The Cognitive Discomfort of Statistical Thinking:

Statistics is conditional, not absolute:

Statistical conclusions describe evidence given ASSUMPTIONS, not biological truth. This conditional logic—models, sampling, and assumptions—feels cognitively unfamiliar and often uncomfortable..

Non-intuitive concepts of statistical error in statistical inference

Type I and Type II errors describe how a decision rule behaves across many hypothetical repetitions under uncertainty, relative to an unseen truth. Statistics therefore does not tell us whether this result is right or wrong, but how risky our decisions would be if we kept applying the same method.

1

Confronting key statistical assumptions

1) The issue of normality (last lecture):

2) The assumption of homogeneity of variances (i.e., heteroscedasticity):

Inference based on standard methods (e.g., p-values from ANOVA or regression) typically assumes homoscedasticity.

When the assumption of equal variances is violated, robust methods like Welch's ANOVA, weighted least squares, mixed models offer more reliable inference by accommodating heteroscedasticity rather than ignoring it.

And parametric test are generally more robust to heteroscedasticity than traditional parametric methods (like OLS), but they are not entirely immune to it.

2

Variance (interesting) / heteroscedasticity (has a bad connotation)

Although ecological data often exhibit non- constant variance, this variation is commonly considered a mere nuisance that violates the model's assumption of homogeneity (i.e. homoscedasticity).

In reality, patterns in variance or heteroscedasticity, can signal ecological, evolutionary and environmental processes.

For example, environmental stress (e.g. temperature increases) can not only change the mean but can also generate more variance in organismal responses.

More than a decade ago, Cleasby and Nakagawa (2011) surveyed and reported that over 95% of published studies in behavioural ecology ignored heteroscedasticity.

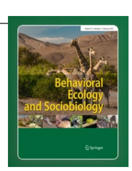
Behav Ecol Sociobiol (2011) 65:2361–2372
DOI 10.1007/s00265-011-1254-7

METHODS

Neglected biological patterns in the residuals

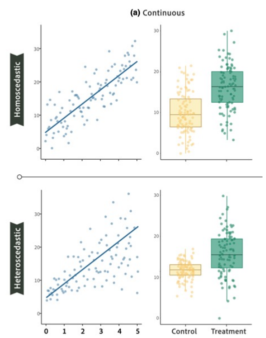
A behavioural ecologist's guide to co-operating with heteroscedasticity

Ian R. Cleasby · Shinichi Nakagawa



3

Variance (interesting) / heteroscedasticity (has a bad connotation)

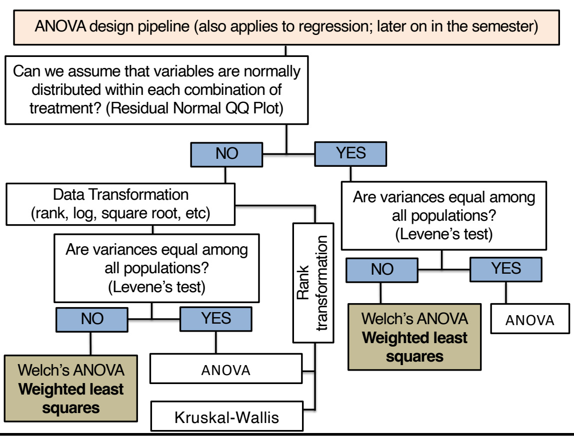


DOI: 10.1111/2041-210x.70203
RESEARCH ARTICLE
Methods in Ecology and Evolution

Location–scale models in ecology and evolution: Heteroscedasticity in continuous, count and proportion data

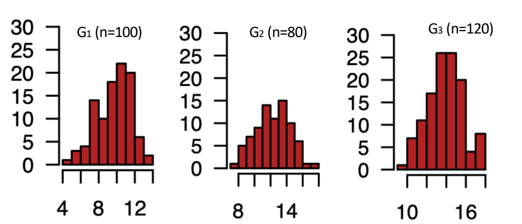
Shinichi Nakagawa[1,2] | Santiago Ortega[1] | Elena Gazzea[3] | Malgorzata Lagisz[1,2] | Anna Lenz[1] | Erick Lundgren[1] | Ayumi Mizuno[1]

4



5

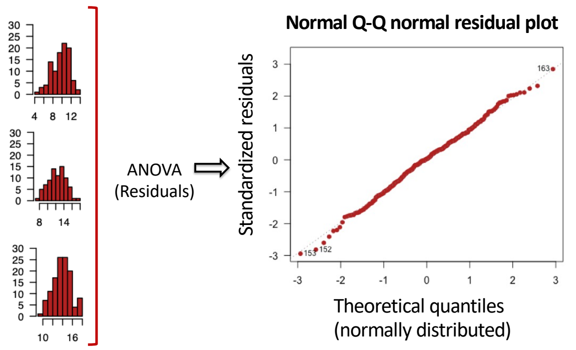
ANOVA design pipeline – we start with simulated data that are normally distributed and homoscedastic, so that the standard ANOVA assumptions are satisfied.



