Lecture 6

Analysis of variance (ANOVA)

方差分析

ANOVA

为了比较两组的均值，我们使用了t检验和非参数检验

- To compare *two* group means, we used *t*-tests and their non-parametric alternatives (Wilcoxon tests)

要比较三组或更多组，策略是：将围绕变量均值的总方差分为组间方差和组内方差

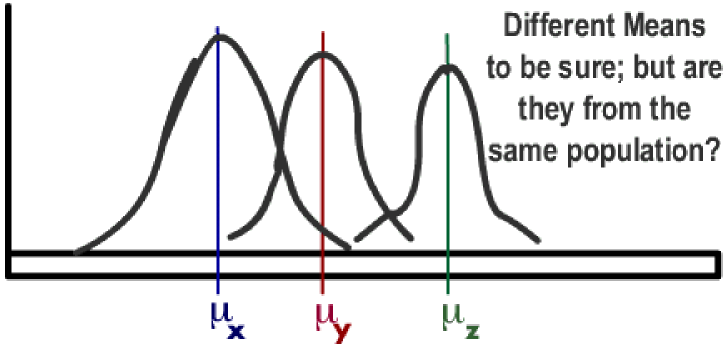
- To compare *three or more* groups, strategy is:
  - to split total variance around mean of our variable into *between-group* and *within-group* variance
  - to test whether ratio between-group/within-group variance is greater than expected by chance; if it is, group means differ

检验组间/组内的比值是否大于随机期望，如果是，则各组均值不同

- This procedure is known as *analysis of variance* (ANOVA)
  - Note: ANOVA assumes that all groups are normally distributed

ANOVA假设所有组均呈正态分布

- ANOVA is still one of the most commonly used techniques in data analysis
  - seasonality (grouping variable: month) 季节性，分组变量：月
  - regional patterns (grouping variable: region) 区域模式，分组变量：区域
  - any difference in group means across more than two groups



ANOVA: between- vs. within- group differences

- If a sample is structured into groups (months, continents, species etc.), each individual case x_i in the sample can be rewritten as:

x = 总平均 + (群平均 - 总平均) + (x - 群平均)

x_i = general mean + (group mean – general mean) + (x_i – group mean)

or

x_i = general mean + (*between-group* difference) + (*within-group* difference)

x = 总平均 + 组间差异 + 组内差异

define a penguin:

Example

- Take a sample of 3000 male heights, divided into 3 nationalities (1000 Dutch, 1000 UK, 1000 Spanish men)
  - General mean height = 182 cm
  - UK mean = 180 cm

- Now take a UK individual with x_1 = 170 cm;
- This can be written as

x_1 = general mean + (UK mean – general mean) + (x_1 – UK mean)
= 182 + (180 – 182) + (170 – 180)
= 182 – 2 – 10 = 170 cm

-12cm的差距分为-2cm的组间差异和-10cm的组内差异

- The difference of -12 cm between x1 (170 cm) and general mean (180 cm) was split into a between-group difference of -2 cm (UK versus general mean) and –10 cm (within group difference between x1 and other UK people)

Between- and within-group differences

- If 总样本差的平方 3000个个体与总均值之间的差值

  S = total sample squared differences

  (between each of the 3000 individuals and general mean)

  S_b= sum of all squared *between-group* differences 组间差异的所有平方和

  (Dutch mean vs. general mean, UK mean vs. general mean, Spanish mean vs. general mean)

  S_w= sum of all squared *within-group* differences 组内差异的所有平方和

  (each Dutch/UK/Spanish individual vs. Dutch/UK/Spanish mean)

- Then:

  $$\boxed{S = S_b + S_w}$$

  i.e. total variation around mean height in sample can be decomposed into *group effect* and *individual effect*

  可以分解为群体效应和个体效应

Between- and within-group variances

- We can calculate the mean of S_b and S_w
  - $M_b = S_b/k\text{-}1$ = between-group variance   组间方差
  - $M_w = S_w/N\text{-}k$ = within-group variance   组内方差

