For a power analysis

Henry Williams

2023-02-04

R Markdown

With my thesis advisor, we wanted to conduct a power analysis in GPower. The parameter (i.e., the ${\eta_p}^2$) that we wanted to use comes from a previous paper in which an interaction between a between-subjects variable (Group) and a within-subjects variable (Contingency) was found, the size of the effect was found to be ${\eta_p}^2$ = .181. Therefore, it seems that we have to use "*ANOVA: Repeated measures, within-between interaction*", hence, we can fill in the ${\eta_p}^2$ = .181. However, a main issue appears when we have to select a value for correlation among repeated measures. In fact, we were lost regarding what one has to fill in for this value. Regarding this correlation among repeated measures, we know that the greater the correlation among repeated measures, the more the power. In fact, an effect size for a within-subjects design cannot be equal to an effect size for a between-subjects design even if they were computed from the same data, because the former takes into account the correlation among repeated measures. As a consequence, if the correlation is strong, it inflates the effect size (e.g., a Cohen's d_D is a merge between the standardized mean difference [i.e., a Cohen's d_p, that is the standardized mean difference that one can find in a between-subject design] and the correlation between repeated measures; Dunlap et al., 1996; see below), offering more power (Lakens, 2013; Cousineau & Goulet-Pelletier, 2021). Now, to deal with such a problem and eliminate the influence of correlation, one can estimate an effect size for a given within-subjects variable in the same way as one would be done with a between-subject variable (regarding Cohen's *d* see Goulet-Pelletier &

Cousineau, 2018; Fitts, 2020; Cousineau & Goulet-Pelletier, 2021). Before continuing, our solution for the power analysis seems to take the more conservative mean to perform our analysis as suggested here:

https://stats.stackexchange.com/questions/44134/correlation-among-repeated-measures-i-need-an-explanation

one can simply fill in 0, but we will see that it can be problematic.

If I correctly understand the issue with GPower, it requires a ${\eta_p}^2$ that does not take into account the correlation among repeated measures (thus, an effect size that was estimated to be comparable across the designs; between a between-subjects variable and a within-subjects variable) and the program will include the correlation among the repeated measures later in order to offer more power, *that why it is demanded*. Hence, we better understand the main issue with a ${\eta_p}^2$ estimated in SPSS, since it directly returns an effect size that is influenced by the correlation among repeated measures (Lakens, 2013).

Let's take a reminder in Lakens (2013):

> *Although ${\eta_p}^2$ is more useful when the goal is to compare effect sizes across studies, it is not perfect, because ${\eta_p}^2$differs when the same two means are compared in a within-subjects design or a between-subjects design. In a within-subjects ANOVA, the error sum of squares can be calculated around the mean of each measurement, but also around the mean of each individual when the measurements are averaged across individuals. This allows researchers to distinguish variability due to individual differences from variability due to the effect in a within-subjects design, whereas this differentiation is not possible in a between-subjects design. As a consequence,*

whenever the two groups of observations are positively correlated, $\eta_p{}^2$ will be larger in a within-subjects design than in a between-subjects design. This is also the reason a within-subjects ANOVA typically has a higher statistical power than a between-subjects ANOVA.

