実験計画法
(May 1, 2013)

- 実験的研究は，どんなものであれ，注意深くデザインされねばならない。
- 実験計画法は，R.A. Fisherがロザムステッドで行った農学研究に始まる。
- 保健医療分野では，この種のデザインは，毒性試験や臨床試験で用量反応関係を分析するために必須。
- もちろん，ヒトを対象にした研究は，実施前に倫理審査を通らねばならず，倫理審査に提出する書類には，適切なサンプルサイズの設定を含む，適切な研究デザインが記述されねばならない。

Fisherの3原則

- 繰り返し（Replication）:各処理について最低2回以上の繰り返し測定が必要。
- 無作為化（Randomization）:実験の順序や区画（農業試験の場合）は無作為に割り付けねばならない。
- 局所コントロール（＝ブロック化：blocking）:大規模な実験の場合，サンプル全体の無作為化は不適切であり，代わりにいくつかのやり方で局所のブロックを作り，各ブロック内で無作為割り付けをすることで，ブロック間変動として，偏りを除去することができる。

実験計画の起源に関わる，ある伝説

- ケンブリッジのある晴れた日，多くの教授がアフタヌーンティーを楽しんでいた。と，ある婦人が，自分はミルクティーを飲めば，それがミルクが先か紅茶が先のどちらで淹れたものか判定できると主張した。
  - 教授たちの間で，彼女の主張を廻って大激論勃発。
- そこでR.A. Fisher曰く「実験したらどうだい？」
- この能力は，ミルクを先に入れて作ったミルクティーと紅茶を先に淹れて作ったミルクティーを何杯か用意して，無作為な順番で飲んで貰えば検定可能。
- 能力の判定条件を考える必要あり。何回必要？

（参考）ミルクと紅茶の順番は本当に味に影響する？

- George Orwell 「完璧な紅茶を淹れる11の法則」
  - 10: ミルクを紅茶に加えるのだ。逆ではいけない。
  - 英国の王立化学会がGeorge Orwellの生誕100年を記念するパーティを開いたとき，Dr. Andrew Stapley (2003)曰く
    - 冷えたミルクをカップの底に入れておいてから，熱い紅茶を注ぐのが良い。こうするとミルクが紅茶を冷ますことができる。逆だと熱い紅茶がミルクの温度を急に上げるのでミルクの風味が損なわれる。
- 情報源: BBC News (http://news.bbc.co.uk/2/hi/uk/3016342.stm)

George Orwell's 11 rules

- 1. Use tea from India or Ceylon (Sri Lanka), not China
- 2. Use a teapot, preferably ceramic
- 3. Warm the pot over direct heat
- 4. Tea should be strong - six spoons of leaves per 1 litre
- 5. Let the leaves move around the pot - no bags or strainers
- 6. Take the pot to the boiling kettle
- 7. Stir or shake the pot
- 8. Drink out of a tall, mug-shaped tea cup
- 9. Don't add creamy milk
- 10. Add milk to the tea, not vice versa
- 11. No sugar!

完璧な一杯の紅茶の真実（続き）

- 本当はどちらが先だとより美味しいのか，試してみた日本人ブロガーがいた。130ccの紅茶と30ccのタカナシ低温殺菌牛乳を使用した（高温殺菌とかロングライフのミルクでは違いが分からないらしい）。
- この人の主観的判断では，「ミルクが先」が美味。
  http://blog.livedoor.jp/teatime312/archives/cat_123365.html

白黒付けるには何杯飲めばいい？

- Correct judge of 1/1 may occur at 50%.
- Correct judges of 2/2 may occur at 25%
- Correct judges of 3/3 may occur at 12.5%
- Correct judges of 4/4 may occur at 6.25%
- Correct judges of 5/5 may occur at 3.125%
  - 本当は判別能力が無いのに偶然5回連続で正解する，3.125% という値は，偶然で片付けるには稀すぎる。通常，この判定基準は5%を切るかどうかにおく。これが有意水準（Fisher流）である。帰無仮説「彼女は判定能力をもっていない」が有意水準5%で棄却される。
- Testing this hypothesis requires at least 5 cups.

有名な実験計画デザイン

- One group pretest-posttest design: paired t-test.
- Completely randomized design: t-test/ANOVA for quantitive data, chi-square test for proportion
- Randomized block design: similar to completely randomized design / considering block's effect
- Latin-square: usually ANOVA
- Crossover design: Matched (paired) analysis of variance (within-subject difference will be zero or not, adjusted by the order of treatment)

One group pretest-posttest design

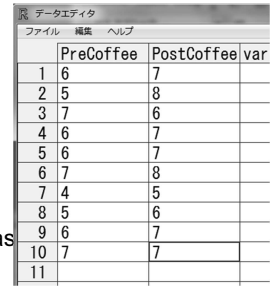
- The design enables a researcher to compute a contrast between means in which the pretest and posttest means are measured with the same precision.
  - Compare serum cortisol levels before and after the surgery in rheumatoid arthritis patients
  - Compare the depression score before and after the sound-therapy in depression patients
  - Compare the simple calculation test score before and after drinking coffee.
- The statisical test is usually paired t-test.
- Null hypothesis: the mean of difference is zero.