- It can be shown that under random sampling of groups from a normal distribution:

$$\boxed{M_b = M_w} \quad \text{or} \quad \boxed{M_b / M_w = 1}$$

In other words: if grouping is random or arbitrary
  - (i.e. if there is no real difference between groups),
- …expected difference between groups
  - (e.g. difference between groups 1 and 2 etc.)
- …is similar to expected difference between individuals
  - (difference between two cases in group 1)

- Why? Because if groups are random, groups have no effect on variable, and on average they randomly deviate as much from each other as an individual randomly deviates from its group
- But if grouping is real (i.e. if it has an effect on variable), between-group variance should exceed what is expected by chance

F-test

用f比率测试mb/mw是否不同于1

- ANOVA uses an *F*-ratio to test whether ratio M_b/M_w differs from 1

$$F = M_b/M_w$$

f分布估计组间方差与组内方差显著不同的概率p=他们的比率不是1

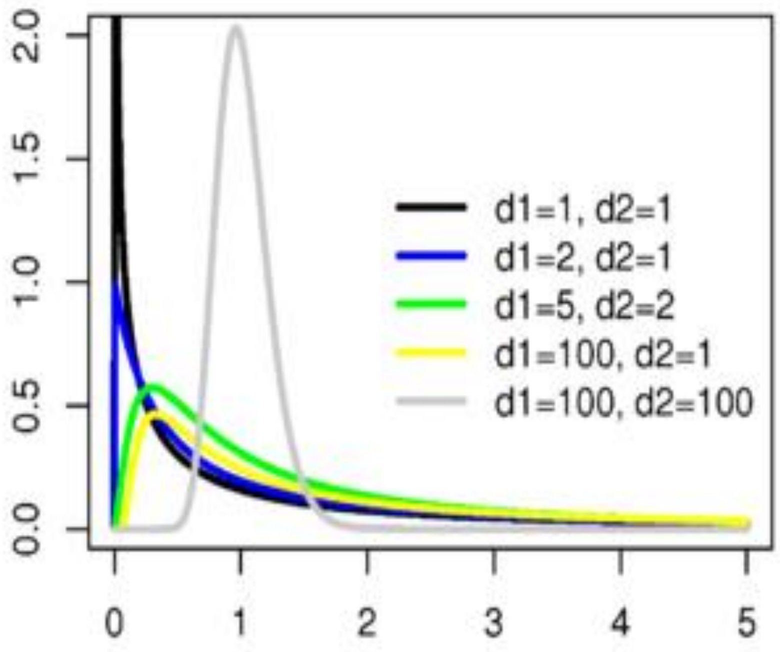
- *F*-distribution estimates the probability *P* that between-group variance is significantly different from within-group variance
  - = the probability that their ratio is not 1

ANOVA is a test of the F ratio

- If ***F* = 1**, group means **do not** differ
  - **null hypothesis: *F does not significantly differ from 1***
  - If $P<0.05$ (95% CI), then $F>1$ and groups differ

- *F*-test is always one-tailed
  - if there is a group effect, $F>1$;
  - if there is no group effect, $F\sim1$

f检验总是单侧的



组间差异相等

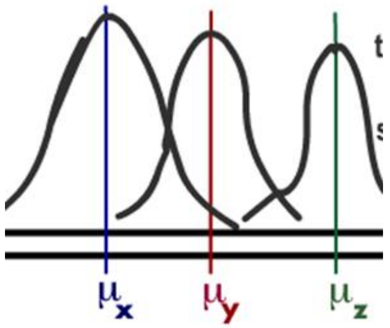
Equality of variances across groups

- In addition to assuming that all groups have a normal distribution, ANOVA and *F*-test may or may not assume that within-group variance is the same for all *k* groups

通过这个检验来检查方差是否相等(通过比较组中观察到的与预期的差异)

- We have to run the *Bartlett's test* to check for equality of variances
  - by comparing observed vs. expected variances from the compared groups

> bartlett.test(variable ~ grouping variable)



Then:

- if variances are similar across groups (=if Bartlett's test is not significant):
  - run ANOVA using ***anova(lm)***
- If variances differ across groups (= (=if Bartlett's test is significant):
  - run ANOVA using ***oneway()*** function

