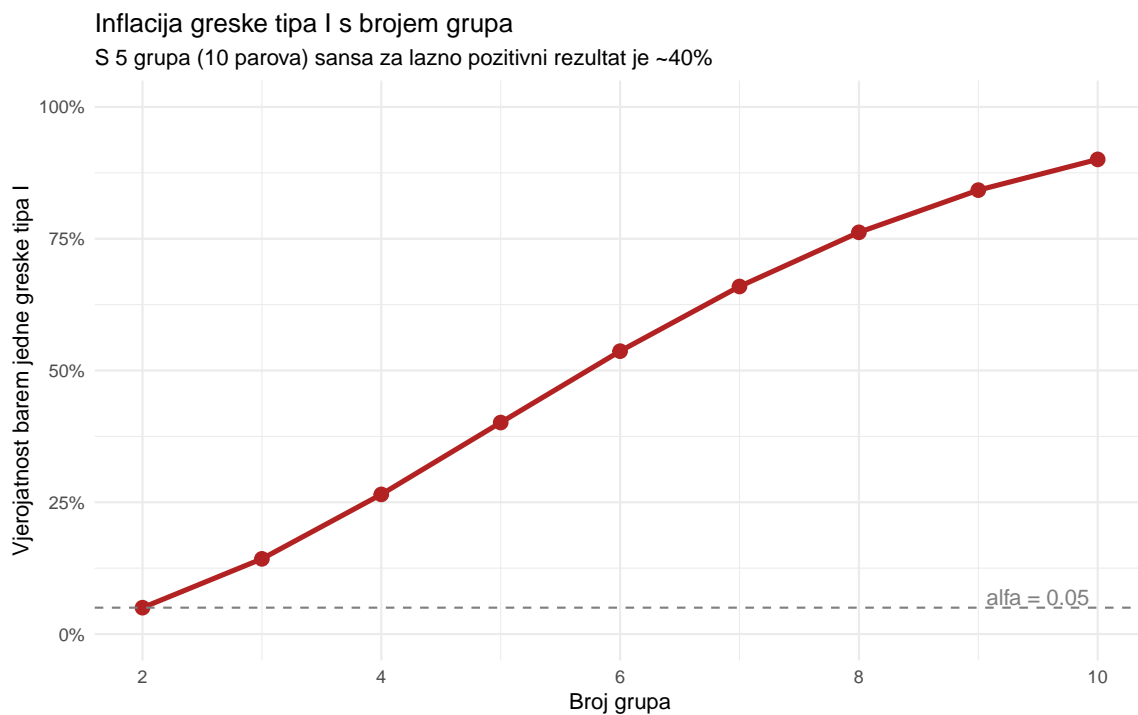
S 10 testova, šansa za barem jedan lažno pozitivni rezultat je oko 40%. To je potpuno neprihvatljivo.

```

# Inflacija greske tipa I s brojem testova
n_grupa <- 2:10
n_parova <- choose(n_grupa, 2)
alpha_inflated <- 1 - (1 - 0.05)^n_parova

tibble(grupe = n_grupa, parovi = n_parova, alpha = alpha_inflated) |>
  ggplot(aes(x = grupe, y = alpha)) +
  geom_line(linewidth = 1.2, color = "firebrick") +
  geom_point(size = 3, color = "firebrick") +
  geom_hline(yintercept = 0.05, linetype = "dashed", color = "grey50") +
  annotate("text", x = 9.5, y = 0.07, label = "alfa = 0.05", color = "grey50") +
  scale_y_continuous(labels = scales::label_percent(), limits = c(0, 1)) +
  labs(
    title = "Inflacija greske tipa I s brojem grupa",
    subtitle = "S 5 grupa (10 parova) sansa za lazno pozitivni rezultat je ~40%",
    x = "Broj grupa",
    y = "Vjerojatnost barem jedne greske tipa I"
  ) +
  theme_minimal()

```



ANOVA rješava ovaj problem — testira sve grupe odjednom jednim testom, održavajući alfa na 0.05.

2 Naši podaci

```
cred <- read_csv("../resources/datasets/news_credibility.csv")
glimpse(cred)
```

```
Rows: 300
Columns: 10
$ participant_id <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, ~
$ news_source    <chr> "TV", "TV", "TV", "TV", "TV", "TV", "TV", "TV", "TV", "~
$ age_group      <chr> "30-44", "18-29", "45-59", "45-59", "45-59", "18-
29", "~
$ education      <chr> "visa", "srednja", "visoka", "srednja", "srednja", "sre~
$ topic          <chr> "zdravlje", "zdravlje", "tehnologija", "politika", "zdr~
$ media_literacy <chr> "visoka", "niska", "srednja", "srednja", "srednja", "vi~
$ credibility    <dbl> 4.2, 5.4, 4.2, 3.3, 5.8, 4.4, 3.8, 4.1, 5.6, 5.7, 5.4, ~
$ trust_general  <dbl> 3.5, 3.8, 4.5, 3.6, 6.3, 4.3, 3.4, 3.6, 7.0, 4.7, 6.1, ~
$ share_intent  <dbl> 2, 3, 2, 2, 2, 2, 3, 2, 2, 5, 2, 2, 3, 2, 3, 3, 3, 1, 3~
$ reading_time  <dbl> 3.9, 4.5, 4.1, 1.9, 4.9, 4.6, 1.7, 2.8, 5.0, 3.6, 3.7, ~
```

```
cred |>
  group_by(news_source) |>
  summarise(
    n = n(),
    M = round(mean(credibility), 2),
    SD = round(sd(credibility), 2),
    Min = min(credibility),
    Max = max(credibility),
    .groups = "drop"
  ) |>
  arrange(desc(M))
```

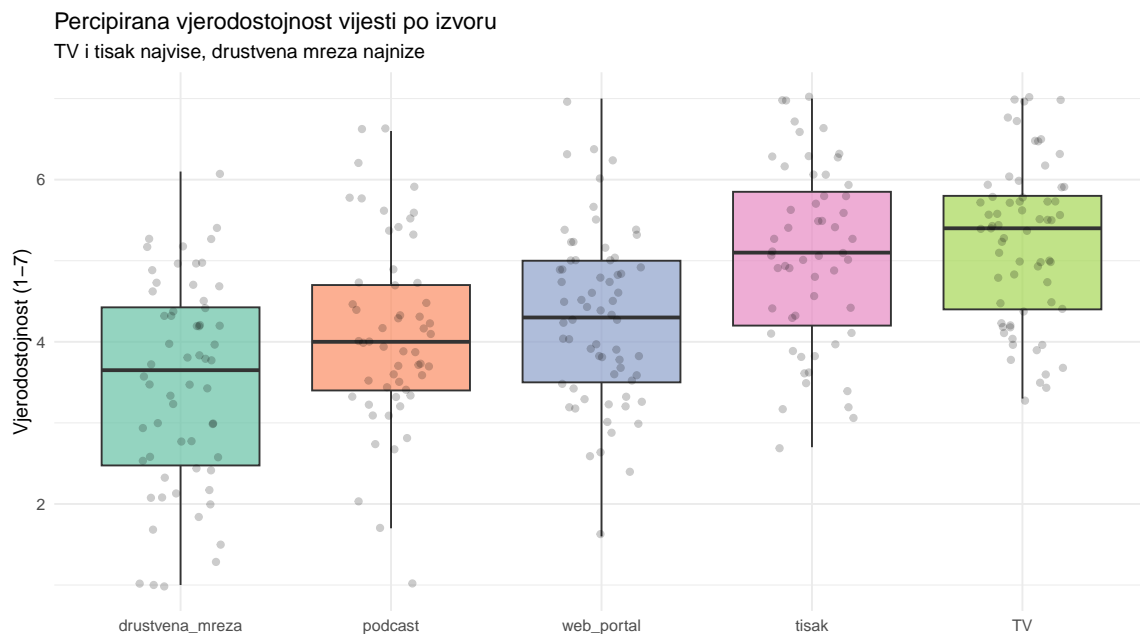
```
# A tibble: 5 x 6
  news_source      n     M   SD  Min  Max
  <chr>          <int> <dbl> <dbl> <dbl> <dbl>
1 TV              65  5.23  0.98  3.3   7
2 tisak           55  5.06  1.12  2.7   7
3 web_portal      65  4.29  1.05  1.6   7
4 podcast         55  4.12  1.17  1     6.6
5 drustvena_mreza 60  3.49  1.27  1     6.1
```

TV i tisak imaju najvišu percipiranu vjerodostojnost ($M > 5$), društvena mreža najnižu ($M < 3.5$). Razlike su očite, ali jesu li statistički značajne kad uzmemo u obzir varijabilnost unutar svake grupe?

```

cred |>
  mutate(news_source = fct_reorder(news_source, credibility)) |>
  ggplot(aes(x = news_source, y = credibility, fill = news_source)) +
  geom_boxplot(alpha = 0.7, outlier.shape = NA) +
  geom_jitter(width = 0.2, alpha = 0.2, size = 1.5) +
  scale_fill_brewer(palette = "Set2") +
  labs(
    title = "Percipirana vjerodostojnost vijesti po izvoru",
    subtitle = "TV i tisak najviše, društvena mreža najniže",
    x = NULL,
    y = "Vjerodostojnost (1-7)"
  ) +
  theme_minimal() +
  theme(legend.position = "none")

```



3 Logika ANOVA-e

ANOVA (Analysis of Variance) uspoređuje varijabilnost IZMEĐU grupa s varijabilnošću UNUTAR grupa.

