

# Sample Size

Power and other methods

# Non-central Chisquare

$$Z_i \stackrel{i.i.d}{\sim} N(0, 1) \Rightarrow \sum_{i=1}^n Z_i^2 \sim \chi^2(n)$$

$$Z_i \stackrel{ind}{\sim} N(\mu_i, 1) \Rightarrow \sum_{i=1}^n Z_i^2 \sim \chi_{nc}^2(n, \phi = \sum_{i=1}^n \mu_i^2)$$

Let  $X_i \stackrel{ind}{\sim} N(\mu_i, \sigma^2)$ . Then

$$Y_i = \frac{X_i}{\sigma} \sim N\left(\frac{\mu_i}{\sigma}, 1\right)$$

$$\sum_{i=1}^n Y_i^2 \sim \chi_{nc}^2\left(n, \phi = \frac{\sum_{i=1}^n \mu_i^2}{\sigma^2}\right)$$

# Non-central F distribution

$$Y_1 \sim \chi_{nc}^2(\nu_1, \phi)$$

$$Y_2 \sim \chi^2(\nu_2)$$

$Y_1$  and  $Y_2$  independent

$$F = \frac{Y_1/\nu_1}{Y_2/\nu_2} \sim F_{nc}(\nu_1, \nu_2, \phi)$$

$$E(F) = \frac{\nu_2(\nu_1 + \phi)}{\nu_1(\nu_2 - 2)} \text{ for } \nu_2 > 2$$

$$Var(F) = \left( \frac{2\nu_2^2}{\nu_1^3(\nu_2 - 2)^2(\nu_2 - 4)} \right) (\phi^2 + (2\phi + \nu_1)(\nu_1 + \nu_2 - 2))$$

for  $\nu_2 > 4$

# General Linear Model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{h}$$

$$F^* = \frac{(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{h})'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}')^{-1}(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{h})}{r \text{MSE}}$$

$$F^* \sim F(r, n - p, \phi), \text{ where}$$

$$\phi = \frac{(\mathbf{L}\boldsymbol{\beta} - \mathbf{h})'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}')^{-1}(\mathbf{L}\boldsymbol{\beta} - \mathbf{h})}{\sigma^2}$$

Note that the sample size is concealed in the non-centrality parameter.

# Power

- When  $L\boldsymbol{\beta} \neq \mathbf{h}$ , the sample size influences the distribution of  $F^*$  through two quantities:
  - The denominator degrees of freedom  $n-p$
  - The non-centrality parameter  $\phi$
- The denominator degrees of freedom also influences the critical value, but the critical value of  $F^*$  settles down to the critical value of a chisquare with  $df=r$ , divided by  $r$ .
- Power goes to one as  $n$  goes to infinity, as long as the null hypothesis is false. That is, the F test is consistent.

# Comparing two means

Suppose we have a random sample of size  $n_1$  from a normal distribution with mean  $\mu_1$  and variance  $\sigma^2$ , and independently, a second random sample from a normal distribution with mean  $\mu_2$  and variance  $\sigma^2$ . We wish to test  $H_0 : \mu_1 = \mu_2$  versus the alternative  $H_a : \mu_1 \neq \mu_2$ .

We'll do it with dummy variable regression, letting  $x_i = 1$  if observation  $i$  is from population one, and  $x_i = 0$  if observation  $i$  is from population two. The model is the usual simple regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$\beta_1 = \mu_1 - \mu_2$$

$$H_0 : \beta_1 = 0$$

$$H_0: \mathbf{L} \boldsymbol{\beta} = \mathbf{h}$$

$$\mathbf{L} = \begin{bmatrix} 0 & 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \text{ and } \mathbf{h} = [0].$$

The  $\mathbf{X}$  matrix has  $n = n_1 + n_2$  rows; the first column is all ones, and the second column has  $n_1$  ones and  $n_2$  zeros.

