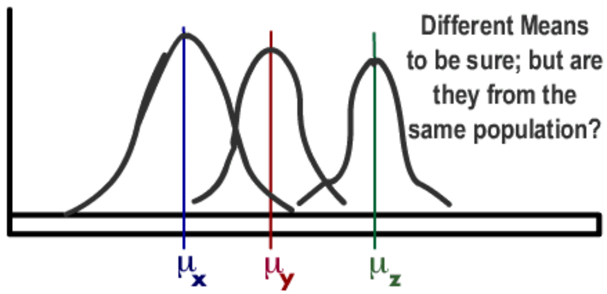
Lecture 6

Analysis of variance (ANOVA)

ANOVA



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

- To compare *three or more* groups, strategy is:
  - to split total variance 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

- This procedure is known as *analysis of variance* (ANOVA)

- Examples of problem addressed by ANOVA:
  - seasonality (grouping variable: month)
  - regional patterns (grouping variable: region)
  - etc

Between- vs. within- group differences

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

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

or

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

Group difference

Individual difference

Example

define a penguin:

- Take a sample of 3000 male heights, divided into 3 nationalities (Dutch, USA, Spanish).
  - General mean = 180 cm
  - USA mean = 178 cm

- Now take a USA individual x_1 = 175 cm;
- This can be written as
  - x_1 = general mean + (USA mean – general mean) + (USA case i – USA mean)
  - = 180 + (178 – 180) + (175 – 178) =
  - = 180 – 2 – 3 = 175 cm

Between- and within-group differences

- If

  S = total sample variation

  (each individual vs. general mean)

  S_b = sum of all squared *between-group* differences

  (Dutch vs. general mean, USA vs. general mean, Spain vs. general mean)

  S_w = sum of all squared *within-group* differences

  (Dutch/US/Spanish individual vs. Dutch/USA/Spanish mean)

- Then:

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

where

- i.e. all variation in sample can be decomposed into *group effect* and *individual effect*

Between- and within-group variances

- It can also be shown if that
  - $M_b = Sb/k\text{-}1$ = between-group variance
  - $M_w = Sw/N\text{-}k$ = within-group variance

- Then, under random sampling into random groups from a normal distribution:

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

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)
- …is similar to expected difference between individuals
  - (difference between two cases from group 1)

- But if grouping is real (i.e. has an effect on variable) then between-group variance should exceed what is expected by chance

F-test

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

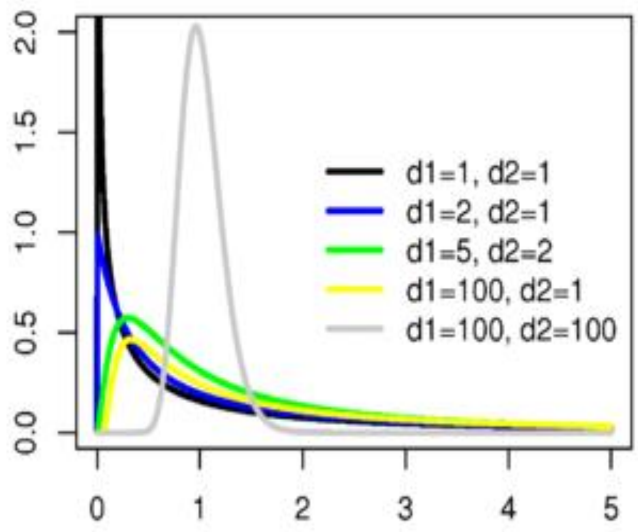
$$F = M_b/M_w$$

- *F*-distribution estimates the probability *P* that between-group variance is significantly different from within-group variance
  - = the probability that their ratio is 1

  Affective factors: sample size and the number of groups

- If ***F* = 1**, group means **do not** differ
  - **null hypothesis: *F*=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*~1



Check first!!!!!!

Equality of variances across groups

- An assumption of ANOVA and *F*-test is that within-group variance is the same for all *k* groups

- We run the ***Bartlett's test*** to check for equality of variances, by comparing observed vs. expected within-group variances from a single normal distribution

    >bartlett.test(variable ~ grouping variable)

- So:

1) run Bartlett's test to check for equality of variances

2) if variances are similar cross groups, run ANOVA using ***anova(lm)***

3) if they differ, run ANOVA using ***oneway*** function

Example: Swedish babies

- Dataset: Swedish Birth Register, with data on all births in Sweden 1982-2005
- Variables:
  - birth year
  - birth weight (variable '*size*')
  - head circumference
  - maternal height
  - pregnancy duration
  - delivery type (natural, caesarean, instrumental)

Head circumference, 2002-05

- Does head circumference in boys change between 2002-05?
  - file *SBR2*, boys from 2002-2005

- Sample sizes are large:

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

- Means look very similar across years:

```
> 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:

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

Variances

- 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
```

- P=0.29 P > 0.05
- →accept null hypothesis
- Conclusion: no differences in variance across years

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

With *anova(lm)*, grouping variable = 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)

- To solve the problem,
  - use function *class()* to check whether your variable is numeric or a factor
  - use function *as.factor()* to force *R* to read numeric variable as a factor – see next example
  - if grouping variable is NOT numeric (i.e. month as Jan, Feb, Mar; or species as human, chimp, gorilla) it is already a factor and you don't need to use *as.factor*