Intuicija je sljedeća. Ako se grupe stvarno razlikuju, prosjeci grupa bit će razbacani daleko jedni od drugih (velika varijabilnost između grupa). Istovremeno, unutar svake grupe

postoji prirodna varijabilnost pojedinaca. Ako je varijabilnost između grupa puno veća od varijabilnosti unutar grupa, onda su razlike između grupa “stvarne” (nije samo šum).

$$F = \frac{\text{varijabilnost IZMEĐU grupa}}{\text{varijabilnost UNUTAR grupa}} = \frac{MS_{between}}{MS_{within}}$$

Ako F je blizu 1, varijabilnost između grupa je podjednaka onoj unutar grupa (nema učinka). Ako je F puno veći od 1, grupe se značajno razlikuju.

3.1 Dekompozicija varijance: SS

Ukupnu varijabilnost podataka rastavimo na dva dijela, gdje je:

$$SS_{total} = SS_{between} + SS_{within}$$

```
# Rucni izracun dekompozicije varijance
grand_mean <- mean(cred$credibility)

# SS_total: ukupno odstupanje svakog opazanja od grand mean
ss_total <- sum((cred$credibility - grand_mean)^2)

# SS_between: odstupanje grupnih prosjeka od grand mean (ponderano s n)
group_stats <- cred |>
  group_by(news_source) |>
  summarise(n = n(), M = mean(credibility), .groups = "drop")

ss_between <- sum(group_stats$n * (group_stats$M - grand_mean)^2)

# SS_within: odstupanje opazanja od njihovog grupnog prosjeka
ss_within <- cred |>
  left_join(group_stats |> select(news_source, M), by = "news_source") |>
  summarise(ss = sum((credibility - M)^2)) |>
  pull(ss)

cat("SS_total: ", round(ss_total, 1), "\n")
```

```
SS_total: 492
```

```
cat("SS_between:", round(ss_between, 1), "\n")
```

```
SS_between: 122.8
```

```
cat("SS_within: ", round(ss_within, 1), "\n")
```

SS_within: 369.2

```
cat("Provjera: ", round(ss_between + ss_within, 1), " (treba biti = SS_total)\n\n")
```

Provjera: 492 (treba biti = SS_total)

```
# Stupnjevi slobode
k <- nrow(group_stats) # broj grupa
N <- nrow(cred)        # ukupno opazanja
df_between <- k - 1
df_within <- N - k

# Mean Squares
ms_between <- ss_between / df_between
ms_within <- ss_within / df_within

# F statistika
f_stat <- ms_between / ms_within
p_val <- pf(f_stat, df_between, df_within, lower.tail = FALSE)

cat("df_between:", df_between, "\n")
```

df_between: 4

```
cat("df_within: ", df_within, "\n")
```

df_within: 295

```
cat("MS_between:", round(ms_between, 2), "\n")
```

MS_between: 30.69

```
cat("MS_within: ", round(ms_within, 2), "\n")
```

MS_within: 1.25

```
cat("F =", round(f_stat, 2), "\n")
```

F = 24.52

```
cat("p =", format(p_val, scientific = TRUE, digits = 3), "\n")
```

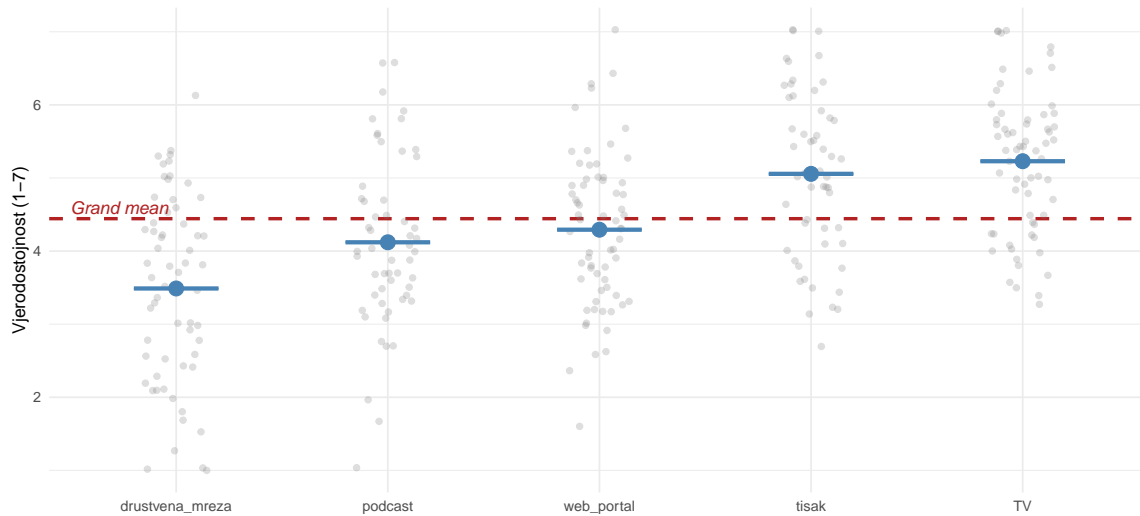
p = 1.55e-17

MS_{between} (varijabilnost između grupa) je mnogo veća od MS_{within} (varijabilnost unutar grupa). F je velik, p je izuzetno mali. Grupe se značajno razlikuju.

```
# Vizualizacija: ukupna vs objašnjena varijanca
cred_plot <- cred |>
  left_join(group_stats |> select(news_source, group_mean = M), by = "news_source") |>
  mutate(news_source = fct_reorder(news_source, group_mean))

cred_plot |>
  ggplot(aes(x = news_source, y = credibility)) +
  geom_hline(yintercept = grand_mean, color = "firebrick", linewidth = 1, linetype = "dash") +
  geom_jitter(width = 0.15, alpha = 0.25, color = "grey50") +
  stat_summary(fun = mean, geom = "point", size = 4, color = "steelblue") +
  stat_summary(fun = mean, geom = "crossbar", width = 0.4, color = "steelblue",
              fun.min = mean, fun.max = mean) +
  annotate("text", x = 0.5, y = grand_mean + 0.15, label = "Grand mean",
          color = "firebrick", hjust = 0, fontface = "italic") +
  labs(
    title = "ANOVA logika: razlike između grupnih prosjeka vs varijabilnost unutar grupa",
    subtitle = "Plavi crossbar = grupni prosjek. Crvena linija = ukupni prosjek. Sive točke = pojedinačni podaci.",
    x = NULL, y = "Vjerodostojnost (1-7)"
  ) +
  theme_minimal()
```

ANOVA logika: razlike između grupnih prosjeka vs varijabilnost unutar grupa
Plavi crossbar = grupni prosjek. Crvena linija = ukupni prosjek. Sive točke = pojedinci.