Example: Swedish babies

- Dataset: Swedish Birth Register, all births 1982-2005

- Variables:
  - birth year
  - birth weight (variable '*size*')
  - head circumference
  - maternal height
  - pregnancy duration
  - delivery type (natural, caesarean, instrumental)

Head circumference, 2002-2005

- Does head circumference (variable *head*) in boys change between 2002-05?
  - file *SBR2*, boys from four years: 2002, 2003, 2004, 2005
- Sample sizes are large:

```
> table(SBR2$year)
 2002   2003   2004  2005
48364 50159 51739 51339
```

- Means look very similar across years: 不同年份平均值看起来很相似

```
> tapply(SBR2$head, SBR2$year, mean, na.rm=T)
      2002       2003       2004       2005
  35.23045   35.31483   35.27654   35.27393
```

- Variances look similar too: 方差看起来也很相似

```
> tapply(SBR2$head, SBR2$year, var, na.rm=T)
     2002       2003       2004       2005
 3.229198   3.235278   3.184767   3.206313
```

Variances

首先，检验组间方差差异是否显著

- First, let's test whether between-group differences in variance are significant:

```
> bartlett.test(SBR2$head ~ SBR2$year)
        Bartlett test of homogeneity of variances
data:  SBR2$head by SBR2$year
Bartlett's K-squared = 3.7236, df = 3, p-value = 0.2929
```

- Null hypothesis: no difference of variance across the four years
  - P=0.29
  - → accept null hypothesis
- Conclusion: no differences in variance across years

= we use function *anova(lm)* to compare groups

分组变量必须是因子

With *anova(lm)*, grouping variable MUST be a factor

- Important: when you run ANOVA with command *anova(lm)*, **grouping variable should always be a factor**, and never a numeric variable!
  - (reason: if your grouping variable is numeric, *R* runs a linear regression instead of ANOVA)

- Use function *class()* to check whether your variable is numeric or a factor
  - if variable is NOT a factor, use *as.factor()* to convert it
    - in your code, Instead of file$variable, write it as as.factor(file$variable)
    - but for permanent change:

    ```
    > file$variable <- as.factor(file$variable)
    ```

    用这个转换成因子类型
  - If grouping variable is already a factor(i.e. month as Jan, Feb, Mar; or species as human, chimp, gorilla), you don’t need to use *as.factor*

Head circumference, 2002-05

- Is head circumference affected by birth year?
- Let's run an ANOVA on head circumference by year

Df = 4 - 1 (总共四年减一)
Sum Sq -> Sb Mean Sq ->Mb 613297->Sw
Fvalue -> Sb/Sw Pr->p value

```
> anova(lm(SBR2$head ~ SBR2$year))
        Analysis of Variance Table
Response: SBR2$head
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
SBR2$year	3	165	55.081	17.[illegible]41	3.987e-11 ***
Residuals	19085[illegible]	613297	3.213		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- null hypothesis: means are similar, i.e F=1
- Result: F=17.1, P<0.05
  →reject null hypothesis

- Conclusion: although small, differences in head circumference are significant

成对t检验，holm校正

Pairwise *t*-tests, Holm correction

ANOVA不会告诉你哪一年是不同的那个

- But ANOVA does not tell you which of the four years is/are the different one/ones!
    - we must run multiple **pairwise *t*-tests** between pairs (2002 vs. 2003, 2002 vs. 2004…) to identify which groups differ

- However, pairwise comparisons cause a problem:
    - because of multiple testing, you increase the chance of finding a ‘significant’ difference by chance
    - example: you have a very small chance of getting 10 heads from 10 coin tosses; but if you try it *many times*, you may eventually get 10 heads by chance!
        - that’s cheating!!!
        - P=0.05 means a chance of 1 in 20 of getting a significant test in *one t*-test; but if you run multiple *t*-tests on new samples, probability of getting $P<0.05$ once keeps increasing

解决办法

- How to solve the problem? A solution is to ***punish multiple testing***
- *Bonferroni correction*: if you run *n* tests on the same variables, you must multiply your test *P* value by *n* (i.e. for 20 tests, P=0.01 becomes P=0.20!)
- *Holm correction:* less stringent, default in R; preferable!

Pairwise *t*-tests, Holm correction

- So let's run the pairwise *t*-tests:

