Powering ANOVA Tests

Analysis of variance, or ANOVA, tests are designed to compare the means of multiple groups. Typically these groups include a control group and a number of treatment groups. Non-treatment factors such as sex may require multi-dimensional grouping of experimental subjects. In this document, we discuss the basic formulas for determining the power (and hence sample size requirements) of ANOVA tests.

The F Distribution

ANOVA hypothesis tests rely on test statistics that have the F distribution. Under the null hypothesis, the statistics have the central F distribution. This distribution has two parameters, called numerator degrees of freedom and denomination degrees of freedom. When the null hypothesis is false, the statistics typically have the noncentral F distribution, which has an additional parameter called the noncentrality parameter. We denote this distribution by $F(x, ndf, ddf, \lambda) = \Pr[F \le x]$. If λ=0, we omit the dependence on λ and refer to this notation as the central F distribution. Often it is required to invert this distribution in the first variable to find critical values, and we use the notation

$x = F^{-1}(p, ndf, ddf, \lambda)$ to solve the equation $p = \Pr[F \le x]$ for *x* when *p* has been specified.

Computing Power

Computing power requires two steps. First we establish the critical region for the null hypothesis. Second we determine the probability that the test statistic lies outside the critical region when the null hypothesis is false. The latter computation requires an a priori estimate of "how false" the null hypothesis is. The so-called "effect size" parameter is a measure of this discrepancy. The list of information we need to determine power is as follows:

1. Significance, α, the probability of rejecting a true null.
2. Number of groups in the experimental design. This may be an array if the treatment regimen is multi-dimensional or if different non-treatment factors must be controlled for.
3. Minimum desired effect size, *f*.
4. Standard deviation of the measurements, σ.
5. Number of samples or replicates per group, *n*.

If the sample size is not known, one may alternatively specify the power 1-β and determine the number of samples per group required to achieve this desired power. The power calculation typically looks like
$1 - F(F^{-1}(ndf, ddf), ndf, ddf, \lambda) = \Pr[\text{non central } F > \text{central } F \text{ specified critical value}]$.

One Way Anova: Power/Effect/Sample considerations

The one-way ANOVA model is given by

$Y_{ik} = \mu_i + \varepsilon_{ik} = \mu + \eta_i + \varepsilon_{ik}, \quad 1 \le i \le a, 1 \le k \le n$

in which:

- μ is the grand mean;
- η_i are the main effects (and it is assumed $\sum_{i=1}^{a} \eta_i = 0$);
- ε_{ik} are iid $N(0, \sigma^2)$ random variables.

This is a balanced design: all groups have the same number of samples, *n*, and the total sample size is *na*. To test hypotheses, we select a significance level α, typically 0.05, but it may be less depending on the situation. The null hypothesis of interest is

- η_i =0 for all *i*.

To test this hypothesis, we compute the F distribution's critical value $F_c = F^{-1}(1-\alpha, a-1, (n-1)a)$. Under the null hypothesis, the F statistic has the central *F* distribution, with *a*-1numerator degrees of freedom and *na*-*a* denominator degrees of freedom. When the null hypothesis is false, the F statistic has a non-central F distribution. The noncentrality parameter λ is given by

$$\lambda = na \frac{f^2}{\sigma^2} \quad ,$$

in which $f = \sqrt{\frac{\sum_{i=1}^{a} (\mu_i - \overline{\mu})^2}{(a-1)}}$ is the "effect size," modeling with 1 simple numerical quantity the distance from the "true" model to the null model of identical means. The null model requires all the μ_i values to be the same, so *f*=0 under the null hypothesis. The power of the ANOVA test is then computed as $1 - \beta = F(F_c, a-1, (n-1)a, naf^2 / \sigma^2)$.

One issue with using R's "pwr" package is that the effect size used therein has a slightly different definition:

$$f' = \sqrt{\frac{\sum_{i=1}^{a} (\mu_i - \overline{\mu})^2}{\sigma^2 a}}$$

is the standardized effect size used in pwr (a usage derived from Cohen, 1988). To compensate for this discrepancy, in using pwr, we must convert

$$f' = \sqrt{\frac{\sum_{i=1}^{a}(\mu_i - \overline{\mu})^2}{\sigma^2 a}} = \frac{f}{\sigma}\sqrt{\frac{a-1}{a}}$$

Two Way Anova: Power/Effect/Sample considerations

The ANOVA model is given by

$Y_{ijk} = \mu + \eta_i + \xi_j + \gamma_{ij} + \varepsilon_{ijk}, \quad 1 \le i \le a, 1 \le j \le b, 1 \le k \le n$

in which:

- μ is the grand mean;
- η_i are the main effects (and it is assumed $\sum_{i=1}^{a}\eta_i = 0$);
- ξ_i are the secondary effects (and it is assumed $\sum_{i=1}^{b}\xi_i = 0$);
- γ_{ij} are the interaction effects (and it is assumed $\sum_{k=1}^{a}\gamma_{kj} = \sum_{l=1}^{b}\gamma_{il} = 0$ for each *i*,.*j*);
- ε_{ijk} are iid $N(0,\sigma^2)$ random variables.

