Power and Sample Size for Linear Regresson[1]

STA442/2101 Fall 2016

---

[1]See last slide for copyright information.

Statistical Power

- Power is the probability of rejecting H_0 when H_0 is false.

- Choose the sample size so as to have decent power.

- Study the distribution of the test statistic when the null hypothesis is false.

Non-central Chisquare

$$Z \sim N(0,1) \;\; \Rightarrow \;\; Z^2 \sim \chi^2(1)$$

$$Z \sim N(\mu,1) \;\; \Rightarrow \;\; Z^2 \sim \chi^2_{nc}(1, \lambda = \mu^2)$$

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

$$Z_i \overset{ind}{\sim} N(\mu_i,1) \;\; \Rightarrow \;\; \sum_{i=1}^{n} Z_i^2 \sim \chi^2_{nc}(n, \lambda = \sum_{i=1}^{n} \mu_i^2)$$

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

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

$$Z_i = \frac{Y_i}{\sigma} \sim N(\frac{\mu_i}{\sigma}, 1)$$

$$\sum_{i=1}^{n} Z_i^2 \sim \chi_{nc}^2(n, \lambda = \frac{\sum_{i=1}^{n} \mu_i^2}{\sigma^2})$$

Non-central F distribution

$W_1 \sim \chi^2_{nc}(\nu_1, \lambda)$
$W_2 \sim \chi^2(\nu_2)$
W_1 and W_2 independent

$$F^* = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F(\nu_1, \nu_2, \lambda)$$

$E(F^*) = \frac{\nu_2(\nu_1+\lambda)}{\nu_1(\nu_2-2)}$ for $\nu_2 > 2$

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

Reduces to the usual central F when $\lambda = 0$.

$$\mathbf{Y} \;=\; \mathbf{X}\boldsymbol{\beta} \;+\; \boldsymbol{\epsilon}$$

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

$$F^* = \frac{(\mathbf{L}\widehat{\boldsymbol{\beta}}-\mathbf{h})^\top (\mathbf{L}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{L}^\top)^{-1}(\mathbf{L}\widehat{\boldsymbol{\beta}}-\mathbf{h})}{r\,MSE}$$

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

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

It's like the proof of $(\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \sim \chi^2(p)$

Recalling $Z_i \overset{ind}{\sim} N(\mu_i, 1) \Rightarrow \sum_{i=1}^n Z_i^2 \sim \chi^2_{nc}(n, \lambda = \sum_{i=1}^n \mu_i^2)$

1. $\widehat{\boldsymbol{\beta}} \sim N_p\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\right)$

2. $\mathbf{L}\widehat{\boldsymbol{\beta}} \sim N_r\left(\mathbf{L}\boldsymbol{\beta}, \sigma^2\mathbf{L}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{L}^\top\right)$

3. $\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{h} \sim N_r\left(\mathbf{L}\boldsymbol{\beta} - \mathbf{h}, \sigma^2\mathbf{L}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{L}^\top\right)$

4. Letting $\mathbf{V} = \sigma^2\mathbf{L}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{L}^\top$,

5. $\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{h} \sim N_r\left(\mathbf{L}\boldsymbol{\beta} - \mathbf{h}, \mathbf{V}\right)$

6. $\mathbf{z} = \mathbf{V}^{-1/2}(\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{h}) \sim N_r\left(\mathbf{V}^{-1/2}(\mathbf{L}\boldsymbol{\beta} - \mathbf{h}), \mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2}\right)$

7. $\mathbf{z} \sim N_r\left(\boldsymbol{\mu}, \mathbf{I}_r\right)$, where $\boldsymbol{\mu} = \mathbf{V}^{-1/2}(\mathbf{L}\boldsymbol{\beta} - \mathbf{h})$

8. $\mathbf{z}^\top\mathbf{z} = \frac{1}{\sigma^2}(\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{h})^\top(\mathbf{L}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{L}^\top)^{-1}(\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{h})$

9. $\mathbf{z}^\top\mathbf{z} = \sum_{j=1}^r Z_j^2 \sim \chi^2_{nc}(n, \lambda = \sum_{j=1}^r \mu_j^2)$

10. $\lambda = \boldsymbol{\mu}^\top\boldsymbol{\mu} = (\mathbf{L}\boldsymbol{\beta} - \mathbf{h})^\top\mathbf{V}^{-1}(\mathbf{L}\boldsymbol{\beta} - \mathbf{h})$

