Analysis of Variance
Part 7: Power and Sample Size

STAT 705: Regression and Analysis of Variance

Overview

- If there truly are differences between treatment means, we want our statistical analysis to detect these differences
- Our ability to do this depends primarily on two things
  1. A properly designed and implemented study
  2. Sample sizes that are large enough to overcome the background variation ('noise') in measurements, so that we accurately estimate the true (population) treatment means
- You can learn more about item 1 in an Experimental Design course
- We will focus on item 2

Criteria for Determining Sample Size

- Select sample size so that the margin of error of the sample mean is a desired value
- Select a sample size so that the LSD (i.e. Fisher's LSD) is a desired value
- Select sample size so that the ANOVA F-test has desired power
  - Power is the probability that the null hypothesis is rejected when it is not true, that is, there <u>are</u> differences among means

Variability in the Populations

- We must know how much variability to expect, i.e., the standard deviation of the population
- Since this is not known, we must either
  - make an intelligent guess as to what it will be, or
  - use the value of the error standard deviation $(\hat{\sigma})$ from a previous study
- We will assume that such information is available from a prior study

Margin of Error Criterion

- Assume that all groups have the same number of observations. We will denote this value as ‘n’, that is, $n = n_1 = n_2 = \ldots = n_t$
- We wish to select a sample size so that the margin of error (MOE) of the sample mean will be a desired value. Let dMOE be the desired MOE.
- To find the sample size, we solve for n in the margin of error equation:

$$\text{dMOE} = t^{*} \times \frac{\hat{\sigma}}{\sqrt{n}}$$

MOE Criterion: Solving for n

- We specify the desired margin of error
- We obtain a value for $\hat{\sigma}$ from a prior study
- We use t* = 2 because the value of t* for a 95% confidence interval is about 2 (except for very small sample sizes)
- The solution is

$$n = \left(t^* \times \frac{\hat{\sigma}}{\text{dMOE}} \right) = \frac{4\hat{\sigma}^2}{(\text{dMOE})^2} = \frac{4\text{MSE}}{(\text{dMOE})^2}$$

Example 1

An analysis of variance was done on the grams of fat in bags of four brands of potato chips. The analysis of variance table is shown below.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	164.09	54.70	11.32	0.0002
Error	18	87.00	4.83		
Corrected Total	21	251.09			

In a future study, we would like to estimate the mean fat content with a desired margin of error of dMOE = 0.5 grams of fat

Example 1, continued

- The value of MSE is 4.83, so we use it in our sample size formula. We have

$$n = \frac{4\,\text{MSE}}{(\text{dMOE})^2} = \frac{4 \times 4.83}{(0.5)^2} = 77.28$$

- Use the next LARGER integer (rounding down would decrease the precision below our desired MOE)
- We would need about 78 potato chip bags <u>per</u> <u>brand</u> to achieve our desired margin of error
- If this sample size is too large, it simply tells us that our expectations for the future study are unrealistic

LSD Criterion

- By specifying a desired value of LSD, we are specifying the difference between sample means that we want to be able to detect. This is sometimes called a ‘meaningful’ difference.
- Let dLSD be the desired value of LSD
- Assuming all of the groups have the same sample size ($n = n_i = n_j$), we solve for n in the equation

$$\text{dLSD} = t^* \times \hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}} = t^* \times \hat{\sigma}\sqrt{\frac{2}{n}}$$

LSD Criterion: Solving for n

- Again, we assume we have an estimate of σ from a previous study and we use $t^* \approx 2$.
- The solution is

$$n = \left(\frac{t^* \hat{\sigma} \sqrt{2}}{dLSD} \right)^2 = \frac{8MSE}{(dLSD)^2}$$

- This is the number of observations we need for each group

Example 2

Continuing with the potato chip example, suppose we would like to have an LSD value of 2.0 for comparing the means. We have MSE = 4.83, so the sample size for each group would be

$$n = \frac{8\text{MSE}}{(\text{dLSD})^2} = \frac{8 \times 4.83}{2^2} = 9.66$$

As before, we round UP to 10.

Thus we would need 10 bags *per brand* to achieve our objective.

Power of the ANOVA F-test

- The power of a statistical test is the probability that the null hypothesis is rejected when it should be (i.e., when the alternative hypothesis is true)
- The power depends on
  - the sample size (larger samples produce greater power)
  - the difference between the true population means (larger differences generate greater power)

Sample Size to Attain a Given Power

- To determine the sample size, we must specify
  - The population standard deviation. As before, we can use an estimate from a previous study.
  - The populations means for which we would like to have a significant F-test
  - The level of significance of the test, usually 5%
  - The desired power of the test, typically between 0.5 and 0.9

Example 3

Suppose we want to do a new study involving four brands of potato chips. From a previous study, we have $\hat{\sigma}$ = 2.20 which we assume to be the population standard deviation in the new study.

We would like to do the F-test at the 5% level of significance and have probability 0.8 of detecting a difference between means if the true population means for the 4 groups are in fact 15, 16, 17 and 18.