```
> pairwise.t.test(SBR2$head, SBR2$year)
Pairwise comparisons using t tests with pooled SD
data:  SBR2$head and SBR2$year
          2002     2003    2004
2003    4.7e-12
2004    0.00039 0.00181
2005    0.00079 0.00120 0.81952

P value adjustment method: holm
```

```
> tapply(SBR2$head, SBR2$year, mean, na.rm=T)
    2002     2003     2004     2005
35.23045 35.31483 35.27654 35.27393
```

- Conclusion: 2004-2005年的统计数据相似
  - years 2004-2005 are statistically similar
  - significant differences are caused by years 2002 and 2003

显著差异是由2002和2003年造成的

Break

不同的方差

Different variances: *oneway* function

用单项检验

- What if group variances are not similar?
  - =run **oneway test,** an ANOVA that does not assume equal variances
  - if group means differ, run pairwise tests *not assuming equal variances*

如果组均值不同，用不假设方差相等的成对检验

- (**note**: *oneway* test returns inflated P values; use it only when necessary)

Different variances: *oneway* function

- Example: in 2005, did head circumference in boys differ by delivery type?
  - file *SBR3* (boys born in 2005)

- Let's compare mean and variance of head circumference by delivery type:

```
> tapply(SBR3$head, SBR3b$delivery, mean, na.rm=T)
Caesarian Instrumen   Natural
35.26178   35.62977   35.23473
> tapply(SBR3$head, SBR3$delivery, var, na.rm=T)
Caesarian Instrumen   Natural
5.182391  3.204711   2.697677
```

- Mean values are roughly similar, but variance is higher in caesarean group

Different variances

- Testing for differences in variance:

```
> bartlett.test(SBR3$head ~ SBR3$delivery)

        Bartlett test of homogeneity of variances

data:  SBR3$head by SBR3$delivery
Bartlett's K-squared = 1717.726, df = 2, p-value < 2.2e-16
```

- Result: *P*~0
  - significant differences in variance across groups
  - we should not run *anova(lm)*; we must run ANOVA with the *oneway*() function

Running *oneway*()

- Testing for differences in mean head circumference:

```
> oneway.test(SBR3$head~ SBR3$delivery)

        One-way analysis of means (not assuming equal variances)

data:  SBR3$head and SBR3$delivery
F = 94.0469, num df = 2.000, denom df = 9208.907, p-value < 2.2e-16
```

- Null hypothesis: means of all groups are equal (F=1)
  - F= 94; P-value ~ 0
  - →null hypothesis rejected
  - Conclusion: differences across delivery types are significant

- So which delivery type(s) cause(s) differences?

Pairwise tests

我们必须进行成对检验，而不是假设方差相等

- We must run pairwise tests *not assuming equal variances*:
  - add argument ***pool.sd=F*** (i.e. no pooling of group variances)

