Multivariate Power Analysis

Aaron R. Caldwell

9/21/2019

Multivariate ANOVA (MANOVA)

A large proportion of research with within-subjects manipulations, or repeated measures, rely upon the "univariate" approach (Maxwell & Delaney). While this approach is valid, when corrections for sphercity are applied, it may not be the most powerful or informative analysis plan. Instead, researchers should consider a multivariate analysis (MANOVA). While the MANOVA is not "assumption free" it does not assume sphercity which makes it a very attractive analytical tool and the preferred method of analysis for some (Maxwell and Delaney).

For a simple one-way repeated measures design, there are some simple guidelines for power analysis set forth by Maxwell and Delaney (pg. 750). All that is needed in the effect size calculated as:

$$d = \frac{\mu_{max} - \mu_{min}}{\sigma}$$

This assumes that each level has a common standard deviation (i.e., there is only 1 `sd` input for the design).

In addition, the non-centrality parameter of the F-statistic can be estimated from Vonesh and Schork (1986) equations as the following:

$$\delta^2 = \frac{n \cdot d^2}{2 \cdot (1 - \rho_{min})}$$

Let us assume we have a `2w` design with `mu = c(0,0.5)`, a common standard deviation of 1 (`sd=1`), and correlation between `a1` and `a4` of `r = .4` and a total sample size of 15 (`n = 15`) participants. Power could then be calculated with the following `R` code.

```
mu = c(0,0.5)
rho = .4
sd = 2
n = 15
d = (max(mu)-min(mu))/sd
noncentrality = ((n*d^2) / (2*(1-min(rho))))
noncentrality
```

```
## [1] 0.78125
```

```
#Critical F
Ft <- qf((1 - .05), 1, 14)
Ft
```

```
## [1] 4.60011
```

```
#Power
power <- (1 - pf(Ft,
                 1,
                 14,
                 noncentrality)) * 100
power
```

```
## [1] 13.07682
```

```r
design_result <- ANOVA_design("2w",
                              n = n,
                              r = rho,
                              sd = sd,
                              mu = mu)

exact_result <- ANOVA_exact(design_result, verbose = FALSE)
```

Table 1: MANOVA Result

	power	pillai_trace	cohen_f	non_centrality
(Intercept)	8.401182	0.0233572	0.1546474	0.3348214
a	13.076818	0.0528541	0.2362278	0.7812500

The problem with this formula for determining power is that it is inexact and makes a number of assumptions (Maxwell and Delaney, ppg. 752). In reality, it can only give a lower bound estimate of power for a given design, and the actual power may be much higher. This is problematic because it could lead to inefficient study design (e.g., determining you need 20 participants when adequate power could be achieved with less participants). This will become increasingly important with designs with multiple levels and possible violations of the assumption of sphericity.

Sphericity Assumption

In all the above examples, you will notice that the correlation is the sample between repeated measures. For most experiments in real life, this will not be the case. The correlations between levels may vary. This is problematic because the univariate (ANOVA) approach assumes sphericity, which means, for all intents and purposes, that the correlations between factor-levels are equal and the standard deviations at each level are equal as well. This assumption is tenuous at best and is typically "adjusted" for by applying a sphericity correction (e.g, Greenhouse-Geisser). How bad can it get? Well, let's simulate an example below.

```r
design_result <- ANOVA_design("4w",
                              n = 29,
                              r = c(.05,.15,.25,.55, .65, .9
                                    ),
                              sd = 1,
                              mu= c(0,0,0,0))

#In order to simulate violations we MUST use ANOVA_power
power_result_s1 <- ANOVA_power(design_result, nsims = nsims, verbose = FALSE)
```

Table 2: Simulated ANOVA Result

	power	effect_size
anova_a	7.16	0.0344525

INSERT ANOVA Result ABOUT HERE

As we can see, the actual type I error rate far exceeds the typical 5%!

Now, let's pour gasoline on the fire and see what happens when we make the sphericity violation worse by varying the standard deviations.

```
design_result <- ANOVA_design("4w",
                              n = 29,
                              r = c(.05,.15,.25,.55, .65, .9
                                    ),
                              sd = c(1,3,5,7),
                              mu= c(0,0,0,0))

#In order to simulate violations we MUST use ANOVA_power
power_result_s2 <- ANOVA_power(design_result, nsims = nsims, verbose = FALSE)
```

Table 3: Simulated ANOVA Result

	power	effect_size
anova_a	8.37	0.0346856

INSERT ANOVA Result ABOUT HERE

These inflated error rates are obviously a problem, which begs the question what do we do about them?