To illustrate the issue, I simulated a 2X2 ANOVA followed by two one-way ANOVAs, though the latter is not optimal.

```
## anova simulation
library(MOTE)
library(MBESS)
library(data.table)
report<-function(anova){
  a<-as.data.table(anova)
  b<-nrow(a)
  test <-
data.table(Factor=factor(b),DFn=integer(b),DFd=integer(b),F=integer(
b),MSE=integer(b),p=integer(b),ges=integer(b), eta=integer(b),
eta.Ll =integer(b), eta.Ul=integer(b))
  test$Factor<-a$ANOVA.Effect
  test$DFn<-a$ANOVA.DFn
  test$DFd<-a$ANOVA.DFd
  test$F<-round(a$ANOVA.F,2)
  test$p<-round(a$ANOVA.p,3)
  test$ges<-round(a$ANOVA.ges, 3)
  loweretasquared <- c()
  upperetasquared <- c()
  for(i in 1:b){
    test$MSE[i]=a$ANOVA.SSd[i]/a$ANOVA.DFd[i]

test$eta[i]=round((a$ANOVA.DFn[i]*a$ANOVA.F[i])/((a$ANOVA.DFn[i]*a$A
NOVA.F[i])+a$ANOVA.DFd[i]),3)
    Lims <- conf.limits.ncf(F.value = a$ANOVA.F[i], conf.level =
0.90, df.1 <- a$ANOVA.DFn[i], df.2 <- a$ANOVA.DFd[i])
    Lower.lim <- Lims$Lower.Limit/(Lims$Lower.Limit + df.1 + df.2 +
1)
    Upper.lim <- Lims$Upper.Limit/(Lims$Upper.Limit + df.1 + df.2 +
1)
    if (is.na(Lower.lim)) {
      Lower.lim <- 0
    }
    if (is.na(Upper.lim)) {
      Upper.lim <- 1
```

```
    }
    loweretasquared <- c(loweretasquared,Lower.lim)
    upperetasquared <- c(upperetasquared,Upper.lim)
  }
  test$eta.Ll <- round(loweretasquared, 3)
  test$eta.Ul <- round(upperetasquared, 3)
  return(test)
}
```

Simulate three bivariate normal distributions

We will generate three measures from bivariate normal distributions:

data ~ N(μ, Σ)

where, $\Sigma = \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix}$, where $\sigma = 1$ and $\rho = 0.7$, across the three simulated samples.

```
library(MASS)
## Generate the data
set.seed(123)


n <- 30

## set the different population means:
mu1 <- c(0.36,0.36)
mu2 <- c(1.4, 1.4)
mu3 <- c(-0.8, 0.8)
## create covariance matrix with correlation = 0.7
# sd = 1 --> var = 1^2; rho = 0.7
cov_matrix <- array(c(1,0.7,0.7,1), dim =c(2,2))
cov_matrix

##      [,1] [,2]
## [1,]  1.0  0.7
## [2,]  0.7  1.0

gp1 <- mvrnorm(n, mu = mu1, Sigma = cov_matrix)
gp2 <- mvrnorm(n, mu = mu2, Sigma = cov_matrix)
gp3 <- mvrnorm(n, mu = mu3, Sigma = cov_matrix)

cor1 <- cor(x = gp1[,1], y = gp1[,2]); print(cor1)

## [1] 0.7771165

cor2 <- cor(x = gp2[,1], y = gp2[,2]); print(cor2)

## [1] 0.6780823

cor3 <- cor(x = gp3[,1], y = gp3[,2]); print(cor3)
```

```
## [1] 0.7879499
```