One-factorial design - 3 groups, normally distributed homoscedastic data ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 4$), varying in means ($\mu_1^2 = 10, \mu_2^2 = 12, \mu_3^2 = 14$)

6

[1] – Can we assume that variables are normally distributed within each combination of treatment? (Residual Normal Q-Q Plot)

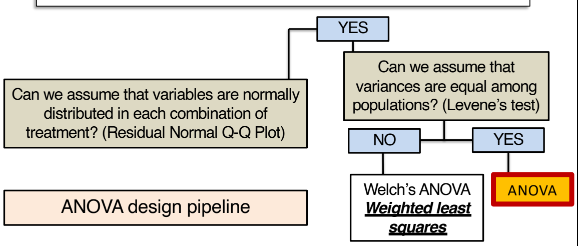


$Y = Factor(G1, G2) + residuals$

7

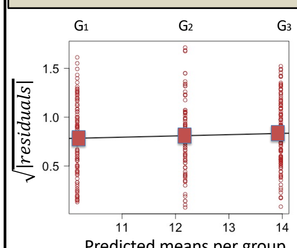
[2] – Can we assume homoscedasticity among populations?
Levene's test indicates that this assumption is met, which is unsurprising given that the data were simulated under this assumption.

```
> leveneTest(values ~ as.factor(groups))

Levene's Test for Homogeneity of Variance (center = median)

        Df  F value  Pr(>F)
group    2    0.223   0.8003
       297
```



8

A graphical representation of residual variation among groups—
a diagnostic alternative to Levene's test that can also be applied to regression models.



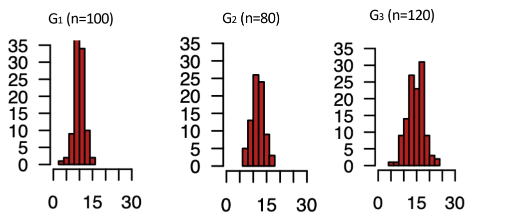
Predicted means per group (squared values) and residuals (observed value – predicted mean; small circles)

A diagnostic plot of the square root of the absolute ANOVA residuals (i.e., deviations from the predicted group means) against the predicted group means should show no systematic pattern—ideally forming a straight line that reflects constant variance, which is the case here.

The Breusch–Pagan test can then be used to formally assess whether any departure from this pattern is statistically significant. We will use this test later to evaluate homoscedasticity in regression models.

9

ANOVA design pipeline – we now turn to simulated data that are normally distributed but heteroscedastic, to illustrate how Weighted Least Squares (WLS) works.

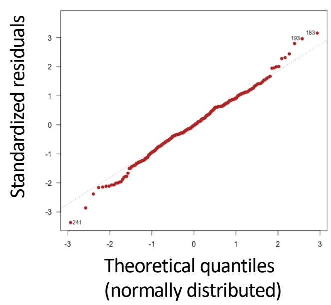


One-factorial design - 3 groups, normally distributed heteroscedastic data ($\sigma_1^2 = 4, \sigma_2^2 = 6.25, \sigma_3^2 = 9$), varying in means ($\mu_1^2 = 10, \mu_2^2 = 12, \mu_3^2 = 14$)

10

[1] – Can we assume that variables are normally distributed within each combination of treatment? (Residual Normal Q-Q Plot)

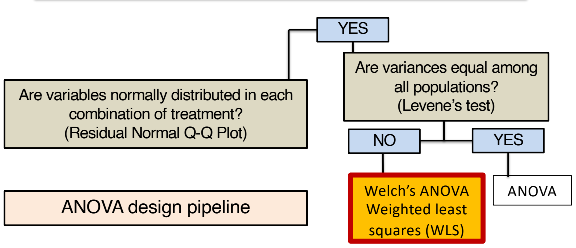
Normal Q-Q normal residual plot



11

Can we assume homoscedasticity among populations? (Levene's test)

```
> leveneTest(values ~ as.factor(groups))

Levene's Test for Homogeneity of Variance (center = median)

       Df  F value   Pr(>F)
group   2   12.295   7.414e-06  ***
      297
```



12

Can we assume homoscedasticity among populations? (Levene's test)

We can understand heteroscedasticity by focusing on how the variance of squared residuals changes among groups.

