Chapter 12

Power Analysis for One-Way ANOVA

Recall that the power of a statistical test is the probability of rejecting H_0 when H_0 is false, and some alternative hypothesis H_1 is true. We saw earlier (Chapter 10) that power for one-sample Z and t tests is a function of the quantity

$$\phi = \frac{(\mu_1 - \mu_0)}{\sigma/\sqrt{n}}, \tag{12.1}$$

where μ_1 and μ_0 are the means under H_1 and H_0, σ is the standard deviation of the observations, and n is the sample size. Anything that increases ϕ increases the power of the test, including greater differences between μ_1 and μ_0, decreasing σ, or increasing the sample size n. Larger values of the Type I error rate α also increase the power of the test, because they make it more likely the test will reject H_0 under any circumstances. Although one-way ANOVA is a more complicated design, we will see that exactly the same factors influence the power of its associated F test.

A power analysis for a one-way ANOVA design is usually conducted before running the experiment or study. This is known as a **prospective power analysis**. We then use the information from this analysis to refine our experimental design, most often the sample sizes needed for each treatment group to yield adequate power. Conversely, a **retrospective power analysis** is one conducted after an experiment or study, using the results from the study in the power calculation. This is a controversial procedure that some statisticians find questionable (Steidl et al. 1997).

Cohen (1988) recommends using a default power value of 0.8 when designing an experiment, if there is no other basis for setting the power. One reason is that achieving higher power values usually requires disproportionately larger sample sizes. He also recommends a power value of 0.8 on the basis of the ratio of Type II (β) to Type I error (α). He suggests that an optimal ratio of β/α is about four, implying that Type I errors are four times more serious than Type II errors. If you use $\alpha = 0.05$ as the Type I error rate, and choose power $= 0.8$, then $\beta = 1 - \text{power} = 0.2$, and so $\beta/\alpha = 4$.

12.1   Power analysis for one-way ANOVA

Suppose we want to design an experiment involving several treatments that has adequate power. Assuming we know the treatments we will apply, the first step in a power analysis is to specify the actual values of the treatment means under H_1, the alternative hypothesis. If the experiment has five treatments, we might speculate that the treatment means take the following values under H_1:

$$H_1 : \mu_1 = 20, \mu_2 = 22, \mu_3 = 22, \mu_4 = 25, \mu_5 = 18. \tag{12.2}$$

For example, these values could be the final weights of fish reared on five different diets. This is the form of H_1 needed by `proc power` (SAS Institute Inc. 2018). We can also express H_1 in terms of the usual model for this design, the fixed effects model of the form

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}. \tag{12.3}$$

In terms of the parameters of this model, H_1 is equivalent to saying

$$H_1 : \alpha_1 = -1.4, \alpha_2 = 0.6, \alpha_3 = 0.6, \alpha_4 = 3.6, \alpha_5 = -3.4, \tag{12.4}$$

where $\alpha_i = \mu_i - \mu$, and μ is the grand mean ($\mu = \sum \mu_i/5 = 107/5 = 21.4$) (Winer et al. 1991; Montgomery 1997).

The null hypothesis in terms of group means would have the form

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu, \tag{12.5}$$

where μ is the grand mean. This is equivalent to the usual null hypothesis for one-way ANOVA, which is $H_0 : \alpha_i = 0$ for all i.

We also need to specify a standard deviation σ for the power analysis. We could potentially estimate σ from similar studies in the literature or through a pilot study. If the paper provides an ANOVA table, we can estimate σ using $\sqrt{MS_{within}} = \sqrt{MS_{error}}$. SAS actually calculates this quantity and labels it `Root MSE` – see previous printouts for `proc glm`. In other situations, you may not know σ precisely but can specify a plausible range of values. Continuing our example, we suppose that previous experiments suggest $\sigma = 3$.

To calculate the power for this example, we also need to specify a sample size n for the treatments. Usually we are interested in determining the power for a range of n values, so we can determine the minimal sample size to needed to reject H_0 with adequate power. Most power analyses assume an equal sample size for each treatment, because this usually yields a higher power than unbalanced designs. We also need to specify the Type I error rate for the overall ANOVA, and $\alpha = 0.05$ is customary.