Create the data frame

```
factor1 <- c(gp1[,1], gp2[,1], gp3[,1])
  factor2 <- c(gp1[,2], gp2[,2], gp3[,2])

id <- 1:90
group <- c(rep(1, 30), rep(2, 30), rep(3, 30))
data <- cbind(id, group, factor1, factor2)
df <- as.data.frame(data); print(df)
```

```
##    id group    factor1      factor2
## 1   1     1 -0.32190190  0.008435869
## 2   2     1  0.26206753  0.033506146
## 3   3     1  1.45037738  2.143738738
## 4   4     1  0.08490589  0.765105162
## 5   5     1  0.16100042  0.797394387
## 6   6     1  1.67450257  2.207921013
## 7   7     1  0.57041236  0.999475130
## 8   8     1 -0.78235053 -0.830307132
## 9   9     1 -0.15474821 -0.391745870
## 10 10     1  0.09647575 -0.198235820
## 11 11     1  1.75760651  1.219488799
## 12 12     1  0.77225797  0.611205942
## 13 13     1  1.21957892  0.239407121
## 14 14     1 -0.37798861  1.302077455
## 15 15     1 -0.62030188  0.315381465
## 16 16     1  2.44243059  1.572474427
## 17 17     1  0.97503209  0.662958836
## 18 18     1 -1.27239659 -1.633866272
## 19 19     1  0.70453900  1.308697383
## 20 20     1 -0.04360344 -0.108180841
## 21 21     1 -0.72259465 -0.526374974
## 22 22     1  0.17009317  0.147980948
## 23 23     1 -0.56932571 -0.602533019
## 24 24     1 -0.84206189  0.218052876
## 25 25     1 -0.12881700 -0.303698459
## 26 26     1 -1.78238094 -0.607727857
## 27 27     1  1.73223087  0.532572108
## 28 28     1  0.27498310  0.727822959
## 29 29     1 -0.73727895 -0.641341867
## 30 30     1  1.43232643  1.599594051
## 31 31     2  1.36522792  2.134792698
## 32 32     2  0.72448743  1.149273888
## 33 33     2  1.00033757  1.185258375
## 34 34     2  0.70410687  0.217732920
## 35 35     2 -0.11512111  0.938835741
## 36 36     2  1.91231912  1.447360042
## 37 37     2  0.96607857  2.660379423
```

```
## 38 38      2  0.85528994  2.042445026
## 39 39      2  2.34157495  2.159002236
## 40 40      2  3.68761580  2.892553585
## 41 41      2  1.22243091  0.672152354
## 42 42      2 -0.82843914 -0.629457878
## 43 43      2  2.42278846  2.231701750
## 44 44      2  0.88075187  0.611546534
## 45 45      2  1.13424768  0.397127111
## 46 46      2  2.36296925  2.328090921
## 47 47      2  1.44144445  0.833460066
## 48 48      2  0.92054501 -0.371437256
## 49 49      2  1.71441465  1.419892451
## 50 50      2  0.91602263  1.627874347
## 51 51      2  1.62814524  1.182483396
## 52 52      2  1.51974741  1.990674548
## 53 53      2  1.68487161  0.431665057
## 54 54      2  2.01560488  1.972566767
## 55 55      2  0.99555589  1.397886979
## 56 56      2  1.58925166  1.822524052
## 57 57      2  2.37030739  2.452163818
## 58 58      2  2.04936188  1.553073140
## 59 59      2  1.42859500  0.770416848
## 60 60      2  2.85579167  2.062504916
## 61 61      3 -0.99662514  1.213554747
## 62 62      3 -1.97137721  0.224320346
## 63 63      3 -1.38093312  0.476389889
## 64 64      3 -0.64556275  0.173352084
## 65 65      3  0.94622057  2.453692977
## 66 66      3 -1.29247146  0.090335243
## 67 67      3 -0.80102921  1.235060605
## 68 68      3 -0.58387874  0.727631442
## 69 69      3 -2.06516817  0.291592166
## 70 70      3 -0.72066826  0.589182651
## 71 71      3  0.12409669  2.539523483
## 72 72      3  0.02261068  0.809921662
## 73 73      3 -0.27392910  0.349958848
## 74 74      3 -2.44477220  1.665726536
## 75 75      3 -2.53155215 -1.254448654
## 76 76      3  0.12753832  1.958544423
## 77 77      3 -2.39318598 -0.300101232
## 78 78      3  0.06956533  1.294830467
## 79 79      3  0.75992671  2.760286332
## 80 80      3 -2.27410307 -0.388304372
## 81 81      3 -0.06957030  1.363596676
## 82 82      3 -1.06702202  0.583553742
## 83 83      3 -2.23625111 -0.662639487
## 84 84      3 -3.02080045  0.227891299
## 85 85      3 -1.98942516 -0.963661631
## 86 86      3 -0.86499410 -0.113949157
```

```
## 87 87      3 -2.16230744 -0.533036676
## 88 88      3 -0.28602075  1.554476604
## 89 89      3  0.96713996  2.905269591
## 90 90      3 -1.80905934 -0.564107599
```

```
library(reshape2)
```

```
##
## Attachement du package : 'reshape2'
```

```
## Les objets suivants sont masqués depuis 'package:data.table':
##
##     dcast, melt
```

```
data_long <- melt(df, id.vars = c("id", "group"), measure.vars =
c("factor1", "factor2"), variable.name = "factor", value.name =
"DV")
```