4 ANOVA u R-u

R koristi funkciju `aov()` za ANOVA-u. Sintaksa je formula pristup: `aov(y ~ grupa, data = ...)`.

```
# Jednosmjerna ANOVA
model <- aov(credibility ~ news_source, data = cred)
summary(model)
```

```
          Df Sum Sq Mean Sq F value Pr(>F)
news_source  4  122.8   30.692   24.52 <2e-16 ***
Residuals 295  369.2    1.252
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA tablica sadrži sve elemente koje smo ručno izračunali, uključujući df, SS (Sum Sq), MS (Mean Sq), F vrijednost i p-vrijednost ($\text{Pr}(>F)$). Rezultat potvrđuje da postoji statistički značajna razlika u percipiranoj vjerodostojnosti ovisno o izvoru vijesti.

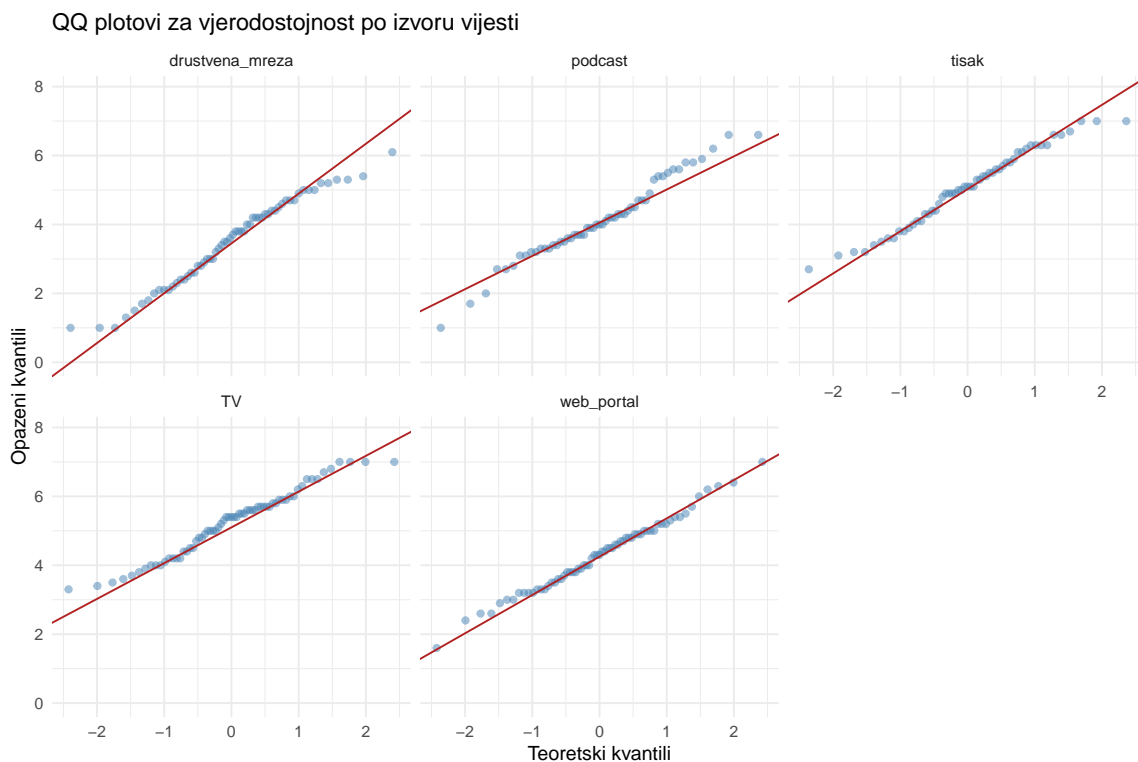
Ali ANOVA je **omnibus test** — govori da se barem dvije grupe razlikuju, ali ne govori KOJE. Za to trebamo post-hoc testove (drugi dio).

5 Pretpostavke ANOVA-e

ANOVA ima tri pretpostavke. Neovisnost opažanja (dizajn istraživanja), normalnost distribucije unutar svake grupe (ili dovoljno velik n) i homogenost varijance (jednake varijance u svim grupama).

5.1 Provjera normalnosti

```
# QQ plotovi po grupi
cred |>
  ggplot(aes(sample = credibility)) +
  stat_qq(color = "steelblue", alpha = 0.5) +
  stat_qq_line(color = "firebrick") +
  facet_wrap(~news_source) +
  labs(
    title = "QQ plotovi za vjerodostojnost po izvoru vijesti",
    x = "Teoretski kvantili",
    y = "Opazeni kvantili"
  ) +
  theme_minimal()
```



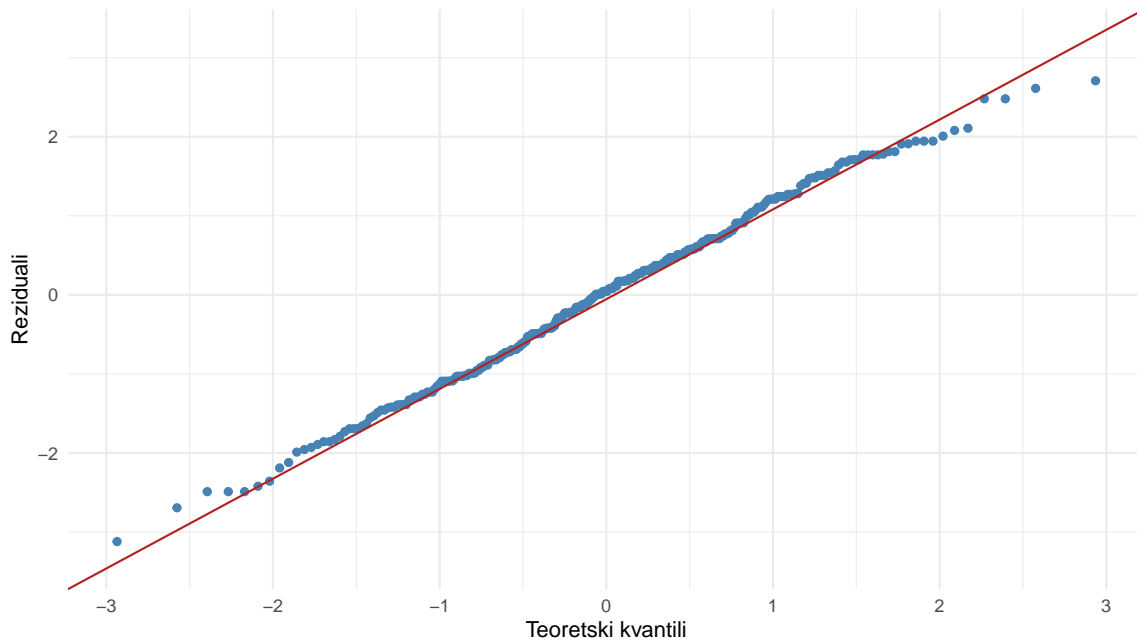
```
# Shapiro-Wilk po grupi
cred |>
  group_by(news_source) |>
  summarise(
    n = n(),
    shapiro_W = round(shapiro.test(credibility)$statistic, 4),
    shapiro_p = round(shapiro.test(credibility)$p.value, 4),
    normalno = shapiro_p >= 0.05,
    .groups = "drop"
  )
```

```
# A tibble: 5 x 5
  news_source      n shapiro_W shapiro_p normalno
  <chr>          <int>   <dbl>   <dbl> <lgl>
1 TV              65     0.970   0.117 TRUE
2 drustvena_mreza 60     0.973   0.198 TRUE
3 podcast         55     0.977   0.367 TRUE
4 tisak           55     0.975   0.309 TRUE
5 web_portal      65     0.992   0.964 TRUE
```

Neke grupe možda ne prolaze Shapiro-Wilk test. Ali s $n > 50$ po grupi, ANOVA je robusna na umjerena odstupanja od normalnosti (CLT). Također možemo provjeriti normalnost **reziduala** modela:

```
# Reziduali modela
tibble(reziduali = residuals(model)) |>
  ggplot(aes(sample = reziduali)) +
  stat_qq(color = "steelblue") +
  stat_qq_line(color = "firebrick") +
  labs(
    title = "QQ plot reziduala ANOVA modela",
    subtitle = "Ako su reziduali normalni, pretpostavka je zadovoljena",
    x = "Teoretski kvantili",
    y = "Reziduali"
  ) +
  theme_minimal()
```

QQ plot reziduala ANOVA modela
Ako su reziduali normalni, pretpostavka je zadovoljena



5.2 Provjera homogenosti varijance: Levenov test

```
# Rucna provjera: omjer najvece i najmanje varijance
var_by_group <- cred |>
  group_by(news_source) |>
  summarise(var = var(credibility), sd = round(sd(credibility), 2), .groups = "drop")
```

```
var_by_group
```

```
# A tibble: 5 x 3
  news_source      var    sd
  <chr>          <dbl> <dbl>
1 TV             0.963  0.98
2 drustvena_mreza 1.63   1.27
3 podcast        1.36   1.17
4 tisak          1.25   1.12
5 web_portal     1.11   1.05
```

```
cat("\nOmjer max/min varijance:", round(max(var_by_group$var) / min(var_by_group$var), 2),
```

```
Omjer max/min varijance: 1.69
```

```
cat("Pravilo palca - ako je omjer < 3, homogenost je prihvatljiva.\n")
```

Pravilo palca - ako je omjer < 3, homogenost je prihvatljiva.

