PROST P600 combined groups analysis using lmer

Joanna Morris

2024-08-19

This script, on the advice of reviewer 1, conducts an ANOVA examining the P600 PROST data with Referentiality, Gender and Anaphor Type as within-subject variables. Gender Identity status will be examined as a post-hoc variable.

Define functions, set parameters and load

Define standard error of mean function

```
sem <- function(x) sd(x)/sqrt(length(x))
```

Set general parameters for `ggplot2`. We will set a general theme using the `theme_set()` function. We will use the 'classic' theme which gives us clean white background rather than the default grey with white grid lines. We will position the legend at the top of the graph rather than at the right side which is the default.

```
theme_set(theme_classic()+
            theme(legend.position = "bottom"))
```

Load the Data

```
prost_2022_singular <- read_csv("prost_2022_singular_p600.csv")
prost_2022_plural <- read_csv("prost_2022_plural_p600.csv")
prost_2024_combined <- bind_rows(list("Gendered Singular" = prost_2022_singular,
                                      "NonGendered Plural" = prost_2022_plural), .id = "Anaphor_Type")
```

Re-order factor levels for *Referentiality* and *Anaphor_Type*

```
prost_2024_combined$Referentiality <- factor(prost_2024_combined$Referentiality,
                                             levels=c('Referential',
                                                      'NonReferential'))
prost_2024_combined$Anaphor_Type <- factor(prost_2024_combined$Anaphor_Type,
                                             levels=c('Gendered Singular',
                                                      'NonGendered Plural'))
levels(prost_2024_combined$Referentiality)
levels(prost_2024_combined$Anaphor_Type)
```

Check ANOVA assumptions

- *No significant outliers in any cell of the design.* This can be checked by visualizing the data using box plot methods and by using the function `identify_outliers()` in the `rstatix` package.

```
library(rstatix)
kable(identify_outliers(prost_2024_combined, diff_score))
```

Anaphor_Type	SubjID	Referentiality	Gender_Status	Group	Baseline	Critical	diff_score	is.outlier	is.extreme
Gendered Singular	203	NonReferential	Gendered	Binary	-2.43645	2.24315	4.67960	TRUE	FALSE
Gendered Singular	207	Referential	Gendered	Binary	3.27080	-0.69125	-3.96205	TRUE	FALSE
Gendered Singular	216	Referential	Gendered	Binary	3.30220	-1.11125	-4.41345	TRUE	FALSE
Gendered Singular	221	NonReferential	Gendered	Binary	0.71770	5.81575	5.09805	TRUE	FALSE
Gendered Singular	305	NonReferential	Gendered	NonBinary	0.12800	6.65140	6.52340	TRUE	FALSE
Gendered Singular	312	NonReferential	Gendered	NonBinary	-1.43700	3.31660	4.75360	TRUE	FALSE
NonGendered Plural	216	Referential	NonGendered	Binary	2.37845	-1.65425	-4.03270	TRUE	FALSE
NonGendered Plural	222	NonReferential	NonGendered	Binary	2.17760	-2.10860	-4.28620	TRUE	FALSE

- *Normality*: the outcome (or dependent) variable should be approximately normally distributed in each cell of the design. This can be checked using the Shapiro-Wilk normality test `shapiro_test()`in the `rstatix` package.

```
kable(shapiro_test(prost_2024_combined, diff_score))
```

variable	statistic	p
diff_score	0.9918709	0.0938187

- *Assumption of sphericity*: the variance of the differences between groups should be equal. This can be checked using the Mauchly's test of sphericity, which is automatically reported when using the R function `anova_test()` in the `rstatix` package.

