SIMPLE PARAMETRIC TESTS

Erik Kusch

erik.kusch@uni-leipzig.de

Behavioural Ecology Research Group
University of Leipzig

26/06/2019

Introduction

Parametric test are those statistical approaches which rely on **assumptions** about the parameters which define a population.

Prominent parametric tests include:

- Pearson correlation (Seminar 9 - Correlation Tests)
- **t-Test**
- **An**alysis **O**f **Va**riance (ANOVA)
- Linear regression
- Multivariate extensions of parametric methods
- ...

Terminology

A reminder about the distinction of parametric and non-parametric tests (taken from Seminar 6):

Non-Parametric Tests

- Less *restrictive*
- Make *little to no assumptions*
- Often a **black box**
- Require *more data*

Parametric Tests

- More *restrictive*
- Make *strict assumptions*
- **Easy to interpret**
- Require *less data*

$\rightarrow$ Parametric tests are **numerous**!

Purpose And Assumptions

t-Test (unpaired)

`t.test(..., paired = FALSE)` in base `R`

Purpose: To identify whether groups of variable values are different from one another.

H_0 *There is no difference in characteristics of the response variable values in dependence of the classes of the predictor variable.*

Assumptions:

- Predictor variable is binary
- Response variable is metric and **normal distributed** within their groups
- Variable values are **independent** (not paired)

$\rightarrow$ Test whether variance of response variable values in groups are equal (`var.test()`) and adjust `t.test()` argument `var.equal` accordingly.

Minimal Working Example

Let's feed data to our `t.test(..., paired = FALSE)` function that holds two groups with clearly differing means:

```
data <- c(rnorm(10, 5, 1), rnorm(10, 10, 1))
factors <- as.factor(rep(c("A", "B"), each = 10))
t.test(data ~ factors, paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  data by factors
## t = -10, df = 20, p-value = 3e-11
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.0 -4.4
## sample estimates:
## mean in group A mean in group B
##             5.1            10.3
```

The output above tells us that the means of our two groups are significantly different.

Purpose And Assumptions

t-Test (paired)

`t.test(..., paired = TRUE)` in base R

Purpose: To identify whether groups of variable values are different from one another.

H_0 *There is no difference in characteristics of the response variable values in dependence of the classes of the predictor variable.*

Assumptions:

- Predictor variable is binary
- Response variable is metric
- *Difference of response variable pairs* is **normal distributed**
- Variable values are **dependent** (paired)

Minimal Working Example

Let's feed data to our `t.test(..., paired = TRUE)` function that holds two connected groups with clearly differing means:

```
data <- c(rnorm(10, 5, 1), rnorm(10, 10, 1))
factors <- as.factor(rep(c("A", "B"), each = 10))
t.test(data ~ factors, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  data by factors
## t = -10, df = 9, p-value = 5e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.9 -4.1
## sample estimates:
## mean of the differences
##                      -5
```

The output above tells us that the means of our two connected groups are significantly different.

Introduction to ANOVA

ANOVAs are used to test whether there is a difference between groups of variable values.

There are **multiple versions of ANOVAs**:

- One-way ANOVA (one predictor variable)
- Two-Way ANOVA (multiple predictor variables)
- MANOVA (multivariate ANOVA/multiple response variables)
- ANCOVA (categorical and continuous predictor variables)
- MANCOVA (multivariate ANCOVA)

Data for ANOVA

We will use the `crabs` data set from the `MASS` package

```r
library(MASS)
data(crabs)
head(crabs)
```

```
##   sp sex index   FL  RW CL CW  BD
## 1  B   M     1  8.1 6.7 16 19 7.0
## 2  B   M     2  8.8 7.7 18 21 7.4
## 3  B   M     3  9.2 7.8 19 22 7.7
## 4  B   M     4  9.6 7.9 20 23 8.2
## 5  B   M     5  9.8 8.0 20 23 8.2
## 6  B   M     6 10.8 9.0 23 26 9.8
```

Purpose And Assumptions

One-Way ANOVA

`anova()` in base `R`

Purpose: To explain the variance of a continuous response variable in relation to one predictor variables.