```
> pairwise.t.test(SBR3$head, SBR3$delivery, pool.sd=F)
        Pairwise comparisons using t tests with non-pooled SD
data:  SBR3$head and SBR3$delivery
          Caesarian  Instrumen
Instrumen  <2e-16        -
Natural      0.26     <2e-16
```

- Conclusion: mean head circumference from instrumental delivery differs from the other two methods
  结论：平均值不同

Caesarian	Instrumen	Natural
35.26178	35.62977	35.23473

- We may conclude that instrumentally delivered boys had larger heads
  - larger-headed babies more likely to require instrumental delivery

非参数替代

Kruskal-Wallis test: Non-parametric alternative

- As with t-tests, we should not run ANOVA (either *anova(lm) or oneway test*) when… 至少有一个组不正态分布或样本量较小
  - distribution of at least one group **is not normal**
  - **sample sizes are small**
    - because you cannot demonstrate that the groups are normally distributed

- *Kruskal-Wallis test* is the non-parametric alternative to ANOVA

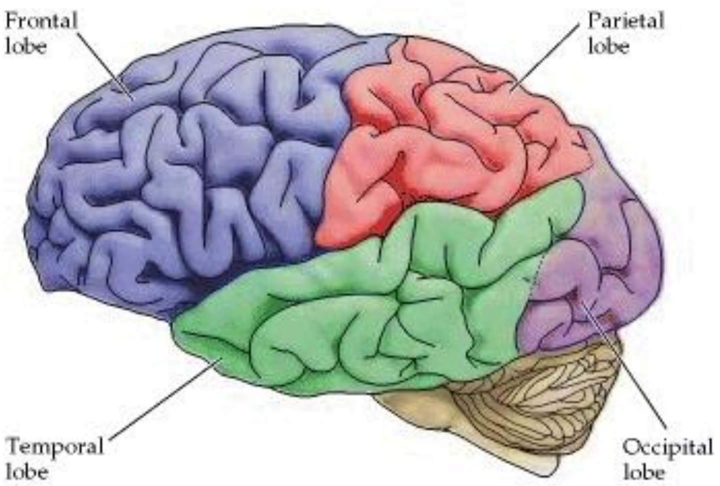
- As the Wilcoxon tests, K-W test is a rank test that calculates between-group squared sums from average ranks rather than original values

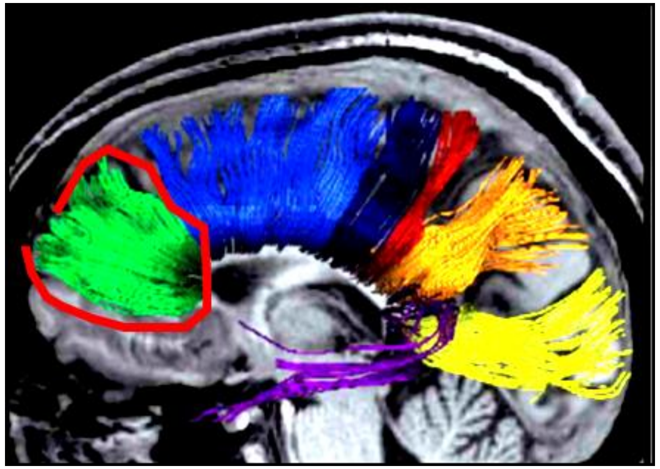
Syntax:

```
> kruskal.test(variable ~ grouping variable)
```

Example: prefrontal cortex size

- Neuroanatomists have argued that human high cognitive abilities are associated with an enlarged *prefrontal cortex* relative to other primates





均值比较

Comparison of means

- So: is the human prefrontal cortex (PFC) larger than in other primates?
  - File: *brain*
- First let's look at PFC size as % of total cerebral cortex (variable *PrebyT*, prefrontal divided by total brain size) across four groups:

```
> tapply(brain$PrebyT, brain$group, mean, na.rm=T)
       ape       Homo       NewW       OldW
0.10193663 0.12721216 0.08929871 0.08236484
```

- It seems PFC is larger in humans (~12.7% of total cerebral cortex)

- Let's test for differences in variances

```
> bartlett.test(brain$PrebyT ~ brain$group)
        Bartlett test of homogeneity of variances
data:  brain$PrebyT by brain$group
Bartlett's K-squared = 1.3772, df = 3, p-value = 0.7109
```

- Conclusion: apparently no significant difference in variance across groups

小样本

Small samples

- But look at sample sizes:

```
> summary(brain$group)
ape   Homo   NewW   OldW
 18     12      8      9
```

- Small sample size is probably the reason Bartlett test returned a high *P*-value
- **Conclusion: do not run Bartlett test or ANOVA when sample size is small**

当样本量较小，不要用bartlett或者ANOVA

- It is safer to run a Kruskal-Wallis test

```
> kruskal.test(brain$PrebyT ~ brain$group)

        Kruskal-Wallis rank sum test

data:  brain$PrebyT by brain$group
Kruskal-Wallis chi-squared = 28.337, df = 3, p-value = 3.086e-06
```

- Result: there are significant differences across primate groups

非参数成对检验

Non-parametric pairwise tests

- But which groups differ?

- Since samples are small, we run pairwise Wilcoxon tests (the non-parametric version of t-tests)