```
# Levenov test (rucna implementacija bazirana na apsolutnim devijacijama)
cred_levne <- cred |>
  left_join(group_stats |> select(news_source, group_median = M), by = "news_source") |>
  mutate(abs_dev = abs(credibility - group_median))

levne_anova <- aov(abs_dev ~ news_source, data = cred_levne)
levne_p <- summary(levne_anova)[[1]][["Pr(>F)"]][1]

cat("Levenov test (ANOVA na apsolutnim devijacijama):\n")
```

Levenov test (ANOVA na apsolutnim devijacijama):

```
cat("p =", round(levne_p, 4), "\n")
```

p = 0.1895

```
cat("Homogenost varijance:", if_else(levne_p >= 0.05, "zadovoljena", "narusena"), "\n")
```

Homogenost varijance: zadovoljena

Ako je Levenov test značajan (varijance nejednake), koristimo **Welchovu ANOVA-u** koja ne pretpostavlja jednake varijance:

```
# Welchova ANOVA (ne pretpostavlja jednake varijance)
oneway.test(credibility ~ news_source, data = cred, var.equal = FALSE)
```

One-way analysis of means (not assuming equal variances)

data: credibility and news_source

F = 23.513, num df = 4.00, denom df = 145.15, p-value = 5.312e-15

```
# Usporedba klasicne i Welchove ANOVA
klasicna <- summary(model)[[1]]
welch <- oneway.test(credibility ~ news_source, data = cred, var.equal = FALSE)

cat("Klasicna ANOVA: F(", df_between, ",", N - k, ") = ",
    round(klasicna$`F value`[1], 2), ", p < 0.001\n", sep = "")
```

Klasična ANOVA: $F(4,295) = 24.52, p < 0.001$

```
cat("Welchova ANOVA: F(", welch$parameter[1], ",", round(welch$parameter[2], 1),
    ") = ", round(welch$statistic, 2), ", p < 0.001\n", sep = "")
```

Welchova ANOVA: $F(4,145.2) = 23.51, p < 0.001$

```
cat("\nOba pristupa daju isti zakljucak.\n")
```

Oba pristupa daju isti zakljucak.

💡 Praktični savjet

Kao i kod t-testa, preporučujemo Welchovu ANOVA-u (`oneway.test(y ~ grupa, var.equal = FALSE)`) kao standardni izbor. Kad su varijance jednake, daje iste rezultate kao klasična ANOVA. Kad su nejednake, daje točnije rezultate. Klasičnu ANOVA-u (`aov()`) koristite kad trebate rezidualne dijagnostike ili post-hoc testove koji zahtijevaju `aov` objekt.

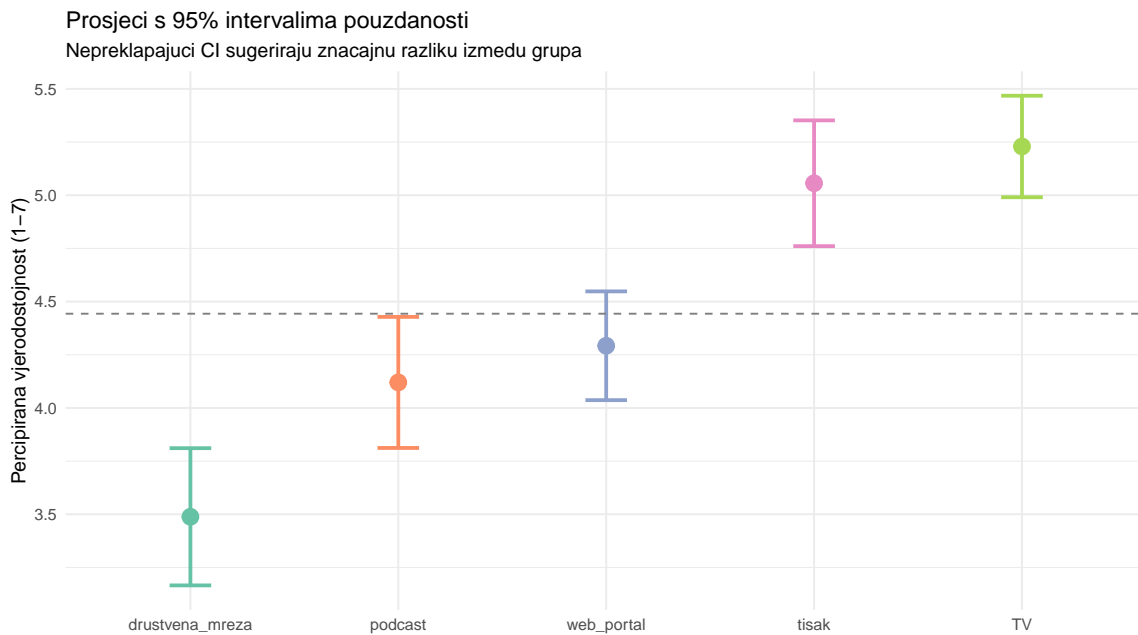
6 Vizualizacija ANOVA rezultata

```
# Prosjeci s 95% CI
cred |>
  group_by(news_source) |>
  summarise(
    M = mean(credibility),
    SE = sd(credibility) / sqrt(n()),
    CI_lo = M - 1.96 * SE,
    CI_hi = M + 1.96 * SE,
    .groups = "drop"
  ) |>
  mutate(news_source = fct_reorder(news_source, M)) |>
  ggplot(aes(x = news_source, y = M, color = news_source)) +
  geom_point(size = 4) +
  geom_errorbar(aes(ymin = CI_lo, ymax = CI_hi), width = 0.2, linewidth = 1) +
  geom_hline(yintercept = grand_mean, linetype = "dashed", color = "grey50") +
  scale_color_brewer(palette = "Set2") +
  labs(
```

```

title = "Prosjeci s 95% intervalima pouzdanosti",
subtitle = "Nepreklapajuci CI sugeriraju znacajnu razliku izmedu grupa",
x = NULL,
y = "Percipirana vjerodostojnost (1-7)"
) +
theme_minimal() +
theme(legend.position = "none")

```



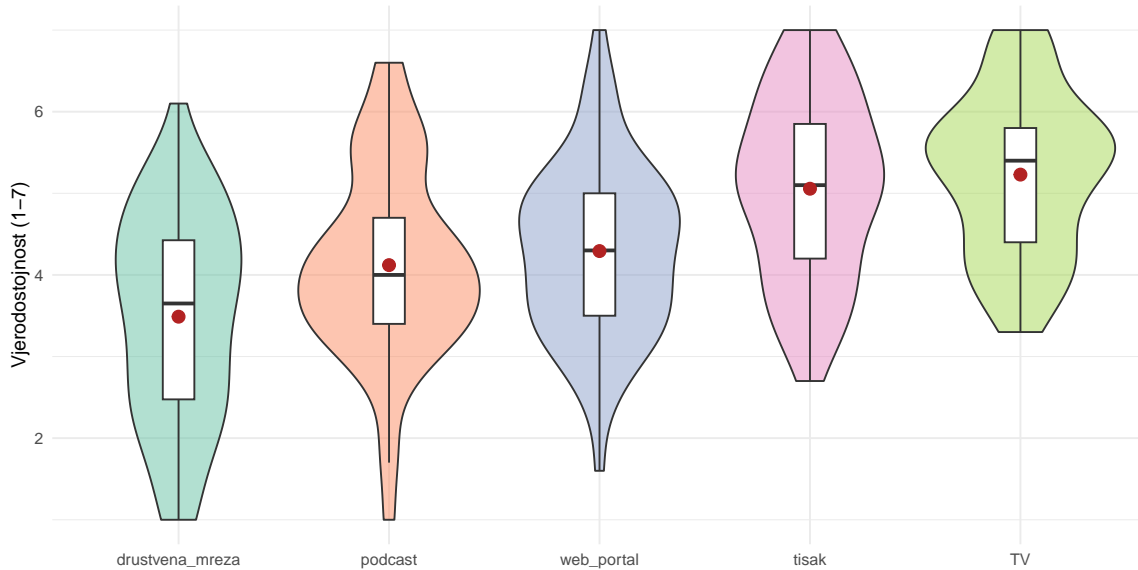
```