H_0 *Variance of response variable values is equal between levels of predictor variable.*

Assumptions:

- Predictor variable is categorical
- Response variable is metric
- *Response variable residuals* are **normal distributed**
- Variance of populations/samples are equal (**homogeneity**)
- Variable values are **independent** (not paired)

Minimal Working Example

Let's test whether body depth (BD) of crabs are varying when grouped by sex:

```r
OneWay <- with(crabs, lm(BD ~ sex))
anova(OneWay)
```

```
## Analysis of Variance Table
##
## Response: BD
##            Df Sum Sq Mean Sq F value Pr(>F)
## sex         1     19    18.8    1.61   0.21
## Residuals 198   2315    11.7
```

As we can see, sex does not make for a statistically significant predictor of crab body depth.

Take note that we do not deal with testing the assumptions here.

Purpose And Assumptions

Two-Way ANOVA

`anova()` in base `R`

Purpose: To explain the variance of a continuous response variable in relation to multiple predictor variables.

H_0 *Variance of response variable values is equal between levels of predictor variables.*

Assumptions:

- Predictor variables are categorical
- Response variable is metric
- *Response variable residuals* are **normal distributed**
- Variance of populations/samples are equal (**homogeneity**)
- Variable values are **independent** (not paired)

Minimal Working Example

Let's test whether body depth (BD) of crabs are varying when grouped by sex and species as well as their interaction:

```
TwoWay <- with(crabs, lm(BD ~ sex * sp))
anova(TwoWay)
```

```
## Analysis of Variance Table
##
## Response: BD
##            Df Sum Sq Mean Sq F value  Pr(>F)
## sex         1     19      19    1.99   0.160
## sp          1    419     419   44.31 2.8e-10 ***
## sex:sp      1     42      42    4.48   0.035 *
## Residuals 196   1854       9
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The output above tells us that species and the interaction effect of sex and species are meaningful for understanding body depth of crabs.

Purpose And Assumptions

ANCOVA

`anova()` in base `R`

Purpose: To explain the variance of a continuous response variable in relation to mixed (continuous and categorical) predictor variables.

H_0 *Adjusted variance and means of response variable values is equal between levels of predictor variables.*

Assumptions:

- Predictor variables are categorical or continuous
- Response variable is metric
- *Response variable residuals* are **normal distributed**
- Variance of populations/samples are equal (**homogeneity**)
- Variable values are **independent** (not paired)
- Relationship between the response and covariate is linear.

Minimal Working Example

Let's test whether body depth (BD) of crabs are varying when grouped by species and the carapace width as a covariate:

```
Ancova <- with(crabs, lm(BD ~ sp * CW))
anova(Ancova)
```

```
## Analysis of Variance Table
##
## Response: BD
##            Df Sum Sq Mean Sq F value   Pr(>F)
## sp          1    419     419  2481.2 < 2e-16 ***
## CW          1   1880    1880 11130.5 < 2e-16 ***
## sp:CW       1      2       2    12.4 0.00054 ***
## Residuals 196     33       0
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The output above tells us that all of our model coefficients are significant.

Variables We Can Use

Response variables (metric)

- Weight
- Height
- Wing Chord
- Nesting Height
- Number of Eggs
- Egg Weight

Predictor variables (categorical)

- Sex (binary)
- Climate (binary)
- Climate (3 levels - Continental, Semi-Coastal, Coastal)
- Home Range (3 levels - Small, Medium, Large)
- Site Index (11 levels)
- Predator Presence/Type (3 levels - Avian vs. Non-Avian vs. None)

Research Questions And Hypotheses

So which of our major research questions (seminar 6) can we answer?

unpaired t-Test

- *Climate Warming/Extremes:* Does sparrow morphology change depend on climate?
- *Sexual Dimorphism:* Does sparrow morphology change depend on Sex?

paired t-Test (suppose a resettling program)

- *Climate Warming/Extremes:* Does sparrow morphology change depend on climate?

One-Way ANOVA

- *Climate Warming/Extremes:* Does sparrow morphology depend on climate?
- *Predation:* Does nesting height depend on predator characteristics?
- *Site-wise variation:* Does sparrow morphology depend on sites?

Two-Way ANOVA

- *Climate Warming/Extremes:* Does sparrow morphology depend on climate and sex?
- *Sexual Dimorphism:* Does sparrow morphology depend on population status and sex?

ANCOVA

- *Climate Warming/Extremes:* Do sparrow characteristics depend on climate and latitude?

*Remember to **check assumptions** and find ways to circumvent violations of assumptions!*