In our experience, most researchers default to using a Greenhouse-Geisser adjustment for sphericity, but this may not be the most statistical efficient way of dealing with violations of sphercity. Further, as we can see from the simulation (`power_result_s2`) the MANOVA maintains the type I error rate at 5%.

Insert MANOVA Result About HERE

MANOVA or Sphericity Adjustment?

In addition to adjusting for sphericity, one could also simply use the multivariate approach to repeated measures. While it is tempting to simply say one approach is superior to another, that is not case when the sample size is small (Maxwell and Delaney, ppg 775). Some general guidelines were proposed by Algina and Keselman (1997):

MANOVA when `levels <= 4, epsilon <= .9, n > levels + 15` and `5 <= levels <= 8, epsilon <= .85, n > levels + 30`.

However, `ANOVA_power` can make the solution easy. Simply simulate, like we have done above, for both the null situation (no differences) and with the hypothesized effect (to determine power), and see which approach best balances type I and II error rates.

As we saw above, the unadjusted repeated measures ANOVA has an elevated type I error rate, but the MANOVA analysis of the same data above approximately preserves the type I error rate. However, how does this perform relative to the sphericity corrections? Let's simulate again, and compare the results of the different corrections.

```
design_result <- ANOVA_design("4w",
                              n = 29,
                              r = c(.05,.15,.25,.55, .65, .9
                              ),
                              sd = c(1,3,5,7),
                              mu= c(0,0,0,0))

power_result_none <- ANOVA_power(design_result, nsims = nsims, verbose = FALSE)
```

Table 4: Simulated MANOVA Result

	power
manova_(Intercept)	5.16
manova_a	5.16

```
power_result_gg <- ANOVA_power(design_result, correction = "GG",
                               nsims = nsims, verbose = FALSE)
```

Table 5: Simulated ANOVA Result

	power	effect_size
anova_a	4.59	0.0339026

```
power_result_hf <- ANOVA_power(design_result, correction = "HF",
                               nsims = nsims, verbose = FALSE)
```

Table 6: Simulated ANOVA Result

	power	effect_size
anova_a	5.12	0.0346905

Both the sphericity corrections, Greenhouse-Geisser (`GG`) and Huynh-Feldt (`HF`), as well as the MANOVA were able to adequately control type I error rate. Howevever, Greenhouse-Geisser seemed to be a tad conservative. We can now directly compare the power of the MANOVA or `HF`-adjusted approach. We can adjust the study design to the alternative, or hypothesized, model with the predicted means of `0, 0.75, 1.5, 3` instead of the null model.

```
design_result_power <- ANOVA_design("4w",
                              n = 29,
                              r = c(.05,.15,.25,.55, .65, .9
                              ),
                              sd = c(1,3,5,7),
                              mu= c(0,0.75,1.5,3))

power_result_hfeffect <- ANOVA_power(design_result_power, correction = "HF",
                                     nsims = nsims, verbose = FALSE)
```

Table 7: Simulated ANOVA Result

	power	effect_size
anova_a	60.48	0.1479382

Table 8: Simulated MANOVA Result

	power
manova_(Intercept)	49.60
manova_a	49.64

Well, it appears the MANOVA based approach has roughly ~10% lower power compared the `HF`-adjusted ANOVA. However, the study still appears to be underpowered so we can increase the sample size. We will continue to evaluate the MANOVA results because, as Algina and Keselman (1997) note, the power disadvantage of MANOVA is diminished with increased sample size.

```r
design_result_power <- ANOVA_design("4w",
                              n = 29,
                              r = c(.05,.15,.25,.55, .65, .9),
                              sd = c(1,3,5,7),
                              mu= c(0,0.75,1.5,3))

power_result_hfeffect <- ANOVA_power(design_result_power, correction = "HF",
                                     nsims = nsims, verbose = FALSE)
```

Table 9: Simulated ANOVA Result

	power	effect_size
anova_a	84.69	0.1385088

Table 10: Simulated MANOVA Result

	power
manova_(Intercept)	74.11
manova_a	79.18

Again, the `HF`-adjusted analysis appears to be more powerful for *this very specific experimental design*. The difference in power between univariate and multivariate ouput is diminished when the sample size is increased.