Head circumference, 2002-05

- So is head circumference affected by birth year?
- Let's run an ANOVA of head circumference by year

```
> anova(lm(SBR2$head ~ as.factor(SBR2$year)))
Analysis of Variance Table
Response: SBR2$head
                    Df    Sum Sq   Mean Sq   F value   Pr(>F)
SBR2$year            3       165    55.081    17.141   3.987e-11 ***
Residuals       190857    613297     3.213
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Degree of freedom

P value

Between group difference

Within group difference

Sample size of all numbers of groups

- 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

Pairwise *t*-tests, Holm correction

- 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 of groups (2002 vs. 2003, 2002 vs. 2004…) to identify differences

- 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 in 10 coin tosses; but if you try it 1,000 times, you may get 10 heads by chance
    - *P*=0.05 means a chance of 1 in 20 of getting a significant test in *one t*-test; but if you run 20 *t*-tests on the same dataset and variables, probability increases of getting one case of $P<0.05$ increases

- 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, preferable!

Pairwise *t*-tests, Holm correction

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

Each one is p value, corrected by pairwise

```
> 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

- Conclusion:
  - years 2004-2005 are statistically similar
  - significant differences across the four years are caused by years 2002 and 2003
  - (ps. no need to use *as.factor* here; we are running t-tests)

Different variances: *oneway* function

- What if 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*

- (note: *oneway* test returns inflated P values, so 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 must run ANOVA with the *oneway* function, not *anova(lm)*

Factor >>> anova(lm)
Numeric >>> oneway.test( )

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
  - P-value ~ 0
  - →null hypothesis rejected
  - Conclusion: differences across delivery types are significant

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

Pairwise tests

- 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

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

Non-parametric alternative

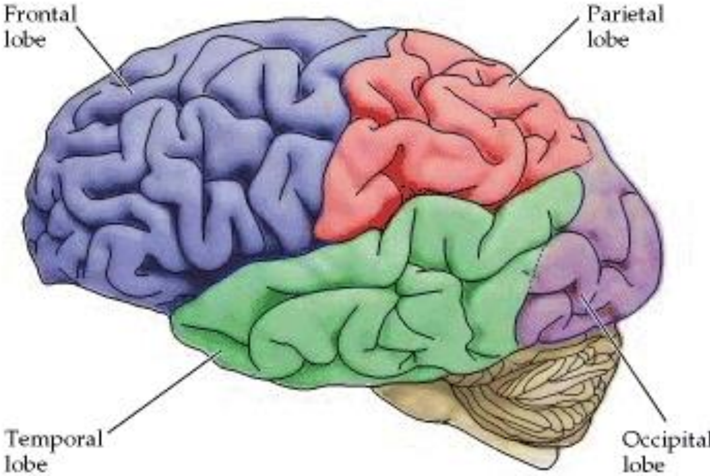
- Second situation where *anova(lm)* should not be used: **when sample sizes are small**
- *Kruskal-Wallis test* is the non-parametric alternative to ANOVA
- K-W test is a rank test that calculates between-group squared sums from average ranks rather than values

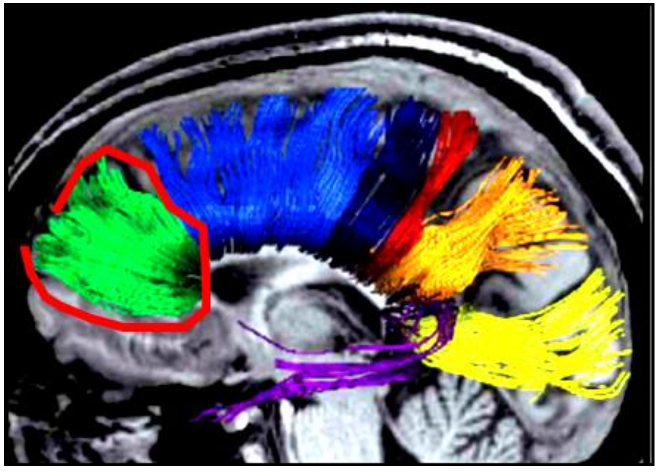
Syntax:

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

Example: prefrontal cortex size

- By comparing brains from different primates, neuroanatomists have argued that human high cognitive abilities are associated with an enlarged *prefrontal cortex*





Example

- So: is the human prefrontal cortex (PFC) enlarged?
  - 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: no significant difference in variance across groups

Example

- You could therefore run *anova(lm)*, but look at sample sizes:

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

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

- It is safer to run a Kruskal-Wallis test

```
> kruskal.test(brain$PrebyT ~ brain$group)
        Kruskal-Wallis rank sum test
data:  Schoenemann$PrebyT by Schoenemann$group
Kruskal-Wallis chi-squared = 28.337, df = 3, p-value = 3.086e-06
```

- Result: significant differences across primate groups

- 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 differ in prefrontal cortex size from the other three groups

Note 1: Two-Way ANOVA

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

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

even tho order is not important, but it relates to how to select out missing data and overlapping data

```
> anova(lm(SBR2$head ~ as.factor(SBR2$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 Two-way ANOVA identifies the two separate effects
(year and delivery type)
```

- Result: both year and delivery have an effect
  - but don't forget to run Bartlett tests first
  - changing order of factors (year+delivery vs. delivery+year) does not change results *when there are no missing values.*

Note 2: Friedman test

Not often seen

- 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 :

If samples are large:

- 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

- Note: Bonferroni correction is very radical! You may want to try p.adj="holm" (Holm correction)

Exercise 1

Using the *SBR2* file

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

class( )

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

tapply(size, delivery)

3) What are the variances in each delivery group?

Tapply

4) Are there significant differences in variance across groups?

Bartlett Significant

5) Are there differences in mean newborn size by delivery type? Which test do you need to run?

Oneway.test( )

6) If so, which groups differ?

Pairwise All significant

Exercise 2

Fake.trypsin file (ISwR library)

nu fac

1) Which type of variable is grp? And grpf?

2) We want to know if there are differences in serum levels of trypsin across groups. Which test do we need to use?

Kruskal

3) Are there differences? Describe the patterns