Note that the average of *all* residuals taken together is always zero.

Square residuals and calculate group residuals variances

$residuals$	$\sqrt{\lvert residuals \rvert}$		
-0.9723056	0.9860556	Group 1	$var(residuals^2)$=0.005018537
-0.8426648	0.9179678		
-0.7130241	0.8444075		
0.1944611	0.4409774	Group 2	$var(residuals^2)$=0.142741
0.9723056	0.9860556		
1.3612278	1.1667167		

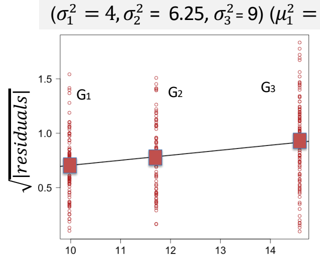
$\sqrt{\lvert residuals \rvert}$=square root of absolute values

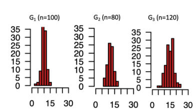
Using here a "tiny" number of observations for demonstration purposes

13

A graphical representation of residual variation among groups— a diagnostic alternative to Levene's test that can also be applied to regression models.

$(\sigma_1^2 = 4, \sigma_2^2 = 6.25, \sigma_3^2 = 9)$ $(\mu_1^2 = 10, \mu_2^2 = 12, \mu_3^2 = 14)$



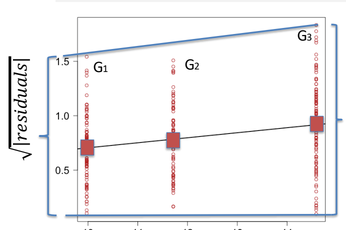


Predicted means per group (squared values) and residuals (observed value – predicted mean; small circles)

14

Can we assume that variances are equal among populations? (alternative to the Levene's test and they way to understand WLS)

$(\sigma_1^2 = 4, \sigma_2^2 = 6.25, \sigma_3^2 = 9)$ $(\mu_1^2 = 10, \mu_2^2 = 12, \mu_3^2 = 14)$



Predicted means per group (squared values) and residuals (observed value – predicted mean; small circles)

If variance were constant, the plot of the square root of the absolute residuals versus predicted group means would appear flat. Instead, the clear pattern here reveals heteroscedasticity.

15

ANOVA is a regression model, i.e., they have predicted values (means) and residuals (observed – means)!

They differ in "design" but not in calculations!

16

Welch's ANOVA (sometimes covered in Intro BioStatistics courses) can handle heteroscedasticity but is limited to single-factor ANOVA designs.

This sets up **Weighted Least Squares** nicely as the general solution when:
you have multiple factors, interactions, or want a unified framework that extends naturally to regression and GLMs.

Two guiding questions:

How does heteroscedasticity influence standard ANOVA?

How can we use Weighted Least Squares (WLS) to handle heteroscedasticity, both for raw data and rank-based ANOVA?

17

ANOVA is a special case of a regression model in which the response variable is continuous and the predictors are categorical, with the categories encoded so that the problem can be treated as a regression.

Let's use a tiny fictional example with 2 groups (control & Group_1)

Response	Factor (predictor)
1.2	control
2.7	control
3.1	control
4.1	Group_1
5.3	Group_1
6.1	Group_1

18

ANOVA as a regression model

Response	Factor (predictor)	Contrast
1.2	control	0
2.7	control	0
3.1	control	0
4.1	Group_1	1
5.3	Group_1	1
6.1	Group_1	1

By coding categorical predictors as contrasts, ANOVA can be treated as a regression problem. Importantly, this does not change the results; the ANOVA derived from the regression model is exactly the same as the standard ANOVA.

19

ANOVA is a regression model where the response variable is continuous and the predictors are categorical.

A tiny example:

```
groups <- c("control", "control", "control", "Group_1", "Group_1", "Group_1")
values <- c(1.2, 2.7, 3.3, 3.1, 4.5, 6.1)
```

Running ANOVA using the R function *aov*:

```
> summary(aov(values~groups))

          Df Sum Sq Mean Sq F value Pr(>F)
groups     1 12.042  12.042   11.94 0.0259 *
Residuals  4  4.033   1.008
```

20

ANOVA is a regression model where the response variable is continuous and the predictors are categorical.