Run the ANOVAs

```
options(contrasts = c("contr.sum", "contr.poly"))
library(ez)
data_long$id <- as.factor(data_long$id)
data_long$group <- as.factor(data_long$group)
data_long$factor <- as.factor(data_long$factor)
data_long$DV <- as.numeric(data_long$DV)
Model <- ezANOVA(data_long,
                dv =. (DV), wid =.(id),
                within =. (factor),
                between =. (group),
                type = 3,
                detailed = T,
                return_aov = T)
report(Model)
```

```
## Warning in as.data.table.list(anova): Item 2 has 3 rows but
longest item has 4;
## recycled with remainder.
```

```
##           Factor DFn DFd     F       MSE p   ges   eta eta.Ll
eta.Ul
## 1:   (Intercept)   1  87 28.63 1.7215142 0 0.224 0.248  0.125
0.362
## 2:         group   2  87 23.19 1.7215142 0 0.319 0.348  0.207
0.450
## 3:        factor   1  87 66.16 0.2406462 0 0.085 0.432  0.300
0.531
## 4:  group:factor   2  87 59.78 0.2406462 0 0.144 0.579  0.458
0.653
```

Regarding the output, the returned ${\eta_p}^2$ are greater than the ${\eta_G}^2$ for our two variables and the interaction as well. The ${\eta_p}^2$ estimates are inflated intrinsically, not only due to correlation among repeated measures since the *group* factor is a between subjects-variable. Lakens (2013) stated that "*For designs where all factors are manipulated between participants, ${\eta_p}^2$ and ${\eta_G}^2$ are identical*". Therefore, one can run a one-way ANOVA on *group*, confirming that both effect sizes are identical.

```
Model2 <- ezANOVA(data_long,
                  dv =. (DV), wid =.(id),
                  between =. (group),
                  type = 3,
                  detailed = T,
                  return_aov = T); report(Model2)

## Warning: Collapsing data to cell means. *IF* the requested
effects are a subset
## of the full design, you must use the "within_full" argument, else
results may be
## inaccurate.

## Coefficient covariances computed by hccm()

## Warning in as.data.table.list(anova): Item 1 has 2 rows but
longest item has 13;
## recycled with remainder.

##          Factor DFn DFd     F       MSE p   ges   eta eta.Ll
eta.Ul
##   1: (Intercept)   1  87 28.63 0.8607571 0 0.248 0.248  0.125
0.362
##   2:       group   2  87 23.19 0.8607571 0 0.348 0.348  0.207
0.450
```

One can perform the same thing on *factor* (though it is not optimal given the fact that one has to carry out a *t*-test), hence, *a part* of the difference between the ${\eta_p}^2$ and the ${\eta_G}^2$ will reflect the impact of correlation among repeated measures on the ${\eta_p}^2$ magnitude. Before, running the last ANOVA, we can just compute the correlation between factor 1 and factor 2 to have an intuitive idea of this impact.

```
corG <- cor(x = factor1, y = factor2); print(corG)

## [1] 0.6746506



Model3 <- ezANOVA(data_long,
                  dv =. (DV), wid =.(id),
                  within =. (factor),
                  type = 3,
                  detailed = T,
                  return_aov = T); report(Model3)

## Warning in as.data.table.list(anova): Item 1 has 2 rows but
longest item has 3;
## recycled with remainder.

##         Factor DFn DFd     F       MSE p   ges   eta eta.Ll
eta.Ul
## 1: (Intercept)   1  89 19.10 2.5800997 0 0.150 0.177  0.071
0.290
## 2:      factor   1  89 28.51 0.5585151 0 0.054 0.243  0.122
0.356
```

Given the fact, again, that "*For designs where all factors are manipulated between participants,* ${\eta_p}^2$ *and* ${\eta_G}^2$*are identical*", a ${\eta_p}^2$ computed with *factor* as a between-subjects variable should equal the ${\eta_G}^2$, hence, the correlation between measures inflated well the ${\eta_p}^2$ depicted above. To be sure about that, we have just to estimate the *t*-statistic when *factor* is treated as a between-subjects variable and as a within-subjects variable and convert them into ${\eta_p}^2$, using ${\eta_p}^2 = \frac{t^2}{t^2+dferror}$

```
tb <- t.test(x = factor1, y = factor2, paired = F, var.equal =
T)$statistic
tw <- t.test(x = factor1, y = factor2, paired = T)$statistic

etab <- tb^2/sum(tb^2, 178); print(etab)

##          t
## 0.05392524

etaw <- tw^2/sum(tw^2, 89); print(etaw)

##         t
## 0.2426018
```

The returned values correspond well to both the $\eta_p{}^2$ and the $\eta_G{}^2$ estimated from the last ANOVA, thus, the returned $\eta_p{}^2$ is inflated *in part* due to the correlation (since t_{within} = d_D*$\sqrt{n}$ = λ, see below) *as in SPSS*.