The power is then calculated using the distribution of the statistic F_s under H_1, called the non-central F distribution (the distribution under H_0 is the F distribution). The non-central F distribution has three parameters, the usual two degrees of freedom plus an additional parameter λ, defined by the formula

$$\lambda = \frac{n \sum_{i=1}^{a} \alpha_i^2}{\sigma^2}, \tag{12.6}$$

where $\alpha_i = \mu_i - \mu$, and $\mu = \sum \mu_i / a$ (Winer et al. 1991, Montgomery 1997). Note that λ is a function of the α_i values, σ, and the sample size n. The non-central F distribution is equal to the F distribution when $\lambda = 0$, which can only happen if there are no treatment effects and $\alpha_i = 0$ for all i. As the value of λ increases, however, the noncentral F distribution will shift to the right, away from the position held by the F distribution. Note the similarity of this quantity with ϕ, which determines the power for one-sample Z and t tests.

Figure 12.1 shows the F and noncentral F distributions for the power analysis example described above, with $a = 5$, the α_i values as specified, and $\sigma = 3$. We also assume for the moment that $n = 5$, and set $\alpha = 0.05$. For this design, we have $df_1 = a - 1 = 5 - 1 = 4$, $df_2 = a(n-1) = 5(5-1) = 20$. For $\alpha = 0.05$, we would reject H_0 if F_s, the test statistic for one-way ANOVA (Chapter 11), exceeded 2.866 (see Table F). We also need to calculate a value of λ for the noncentral F distribution. We have

$$\lambda = \frac{5\left[(-1.4)^2 + 0.6^2 + 0.6^2 + 3.6^2 + (-3.4)^2\right]}{3^2} = 15.111. \tag{12.7}$$

We see that the noncentral F lies to the right of the F distribution, because λ is fairly large in this example. What is the power of the test? It is the area of the noncentral F distribution lying to the right of 2.866, because this is the probability that F_s will exceed 2.866 under H_1, i.e., the probability of rejecting H_0 if it is false and H_1 is true.

What would happen to the power for other values of n or σ, or for that matter smaller or larger differences among groups under H_1 (implying smaller or larger α_i values)? Any change that increases the value of λ will increase the power of the test, because it reduces the amount of overlap between the two distributions. Examining λ, we see that larger n, larger differences among groups, and smaller σ values would all increase λ and so increase the power of the test. Larger α (Type I error rate) values also increase the power of the test, because they reduce the acceptance and increase the rejection region size. Sample size n also has an effect on power through the acceptance region – larger n reduces its upper boundary through its effect on $df_2 = a(n-1)$. Fig. 12.2 shows the F and noncentral F distributions for the power example, now using $n = 8$. Note how the overlap between the two distributions is reduced for larger n, increasing the power of the test. See Table 12.1 for a summary of how these factors affect power and β.

Table 12.1: Effects on power and the Type II error rate β of changes in various parameters. The arrows indicate if a particular quantity increases or decreases.

Parameter	Direction	λ	power	β
α_i values	↑	↑	↑	↓
n	↑	↑	↑	↓
σ	↑	↓	↓	↑
α	↑	no change	↑	↓

The effect of n on λ implies that a sufficient large sample size can generate adequate power, even when the α_i values are small or σ is large. Thus, large sample sizes should make it possible to detect small treatment effects, and can also compensate for noisy data.
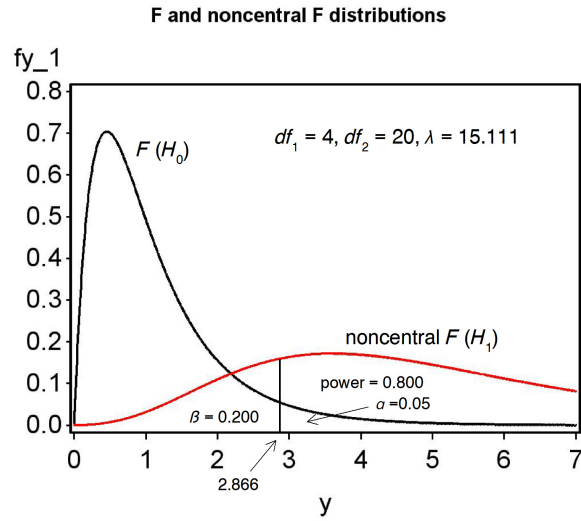
F and noncentral F distributions

Figure 12.1: The F and noncentral F distributions for the power example, using $n = 5$.