# Violin plot s prosjekom
cred |>
mutate(news_source = fct_reorder(news_source, credibility)) |>
ggplot(aes(x = news_source, y = credibility, fill = news_source)) +
geom_violin(alpha = 0.5) +
geom_boxplot(width = 0.15, fill = "white", outlier.shape = NA) +
stat_summary(fun = mean, geom = "point", size = 3, color = "firebrick") +
scale_fill_brewer(palette = "Set2") +
labs(
title = "Distribucija vjerodostojnosti po izvoru",
subtitle = "Violin = oblik distribucije. Crvena tocka = prosjek. Bijeli box = medijan",
x = NULL,
y = "Vjerodostojnost (1-7)"
) +
theme_minimal() +
theme(legend.position = "none")

```

Distribucija vjerodostojnosti po izvoru

Violin = oblik distribucije. Crvena točka = prosjek. Bijeli box = medijan i IQR.



i Podsjetnik

U prvom dijelu naučili smo logiku ANOVA-e ($F = \text{varijabilnost između} / \text{varijabilnost unutar}$), dekompoziciju varijance ($SS_{\text{total}} = SS_{\text{between}} + SS_{\text{within}}$), pretpostavke (normalnost, homogenost varijance) i osnovno provođenje `aov()` i `oneway.test()`. ANOVA je pokazala da se grupe značajno razlikuju ($p < 0.001$). Sada utvrđujemo KOJE se grupe razlikuju.

7 Post-hoc testovi: Tukey HSD

ANOVA je omnibus test — govori da se barem dvije grupe razlikuju, ali ne govori koje. **Post-hoc testovi** uspoređuju sve parove grupa uz kontrolu greške tipa I.

Najčešći post-hoc test je **Tukey HSD** (Honestly Significant Difference). On testira sve parove istovremeno i prilagođava p-vrijednosti tako da ukupna greška tipa I ostane na $\alpha = 0.05$.

```
cred <- read_csv("../resources/datasets/news_credibility.csv")
model <- aov(credibility ~ news_source, data = cred)

# Tukey HSD
tukey_rez <- TukeyHSD(model)
tukey_rez
```

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = credibility ~ news_source, data = cred)

```
$news_source
```

	diff	lwr	upr	p adj
podcast-drustvena_mreza	0.6316667	0.05842531	1.2049080	0.0225724
tisak-drustvena_mreza	1.5680303	0.99478894	2.1412717	0.0000000
TV-drustvena_mreza	1.7408974	1.19114360	2.2906513	0.0000000
web_portal-drustvena_mreza	0.8039744	0.25422052	1.3537282	0.0007187
tisak-podcast	0.9363636	0.35079309	1.5219342	0.0001538
TV-podcast	1.1092308	0.54663279	1.6718287	0.0000013
web_portal-podcast	0.1723077	-0.39029029	0.7349057	0.9177070
TV-tisak	0.1728671	-0.38973085	0.7354651	0.9168041
web_portal-tisak	-0.7640559	-1.32665392	-0.2014580	0.0021451
web_portal-TV	-0.9369231	-1.47556963	-0.3982765	0.0000279

```
# Preglednija tablica
tukey_df <- as_tibble(tukey_rez$news_source, rownames = "par") |>
  mutate(
    razlika = round(diff, 2),
    CI_lo = round(lwr, 2),
    CI_hi = round(upr, 2),
    p = round(`p adj`, 4),
    znacajno = p < 0.05
  ) |>
  select(par, razlika, CI_lo, CI_hi, p, znacajno) |>
  arrange(p)
```

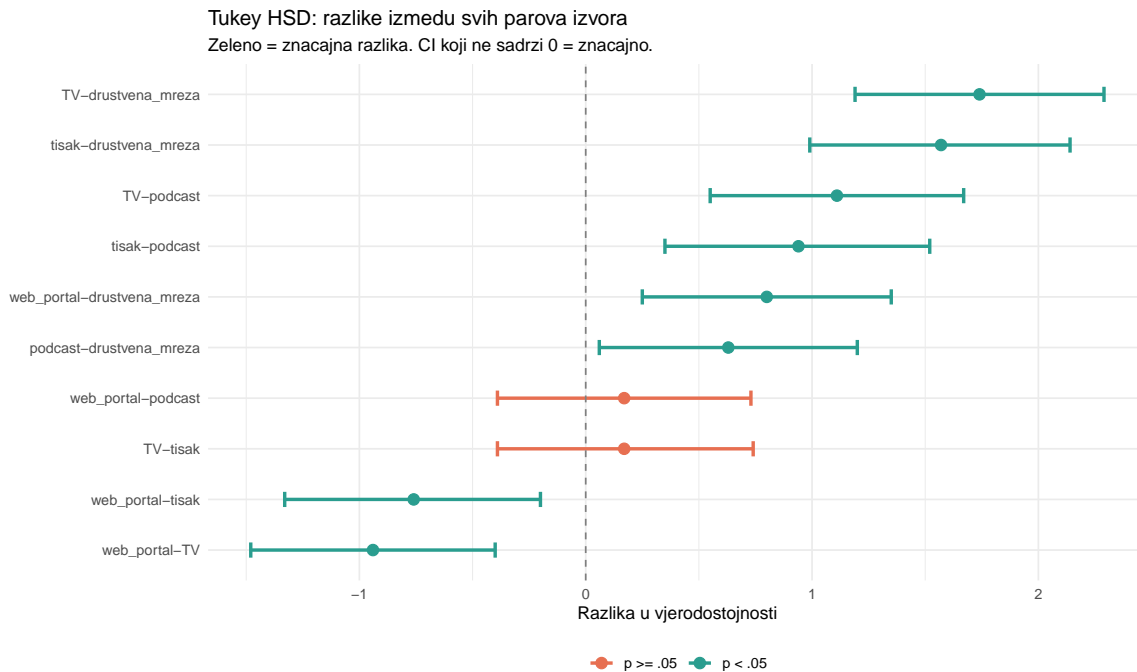
```
tukey_df
```

```
# A tibble: 10 x 6
```

par	razlika	CI_lo	CI_hi	p	znacajno
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<lgl>
1 tisak-drustvena_mreza	1.57	0.99	2.14	0	TRUE
2 TV-drustvena_mreza	1.74	1.19	2.29	0	TRUE
3 TV-podcast	1.11	0.55	1.67	0	TRUE
4 web_portal-TV	-0.94	-1.48	-0.4	0	TRUE
5 tisak-podcast	0.94	0.35	1.52	0.0002	TRUE
6 web_portal-drustvena_mreza	0.8	0.25	1.35	0.0007	TRUE
7 web_portal-tisak	-0.76	-1.33	-0.2	0.0021	TRUE
8 podcast-drustvena_mreza	0.63	0.06	1.2	0.0226	TRUE
9 TV-tisak	0.17	-0.39	0.74	0.917	FALSE
10 web_portal-podcast	0.17	-0.39	0.73	0.918	FALSE

Tablica pokazuje razliku prosjeka za svaki par, 95% CI za tu razliku i prilagođenu p-vrijednost. Značajni parovi ($p < 0.05$) su oni gdje CI ne uključuje nulu.

```
# Vizualizacija Tukey rezultata
tukey_df |>
  mutate(par = fct_reorder(par, razlika)) |>
  ggplot(aes(y = par)) +
  geom_errorbarh(aes(xmin = CI_lo, xmax = CI_hi, color = znacajno),
                height = 0.3, linewidth = 1) +
  geom_point(aes(x = razlika, 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 = "Tukey HSD: razlike između svih parova izvora",
    subtitle = "Zeleno = značajna razlika. CI koji ne sadrži 0 = značajno.",
    x = "Razlika u vjerodostojnosti",
    y = NULL,
    color = NULL
  ) +
  theme_minimal() +
  theme(legend.position = "bottom")
```



Iz grafa se jasno vide obrasci. Društvena mreža ima značajno nižu vjerodostojnost od svih drugih izvora. TV i tisak nemaju značajnu razliku međusobno (oba su visoki). Podcast i web portal su u sredini.

7.1 Compact letter display

Česta praksa u izvještavanju je grupiranje izvora koji se NE razlikuju značajno. To zovemo “compact letter display” (CLD).

