The General Linear Model – ANOVA Part 1 (Video 2)

Dr Andrew J. Stewart

E: drandrewjstewart@gmail.com
T: @ajstewart_lang
G: ajstewartlang

The Packages

There is a built in ANOVA function in R, called `aov()`, but you shouldn't use it - it's not easy to build models for repeated measures design, and the default is to use Type I Sums of Squares in the model output. This can make things tricky when you have factorial designs with interactions.

```
library(tidyverse) #load the tidyverse packages

library(afex) #load afex for running ANOVA

library(emmeans) #load emmeans for running pairwise comparisons
```

Example ANOVA

We have 45 participants in a between participants design where we are interested in the effect of beverage consumed on ability on a motor task. Our experimental factor (beverage type) has 3 levels. These are Water vs. Single Espresso vs. Double Espresso, and Ability is our DV measured on a continuous scale.

```
my_data <-
read_csv("https://raw.githubusercontent.com/ajstewartlang/11_glm_anova_pt1/master/data/cond.csv")
```

Example ANOVA

```
head(my_data)

# A tibble: 6 x 3
  Participant Condition Ability
        <dbl> <chr>       <dbl>
1           1 Water        4.82
2           2 Water        5.41
3           3 Water        5.73
4           4 Water        4.36
5           5 Water        5.47
6           6 Water        5.50
```

We need to ensure that R recognises that our Condition variable is a factor so let’s change that.

Tidying our Data

```
my_data_tidied <- my_data %>%
  mutate(Condition = factor(Condition))
head(my_data_tidied)

# A tibble: 6 x 3
  Participant Condition Ability
        <dbl> <fct>       <dbl>
1           1 Water        4.82
2           2 Water        5.41
3           3 Water        5.73
4           4 Water        4.36
5           5 Water        5.47
6           6 Water        5.50
```

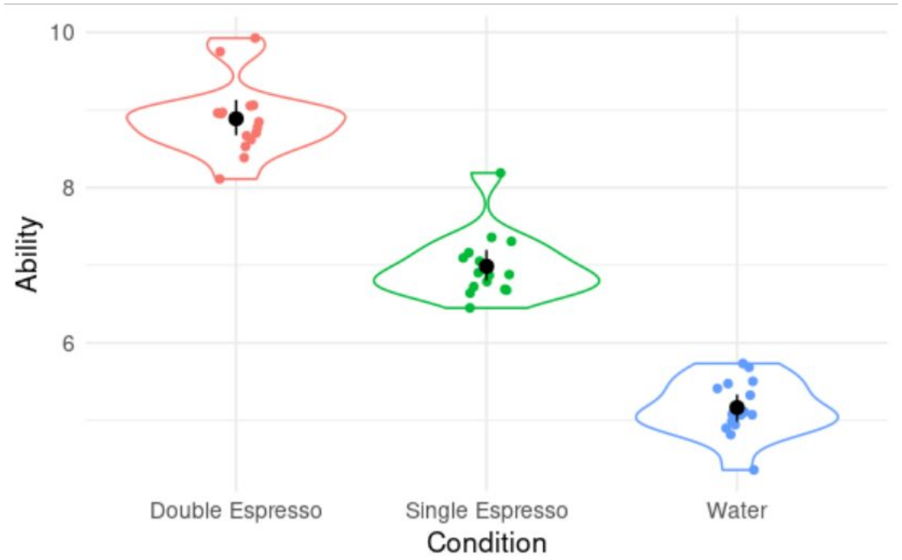
Summarising our Data

```
my_data_tidied %>%
  group_by(Condition) %>%
  summarise(mean = mean(Ability), sd = sd(Ability))

# A tibble: 3 x 3
  Condition          mean    sd
  <fct>             <dbl> <dbl>
1 Double Espresso    8.89 0.467
2 Single Espresso  6.99 0.419
3 Water              5.17 0.362
```

Visualising our Data

```
my_data_tidied %>%
  ggplot(aes(x = Condition, y = Ability,
colour = Condition)) +
  geom_violin() +
  geom_jitter(width = .1) +
  guides(colour = FALSE) +
  stat_summary(fun.data = "mean_cl_boot",
colour = "black") +
  theme(text = element_text(size = 13)) +
  theme_minimal()
```



Modelling our Data

```
model <- aov_4(Ability ~ Condition + (1 | Participant), data = my_data_tidied)

summary(model)

Anova Table (Type 3 tests)

Response: Ability
          num Df den Df     MSE      F     ges    Pr(>F)
Condition      2     42 0.17484 297.05 0.93397 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Our Effect Size Measure

The effect size is measured by ges which stands for generalised effect size or generalised eta squared. For designs with more than one factor it can be a useful indicator of how much variance in the dependent variable can be explained by each factor (plus any interactions between factors).

```
Anova Table (Type 3 tests)

Response: Ability
          num Df den Df     MSE      F     ges    Pr(>F)
Condition      2     42 0.17484 297.05 0.93397 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpreting our Model

To determine what's driving the effect we can use `emmeans::emmeans()` to run pairwise comparisons (note, default is Tukey correction).

```
emmeans(model, pairwise ~ Condition)
$emmeans
 Condition        emmean    SE df lower.CL upper.CL
 Double Espresso    8.89 0.108 42     8.67     9.10
 Single Espresso    6.99 0.108 42     6.77     7.20
 Water              5.17 0.108 42     4.95     5.38

Confidence level used: 0.95

$contrasts
 contrast                          estimate    SE df t.ratio p.value
 Double Espresso - Single Espresso     1.90 0.153 42 12.453  <.0001
 Double Espresso - Water               3.72 0.153 42 24.372  <.0001
 Single Espresso - Water               1.82 0.153 42 11.920  <.0001

P value adjustment: tukey method for comparing a family of 3 estimates
```

Interpreting our Model

We found a significant effect of Beverage type ($F\ (2,42) = 297.05$, $p < .001$, generalised $\eta^2 = .93$). Tukey comparisons revealed that the Water group performed significantly worse than the Single Espresso Group ($p < .001$), that the Water group performed significantly worse than the Double Espresso Group ($p < .001$), and that the Single Espresso Group performed significantly worse than the Double Espresso Group ($p < .001$).

In other words, drinking some coffee improves motor performance relative to drinking water, and drinking a lot of coffee improves motor performance even more.