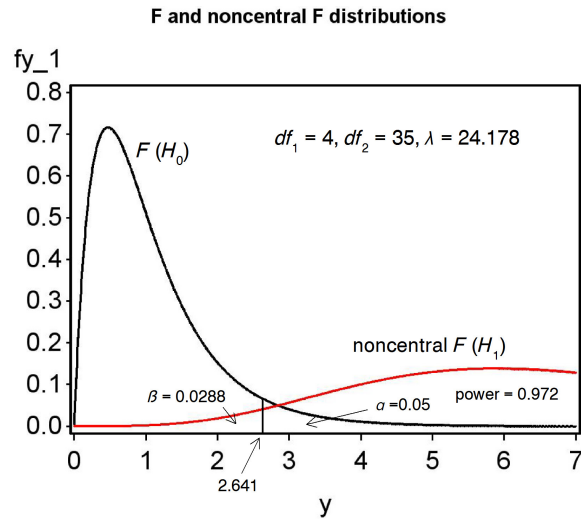**F and noncentral F distributions**

Figure 12.2: The F and noncentral F distributions for the power example, for $n = 8$.

12.2   Power analysis - SAS Demo

SAS makes power analysis relatively easy and provides specific methods for one-way ANOVA and many other designs. Consider our previous example involving five different treatments. We are interested in determining the power of a one-way ANOVA, when the following alternative hypothesis is true:

$$H_1 : \mu_1 = 20, \mu_2 = 22, \mu_3 = 22, \mu_4 = 25, \mu_5 = 18. \qquad (12.8)$$

We need another piece of information for the power analysis, the value of σ. From preliminary studies or a previously published paper, we estimate that $\sigma = 3$. We also specify the Type I error rate, setting $\alpha = 0.05$.

This is everything required to carry out a power analysis using `proc power` (SAS Institute Inc. 2018). We first specify that we want a power analysis for one-way ANOVA using the option `onewayanova`. The means for each treatment group are specified using the `groupmeans` option, with the means listed in parentheses. See program listing below.

The values of σ and α are similarly specified using the `stddev` and `alpha` options. We are interested in determining the power for a range of n values, the sample size per group. This is specified using the `npergroup` option. You can either give a list of n values or use the syntax `x to y by z` to specify a sequence of values. The `power` option is specified as a missing value (a period), because we want SAS to solve for power as a function of sample size per group. The `plot` command generates a plot of power vs. sample size.

We see that power increases rapidly with sample size per group (n), from both the SAS output and graph (Fig. 12.3, 12.4). A power value of 0.8 is achieved for $n = 5$ in this example.

--------------------------------- SAS Program ---------------------------------

```
* oneway_power.sas;
title 'Power Analysis for One-Way Anova';
proc power;
    onewayanova
        groupmeans = (20 22 22 25 18)
        stddev = 3
        alpha = 0.05
        npergroup = 2 to 20 by 1
        power = . ;
    plot x=n;
run;
quit;
```

Power Analysis for One-Way ANOVA

The POWER Procedure
Overall F Test for One-Way ANOVA

Fixed Scenario Elements	
Method	Exact
Alpha	0.05
Group Means	20 22 22 25 18
Standard Deviation	3

Computed Power		
Index	N per Group	Power
1	2	0.222
2	3	0.456
3	4	0.657
4	5	0.800
5	6	0.891
6	7	0.944
7	8	0.972
8	9	0.987
9	10	0.994
10	11	0.997
11	12	0.999
12	13	>.999
13	14	>.999
14	15	>.999
15	16	>.999
16	17	>.999
17	18	>.999
18	19	>.999
19	20	>.999

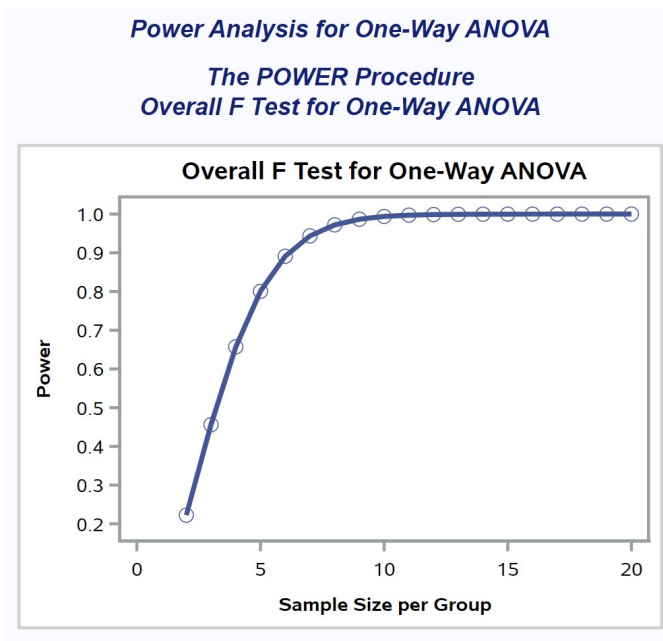Figure 12.3: `one-way_power.sas - proc power`

Figure 12.4: `one-way_power.sas - proc power`

12.3   Power analysis continued - SAS demo

It is often worthwhile to compare power curves for different values of σ and α, to see how these influence power. We can obtain this from `proc power` by specifying several different values of these parameters. We will examine the results for $\alpha = 0.05$ vs. 0.01 and $\sigma = 3$ vs. 6. These are requested by listing both values under the `alpha` and `stddev` statements. See program and output below.