Analysis using `rstatix()`

```
fitted.model.0 <- anova_test(data = prost_2024_combined,
                      dv = diff_score,
                      wid = SubjID,
                      within = c(Referentiality, Gender_Status, Anaphor_Type))
kable(fitted.model.0)
```

Effect	DFn	DFd	F	p	p<.05	ges
Referentiality	1	37	6.212	1.70e-02	*	0.019000
Gender_Status	1	37	2.094	1.56e-01		0.008000
Anaphor_Type	1	37	0.317	5.77e-01		0.000985
Referentiality:Gender_Status	1	37	0.136	7.15e-01		0.000602
Referentiality:Anaphor_Type	1	37	28.976	4.30e-06	*	0.080000
Gender_Status:Anaphor_Type	1	37	0.410	5.26e-01		0.001000
Referentiality:Gender_Status:Anaphor_Type	1	37	11.701	2.00e-03	*	0.036000

Analysis using EZANOVA

```
library(ez)
fitted.model.1 <- ezANOVA(data = prost_2024_combined
             , dv = diff_score
             , wid = SubjID
             , within = .(Referentiality, Gender_Status, Anaphor_Type)
             , type = 3
             , return_aov = F
             )
pander(fitted.model.1)
```

- **ANOVA**:

Table 4: Table continues below

	Effect	DFn	DFd	F
2	Referentiality	1	37	6.212
3	Gender_Status	1	37	2.094
4	Anaphor_Type	1	37	0.3171
5	Referentiality:Gender_Status	1	37	0.1358
6	Referentiality:Anaphor_Type	1	37	28.98

	Effect	DFn	DFd	F
7	Gender__Status:Anaphor__Type	1	37	0.4098
8	Referentiality:Gender__Status:Anaphor__Type	1	37	11.7

	p	p<.05	ges
2	0.0173	*	0.01894
3	0.1563		0.008441
4	0.5768		0.0009849
5	0.7146		0.0006023
6	4.299e-06	*	0.08024
7	0.526		0.001088
8	0.001537	*	0.03579

Analysis using lmer

```
library(lme4)
library(car)
fitted.model.2 <- lmer(diff_score ~ Referentiality * Gender_Status * Anaphor_Type + (1|SubjID), data=prost_2024_combined)
kable(Anova(fitted.model.2))
```

	Chisq	Df	Pr(>Chisq)
Referentiality	5.7358431	1	0.0166221
Gender__Status	2.5294548	1	0.1117388
Anaphor__Type	0.2929198	1	0.5883555
Referentiality:Gender__Status	0.1790786	1	0.6721663
Referentiality:Anaphor__Type	25.9197219	1	0.0000004
Gender__Status:Anaphor__Type	0.3236341	1	0.5694318
Referentiality:Gender__Status:Anaphor__Type	11.0278606	1	0.0008975

Post-hoc tests

If there is a significant three-way interaction effect, you can decompose it into:

- *Simple two-way interaction*: run two-way interaction at each level of third variable,
- *Simple simple main effect*: run one-way model at each level of second variable, and/or
- *Simple simple pairwise comparisons*: run pairwise or other post-hoc comparisons if necessary.

Compute simple two-way interaction

You are free to decide which two variables will form the simple two-way interactions and which variable will act as the third (moderator) variable. In the following R code, we have considered the simple two-way interaction of Referentiality*Gender Status at each level of Anaphor Type

It is recommended to adjust the p-value for multiple testing (Bonferroni correction) by dividing the current α-level you declare statistical significance at (i.e., p < 0.05) by the number of simple two-way interaction you are computing (i.e., 2). Thus two-way interaction as statistically significant when $p < 0.025$ (i.e., $p < 0.05/2$).

```
prost_2024_combined <- prost_2024_combined |> ungroup() |> group_by(Anaphor_Type)

kable(two.way <- prost_2024_combined |>
  anova_test(dv = diff_score, wid = SubjID, within = c(Referentiality, Gender_Status)))
```

Anaphor__Type	Effect	DFn	DFd	F	p	p<.05	ges
Gendered Singular	Referentiality	1	37	24.535	1.63e-05	*	0.160
Gendered Singular	Gender__Status	1	37	2.082	1.57e-01		0.016
Gendered Singular	Referentiality:Gender__Status	1	37	5.367	2.60e-02	*	0.028
NonGendered Plural	Referentiality	1	37	5.378	2.60e-02	*	0.024
NonGendered Plural	Gender__Status	1	37	0.594	4.46e-01		0.003
NonGendered Plural	Referentiality:Gender__Status	1	37	4.739	3.60e-02	*	0.045

Can also compute these using lmer

```
library(lme4)
fitted.model.2a <- lmer(diff_score ~ Referentiality * Gender_Status  + (1|SubjID), data=filter(prost_2024_combined,Anaphor_Type == "Gendered Singular" ))
kable(Anova(fitted.model.2a))
```