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} = \begin{bmatrix} n_1 + n_2 & n_1 \\ n_1 & n_1 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{n_2} & -\frac{1}{n_2} \\ -\frac{1}{n_2} & \frac{1}{n_1} + \frac{1}{n_2} \end{bmatrix}$$

# Non-centrality parameter is

$$\begin{aligned}\phi &= \frac{\beta_1^2}{\sigma^2} \left( \frac{n_1 n_2}{n_1 + n_2} \right) \\ &= n \left( \frac{\mu_1 - \mu_2}{\sigma} \right)^2 \frac{n_1}{n} \frac{n_2}{n} \\ &= n f (1 - f) \left( \frac{\mu_1 - \mu_2}{\sigma} \right)^2 \\ &= n f (1 - f) d^2\end{aligned}$$

$f = n_1/n$  , the proportion of observations in treatment 1



# Non-centrality Parameter

$$\phi = nf(1 - f)d^2, \text{ where } f = \frac{n_1}{n} \text{ and } d = \frac{|\mu_1 - \mu_2|}{\sigma}$$

- $d$  is called *effect size*. The effect size specifies how wrong the null hypothesis is, by expressing the absolute difference between means in units of the common within-cell standard deviation.
- The non-centrality parameter (and hence, power) depends on the three parameters  $\mu_1$ ,  $\mu_2$  and  $\sigma^2$  *only* through the effect size  $d$ .
- Power depends on sample size, effect size and an aspect of design – allocation of relative sample size to treatments. Equal sample sizes yield the highest power in the 2-sample case.

# How to proceed

- Pick an effect size you'd like to be able to detect. It should be just over the boundary of interesting and meaningful.
- Pick a desired power – a probability with which you'd like to be able to detect the effect by rejecting the null hypothesis.
- Start with a fairly small  $n$  and calculate the power. Increase the sample size until the desired power is reached.

## For the 2-sample comparison

- Suppose we want to be able to detect a half standard deviation difference between means with power = 0.80 at the alpha = 0.05 significance level.
- Definitely use equal sample sizes.
- $\Phi = n f (1-f) d^2 = n * \frac{1}{2} * \frac{1}{2} * (\frac{1}{2})^2 = n/16$

# Two sample test with R

```
> n <- seq(from=120,to=140,by=2) ; phi <- n/16 ; ddf <- n-2  
> fcrit = qf(.95,1,ddf) # Critical value for each n  
> cbind(n,1-pf(fcrit,1,ddf,phi))
```

```
      n  
[1,] 120 0.7752659  
[2,] 122 0.7820745  
[3,] 124 0.7887077  
[4,] 126 0.7951683  
[5,] 128 0.8014596  
[6,] 130 0.8075844  
[7,] 132 0.8135460  
[8,] 134 0.8193475  
[9,] 136 0.8249920  
[10,] 138 0.8304825  
[11,] 140 0.8358223
```

# One Factor ANOVA (r means)

$$\phi = \frac{\sum_{k=1}^r n_k (\mu_k - \mu_{\cdot})^2}{\sigma^2} \quad \text{where } \mu_{\cdot} = \sum_{k=1}^r \frac{n_k}{n} \mu_k$$

$$\phi = n \frac{\sum_{k=1}^r \frac{n_k}{n} (\mu_k - \mu_{\cdot})^2}{\sigma^2} = n \frac{\sum_{k=1}^r f_k (\mu_k - \mu_{\cdot})^2}{\sigma^2}$$

$$= n \sum_{k=1}^r f_k \left( \frac{\mu_k - \mu_{\cdot}}{\sigma} \right)^2$$

Notice how division and multiplication by  $n$  allows us to write the noncentrality parameter as the product of the sample size and a quantity that is either a constant or settles down rapidly to a constant as the sample size increases.

$$\phi = n \frac{\sum_{k=1}^r \frac{n_k}{n} (\mu_k - \mu_{.})^2}{\sigma^2} = n \frac{\sum_{k=1}^r f_k (\mu_k - \mu_{.})^2}{\sigma^2}$$

- In the two-sample case, the non-centrality parameter was the product of sample size, effect size and the configuration of relative sample sizes.
- Once we get beyond two groups, effect and design are mixed together in a way that's impossible to separate.
- For a fixed sample size, phi (and hence power) is maximized by splitting the sample equally between the two treatments whose means are farthest apart, and giving zero observations to the other treatments.
- Reluctantly, we will still call  $\phi$   $n$  times “effect size.”
  - Even though it does not reduce to what we called effect size before, if  $r = 2$ . It's  $d^2/4$ .
  - And it is “size” in a metric strongly influenced by the allocation of relative sample size to treatments.