```
# Rucna interpretacija Tukey rezultata u grupe
# Grupe koje se NE razlikuju znacajno dijele isto slovo
group_stats <- cred |>
  group_by(news_source) |>
  summarise(M = round(mean(credibility), 2), .groups = "drop") |>
  arrange(desc(M))

cat("Grupiranje izvora (isti skup slova = nema znacajne razlike):\n\n")
```

Grupiranje izvora (isti skup slova = nema znacajne razlike):

```
# Na temelju Tukey rezultata:
znacajni_parovi <- tukey_df |> filter(znacajno) |> pull(par)
cat("Znacajno razliciti parovi:\n")
```

Znacajno razliciti parovi:

```
for (p in znacajni_parovi) cat(" ", p, "\n")
```

```
tisak-drustvena_mreza
TV-drustvena_mreza
TV-podcast
web_portal-TV
tisak-podcast
web_portal-drustvena_mreza
web_portal-tisak
podcast-drustvena_mreza
```

8 Veličina učinka: eta-kvadrat

Kao i kod t-testa, p-vrijednost ne govori koliko je učinak velik. **Eta-kvadrat** (η^2) je mjera veličine učinka za ANOVA-u:

$$\eta^2 = \frac{SS_{between}}{SS_{total}}$$

Eta-kvadrat je proporcija ukupne varijabilnosti koja je objašnjena grupnom pripadnošću. Interpretira se kao R-kvadrat u regresiji.

```
# Iz ANOVA tablice
anova_table <- summary(model)[[1]]
ss_between <- anova_table$`Sum Sq`[1]
ss_within <- anova_table$`Sum Sq`[2]
ss_total <- ss_between + ss_within

eta2 <- ss_between / ss_total

cat("SS_between:", round(ss_between, 1), "\n")
```

SS_between: 122.8

```
cat("SS_total: ", round(ss_total, 1), "\n")
```

SS_total: 492

```
cat("Eta-kvadrat:", round(eta2, 3), "\n")
```

Eta-kvadrat: 0.25

```
cat("Interpretacija:", round(eta2 * 100, 1), "% varijabilnosti u vjerodostojnosti\n")
```

Interpretacija: 25 % varijabilnosti u vjerodostojnosti

```
cat("je objasnjeno izvorom vijesti.\n")
```

je objasnjeno izvorom vijesti.

8.1 Interpretacija eta-kvadrata

Cohen (1988) je predložio sljedeće smjernice za interpretaciju eta-kvadrata.

```
tribble(
  ~eta2, ~interpretacija,
  "0.01", "Mali ucinak",
  "0.06", "Srednji ucinak",
  "0.14", "Veliki ucinak"
)
```

```
# A tibble: 3 x 2
  eta2 interpretacija
  <chr> <chr>
1 0.01 Mali ucinak
2 0.06 Srednji ucinak
3 0.14 Veliki ucinak
```

Naš eta2 = 0.25 je veliki učinak. Izvor vijesti objašnjava značajan dio varijabilnosti u percipiranoj vjerodostojnosti.

8.2 Omega-kvadrat: manje pristrana mjera

Eta-kvadrat je pristran (precjenjuje veličinu učinka u populaciji). **Omega-kvadrat** je manje pristrana alternativa:

$$\omega^2 = \frac{SS_{between} - (k - 1) \cdot MS_{within}}{SS_{total} + MS_{within}}$$

```
ms_within <- anova_table$`Mean Sq`[2]
k <- length(unique(cred$news_source))

omega2 <- (ss_between - (k - 1) * ms_within) / (ss_total + ms_within)

cat("Eta-kvadrat: ", round(eta2, 3), "\n")
```

Eta-kvadrat: 0.25

```
cat("Omega-kvadrat:", round(omega2, 3), "\n")
```

Omega-kvadrat: 0.239

```
cat("Razlika je mala za velike uzorke, ali omega2 je tocnija procjena.\n")
```

Razlika je mala za velike uzorke, ali omega2 je tocnija procjena.

Praktični savjet

Izvijestite eta-kvadrat (jer je poznatiji) ili omega-kvadrat (jer je manje pristran). Za objavljivanje u časopisima, sve češće se traži omega-kvadrat. U praksi, za $n > 50$ po grupi, razlika je minimalna.

9 Planirane usporedbe

Ponekad unaprijed znamo koje usporedbe nas zanimaju. Umjesto da uspoređujemo sve parove (Tukey), možemo specificirati **planirane usporedbe** (contrasts). Ovo je snažniji pristup jer testira manje hipoteza.

```
# Planirana usporedba 1: tradicionalni (TV + tisak) vs digitalni (portal + mreza + podcast)
cred <- cred |>
  mutate(tip_medija = if_else(
    news_source %in% c("TV", "tisak"), "tradicionalni", "digitalni"
  ))

t.test(credibility ~ tip_medija, data = cred)
```

Welch Two Sample t-test

```
data:  credibility by tip_medija
t = -8.9741, df = 278.64, p-value < 2.2e-16
alternative hypothesis: true difference in means between group digitalni and group tradici
95 percent confidence interval:
 -1.436806 -0.919861
sample estimates:
 mean in group digitalni mean in group tradicionalni
                3.971667                5.150000
```

```
# Cohenov d za ovu usporedbu
trad <- cred |> filter(tip_medija == "tradicionalni") |> pull(credibility)
dig <- cred |> filter(tip_medija == "digitalni") |> pull(credibility)
s_p <- sqrt(((length(trad)-1)*sd(trad)^2 + (length(dig)-1)*sd(dig)^2) / (length(trad)+length(dig)))
d_td <- (mean(trad) - mean(dig)) / s_p

cat("Tradicionalni M:", round(mean(trad), 2), "\n")
```

Tradicionalni M: 5.15

```
cat("Digitalni M:", round(mean(dig), 2), "\n")
```

Digitalni M: 3.97

```
cat("Cohenov d:", round(d_td, 2), "(veliki ucinak)\n")
```

Cohenov d: 1.03 (veliki ucinak)

Planirana usporedba (tradicionalni vs digitalni) daje jasnu sliku — tradicionalni mediji imaju značajno višu percipiranu vjerodostojnost od digitalnih.

10 Kruskal-Wallisov test

Kruskal-Wallisov test je neparametrijska alternativa jednosmjernoj ANOVA-i. Koristi rangove umjesto izvornih vrijednosti i ne pretpostavlja normalnost.

```
# Kruskal-Wallis test
kw_test <- kruskal.test(credibility ~ news_source, data = cred)
kw_test
```

```
Kruskal-Wallis rank sum test
```

```
data:  credibility by news_source
Kruskal-Wallis chi-squared = 70.893, df = 4, p-value = 1.471e-14
```

```
# Usporedba ANOVA i Kruskal-Wallis
cat("Klasicna ANOVA:    F = ", round(summary(model)[[1]]$`F value`[1], 2),
    ", p < 0.001\n", sep = "")
```

```
Klasicna ANOVA:    F = 24.52, p < 0.001
```

```
cat("Kruskal-Wallis:    H = ", round(kw_test$statistic, 2),
    ", p < 0.001\n", sep = "")
```

```
Kruskal-Wallis:    H = 70.89, p < 0.001
```

```
cat("\nOba testa daju isti zakljucak.\n")
```

```
Oba testa daju isti zakljucak.
```

10.1 Post-hoc za Kruskal-Wallis: Dunn test

Kruskal-Wallis test je omnibus. Za identifikaciju specifičnih razlika koristimo **Dunnov test** (neparametrijski ekvivalent Tukeyu).