	Chisq	Df	Pr(>Chisq)

	Chisq	Df	Pr(>Chisq)
Referentiality	28.130117	1	0.0000001
Gender__Status	2.340409	1	0.1260565
Referentiality:Gender__Status	4.214542	1	0.0400789

```
fitted.model.2b <- lmer(diff_score ~ Referentiality * Gender_Status  + (1|SubjID), data=filter(prost_2024_combined,Anaphor_Type == "NonGendered Plural" ))
kable(Anova(fitted.model.2b))
```

	Chisq	Df	Pr(>Chisq)
Referentiality	3.595763	1	0.0579270
Gender__Status	0.516181	1	0.4724753
Referentiality:Gender__Status	6.933698	1	0.0084587

Compute simple simple main effects using `anova_test()` function in the `rstatix` package

A statistically significant simple two-way interaction can be followed up with simple simple main effects.

Group the data by Anaphor__Type and Gender__Status, and analyze the simple main effect of Referentiality. The Bonferroni adjustment will be considered leading to statistical significance being accepted at the p < 0.025 level (that is 0.05 divided by the number of tests (here 2).

```
# Effect of Referentiality at each Anaphor_Type X Gender_Status cell
kable(ref.effect <- prost_2024_combined |>
  group_by(Anaphor_Type, Gender_Status) |>
  anova_test(dv = diff_score, wid = SubjID, within = Referentiality) )
```

Anaphor__Type	Gender__Status	Effect	DFn	DFd	F	p	p<.05	ges
Gendered Singular	Gendered	Referentiality	1	37	23.359	2.36e-05	*	0.233
Gendered Singular	NonGendered	Referentiality	1	37	6.833	1.30e-02	*	0.082
NonGendered Plural	Gendered	Referentiality	1	37	11.333	2.00e-03	*	0.143
NonGendered Plural	NonGendered	Referentiality	1	37	0.221	6.41e-01		0.003

Compute simple simple main effects using `t.test()` function

Create 4 separate dataframes for each test in order to do paired test. Below I run an unpaired test after each paired test just to see what how different they are..

```
singular.gendered <- prost_2022_singular |> filter(Gender_Status == "Gendered")
singular.nongendered <- prost_2022_singular |> filter(Gender_Status == "NonGendered")
plural.gendered <- prost_2022_plural |> filter(Gender_Status == "Gendered")
plural.nongendered <- prost_2022_plural |> filter(Gender_Status == "NonGendered")
```

"Some woman. . . himself" vs. "Mary. . . himself"

```
pander(t.test(singular.gendered$diff_score[singular.gendered$Referentiality == "Referential"],
              singular.gendered$diff_score[singular.gendered$Referentiality == "NonReferential"],
              paired = TRUE))
```

Table 11: Paired t-test: `singular.gendered$diff_score[singular.gendered$Referentiality == "Referential"]` and `singular.gendered$diff_score[singular.gendered$Referentiality == "NonReferential"]`

Test statistic	df	P value	Alternative hypothesis	mean difference
-4.833	37	2.36e-05 * * *	two.sided	-1.893

```
pander(t.test(diff_score ~ Referentiality,
              filter(prost_2024_combined,
                     Anaphor_Type == "Gendered Singular" & Gender_Status == "Gendered" )))
```

Table 12: Welch Two Sample t-test: `diff_score` by `Referentiality` (continued below)

Test statistic	df	P value	Alternative hypothesis
-4.747	72.97	1.002e-05 * * *	two.sided

mean in group Referential	mean in group NonReferential
-0.4045	1.489

mean in group Referential	mean in group NonReferential

"Someone... himself" vs. "The participant... himself"

```
pander(t.test(singular.nongendered$diff_score[singular.nongendered$Referentiality == "Referential"],
              singular.nongendered$diff_score[singular.nongendered$Referentiality == "NonReferential"],
              paired = TRUE))
```

Table 14: Paired t-test: `singular.nongendered$diff_score[singular.nongendered$Referentiality == "Referential"]` and `singular.nongendered$diff_score[singular.nongendered$Referentiality == "NonReferential"]`