# Example

Suppose we have four treatments, and that the four population treatment means are equally spaced, one-quarter of a standard deviation apart. We'd like to be able to detect the differences among treatment means with probability 0.80, using the conventional significance level of  $\alpha = 0.05$ . We'll use equal sample sizes.

Without loss of generality, we'll let the four population treatment means be 0, 0.25, 0.50 and 0.75. Using R as a calculator, and remembering that the `var` function divides by the number of observations minus one, we'll calculate the effect size as

```
> 3 * var(c(0, .25, .5, .75)) / 4  
[1] 0.078125
```

```

fpow2 <- function(p,r,effsize,wantpow=0.80,alpha=0.05)
#####
# Power for the general multiple regression model, testing H0: L Beta = h  #
#   p      is the number of beta parameters                               #
#   r      Number rows in the L matrix = numerator df                    #
#   effsize is ncp/n, a squared distance between L Beta and h           #
#   wantpow is the desired power, default = 0.80                        #
#   alpha   is the significance level, default = 0.05                   #
# Returns sample size needed for desired power.                          #
#####
{
  pow <- 0 ; nn <- p+1 ; oneminus <- 1 - alpha
  while(pow < wantpow)
  {
    nn <- nn+1
    phi <- nn * effsize
    ddf <- nn-p
    pow <- 1 - pf(qf(oneminus,r,ddf),r,ddf,phi)
  }#End while
  fpow2 <- nn
  fpow2 # Returns needed n
}      # End of function fpow2

```

```

> source("http://www.utstat.toronto.edu/~brunner/appliedf12/data/fpow2.R")
> fpow2(p=4,r=3,effsize=0.078125)
[1] 144

```



# The Substitution Method

- Does  $\phi = \frac{\sum_{k=1}^r n_k (\mu_k - \mu.)^2}{\sigma^2}$  look familiar?
- It's the standard elementary formula for the Between-Groups sum of squares in a one-way ANOVA, except with  $\mu$  values substituted for sample means.
- This happens because the general formulas for F and  $\phi$  are so similar.

$$F^* = \frac{(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{h})'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}')^{-1}(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{h})}{r \text{ MSE}}$$

$$\phi = \frac{(\mathbf{L}\boldsymbol{\beta} - \mathbf{h})'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}')^{-1}(\mathbf{L}\boldsymbol{\beta} - \mathbf{h})}{\sigma^2}$$

- Any re-expression of the numerator of  $F^*$  in terms of the sample cell means corresponds to a re-expression of the numerator of  $\phi$  in terms of population cell means.
- So, to obtain a formula for the non-centrality parameter, all you have to do is locate a convenient formula for the F-test of interest. In the expression for the numerator sum of squares, replace sample cell means by population cell means. Then divide by  $\sigma^2$ . The result is a formula for the non-centrality parameter.
- This applies to any F-test in any fixed effects factorial ANOVA.
- See Scheffé (1959), page 39 for a more general version of this rule.

# Example: a 2-factor design

	Level of B		
Level of A	1	2	Average
1	$\mu_{11}$	$\mu_{12}$	$\mu_{1.}$
2	$\mu_{21}$	$\mu_{22}$	$\mu_{2.}$
3	$\mu_{31}$	$\mu_{32}$	$\mu_{3.}$
Average	$\mu_{.1}$	$\mu_{.2}$	$\mu_{..}$

For equal sample sizes I found the formula,

$$SSAB = n \sum_i \sum_j (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

Which yields

$$\phi = \frac{n}{6} \frac{\sum_{i=1}^3 \sum_{j=1}^2 (\mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..})^2}{\sigma^2}$$

Different n!

$$\phi = \frac{n}{6} \frac{\sum_{i=1}^3 \sum_{j=1}^2 (\mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..})^2}{\sigma^2}$$

- What is a meaningful effect size? As far as I can tell, the only solution is to make up a meaningful effect, and apply the formula to it.
- In general, special purpose formulas may yield insight, but maybe not.
- Locating a special-purpose formula can be time consuming.
- You have to be sure of the notation, too.
- It can require some calculator work or a little programming. Errors are possible.
- Often, a matrix approach is better, especially if you have to make up an effect and calculate its size anyway.