11. $\lambda = \frac{(\mathbf{L}\boldsymbol{\beta} - \mathbf{h})^\top(\mathbf{L}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{L}^\top)^{-1}(\mathbf{L}\boldsymbol{\beta} - \mathbf{h})}{\sigma^2}$

Power increases to one with sample size when H_0 is false

- The non-central F distribution is *stochastically increasing* in λ.
- Stochastically increasing means $Pr\{F > x\}$ is an increasing function of λ for each fixed $x > 0$.
- The sample size is concealed in the non-centrality parameter.

$$\lambda = n \, \frac{(\mathbf{L\boldsymbol{\beta}} - \mathbf{h})^\top (\mathbf{L}(\frac{1}{n}\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{L}^\top)^{-1}(\mathbf{L\boldsymbol{\beta}} - \mathbf{h})}{\sigma^2}$$

- So the non-centrality parameter increases with n unless $\mathbf{X}$ is crazy.
- This means $Pr\{F > x\}$ is an increasing function of n.
- The critical value x is a function of $n - p$, but it's a *decreasing* function.
- Finally, $\lim_{\lambda \to \infty} Pr\{F > x\} = 1$.
- All this means that power goes to one as n goes to infinity, as long as the null hypothesis is false. That is, the F test is consistent.

Suppose we have a random sample of size n_1 from a normal distribution with mean μ_1 and variance σ^2, and independently, a second random sample from a normal distribution with mean μ_2 and variance σ^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. Use cell means coding:

$$Y_i = \mu_1 x_i + \mu_2 (1 - x_i) + \epsilon_i$$

$$
\begin{array}{ccc}
\mathbf{L} & \boldsymbol{\beta} & = \mathbf{h} \\
\begin{pmatrix} 1 & -1 \end{pmatrix} & \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} & = 0
\end{array}
$$

$$
\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix} \qquad (\mathbf{X}^\top \mathbf{X})^{-1} = \begin{pmatrix} 1/n_1 & 0 \\ 0 & 1/n_2 \end{pmatrix}
$$

$$
\begin{aligned}
\lambda &= \frac{1}{\sigma^2} (\mathbf{L}\boldsymbol{\beta} - \mathbf{h})^\top (\mathbf{L}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{L}^\top)^{-1} (\mathbf{L}\boldsymbol{\beta} - \mathbf{h}) \\
&= \frac{1}{\sigma^2} (\mu_1 - \mu_2) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1} (\mu_1 - \mu_2)
\end{aligned}
$$

Non-centrality parameter

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

where $f = \frac{n_1}{n}$, the fraction of observations in group one.

Effect size

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

- 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 μ_1, μ_2 and σ^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.

- 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.
- $\lambda = nf(1-f)d^2 = n\left(\frac{1}{2}\right)\left(1 - \frac{1}{2}\right)\left(\frac{1}{2}\right)^2 = n/16$

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

$n_1 = n_2 = 64$

One Factor ANOVA (p means)

$$\lambda = \frac{\sum_{k=1}^{p} n_k(\mu_k - \mu_.)^2}{\sigma^2} \quad \text{where } \mu_. = \sum_{k=1}^{p} \frac{n_k}{n}\mu_k$$

$$
\begin{aligned}
\lambda &= \frac{\sum_{k=1}^{p} n_k(\mu_k - \mu_.)^2}{\sigma^2} \\
&= n \frac{\sum_{k=1}^{p} \frac{n_k}{n}(\mu_k - \mu_.)^2}{\sigma^2} \\
&= n \frac{\sum_{k=1}^{p} f_k(\mu_k - \mu_.)^2}{\sigma^2} \\
&= n \sum_{k=1}^{p} f_k \left(\frac{\mu_k - \mu_.}{\sigma} \right)^2
\end{aligned}
$$

Notice how division and multiplication by n helps.