Test statistic	df	P value	Alternative hypothesis	mean difference
-2.614	37	0.01286 *	two.sided	-0.8365

```
pander (t.test(diff_score ~ Referentiality,
               filter(prost_2024_combined,
                      Anaphor_Type == "Gendered Singular" & Gender_Status == "NonGendered")))
```

Table 15: Welch Two Sample t-test: `diff_score` by `Referentiality` (continued below)

Test statistic	df	P value	Alternative hypothesis
-2.571	67.79	0.01233 *	two.sided

mean in group Referential	mean in group NonReferential
-0.2699	0.5666

"Some woman... themselves" vs. "Mary..themselves"

```
pander(t.test(plural.gendered$diff_score[plural.gendered$Referentiality == "Referential"],
              plural.gendered$diff_score[plural.gendered$Referentiality == "NonReferential"],
              paired = TRUE))
```

Table 17: Paired t-test: `plural.gendered$diff_score[plural.gendered$Referentiality == "Referential"]` and `plural.gendered$diff_score[plural.gendered$Referentiality == "NonReferential"]`

Test statistic	df	P value	Alternative hypothesis	mean difference
3.366	37	0.001787 * *	two.sided	1.174

```
pander(t.test(diff_score ~ Referentiality,
              filter(prost_2024_combined,
                     Anaphor_Type == "NonGendered Plural" & Gender_Status == "Gendered")))
```

Table 18: Welch Two Sample t-test: `diff_score` by `Referentiality` (continued below)

Test statistic	df	P value	Alternative hypothesis
3.51	73.53	0.0007689 * * *	two.sided

mean in group Referential	mean in group NonReferential
0.9267	-0.2474

"Someone... themselves" vs. "The participant... themselves"

```
pander(t.test(plural.nongendered$diff_score[plural.gendered$Referentiality == "Referential"],
              plural.nongendered$diff_score[plural.gendered$Referentiality == "NonReferential"],
              paired = TRUE))
```

Table 20: Paired t-test: `plural.nongendered$diff_score[plural.gendered$Referentiality == "Referential"]` and `plural.nongendered$diff_score[plural.gendered$Referentiality == "NonReferential"]`

Test statistic	df	P value	Alternative hypothesis	mean difference
-0.4705	37	0.6407	two.sided	-0.191

```r
pander(t.test(diff_score ~ Referentiality,
              filter(prost_2024_combined,
                     Anaphor_Type == "NonGendered Plural" & Gender_Status == "NonGendered" )))
```

Table 21: Welch Two Sample t-test: `diff_score` by `Referentiality` (continued below)

Test statistic	df	P value	Alternative hypothesis
-0.482	73.91	0.6312	two.sided

mean in group Referential	mean in group NonReferential
0.0579	0.2489

Condition Means

Significant Effects: **Referentiality; Referentiality x Anaphor Type; Referentiality X Gender Status X Anaphor Type**

```r
kable(singular_means1 <- prost_2024_combined |>
  group_by(Referentiality) |>
  summarise(Mean = mean(diff_score),
            SE = sem(diff_score),
            SD = sd(diff_score),
            Max = max(diff_score),
            Min = min(diff_score)), digits = 2)
```

Referentiality	Mean	SE	SD	Max	Min
Referential	0.08	0.13	1.59	4.45	-4.41
NonReferential	0.51	0.14	1.76	6.52	-4.29

```r
kable(singular_means1 <- prost_2024_combined |>
  group_by(Referentiality, Anaphor_Type) |>
  summarise(Mean = mean(diff_score),
            SE = sem(diff_score),
            SD = sd(diff_score),
            Max = max(diff_score),
            Min = min(diff_score)), digits = 2)
```

Referentiality	Anaphor_Type	Mean	SE	SD	Max	Min
Referential	Gendered Singular	-0.34	0.16	1.42	4.15	-4.41
Referential	NonGendered Plural	0.49	0.19	1.66	4.45	-4.03
NonReferential	Gendered Singular	1.03	0.20	1.78	6.52	-3.33
NonReferential	NonGendered Plural	0.00	0.18	1.60	3.52	-4.29

```r
kable(singular_means2 <- prost_2024_combined |>
  group_by( Anaphor_Type, Gender_Status,  Referentiality) |>
  summarise(Mean = mean(diff_score),
            SE = sem(diff_score),
            SD = sd(diff_score),
            Max = max(diff_score),
            Min = min(diff_score)), digits = 2)
```