Effect sizes and correlation among repeated measures

Above, I described that the returned $\eta_p{}^2$ is inflated due to correlation among repeated measures (regardless of whether it was computed from a variance table or from the *t* statistics; as a precision: $F = t^2$). In fact, consider a within-subjects variable, as highlighted earlier $\eta_p{}^2 = \frac{t^2_{within}}{t^2_{within} + dferror}$, thus, given that t_{within} = d_D*$\sqrt{n}$, which is the noncentrality parameter λ (Fitts, 2021, 2022), where $d_D = \frac{D}{S_D}$, where $D = m_x - m_y$, and *sd* = the standard deviation of the difference scores, such that $\frac{D}{S_D} * \sqrt{n} \sim t_{n-1}\ (\frac{\Delta}{\sigma_D} * \sqrt{n} = \lambda)$, where $\frac{\Delta}{\sigma_D} = \delta_D$, where $\delta_D = \frac{\mu_x - \mu_y}{\sqrt{\sigma_x^2 + \sigma_y^2 - 2 * \rho * \sigma_x * \sigma_y}}$ (Cousineau & Goulet-Pelletier, 2021), hence, the standardizer for the sample is equivalent, so that $S_D = \sqrt{S_1^2 + S_2^2 - 2 * r * S_1 * S_2}$, where *r* is the Pearson correlation coefficient, we better understand the relationship between the correlation coefficient and the Cohen's d_D (see Cousineau & Goulet-Pelletier, 2021) and so with the returned $\eta_p{}^2$. In that, "*As the correlation between measures increases, the standard deviation of the difference scores decreases*" (Lakens, 2013; see also Dunlap et al., 1996), thus a strong correlation will inflate the standardized mean difference, while a weak correlation will deflate it, thus, a correlation = 0.5 will return a Cohen's d_D that equals Cohen's d_p. However, this is not because that the correlation among repeated measures equals 0.5, making Cohen's d_D indistinguishable from Cohen's d_p that this will make $\eta_p{}^2$ indistinguishable from $\eta_G{}^2$ or a $\eta_p{}^2$ estimated from a between-subjects variable, at minimum it seems true for the latter effect size when "*For designs where all factors are manipulated between participants, $\eta_p{}^2$ and $\eta_G{}^2$are identical* " (In other scenarios I don't know). Let's perform two analyses, in which, we will set different parameters:

Δ = 15, σ = 15, ρ = 0.5 and ρ = 0

```
## Compute the different Cohen's d

# Cohen's d_p
dp <- Delta/sdp; dp

## [1] 1

# Cohen's d_av = Cohen's d_p

dav <- Delta/sdav; dav

## [1] 1

# Cohen's d_D with Rho = 0.5

dDHalf <- Delta/sdD[2]; dDHalf

## [1] 1

# Cohen's d_D with Rho = 0

dDZero <- Delta/sdD[1]; dDZero

## [1] 0.7071068

library(psych)
ms <- c(n, n)
hm <- harmonic.mean(ms)

## compute the t-statistics
# between-subjects design
lambda_dp <- dp* sqrt(hm/2)

# within-subjects design, when Rho = 0.5
lambda_dDH <- dDHalf*sqrt(n)

# within-subjects design, when Rho = 0
lambda_dDZ <- dDZero*sqrt(n)

## compute the degrees of freedom
dfp <- sum(n,n)-2
dfD <- n-1
```

```r
## return the partial etas squared

eta <- function(t, df) {
  partial <- t^2/sum(t^2, df)
  return(partial)
}

# eta for a between subjects-variable that equals the generalized
one
etaB <- eta(lambda_dp, dfp); etaB

## [1] 0.203252

# eta for a within subjects-variable when Rho = 0.5
etawH <- eta(lambda_dDH, dfD); etawH

## [1] 0.5050505

# eta for a within subjects-variable when Rho = 0
etawZ <- eta(lambda_dDZ, dfD); etawZ

## [1] 0.3378378
```