Example of paired t-test

- We can use “survey” dataframe of MASS package in R (EZR), whereas it is the result of cross-sectional study.
- In EZR, select [File] and [Read data included in package], then select [MASS] and [survey].
- The “survey” contains the responses of 237 students at the University of Adelaide to a number of questions (Venables and Ripley, 1999). Variables include the span (distance from tip of thumb to tip of little finger of spread hand) of writing hand as [Wr.Hnd] and that of non-writing hand as [NW.Hnd].
- Select [Statistical Analysis], [continuous variables], then [paired t-test]. Select [NW.Hnd] at left panel and [Wr.Hnd] at right panel and click [OK]. That's all.

Exercise

- Compare the results of simple calculation test before and after drinking coffee.
- In EZR, at first, making data: select [File], [New data], then enter the data as right screen-capture.
- Conducting paired t-test can be done in similar manner as NW.Hnd-Wr.Hnd
- [t = -2.862, df = 9, p-value = 0.01872] mean significant difference.



	PreCoffee	PostCoffee	var
1	6	7	
2	5	8	
3	7	6	
4	6	7	
5	6	7	
6	7	8	
7	4	5	
8	5	6	
9	6	7	
10	7	7	
11			

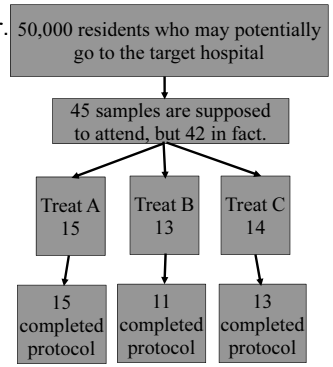
Pararell group design (=Completely randomized design)

- Very simple. The subjects who signed informed consent are completely randomly (not haphazardly) assigned to one of the several treatment (exposure).
- There are several randomization methods. Fleiss JL (1986) “The design and analysis of clinical experiments” recommends to use “random permutation tables” instead of “random number tables”.
- However, now we can use computer software. If we want to assign 45 subjects into 3 treatments, type matrix(sample(1:45, 45, replace=FALSE),3) in R.

```
> matrix(sample(1:45, 45, replace=FALSE),3) -> x
> x[1,]
 [1] 34 13 37 17  3 11  4 23 18 39 42 40 36  8 25
> t(apply(x,1,sort))
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
[1,]    3    4    8   11   13   17   18   23   25   34   36   37   39   40   42
[2,]    1    6    7   10   14   15   16   19   26   29   31   33   41   43   44
[3,]    2    5    9   12   20   21   22   24   27   28   30   32   35   38   45
```

How to describe?

- Dropouts sometimes occur.
- The design is usually shown as diagram (right).
- In this diagram, the quantitive data of the subjects can be compared by one-way ANOVA , proportion by chi-square test.
- Unbalanced sample size may reduce the statistical power.



Randomized block(s) design （乱塊法）

- Due to incompleted study, completely randomized design may lead to unbalanced sample sizes among groups.
- If the sample size for each treatment is 15 and the kind of treatment is 3, randomized block design randomly select one of 6 possible blocks ({A, B, C}, {A, C, B}, {B, C, A}, {B, A, C}, {C, A, B}, {C, B, A}) 15 times. By doing so, if the study may suspend in the middle, the sample size difference is at most 1 among groups. Description and analysis can be similar to complete randomization, but the analysis considering blocks' effect is also possible.
- Another method to keep size balance is “Minimization design”. It minimize the sample size difference at each time of sampling.

Factorial design

- Example of 2x2 factorial design
- McMaster et al. (1985): a randomized trial to evaluate breast self-examination teaching materials as leaflets or tape/slides.
- The treatment groups were designed as four pararell groups as
  1. No leaflets nor tape/slides given (control)
  2. Leaflets displayed
  3. Tape/slides program
  4. Both given
- The effect of teaching can be evaluated using ANOVA: two kinds of materials can be evaluated.

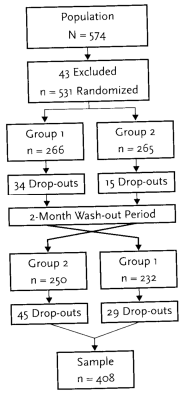
Latin-square design（ラテン方格法）

- When the experiment have one treatment (A) with p≧2 levels, 2 nuisance variables (B, C) each with p levels, this design is useful. The name is originated from ancient puzzle.
- Assume p is 3. Latin square is shown as right. Group 1 gets the combination of treatment a1b1c1 for n1 individuals. Following n2 individuals get treatment a1b2c3 as group 2.
- By doing so, the effects of B and C on the outcome measure can be controlled in ANOVA.

	c1	c2	c3
b1	a1	a2	a3
b2	a2	a3	a1
b3	a3	a1	a2

Crossover design（クロスオーバー法）

- Subjects will be serially treated by 2 kinds of intervention with proper interval (wash-out period to avoid carry-over) in different order.
- (Example) Hilman BC et al. “Intracutaneous Immune Serum Globulin Therapy in Allergic Children.”, JAMA. 1969; 207(5): 902-906.



Types of outcome measure

- Superiority trials: The effect of new treatment is significantly better than control or not
- Equivalence trials: The effect of new treatment is similar to control or not
- Non-inferiority trials: Special case of equivalence trials.