Anaphor_Type	Gender_Status	Referentiality	Mean	SE	SD	Max	Min
Gendered Singular	Gendered	Referential	-0.40	0.26	1.63	4.15	-4.41
Gendered Singular	Gendered	NonReferential	1.49	0.30	1.84	6.52	-1.66
Gendered Singular	NonGendered	Referential	-0.27	0.19	1.18	2.54	-2.50
Gendered Singular	NonGendered	NonReferential	0.57	0.26	1.62	4.02	-3.33
NonGendered Plural	Gendered	Referential	0.93	0.25	1.52	4.45	-2.35
NonGendered Plural	Gendered	NonReferential	-0.25	0.23	1.40	3.52	-2.77
NonGendered Plural	NonGendered	Referential	0.06	0.28	1.70	3.52	-4.03
NonGendered Plural	NonGendered	NonReferential	0.25	0.29	1.76	3.18	-4.29

Visualization: Box plots with p-values

Compute simple simple main effects with Bonferroni adjustment using `pwc()` function in the `rstatix`

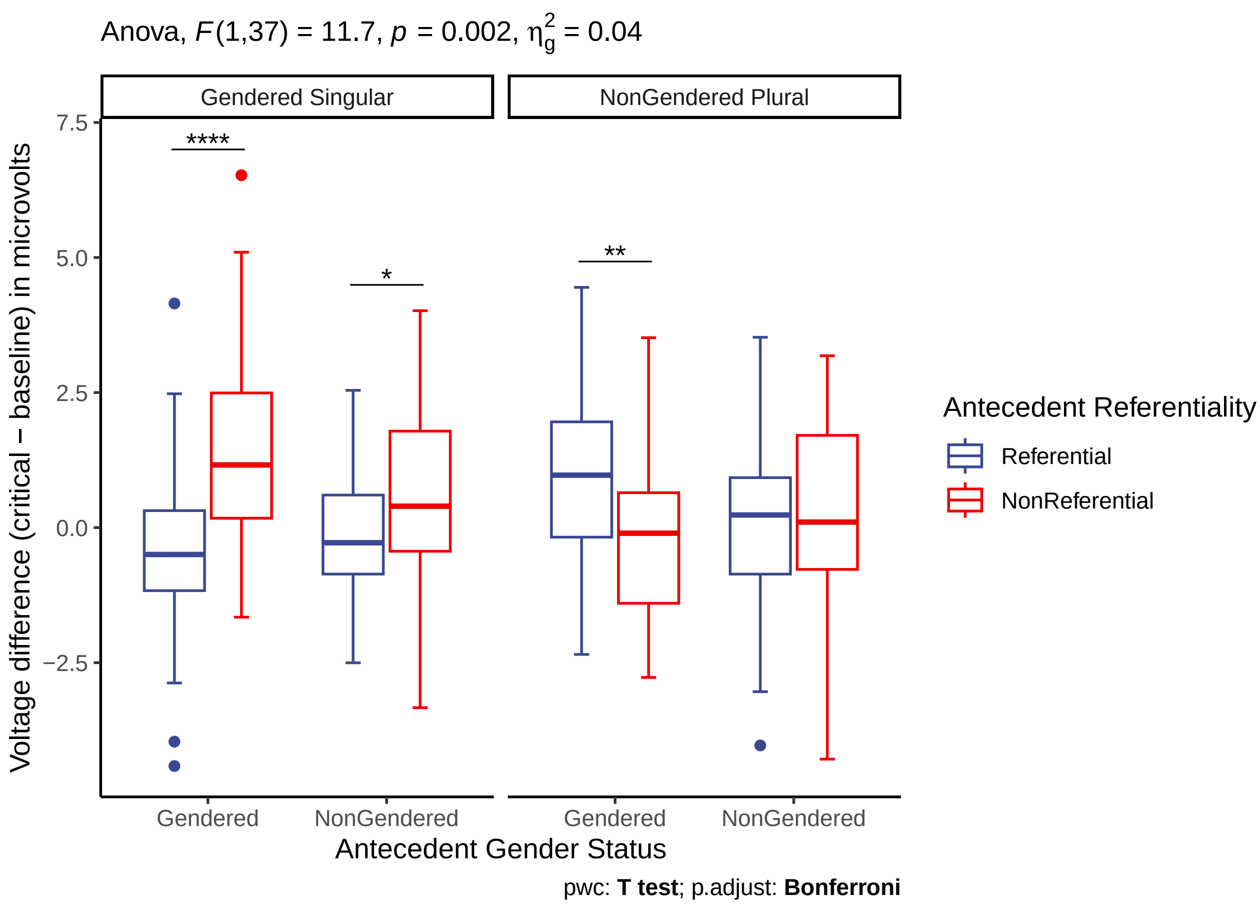
```r
# Pairwise comparisons
pwc <- prost_2024_combined |>
  group_by(Anaphor_Type, Gender_Status) |>
  pairwise_t_test(diff_score ~ Referentiality, paired = TRUE, p.adjust.method = "bonferroni") |>
  select(-p, -.y., -n2)
kable(pwc)
```