This is a balanced design: all groups have the same number of samples, *n*, and the total sample size is *nab*. To test hypotheses, we select a significance level α, typically 0.05, but it may be less depending on the situation. The three hypotheses of interest are

- η_i =0 for all *i*.
- ξ_j =0 for all *j*.
- γ_{ij} =0 for all *i,j*.

Testing each hypothesis, we will need the critical F value for the test

1. $F_c = F^{-1}(1-\alpha, a-1, (n-1)ab)$
2. $F_c = F^{-1}(1-\alpha, b-1, (n-1)ab)$
3. $F_c = F^{-1}(1-\alpha, (a-1)(b-1), (n-1)ab)$

For each hypothesis, we have an effect size *f*. In each case, *f* denotes the discrepancy from null model:

$$f_M^2 = \frac{\sum_{i=1}^{a}(\eta_i - \overline{\eta})^2}{a-1} \quad or \quad f_S^2 = \frac{\sum_{i=1}^{b}(\xi_i - \overline{\xi})^2}{b-1} \quad or \quad f_I^2 = \frac{\sum_{j=1}^{b}\sum_{i=1}^{a}(\gamma_{ij} - \overline{\gamma})^2}{(a-1)(b-1)}.$$

For each hypothesis, we compute the power using the formulae

1. $1-\beta = F(F_c, a-1, (n-1)ab, n(a-1)bf_M^2/\sigma^2)$
2. $1-\beta = F(F_c, b-1, (n-1)ab, na(b-1)f_S^2/\sigma^2)$
3. $1-\beta = F(F_c, (a-1)(b-1), (n-1)ab, n(a-1)(b-1)f_I^2/\sigma^2)$

These relationships may also be used to determine the sample size per group, given the power 1-β and the effect size *f*. Alternatively, one may determine the minimal detectable effect size *f* for a given power and sample size.

To treat the case of no interactions, the model becomes

$$Y_{ijk} = \mu + \eta_i + \xi_j + \varepsilon_{ijk}, \quad 1 \le i \le a, 1 \le j \le b, 1 \le k \le n$$

in which

- μ is the grand mean;
- η_i are the main effects (and it is assumed $\sum_{i=1}^{a}\eta_i = 0$);
- ξ_i are the secondary effects (and it is assumed $\sum_{i=1}^{b}\xi_i = 0$);
- ε_{ijk} are iid $N(0,\sigma^2)$ random variables.

Here we have only *a*+*b*-1 parameters instead of *ab* parameters. Only the first two hypotheses are of interest.

1. $F_c = F^{-1}(1-\alpha, a-1, nab-(a+b-1))$
2. $F_c = F^{-1}(1-\alpha, b-1, nab-(a+b-1))$

The power/effect/sample computation is determined, for each test, by

1. $1-\beta = F(F_c, a-1, nab-(a+b-1), n(a-1)bf_M^2/\sigma^2)$
2. $1-\beta = F(F_c, b-1, nab-(a+b-1), na(b-1)f_S^2/\sigma^2)$

Generally speak, it is important to include interaction effects, and I would not recommend leaving them out *a priori*.

Power for Three Way ANOVA

The 3-way ANOVA model with complete interaction modeling is given by

$$Y_{ijkm} = \mu + \eta_i + \xi_j + \zeta_k + \gamma_{ij} + \kappa_{jk} + \nu_{ik} + \tau_{ijk} + \varepsilon_{ijkm}, \quad 1 \le i \le a, 1 \le j \le b, 1 \le k \le c, 1 \le m \le n$$

in which:

- μ is the grand mean;
- η_i are the main effects (and it is assumed $\sum_{i=1}^{a} \eta_i = 0$);
- ξ_i are the secondary effects (and it is assumed $\sum_{i=1}^{b} \xi_i = 0$);
- ζ_i are the tertiary effects (and it is assumed $\sum_{i=1}^{c} \zeta_i = 0$);
- γ_{ij} are the main-secondary interaction effects (and it is assumed $\sum_{k=1}^{a} \gamma_{kj} = \sum_{l=1}^{b} \gamma_{il} = 0$ for each $i,.j$);
- κ_{ij} are the secondary-tertiary interaction effects (and it is assumed $\sum_{l=1}^{b} \kappa_{lj} = \sum_{l=1}^{c} \kappa_{il} = 0$ for each $i,.j$);
- ν_{ij} are the main-tertiary interaction effects (and it is assumed $\sum_{l=1}^{a} \nu_{lj} = \sum_{l=1}^{c} \nu_{il} = 0$ for each $i,.j$);
- τ_{ijk} are the three-way interaction effects with even more complicated constraints.
- ε_{ijkm} are iid $N(0,\sigma^2)$ random variables.

This is a balanced design: all groups have the same number of samples, *n*, and the total sample size is *nab*. To test hypotheses, we select a significance level α, typically 0.05, but it may be less depending on the situation. There are seven hypotheses of interest:

- η_i =0 for all *i*.
- ξ_i =0 for all *j*.
- ζ_i =0 for all *i*.
- γ_{ij} =0 for all *i,j*.
- ν_{ij} =0 for all *l,j*.
- κ_{ij} =0 for all *i,j*.
- τ_{ijk} =0 for all *i,j,k*.

Testing each hypothesis, we will need the critical F value for the test

1. $F_c = F^{-1}(1-\alpha, a-1, (n-1)abc)$
2. $F_c = F^{-1}(1-\alpha, b-1, (n-1)abc)$
3. $F_c = F^{-1}(1-\alpha, c-1, (n-1)abc)$
4. $F_c = F^{-1}(1-\alpha, (a-1)(b-1), (n-1)abc)$
5. $F_c = F^{-1}(1-\alpha, (a-1)(c-1), (n-1)abc)$
6. $F_c = F^{-1}(1-\alpha, (b-1)(c-1), (n-1)abc)$
7. $F_c = F^{-1}(1-\alpha, (a-1)(b-1)(c-1), (n-1)abc)$

For each hypothesis, we have an effect size *f*. In each case, *f* denotes the distance from the null model:

$$f_M^2 = \frac{\sum_{i=1}^{a}(\eta_i - \overline{\eta})^2}{a-1} \quad or \quad f_S^2 = \frac{\sum_{i=1}^{b}(\xi_i - \overline{\xi})^2}{b-1} \quad or \quad f_T^2 = \frac{\sum_{i=1}^{c}(\varsigma_i - \overline{\varsigma})^2}{c-1}.$$

$$f_{MS}^2 = \frac{\sum_{j=1}^{b}\sum_{i=1}^{a}(\gamma_{ij} - \overline{\gamma})^2}{(a-1)(b-1)} \quad or \quad f_{MT}^2 = \frac{\sum_{j=1}^{c}\sum_{i=1}^{a}(\nu_{ij} - \overline{\nu})^2}{(a-1)(c-1)} \quad or \quad f_{ST}^2 = \frac{\sum_{j=1}^{c}\sum_{i=1}^{b}(\kappa_{ij} - \hat{\kappa})^2}{(b-1)(c-1)}$$

$$f_{MST}^2 = \frac{\sum_{k=1}^{c}\sum_{j=1}^{b}\sum_{i=1}^{a}(\tau_{ij} - \overline{\tau})^2}{(a-1)(b-1)(c-1)}$$

For each hypothesis, we compute the power using the formulae

1. $1-\beta = F(F_c, a-1, (n-1)abc, n(a-1)bcf_M^2/\sigma^2)$
2. $1-\beta = F(F_c, b-1, (n-1)abc, na(b-1)cf_S^2/\sigma^2)$
3. $1-\beta = F(F_c, c-1, (n-1)abc, nab(c-1)f_T^2/\sigma^2)$
4. $1-\beta = F(F_c, (a-1)(b-1), (n-1)abc, n(a-1)(b-1)cf_{MS}^2/\sigma^2)$
5. $1-\beta = F(F_c, (b-1)(c-1), (n-1)abc, na(b-1)(c-1)f_{ST}^2/\sigma^2)$
6. $1-\beta = F(F_c, (a-1)(c-1), (n-1)abc, n(a-1)b(c-1)f_{MT}^2/\sigma^2)$
7. $1-\beta = F(F_c, (a-1)(b-1)(c-1), (n-1)abc, n(a-1)(b-1)(c-1)f_{MST}^2/\sigma^2)$

   These relationships may also be used to determine the sample size per group, given the power 1-β and the effect size *f*. Alternatively, one may determine the minimal detectable effect size *f* for a given power and sample size.