We better see that the ${\eta_p}^2$ is always inflated compared to the ${\eta_G}^2$, even when ρ = 0 and ρ = 0.5. The following function use the formula provided by Lakens (2013) to convert ${\eta_p}^2 SPSS$ to ${\eta_p}^2 GPower$ for a power analysis.

```r
Spss_to_GP <-function(eta, N, k, m, Rho) {
  fsq <- eta/(1-eta)
  Gpower <- fsq*((N-k)/N)*((m-1)/m)*(1-Rho)
  etaGP <- GPower/sum(1, 1, GPower)
  return(test)
}
```

At this moment using this formula for the two latter ${\eta_p}^2$ (i.e., from a within-subject variable when ρ = 0.5 and ρ = 0) returns well the former (i.e., the effect size for a within-subjects variable that is estimated in the same way as for a between-subjects variable). Thus, ${\eta_p}^2 GPower \approx {\eta_G}^2$. The main issue is that the ${\eta_p}^2 GPower$ is not only free from the correlation among repeated measures (remember how to compute the error sum of squares for within-subjects designs). As a consequence, it seems to respond to another issue regarding the ${\eta_p}^2$ estimated with SPSS: why use specific conversion from a ${\eta_p}^2$ estimated in

SPSS to $\eta_p{}^2$ for GPower (see Lakens, 2013) given the fact that one can use the former and fill in 0 for the value in "correlation among repeated measures"? It seems a reasonable idea, but it appears to be a trap. In fact, I have just demonstrated that a $\eta_p{}^2$ for a within-subjects-variable even with no correlation is inflated compared to a $\eta_p{}^2$that is estimated to be comparable to a $\eta_p{}^2$ estimated for a between-subjects variable (≈ $\eta_G{}^2$), and GPower requires the latter. So, even with a null correlation, using a $\eta_p{}^2$ for a within-subjects variable should not be used when selecting options: as in GPower and fill in ρ = 0 is not well-suited (see below).

Resolve the main issue

Now, let 's take the illustration results in Lakens (2013), in which he compared two sampled data, consider the following parameters: $r_{xy} = 0.726$; $\eta_p{}^2 = 0.71$ (within-subjects variable); $\eta_G{}^2 = \eta_p{}^2 = 0.26$ (between-subjects variable). Following Lakens (2013) and the lines cited above, if in Gpower we choose "*Options: as in GPower 3.0*", we have to use $\eta_p{}^2 =$ 0.26 (which is identical to the $\eta_G{}^2$), and $r = 0.726$ for "correlation among repeated measures", theoretically, it should return similar results to what we would find if we choose "Options: as in SPSS", and that we use $\eta_p{}^2 = 0.71$. To test this prediction, I choose two scenarios for ANOVA: repeated measures, within factors:

1) 1 group, and 3 repeated measurements

2) 2 group, and 2 repeated measurements

Options: as in GPower 3.0.

F tests – ANOVA: Repeated measures, within factors
Analysis: A priori: Compute required sample size
Input: Effect size f = 0.5927490

	α err prob	=	0.05
	Power (1−β err prob)	=	0.95
	Number of groups	=	1
	Number of measurements	=	3
	Corr among rep measures	=	0.726
	Nonsphericity correction ϵ	=	1
Output:	Noncentrality parameter λ	=	23.0814773
	Critical F	=	4.1028210
	Numerator df	=	2.0000000
	Denominator df	=	10.0000000
	Total sample size	=	6
	Actual power	=	0.9633994

Options: as in SPSS.

F tests – ANOVA: Repeated measures, within factors

Analysis:	A priori: Compute required sample size		
Input:	Effect size f(U)	=	1.5646967
	α err prob	=	0.05
	Power (1−β err prob)	=	0.95
	Number of groups	=	1
	Number of measurements	=	3
	Nonsphericity correction ϵ	=	1
Output:	Noncentrality parameter λ	=	24.4827576
	Critical F	=	4.1028210
	Numerator df	=	2.0000000
	Denominator df	=	10.0000000
	Total sample size	=	6
	Actual power	=	0.9718118

Options: as in GPower 3.0.