Anaphor_Type	Gender_Status	group1	group2	n1	statistic	df	p.adj	p.adj.signif
Gendered Singular	Gendered	Referential	NonReferential	38	-4.8331203	37	2.36e-05	****
Gendered Singular	NonGendered	Referential	NonReferential	38	-2.6140474	37	1.30e-02	*
NonGendered Plural	Gendered	Referential	NonReferential	38	3.3664471	37	2.00e-03	**
NonGendered Plural	NonGendered	Referential	NonReferential	38	-0.4705374	37	6.41e-01	ns

```
prost_2024_combined <- ungroup(prost_2024_combined)
library(ggplot2)
library(ggsci)
library(ggpubr)

bxp2 <- ggplot(data = prost_2024_combined,
       mapping = aes(x = Gender_Status, y = diff_score, colour = Referentiality)) +
  geom_boxplot(staplewidth = .25) +
  facet_wrap(~ Anaphor_Type, ncol = 2) +
  labs(y = "Voltage difference (critical - baseline) in microvolts", x = "Gender Status") +
  theme_classic() +
  scale_color_aaas() +
  scale_fill_aaas(alpha = 0.3)


pwc <- pwc |> add_xy_position(x = "Gender_Status")
pwc.filtered <- pwc |>
  filter(Anaphor_Type == "Gendered Singular", Gender_Status == "Gendered")
bxp2 +
  stat_pvalue_manual(pwc, tip.length = 0, hide.ns = TRUE) +
  labs(subtitle = get_test_label(fitted.model.0, detailed = TRUE), caption = get_pwc_label(pwc)) +
  xlab("Antecedent Gender Status") +
  labs(colour="Antecedent Referentiality" )
```



Visualization: Bar chart

```r
summary_data_2 <- prost_2024_combined |>
  group_by(Anaphor_Type,Gender_Status, Referentiality) |>
  summarise(mean = mean(diff_score), se = sem(diff_score))|>
  mutate(lwr = mean - (1.96*se), upr = mean + (1.96*se))
summary_data_2$label = c("John...herself","some man...herself","the participant...herself", "someone...herself", "John...themselves", "some man...

(plotA <- ggplot(summary_data_2, aes(x = Gender_Status,
                         y = mean,
                         ymin = lwr,
                         ymax = upr,
                         fill = Referentiality,
                         colour = Referentiality,
                         label = label)) +
    geom_bar(position=position_dodge(), stat="identity") +
    geom_errorbar(width = .15, position = position_dodge(.9), colour = "black")  +
    geom_label_repel(show.legend = FALSE, cex = 2, colour = "black") +
    ylab("Voltage difference (critical - baseline) in microvolts") +
    xlab("Antecedent Gender Status") +
    theme(legend.text = element_text(size=10))+
    theme(legend.title = element_blank()) +
    facet_wrap(~ Anaphor_Type, ncol = 2) +
    scale_color_npg() +
    scale_fill_npg(alpha = 0.8))
```