# Cell means dummy variable coding: $r$ indicators and no intercept

$$E[Y | \mathbf{X} = \mathbf{x}] = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Drug	$x_1$	$x_2$	$x_3$	$\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$
A	1	0	0	$\mu_1 = \beta_1$
B	0	1	0	$\mu_2 = \beta_2$
Placebo	0	0	1	$\mu_3 = \beta_3$

Test contrasts of the means:  $H_0: \mathbf{L} \boldsymbol{\beta} = \mathbf{0}$

# For designs with more than one factor

- Use cell means coding with one indicator for each treatment combination.
- All the usual tests are tests of contrasts.
- Use

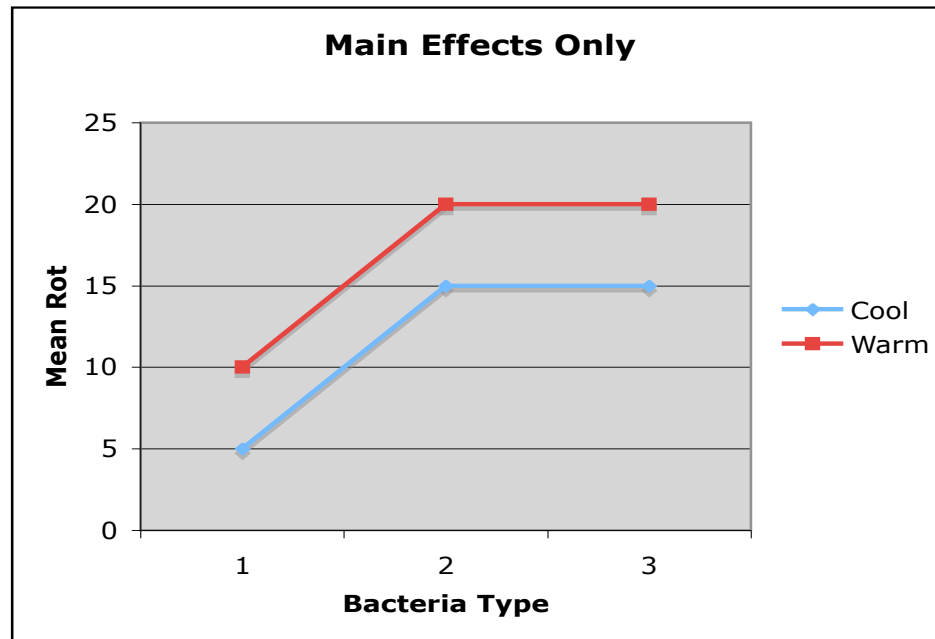
$$\phi = \frac{(\mathbf{L}\boldsymbol{\beta} - \mathbf{h})' (\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}')^{-1} (\mathbf{L}\boldsymbol{\beta} - \mathbf{h})}{\sigma^2}$$

# Testing Contrasts

	Bacteria Type			
Temp	1	2	3	
1=Cool	$\mu_{1,1}$	$\mu_{1,2}$	$\mu_{1,3}$	$\frac{\mu_{1,1} + \mu_{1,2} + \mu_{1,3}}{3}$
2=Warm	$\mu_{2,1}$	$\mu_{2,2}$	$\mu_{2,3}$	$\frac{\mu_{2,1} + \mu_{2,2} + \mu_{2,3}}{3}$
	$\frac{\mu_{1,1} + \mu_{2,1}}{2}$	$\frac{\mu_{1,2} + \mu_{2,2}}{2}$	$\frac{\mu_{1,3} + \mu_{2,3}}{2}$	$\mu$