```
> pairwise.wilcox.test(brain$PrebyT, brain$group)
        Pairwise comparisons using Wilcoxon rank sum test
data:  brain$PrebyT and brain$group

        ape      Homo     NewW
Homo    0.00031  -        -
NewW    0.12352  7.9.e-05  -
OldW    0.00462  4.1e-05  0.27659

P value adjustment method: holm
```

- Conclusion: humans have larger prefrontal cortex sizes than other primate groups

双向方差分析

Note 1: Two-Way ANOVA

- You may want to simultaneously analyse the effect of two grouping factors

您可能希望同时分析两个分组因素的影响

- For example, you can test at the same time whether newborn head circumference is affected by *year* and *delivery type*:

```
> anova(lm(SBR2$head ~ SBR2$year + SBR2$delivery))
Analysis of Variance Table

Response: SBR2$head
                 Df  Sum Sq   Mean Sq   F value   Pr(>F)
SBR2$year         3     165     55.08    17.213   3.587e-11 ***
SBR2$delivery     2    2568   1283.81   401.197   < 2.2e-16 ***
Residuals    190855 610729      3.20
```

- Result: both year and delivery have an effect
  - but don't forget to run Bartlett tests first 先做bar t let t测试
  - changing order of factors (year+delivery vs. delivery+year) does not change results only *when there are no missing values*
  - You may run ANOVA with interaction year*delivery too (not significant in this case; see code file)

Note 2: Friedman test

双向方差分析的非参数代替方法

- The Friedman test is the non-parametric alternative to two-way ANOVA

- Syntax:

```
> Friedman.test(variable ~ grouping|grouping2, data=datafile)
```

Summary: Selecting your test

- To compare one variable across > 3 groups :   比较一个变量大于三个组时

If samples are large:

- Check if distribution is normal across groups (visual inspection, Shapiro-Wilks test)
- If distribution is normal, check for equality of variances (Bartlett’s test)
  - if variances are similar:
    - anova(lm)
      - don’t forget *as.factor* if needed
    - if group means differ, pairwise t-test with Holm correction
  - if variances differ
    - *oneway()*
    - if group means differ, pairwise t-test not assuming equal variances, Holm correction

If samples are small:

- Kruskal-Wallis test
  - if group means differ, pairwise Wilcoxon tests, Holm correction

Exercise 1

```
> class(SBR2$size)
[1] "integer"
```

deliver是factor

```
> levels(SBR2$deliver)
```

Using the *SBR2* file

1) What type of variable is *size* (birth weight): numeric or factor?

2) Which are the mean newborn sizes by delivery type?

```
> tapply(SBR2$size, SBR2$delivery, mean, na.rm=T)
Caesarian Instrumen   Natural
 3421.072  3651.949  3618.733
```

3) Look at histograms of size for each delivery type: do they look normal?)

```
> table(SBR2$delivery)
```

(tip: use code

> tapply(SBR2$size, SBR2$delivery, Shapiro.test)

```
> tapply(SBR2$size, SBR2$delivery, hist, breaks = seq(0, 8000, 100))
```

出来是三个图，每一种delivery一个图

Which will not work. Why?

shapiro test的样本必须在3-5000之间

Then use this:

> tapply(SBR2$size, SBR2$delivery, function(x) shapiro.test(sample(x, 4999)))

从每一种delivery中随机选5000个size

4) Calculate variance in each delivery group

```
> tapply(SBR2$size, SBR2$delivery, var, na.rm=T)
```

不能用shapiro test，不能准确的知道他到底正态不正态
假设正态做下面的题

5) Are there significant differences in variance across groups?

```
bartlett.test(SBR2$size ~ SBR2$delivery)
oneway.test(SBR2$size ~ SBR2$delivery)
pairwise.t.test(SBR2$size, SBR2$delivery, pool.sd=F)
```

6) Is there a difference in mean newborn size by delivery type? Which test do you need to run?

用非正态和正态结果一样的

7) If so, which groups differ? How do they differ?