F tests – ANOVA: Repeated measures, within factors

Analysis:	A priori: Compute required sample size		
Input:	Effect size f	=	0.5927490
	α err prob	=	0.05
	Power (1−β err prob)	=	0.95
	Number of groups	=	2
	Number of measurements	=	2
	Corr among rep measures	=	0.726
	Nonsphericity correction ϵ	=	1
Output:	Noncentrality parameter λ	=	20.5168687
	Critical F	=	5.9873776
	Numerator df	=	1.0000000
	Denominator df	=	6.0000000
	Total sample size	=	8
	Actual power	=	0.9620289

Options: as in SPSS.

F tests – ANOVA: Repeated measures, within factors			
Analysis:	A priori: Compute required sample size		
Input:	Effect size f(U)	=	1.5646967
	α err prob	=	0.05
	Power (1−β err prob)	=	0.95
	Number of groups	=	2
	Number of measurements	=	2
	Nonsphericity correction ϵ	=	1
Output:	Noncentrality parameter λ	=	19.5862061
	Critical F	=	5.3176551
	Numerator df	=	1.0000000
	Denominator df	=	8.0000000
	Total sample size	=	10
	Actual power	=	0.9707300

Results seem comparable, the little gap between results could be due to a history of approximate values. In the following tests, we will take back the results from our simulated data and perform two power analyses, in which we will use the effect sizes estimated for the interaction between *factor* and *group*. I also imagine that for this interaction, the returned ${\eta_G}^2$ is equivalent to the ${\eta_p}^2$ that one would find for an interaction between two between-subjects variables (I think). As a consequence, if we fill in the ${\eta_G}^2$ and the correlation among repeated measures, the total sample size needed may be equal the total sample size estimated with options: as in SPSS when we use the ${\eta_p}^2$.

Options: as in GPower 3.0.

F tests – ANOVA: Repeated measures, within-between interaction			
Analysis:	A priori: Compute required sample size		
Input:	Effect size f	=	0.4101516
	α err prob	=	0.05
	Power (1−β err prob)	=	0.95
	Number of groups	=	3
	Number of measurements	=	2
	Corr among rep measures	=	0.6746506
	Nonsphericity correction ϵ	=	1
Output:	Noncentrality parameter λ	=	21.7164134
	Critical F	=	3.5545571
	Numerator df	=	2.0000000
	Denominator df	=	18.0000000
	Total sample size	=	21
	Actual power	=	0.9759238

Options: as in SPSS.

F tests – ANOVA: Repeated measures, within-between interaction

Analysis:	A priori: Compute required sample size		
Input:	Effect size f(U)	=	1.1727305
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Number of groups	=	3
	Number of measurements	=	2
	Nonsphericity correction ϵ	=	1
Output:	Noncentrality parameter λ	=	20.6294524
	Critical F	=	3.6823203
	Numerator df	=	2.0000000
	Denominator df	=	15.0000000
	Total sample size	=	18
	Actual power	=	0.9638169

As we see results are fairly comparable. Apparently, one can choose the option as in GPower because it returns a greater a priori sample size which is desirable to better increase the power. The little gap between both results could be due to the fact that ${\eta_G}^2$ is smaller than the ${\eta_p}^2$ (though theoretically it should not be the case, cf. the citation from Lakens, but here I'm not sure because it is applied for an interaction), Hence, one can use the formula provided in Lakens (2013) to convert the ${\eta_p}^2 SPSS$ into ${\eta_p}^2 GPower$, use the latter in GPower, and theoretically it should return similar results.

F tests – ANOVA: Repeated measures, within-between interaction

Analysis:	A priori: Compute required sample size		
Input:	Effect size f	=	0.4650467
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Number of groups	=	3
	Number of measurements	=	2
	Corr among rep measures	=	0.6746506
	Nonsphericity correction ϵ	=	1
Output:	Noncentrality parameter λ	=	23.9301612
	Critical F	=	3.6823203
	Numerator df	=	2.0000000
	Denominator df	=	15.0000000
	Total sample size	=	18
	Actual power	=	0.9821984

Bingo, it returns nearly similar outputs. In my opinion the three options are fairly valid.