- Differences between marginal means are definitely contrasts
- Interactions are also sets of contrasts

# Interactions are sets of Contrasts



- $H_0 : \mu_{1,1} - \mu_{2,1} = \mu_{1,2} - \mu_{2,2} = \mu_{1,3} - \mu_{2,3}$
- $H_0 : \mu_{1,2} - \mu_{1,1} = \mu_{2,2} - \mu_{2,1}$  and  
 $\mu_{1,3} - \mu_{1,2} = \mu_{2,3} - \mu_{2,2}$



# With cell means coding

- Assume there are  $p$  treatment combinations.
- The  $\mathbf{X}$  matrix has exactly one 1 in each row, and all the rest zeros.
- There are  $n_j$  ones in each column.

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n_r \end{bmatrix}$$

# Multiplying and dividing by n

$$\phi = n \times \left( \frac{\mathbf{L}\boldsymbol{\beta} - \mathbf{h}}{\sigma} \right)' \left( \mathbf{L} \begin{bmatrix} 1/f_1 & 0 & \cdots & 0 \\ 0 & 1/f_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/f_r \end{bmatrix} \mathbf{L}' \right)^{-1} \left( \frac{\mathbf{L}\boldsymbol{\beta} - \mathbf{h}}{\sigma} \right)$$

- $f_1, \dots, f_r$  are relative sample sizes:  $f_j = n_j/n$
- As usual, the non-centrality parameter is sample size times a quantity that we reluctantly call effect size.
- $\mathbf{L}\boldsymbol{\beta}$  is an *effect* -- a particular way in which the null hypothesis is wrong. It is naturally expressed in units of the common within-treatment standard deviation  $\sigma$ , and in general there is no reasonable way to avoid it.
- Almost always,  $\mathbf{h} = \mathbf{0}$ .

# To actually do a power analysis

- All you need is a vector of relative sample sizes,
- The contrast matrix  $\mathbf{L}$
- And a vector of numbers representing the differences between  $\mathbf{L}\boldsymbol{\beta}$  and  $\mathbf{h}$  in units of  $\sigma$ .

$$\phi = n \times \left( \frac{\mathbf{L}\boldsymbol{\beta} - \mathbf{h}}{\sigma} \right)' (\mathbf{L} \begin{bmatrix} 1/f_1 & 0 & \cdots & 0 \\ 0 & 1/f_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/f_r \end{bmatrix} \mathbf{L}')^{-1} \left( \frac{\mathbf{L}\boldsymbol{\beta} - \mathbf{h}}{\sigma} \right)$$

# Recall the two-factor interaction

	Level of B		
Level of A	1	2	Average
1	$\mu_{11}$	$\mu_{12}$	$\mu_{1.}$
2	$\mu_{21}$	$\mu_{22}$	$\mu_{2.}$
3	$\mu_{31}$	$\mu_{32}$	$\mu_{3.}$
Average	$\mu_{.1}$	$\mu_{.2}$	$\mu_{..}$

$$H_0 : \mu_{11} - \mu_{12} = \mu_{21} - \mu_{22} = \mu_{31} - \mu_{32}$$

$$\mathbf{L} = \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \\ \mu_{31} \\ \mu_{32} \end{bmatrix}$$

# An example

Suppose that for  $A=1$  and  $A=2$ , the population mean of  $Y$  is a quarter of a standard deviation higher for  $B=2$ , but if  $A=3$ , the population mean of  $Y$  is a quarter of a standard deviation higher for  $B=1$ . Of course there are infinitely many sets of means satisfying these constraints, even if they are expressed in standard deviation units. But they will all have the same effect size. One such pattern is the following.

	Level of B		
Level of A	1	2	Average
1	0.000	0.250	0.125
2	0.000	0.250	0.125
3	0.000	-0.250	-0.125
Average	0.000	0.083	0.042

Cell sample sizes are all equal, and we want to be able to detect an effect of this magnitude with probability at least 0.80.