```
# Pairwise Wilcoxon s Bonferroni korekcijom (ugraden u R)
pw_wilcox <- pairwise.wilcox.test(cred$credibility, cred$news_source,
                                  p.adjust.method = "BH")
pw_wilcox
```

Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data: cred\$credibility and cred\$news_source

	drustvena_mreza	podcast	tisak	TV
podcast	0.02530	-	-	-
tisak	3.0e-08	0.00019	-	-
TV	7.3e-11	1.8e-06	0.40500	-
web_portal	0.00128	0.40500	0.00048	3.4e-06

P value adjustment method: BH

11 Potpuna analiza: izvještaj

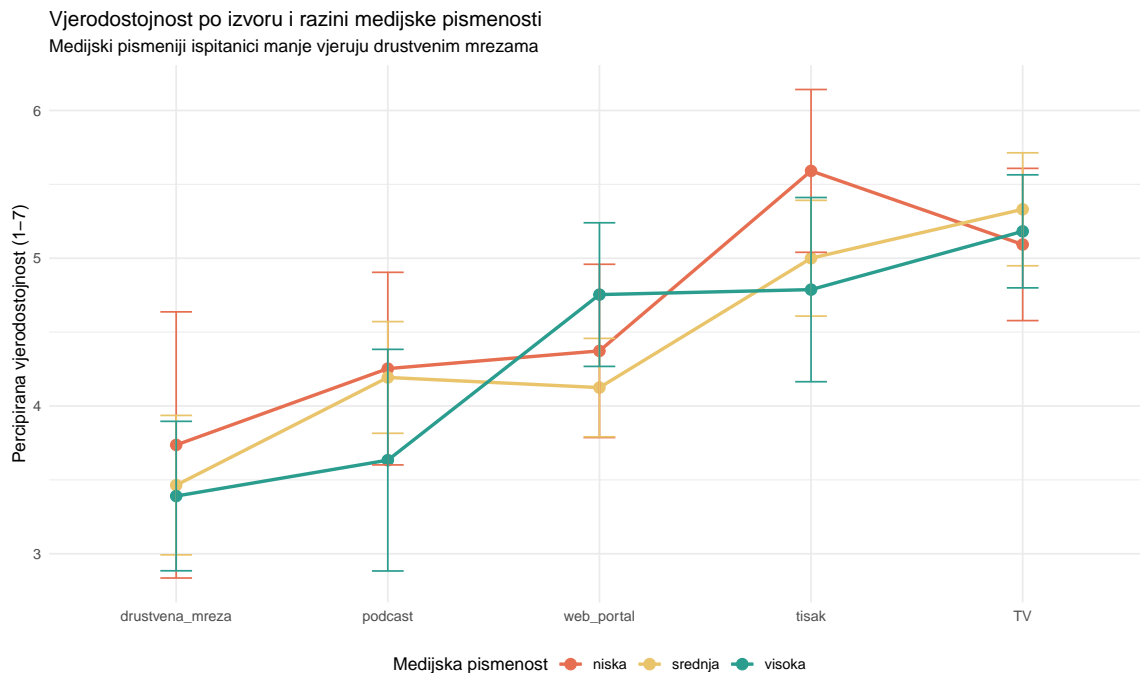
```
# Kompletna opisna statistika
cred |>
  group_by(news_source) |>
  summarise(
    n = n(),
    M = round(mean(credibility), 2),
    SD = round(sd(credibility), 2),
    Median = median(credibility),
    SE = round(sd(credibility) / sqrt(n()), 2),
    .groups = "drop"
  ) |>
  arrange(desc(M))
```

```
# A tibble: 5 x 6
  news_source      n      M    SD Median  SE
  <chr>          <int> <dbl> <dbl> <dbl> <dbl>
1 TV              65  5.23  0.98  5.4  0.12
2 tisak           55  5.06  1.12  5.1  0.15
3 web_portal      65  4.29  1.05  4.3  0.13
4 podcast         55  4.12  1.17  4    0.16
5 drustvena_mreza 60  3.49  1.27  3.65 0.16
```

```

# Moderacija: medijska pismenost
cred |>
  mutate(
    news_source = fct_reorder(news_source, credibility),
    media_literacy = factor(media_literacy, levels = c("niska", "srednja", "visoka"))
  ) |>
  group_by(news_source, media_literacy) |>
  summarise(M = mean(credibility), SE = sd(credibility)/sqrt(n()), .groups = "drop") |>
  ggplot(aes(x = news_source, y = M, color = media_literacy, group = media_literacy)) +
  geom_line(linewidth = 1) +
  geom_point(size = 3) +
  geom_errorbar(aes(ymin = M - 1.96 * SE, ymax = M + 1.96 * SE), width = 0.15) +
  scale_color_manual(values = c("niska" = "#e76f51", "srednja" = "#e9c46a", "visoka" = "#20998c")) +
  labs(
    title = "Vjerodostojnost po izvoru i razini medijske pismenosti",
    subtitle = "Medijski pismeniji ispitanici manje vjeruju društvenim mrežama",
    x = NULL,
    y = "Percepivana vjerodostojnost (1-7)",
    color = "Medijska pismenost"
  ) +
  theme_minimal() +
  theme(legend.position = "bottom")

```



```
# ANOVA unutar svake dobne skupine
cred |>
  group_by(age_group) |>
  summarise(
    F_val = round(summary(aov(credibility ~ news_source))[[1]]$`F value`[1], 2),
    p = summary(aov(credibility ~ news_source))[[1]]$`Pr(>F)`[1],
    eta2 = {
      a <- summary(aov(credibility ~ news_source))[[1]]
      round(a$`Sum Sq`[1] / sum(a$`Sum Sq`), 3)
    },
    .groups = "drop"
  ) |>
  mutate(p = format(p, scientific = TRUE, digits = 2))
```

```
# A tibble: 4 x 4
  age_group F_val p      eta2
<chr>      <dbl> <chr>  <dbl>
1 18-29      6.36 1.6e-04 0.233
2 30-44     10.1 6.5e-07 0.286
3 45-59      6.2 2.8e-04 0.276
4 60+        3.49 1.9e-02 0.317
```

Efekt izvora na vjerodostojnost je značajan u svim dobnim skupinama, ali eta-kvadrat varira. Ovo sugerira da je obrazac (TV > tisak > portal > podcast > mreža) konzistentan, ali jačina razlika može varirati po dobi.

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

```
=====
```

```
cat("  APA IZVJESTAJ: PERCIPIRANA VJERODOSTOJNOST PO IZVORU\n")
```

```
  APA IZVJESTAJ: PERCIPIRANA VJERODOSTOJNOST PO IZVORU
```

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

```
=====
```

```
cat("Jednosmjerna ANOVA pokazala je statisticki znacajnu razliku u\n")
```

```
Jednosmjerna ANOVA pokazala je statisticki znacajnu razliku u
```

```
cat("percipiranoj vjerodostojnosti vijesti ovisno o izvoru,\n")
```

percipiranoj vjerodostojnosti vijesti ovisno o izvoru,

```
cat("F(", k-1, ", ", ", nrow(cred)-k, ") = ",  
    round(summary(model)[[1]]$`F value`[1], 2),  
    ", p < .001, eta2 = ", round(eta2, 2), ".\n\n", sep = "")
```

F(4, 295) = 24.52, p < .001, eta2 = 0.25.

```
cat("Tukey HSD post-hoc testovi pokazali su da:\n")
```

Tukey HSD post-hoc testovi pokazali su da:

```
znac <- tukey_df |> filter(znacajno)  
for (i in 1:nrow(znac)) {  
  r <- znac[i, ]  
  cat(" ", r$par, ": razlika = ", r$razlika,  
      ", 95% CI [", r$CI_lo, ", ", r$CI_hi, "], p = ",  
      if_else(r$p < 0.001, "< .001", as.character(round(r$p, 3))), "\n", sep = "")  
}
```

tisak-drustvena_mreza: razlika = 1.57, 95% CI [0.99, 2.14], p = < .001
TV-drustvena_mreza: razlika = 1.74, 95% CI [1.19, 2.29], p = < .001
TV-podcast: razlika = 1.11, 95% CI [0.55, 1.67], p = < .001
web_portal-TV: razlika = -0.94, 95% CI [-1.48, -0.4], p = < .001
tisak-podcast: razlika = 0.94, 95% CI [0.35, 1.52], p = < .001
web_portal-drustvena_mreza: razlika = 0.8, 95% CI [0.25, 1.35], p = < .001
web_portal-tisak: razlika = -0.76, 95% CI [-1.33, -0.2], p = 0.002
podcast-drustvena_mreza: razlika = 0.63, 95% CI [0.06, 1.2], p = 0.023

```
cat("\nTV (M = ", round(mean(cred$credibility[cred$news_source == "TV"]), 2),  
    ") i tisak (M = ", round(mean(cred$credibility[cred$news_source == "tisak"]), 2),  
    ") imaju\nnajvisu vjerodostojnost. Drustvena mreza (M = ",  
    round(mean(cred$credibility[cred$news_source == "drustvena_mreza"]), 2),  
    ") ima najnizu.\n", sep = "")
```

TV (M = 5.23) i tisak (M = 5.06) imaju najvisu vjerodostojnost. Drustvena mreza (M = 3.49) ima najnizu.