The current experiment

For my experiment, we want to use the effect size estimated for the interaction between “*Group*” and “*Contingency*”, that is, an interaction between a between-subject variable and a within-subject variable. However, *I don’t know whether the correlation between high and low contingency trials are taken into account the effect size of interest*. It appears that this is the case since using the same descriptive statistics in a simulation and play with different correlation coefficient returns different ${\eta_p}^2$ for an interaction in a same design, see the following link:

Chapter 5 Mixed ANOVA | Power Analysis with Superpower (aaroncaldwell.us)

So, from the paper of interest we have the following estimates:

$r = 0.8513636$; ${\eta_p}^2 SPSS = 0.181$; ${\eta_p}^2 GPower = 0.01574769$; ${\eta_G}^2 = 0.01386743$

and we can perform the power analysis using three different means:

Options: as in GPower 3.0, with ${\eta_p}^2 GPower$

F tests – ANOVA: Repeated measures, within-between interaction

Analysis:	A priori: Compute required sample size		
Input:	Effect size f	=	0.1264897
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Number of groups	=	3
	Number of measurements	=	2
	Corr among rep measures	=	0.8513636
	Nonsphericity correction ϵ	=	1
Output:	Noncentrality parameter λ	=	16.1464260
	Critical F	=	3.1239074
	Numerator df	=	2.0000000
	Denominator df	=	72.0000000
	Total sample size	=	75
	Actual power	=	0.9504779

Options: as in GPower 3.0, with ${\eta_G}^2$

F tests – ANOVA: Repeated measures, within–between interaction			
Analysis:	A priori: Compute required sample size		
Input:	Effect size f	=	0.1185852
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Number of groups	=	3
	Number of measurements	=	2
	Corr among rep measures	=	0.8513636
	Nonsphericity correction ϵ	=	1
Output:	Noncentrality parameter λ	=	16.4620930
	Critical F	=	3.1051566
	Numerator df	=	2.0000000
	Denominator df	=	84.0000000
	Total sample size	=	87
	Actual power	=	0.9553120

Options: as in SPSS.

F tests – ANOVA: Repeated measures, within–between interaction			
Analysis:	A priori: Compute required sample size		
Input:	Effect size f(U)	=	0.4701077
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Number of groups	=	3
	Number of measurements	=	2
	Nonsphericity correction ϵ	=	1
Output:	Noncentrality parameter λ	=	16.5750937
	Critical F	=	3.1186421
	Numerator df	=	2.0000000
	Denominator df	=	75.0000000
	Total sample size	=	78
	Actual power	=	0.9557629

Thus, according to the power analysis at a total of 80 participants seems desirable for my following experiment.

A last power analysis can also be conducted to better illustrate that one has not to use: Set the following parameters: ${\eta_p}^2 = 0.181$ and ρ= 0.

F tests – ANOVA: Repeated measures, within–between interaction			
Analysis:	A priori: Compute required sample size		
Input:	Effect size f	=	0.4701077
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Number of groups	=	3
	Number of measurements	=	2
	Corr among rep measures	=	0
	Nonsphericity correction ϵ	=	1
Output:	Noncentrality parameter λ	=	17.2380975

Critical F	=	3.2594463
Numerator df	=	2.0000000
Denominator df	=	36.0000000
Total sample size	=	39
Actual power	=	0.9547314

References

Cousineau, D., & Goulet-Pelletier, J. C. (2021). A study of confidence intervals for Cohen's dp in within-subject designs with new proposals. *The Quantitative Methods for Psychology*, *17*(1), 51-75.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological methods*, *1*(2), 170.

Fitts, D. A. (2020). Commentary on "A review of effect sizes and their confidence intervals, Part I: The Cohen's d family": The degrees of freedom for paired samples designs. *The Quantitative Methods for Psychology*, *16*(4), 250-261.

Fitts, D. A. (2021). Expected and empirical coverages of different methods for generating noncentral t confidence intervals for a standardized mean difference. *Behavior Research Methods*, 1-18.

Goulet-Pelletier, J. C., & Cousineau, D. (2018). A review of effect sizes and their confidence intervals, Part I: The Cohen'sd family. *The Quantitative Methods for Psychology*, *14*(4), 242-265.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*, *4*, 863.