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

- 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, λ (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 λ/n "effect size"
  - Even though it does not reduce to what we called effect size before, if $p = 2$. It's $d^2/4$.
  - And it is "size" in a metric strongly influenced by the allocation of relative sample size to treatments.

$$\lambda \;=\; n \sum_{k=1}^{p} f_k \left(\frac{\mu_k - \mu_.}{\sigma} \right)^2$$

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

Example using $\lambda = n \sum_{k=1}^{p} f_k \left(\frac{\mu_k - \mu_.}{\sigma} \right)^2$

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 $\alpha = 0.05$. 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
```

An R function

```
fpow3 <- 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                     #
##############################################################################
    {
    pow <- 0 ; nn <- p+1 ; oneminus <- 1 - alpha
    while(pow < wantpow)
        {
        nn <- nn+1
        lambda <- nn * effsize
        ddf <- nn-p
        pow <- 1 - pf(qf(oneminus,r,ddf),r,ddf,lambda)
        }#End while
    return(nn)
    }       # End of function fpow3
```

Finding sample size with `fpow3`

```
> source("http://www.utstat.utoronto.ca/~brunner/Rfunctions/fpow3.txt")
> fpow3(p=4,r=3,effsize=0.078125)
[1] 144
```

- Does the numerator of $\lambda = \frac{\sum_{k=1}^{p} n_k (\mu_k - \mu_{\cdot})^2}{\sigma^2}$ look familiar?
- Its the standard elementary formula for the Between-Groups sum of squares in a one-way ANOVA, except with μ values substituted for sample means.
- This happens because the general formulas for F and λ are so similar.

$$
\begin{aligned}
F^* &= (\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{h})^\top (\mathbf{L}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{L}^\top)^{-1}(\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{h})/(r\,MSE) \\
\lambda &= (\mathbf{L}\boldsymbol{\beta} - \mathbf{h})^\top (\mathbf{L}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{L}^\top)^{-1}(\mathbf{L}\boldsymbol{\beta} - \mathbf{h})/\sigma^2
\end{aligned}
$$

- Any re-expression of the numerator of F^* in terms of the sample cell means corresponds to a re-expression of the numerator of λ 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 σ^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.

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

For equal sample sizes, I found the formula

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

Which yields

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

Different n !

How to use $\lambda = \frac{n}{6} \frac{\sum_{i=1}^{3} \sum_{j=1}^{2} (\mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..})^2}{\sigma^2}$?

- 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.

- Use cell means coding with one indicator for each treatment combination.
- There are p treatment combinations.
- All the usual tests are tests of contrasts.
- The $\mathbf{X}$ matrix has exactly one 1 in each row, and all the rest zeros.
- There are n_j ones in column j.

$$\mathbf{X}^{\top}\mathbf{X} = \begin{pmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n_p \end{pmatrix} = n \begin{pmatrix} f_1 & 0 & \cdots & 0 \\ 0 & f_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f_p \end{pmatrix}$$

$$\lambda = n \times (\frac{\mathbf{L}\boldsymbol{\beta} - \mathbf{h}}{\sigma})^\top (\mathbf{L} \begin{pmatrix} 1/f_1 & 0 & \cdots & 0 \\ 0 & 1/f_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/f_r \end{pmatrix} \mathbf{L}^\top)^{-1}(\frac{\mathbf{L}\boldsymbol{\beta} - \mathbf{h}}{\sigma})$$

- $f_1, \ldots f_p$ 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.
- $\mathbf{L}\boldsymbol{\beta} - \mathbf{h}$ is an *effect* – a particular way in which the null hypothesis is wrong.
- The effect is naturally expressed in units of the common within-treatment standard deviation σ, and in general this is hard to avoid.
- 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 σ.

$$\lambda = n \times (\frac{\mathbf{L}\boldsymbol{\beta} - \mathbf{h}}{\sigma})^{\top} (\mathbf{L} \begin{pmatrix} 1/f_1 & 0 & \cdots & 0 \\ 0 & 1/f_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/f_r \end{pmatrix} \mathbf{L}^{\top})^{-1} (\frac{\mathbf{L}\boldsymbol{\beta} - \mathbf{h}}{\sigma})$$

Recall the two-factor interaction

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

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

$$\mathbf{L} = \left(\begin{array}{cccccc} 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{array} \right) \qquad \boldsymbol{\beta} = \left(\begin{array}{c} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \\ \mu_{31} \\ \mu_{32} \end{array} \right)$$

A numerical 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

- $\frac{\mathbf{L}\boldsymbol{\beta}-\mathbf{h}}{\sigma} = \begin{pmatrix} 0 \\ 0.5 \end{pmatrix}$
- 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.

Copyright Information

This slide show was prepared by Jerry Brunner, Department of Statistics, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely. The LaTeX source code is available from the course website:
`http://www.utstat.toronto.edu/~brunner/oldclass/appliedf16`