The effects of α and σ on power can be readily seen in Fig. 12.6. Lower α reduces the power of the test across all sample sizes, because it makes it harder to reject H_0 under any circumstances. Larger values of σ also decrease the power at all sample sizes. The larger the value of σ, the more variable the data, and the harder it is for the statistical test to distinguish between the null and alternative hypotheses. Adequate power is only obtained for a larger sample size.

```
──────────────────────── SAS Program ────────────────────────

* oneway_power2.sas;
title 'Power Analysis for One-Way Anova';
proc power;
    onewayanova
        groupmeans = (20 22 22 25 18)
        stddev = 3 6
        alpha = 0.05 0.01
        npergroup = 2 to 20 by 1
        power = . ;
    plot x=n;
run;
quit;
```

Power Analysis for One-Way ANOVA

The POWER Procedure
Overall F Test for One-Way ANOVA

Fixed Scenario Elements	
Method	Exact
Group Means	20 22 22 25 18

Computed Power				
Index	Alpha	Std Dev	N per Group	Power
1	0.05	3	2	0.222
2	0.05	3	3	0.456
3	0.05	3	4	0.657
4	0.05	3	5	0.800
5	0.05	3	6	0.891
6	0.05	3	7	0.944
7	0.05	3	8	0.972
8	0.05	3	9	0.987
9	0.05	3	10	0.994
10	0.05	3	11	0.997

etc.
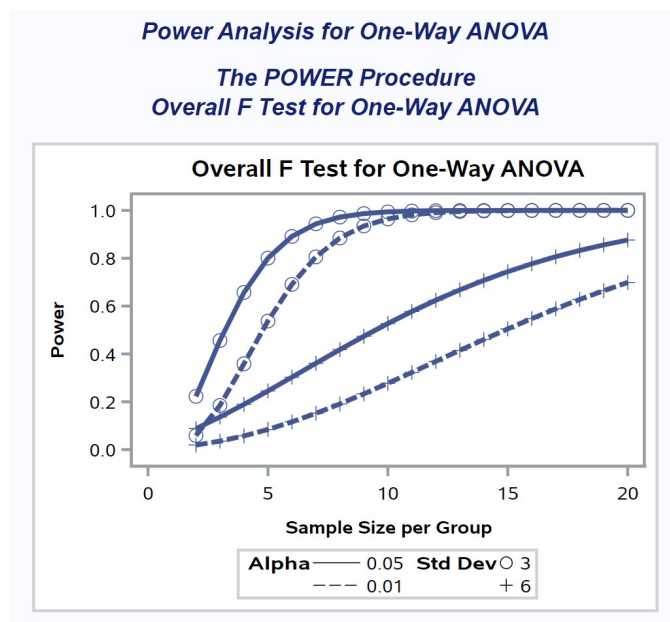
Figure 12.5: `one-way_power2.sas - proc power`

Figure 12.6: one-way_power2.sas - proc power

12.4 Power analysis continued - SAS demo

The SAS procedure `power` can be used to directly find the sample size n for a power of 0.8. One way is to simply read the value of n from a power vs. sample size graph, choosing the smallest n that gives power greater than or equal to 0.8. Returning to the first output we generated using `power` with $\alpha = 0.05$ and $\sigma = 3$, we see that for $n = 5$ the power exactly equals 0.8, so this is our sample size.

Alternately, you can set a power value of 0.8 and have `proc power` find the sample size. We first set the power option equal to 0.8 in the program, then change the `npergroup` option to a missing value, which tells `power` to solve for it. See program below and attached SAS output. SAS indicates that a sample size of $n = 5$ would give power $= 0.8$. This is the same result as obtained earlier by inspecting the power curve. For this particular example, there was a value of n that gave exactly the required power. More often, `power` will provide an n that guarantees power ≥ 0.8, not exactly 0.8.