```
cat("Medijska pismenost moderira efekt: visoko pismeni\n")
```

Medijska pismenost moderira efekt: visoko pismeni

```
cat("ispitanici manje vjeruju vijestima s drustvenih mreza.\n")
```

ispitanici manje vjeruju vijestima s drustvenih mreza.

12 Dijagram odlučivanja: ANOVA ili nešto drugo?

```
tribble(  
  ~situacija, ~test,  
  "Dvije nezavisne grupe", "Nezavisni t-test (tjedan 12)",  
  "Dvije uparene grupe", "Upareni t-test (tjedan 12)",  
  "Tri+ nezavisne grupe, normalni podaci", "Jednosmjerna ANOVA + Tukey HSD",  
  "Tri+ nezavisne grupe, nejednake varijance", "Welchova ANOVA + Games-Howell",  
  "Tri+ nezavisne grupe, nenormalni/ordinalni", "Kruskal-Wallis + Dunn",  
  "Tri+ uparene grupe", "Repeated measures ANOVA (napredno)",  
  "Dva faktora istovremeno", "Dvosmjerna ANOVA (napredno)",  
  "Kategoricka x kategoricka varijabla", "Hi-kvadrat test (tjedan 11)"  
)
```

```
# A tibble: 8 x 2
```

situacija <chr>	test <chr>
1 Dvije nezavisne grupe	Nezavisni t-test (tjedan 12)
2 Dvije uparene grupe	Upareni t-test (tjedan 12)
3 Tri+ nezavisne grupe, normalni podaci	Jednosmjerna ANOVA + Tukey HSD
4 Tri+ nezavisne grupe, nejednake varijance	Welchova ANOVA + Games-Howell
5 Tri+ nezavisne grupe, nenormalni/ordinalni	Kruskal-Wallis + Dunn
6 Tri+ uparene grupe	Repeated measures ANOVA (napredno)
7 Dva faktora istovremeno	Dvosmjerna ANOVA (napredno)
8 Kategoricka x kategoricka varijabla	Hi-kvadrat test (tjedan 11)

! Ključni zaključci

1. ANOVA uspoređuje prosjeke tri ili više grupa jednim testom, kontrolirajući grešku tipa I. Višestruki t-testovi inflacioniraju alfa i nisu prihvatljivi.
2. F statistika = $MS_{\text{between}} / MS_{\text{within}}$. Velik F znači da su razlike između grupa veće od varijabilnosti unutar grupa.
3. Dekompozicija varijance — $SS_{\text{total}} = SS_{\text{between}} + SS_{\text{within}}$. ANOVA testira je li SS_{between} dovoljno velik u odnosu na SS_{total} .
4. ANOVA je omnibus test — govori da razlika postoji, ali ne govori KOJE se grupe razlikuju. Post-hoc testovi identificiraju specifične parove.
5. Tukey HSD (`TukeyHSD(model)`) uspoređuje sve parove uz kontrolu ukupne greške tipa I. Daje razlike, CI i prilagođene p-vrijednosti za svaki par.
6. Eta-kvadrat = $SS_{\text{between}} / SS_{\text{total}}$. Proporcija varijabilnosti objašnjena grupom. Smjernice za interpretaciju — 0.01 mali, 0.06 srednji, 0.14 veliki učinak. Omega-kvadrat je manje pristrana alternativa.
7. Pretpostavke uključuju nezavisnost, normalnost unutar grupa (ili $n > 30$) i homogenost varijance. Welchova ANOVA (`oneway.test(var.equal = FALSE)`) ne zahtijeva jednake varijance.
8. Kruskal-Wallisov test (`kruskal.test()`) je neparametrijska alternativa. Koristi rangove. Post-hoc: `pairwise.wilcox.test()` s BH korekcijom.
9. Planirane usporedbe (contrasts) su snažniji pristup kad unaprijed znate koje grupe želite usporediti (npr. tradicionalni vs digitalni mediji).
10. APA format za ANOVA-u uključuje $F(df_{\text{between}}, df_{\text{within}}) = \text{vrijednost}, p, \eta^2$, kao i post-hoc razlike s CI.
11. Moderacijska analiza (ANOVA po podgrupama treće varijable) otkriva je li obrazac konzistentan ili varira ovisno o nekoj trećoj varijabli (npr. medijska pismenost, dob).
12. Za dva faktora istovremeno (npr. izvor x dob) koristite dvosmjernu ANOVA-u. Za uparena mjerenja kroz tri+ uvjeta koristite repeated measures ANOVA-u. Ovo su napredni pristupi izvan dosega ovog kolegija.

13 Zadaci za pripremu

1. Učitajte `news_credibility.csv`. Provedite jednosmjernu ANOVA-u za varijablu `trust_general` po `news_source`. Izračunajte eta-kvadrat i provedite Tukey HSD. Koji se parovi izvora značajno razlikuju?
2. Provedite Kruskal-Wallisov test za `share_intent` po `news_source`. Usporedite rezultat s klasičnom ANOVA-om. Jesu li zaključci konzistentni?
3. Testirajte planiranu usporedbu gdje trebate odrediti razlikuje li se `credibility` između tri razine medijske pismenosti (`media_literacy`). Koji par se razlikuje prema Tukey HSD testu?

14 Dodatno čitanje

Obavezno

Navarro, D. (2018). *Learning Statistics with R*, Chapter 14 (Comparing Several Means). Besplatno dostupno na learningstatisticswithr.com. Pokriva jednosmjernu ANOVA-u, post-hoc testove i veličinu učinka.

Preporučeno

Field, A. (2018). *Discovering Statistics Using R*. SAGE. Poglavlje 10. Detaljna obrada ANOVA-e s dijagnostikom i vizualizacijama.

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2018). *Designing Experiments and Analyzing Data* (3rd edition). Routledge. Referentni udžbenik za eksperimentalne dizajne i ANOVA-u.

15 Pojmovnik

Pojam	Objašnjenje
ANOVA	Analysis of Variance. Omnibus test za usporedbu prosjeka tri ili više grupa.
Jednosmjerna ANOVA	ANOVA s jednim faktorom (jednom nezavisnom varijablom).
F statistika	Omjer $MS_{\text{between}} / MS_{\text{within}}$. $F \gg 1$ sugerira značajne razlike između grupa.

Pojam	Objašnjenje
SS (Sum of Squares)	Mjera varijabilnosti. $SS_{total} = SS_{between} + SS_{within}$.
MS (Mean Square)	SS / df . $MS_{between} = SS_{between} / (k-1)$. $MS_{within} = SS_{within} / (N-k)$.
Omnibus test	Test koji detektira da razlika postoji negdje, ali ne govori gdje specifično.
Post-hoc test	Test koji se provodi NAKON značajne ANOVA-e za identifikaciju specifičnih razlika.
Tukey HSD	Post-hoc test koji uspoređuje sve parove grupa uz kontrolu ukupne greške tipa I.
Eta-kvadrat	$SS_{between} / SS_{total}$. Proporcija varijabilnosti objašnjena grupnom pripadnošću. 0.01/0.06/0.14 mali/srednji/veliki.
Omega-kvadrat	Manje pristrana alternativa eta-kvadratu. Preporučena za publikacije.
Inflacija alfa	Porast greške tipa I pri višestrukim usporedbama. ANOVA to kontrolira.
Homogenost varijance	Pretpostavka jednakih varijanci u svim grupama. Levenov test je provjerava.
Welchova ANOVA	<code>oneway.test(var.equal = FALSE)</code> . ANOVA koja ne zahtijeva jednake varijance.
Kruskal-Wallisov test	Neparametrijska alternativa jednosmjernoj ANOVA-i. Koristi rangove. <code>kruskal.test()</code> .
Dunnov test	Post-hoc test za Kruskal-Wallis. Neparametrijski ekvivalent Tukeyu.
Planirane usporedbe	Unaprijed definirane specifične usporedbe. Snažniji od post-hoc testova jer testiraju manje hipoteza.
<code>aov()</code>	R funkcija za klasičnu ANOVA-u. <code>aov(y ~ grupa, data = ...)</code> .
<code>TukeyHSD()</code>	R funkcija za Tukey post-hoc test. Prima <code>aov()</code> objekt.
<code>oneway.test()</code>	R funkcija za Welchovu ANOVA-u. <code>var.equal = FALSE</code> za robusnu verziju.
<code>kruskal.test()</code>	R funkcija za Kruskal-Wallisov test. Sintaksa kao <code>aov()</code> .
<code>pairwise.wilcox.test()</code>	R funkcija za pairwise Wilcoxon testove s korekcijom p-vrijednosti.