Example 3, continued

- We used SAS to solve this problem

```
proc power;
  onewayanova
   groupmeans = 15 | 16 | 17 | 18
   stddev = 2.20
   alpha = 0.05
   npergroup = .
   power = 0.5 0.6 0.7 0.8 0.9
   ;
run;
```

Assumed population treatment means, separated by vertical bars

A dot indicates this is what we want Proc Power to calculate

SAS will calculate the sample size needed to achieve each value of power

- Additional details about ‘proc power’ are in a supplemental handout to this lesson

SAS Proc Power Output

Fixed Scenario Elements	
Method	Exact
Alpha	0.05
Group Means	15 16 17 18
Standard Deviation	2.2

Computed N Per Group			
Index	Nominal Power	Actual Power	N Per Group
1	0.5	0.529	7
2	0.6	0.603	8
3	0.7	0.725	10
4	0.8	0.817	12
5	0.9	0.906	15

The first table contains the specifications we provided.

The results are in the second table. For example, with 12 per group, we achieve power 0.817.

The actual power is the power we get when the sample size is rounded up to an integer. The nominal power is the target power we specified. It is usually not possible to achieve the nominal power exactly, since that would require a fractional sample size.

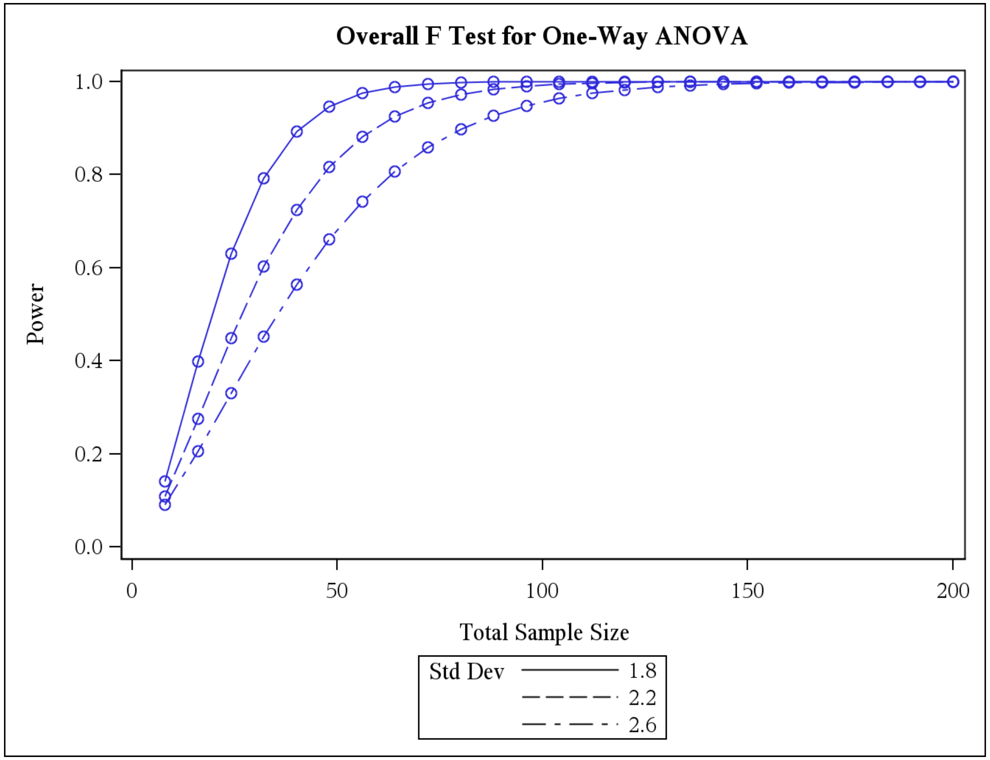
Power Curves

Power curves allow us to compare power and sample size for various scenarios

```
proc power;
 onewayanova test=overall
 groupmeans=15 | 16 | 17 | 18
 stddev= 1.8 to 2.6 by 0.4
 alpha = 0.05
 ntotal = 8 to 200 by 4
 power=.;
 plot  x=n min=8 max=200;
 run;
```

- We can specify ranges of values for
  - population standard deviation (**stddev**)
  - significance level (**alpha**)
  - total sample size (**ntotal**)
- The **plot** statement generates the graph on the next page

Power Curve for Bean Data



- Suppose we want power 0.8 to detect the differences in samples means that were specified on the previous slide
- We would need about 40 observations total (10 per group) if the population standard deviation is 1.8
- Need about 48 (12 per group) if the std. dev. is 2.2
- Need about 60 (15 per group) if the std. dev. is 2.6

There are other forms of power curves.
More examples are in the handout provided with this lesson.

What You Should Know

- How to calculate ‘by hand’ the sample size needed to achieve a desired LSD or a desired margin of error
- Understand what power is and how it depends on
  - the magnitude of the differences in means that we want to be able to detect
  - the underlying variability in the population values
  - the sample size
- Be able to use SAS Proc Power to calculate the power for specific scenarios
- Be able to interpret a power curve