──────────────── SAS Program ────────────────

```
* oneway_power3.sas;
title 'Power Analysis for One-Way ANOVA';
proc power;
    onewayanova
        groupmeans = (20 22 22 25 18)
        stddev = 3
        alpha = 0.05
        npergroup = .
        power = 0.8;
run;
quit;
```
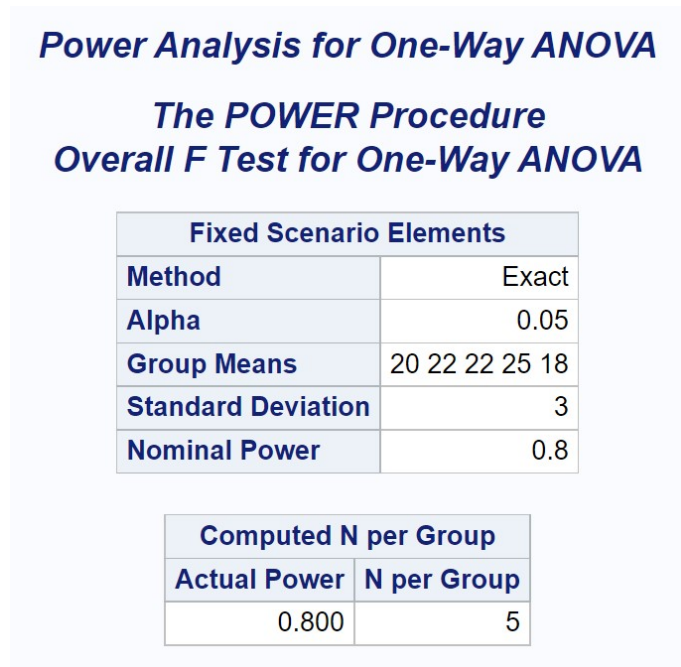---

Power Analysis for One-Way ANOVA

The POWER Procedure
Overall F Test for One-Way ANOVA

Fixed Scenario Elements	
Method	Exact
Alpha	0.05
Group Means	20 22 22 25 18
Standard Deviation	3
Nominal Power	0.8

Computed N per Group	
Actual Power	N per Group
0.800	5

Figure 12.7: `one-way_power3.sas - proc power`

12.5 References

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences, Second Edition.* Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.

Montgomery, D. C. (1997) *Design and Analysis of Experiments.* John Wiley & Sons, New York, NY.

SAS Institute Inc. (2018) *SAS/STAT 15.1 Users Guide* SAS Institute Inc., Cary, NC

Steidl, R. J., Hayes, J. P. & Schauber, E. (1997) Statistical power analysis in wildlife research. *Journal of Wildlife Management* 61: 270-279.

Winer, B. J., Brown, D. R. & Michels, K. M. (1991) *Statistical Principles in Experimental Design.* McGraw-Hill, Inc., Boston, MA.

12.6   Problems

1. Suppose you want to compare the effect of four different diets on the weight of prawns reared in aquaculture ponds. There is a standard diet (S) and three other diets (A, B and C) that will be fed to prawns in replicate ponds. Relative to diet S, you would like to see a 20% increase in weight on diet A, a 20% increase on diet B, and a 30% increase on diet C. If the mean weight on diet S is 100 g, this translates into the following alternative hypothesis:

$$H_1 : \mu_S = 100, \mu_A = 120, \mu_B = 120, \mu_C = 130. \qquad (12.9)$$

From previous studies the researchers estimate that $\sigma = 22$. Assume a Type I error rate of $\alpha = 0.05$.

   (a) Use SAS and `proc power` to determine the sample size per treatment (number of ponds) necessary to give power ≥ 0.8. Attach your SAS program and output.

   (b) Repeat the same analysis for $\alpha = 0.01$. How does this change in the Type I error rate affect the sample size? Why?

2. Suppose you want to compare the effect of five different diets on the weight of fish reared in aquaculture. There is a control diet (C) and four other diets (D1, D2, D3, and D4). Relative to diet S, you would like to see a 10% increase in weight on diet D1, a 15% increase on diet D2, and 20% increases on diets D3 and D4. If the weight on the control diet C is 100 g, this translates into the following alternative hypothesis:

$$H_1 : \mu_C = 100, \mu_{D1} = 110, \mu_{D2} = 115, \mu_{D3} = 120, \mu_{D4} = 120. \quad (12.10)$$

Previous studies suggest that $\sigma = 10$. Assume a Type I error rate of $\alpha = 0.05$.

   (a) Use SAS to determine the sample size per treatment necessary to give power ≥ 0.8. Attach your program and output.

   (b) Repeat the same analysis for the following alternative hypothesis:

$$H_1 : \mu_C = 100, \mu_{D1} = 105, \mu_{D2} = 108, \mu_{D3} = 110, \mu_{D4} = 110. \qquad (12.11)$$

How does this change affect the sample size? Why?