A tiny example:

```
groups <- c("control", "control", "control", "Group_1", "Group_1", "Group_1")
values <- c(1.2, 2.7, 3.3, 3.1, 4.5, 6.1)
```

Running ANOVA using the R function *aov*:

```
> summary(aov(values~groups))

          Df Sum Sq Mean Sq F value Pr(>F)
groups     1 12.042  12.042   11.94 0.0259 *
Residuals  4  4.033   1.008
```

Running ANOVA using the R function *lm* (linear model = regression) setting group as a ***factor***:

```
> anova(lm(values~factor(groups)))
Analysis of Variance Table
Response: values
               Df  Sum Sq Mean Sq F value  Pr(>F)
factor(groups)  1 12.0417 12.0417  11.942 0.02592 *
Residuals       4  4.0333  1.0083
```

21

Let's quickly revisit a simple regression model from Intro Stats. We'll come back to regression in greater depth later in the Multiple Regression module.

$Y = \beta_0 + \beta_1 X + e$

- e represents the vector of residual values.

$\beta = (X^T X)^{-1} X^T Y$

- Slope and intercept estimated by one single operation via Ordinary Least Squares (OLS).

22

Simple regression model

$Y = \beta_0 + \beta_1 X + e$

- e represents the vector of residual values.

$\beta = (X^T X)^{-1} X^T Y$

- Slope and intercept estimated by one single operation via Ordinary Least Squares (OLS).

$\hat{Y} = \beta_0 + \beta_1 X$

- $\hat{Y}$ is called Y-hat and is a vector containing predicted values.

23

Simple regression model

$Y = \beta_0 + \beta_1 X + e$

- e represents the vector of residual values.

$\beta = (X^T X)^{-1} X^T Y$

- Slope and intercept estimated by one single operation via Ordinary Least Squares (OLS).

$\hat{Y} = \beta_0 + \beta_1 X$

- $\hat{Y}$ is called Y-hat and is a vector containing predicted values.

$e = Y - \hat{Y}$

- e represents the vector of residual values.

24

ANOVA as a regression model

$Y = \beta_0 + \beta_1 X + e$

$\hat{Y} = \beta_0 + \beta_1 X$

$\beta = (X^T X)^{-1} X^T Y$

$\beta_0 = 2.333 \therefore \beta_1 = 2.833$

back to our tiny example

X

Response (Y)	Constant (β_0)	Predictor (β_1)
1.2	1	0
2.7	1	0
3.1	1	0
4.1	1	1
5.3	1	1
6.1	1	1

25

ANOVA as a regression model

$Y = \beta_0 + \beta_1 X + e$

$\hat{Y} = \beta_0 + \beta_1 X$

$\beta = (X^T X)^{-1} X^T Y$

$\beta_0 = 2.333 \therefore \beta_1 = 2.833$

$\hat{Y} = 2.333 + 2.833 X_1$

$e = Y - \hat{Y}$

- $\hat{Y}$ is called Y-hat and represents the vector of predicted values.
- e represents the vector of residual values.

Response (Y)	Constant (β_0)	Predictor X_1 (β_1)	$\hat{Y}$	e
1.2	1	0	2.33	-1.13
2.7	1	0	2.33	0.37
3.1	1	0	2.33	0.77
4.1	1	1	5.17	-1.07
5.3	1	1	5.17	0.13
6.1	1	1	5.17	0.93

26

ANOVA as a regression model

Response (Y)		Constant (β_0)	Predictor (β_1)	$\hat{Y}$	e
1.2		1	0	2.33	-1.13
2.7	$\bar{X}$	1	0	2.33	0.37
3.1		1	0	2.33	0.77
4.1		1	1	5.17	-1.07
5.3	$\bar{X}$	1	1	5.17	0.13
6.1		1	1	5.17	0.93

In ANOVAs, predicted values are the predicted mean values per group

27

ANOVA as a regression model

Response (Y)		Constant (β_0)	Predictor (β_1)	$\hat{Y}$	e
1.2		1	0	2.33	-1.13
2.7	$\bar{X}$	1	0	2.33	0.37
3.1		1	0	2.33	0.77
4.1		1	1	5.17	-1.07
5.3	$\bar{X}$	1	1	5.17	0.13
6.1		1	1	**5.17**	**0.93**

$$e_6 = Y_6 - \hat{Y} = 6.10 - 5.17 = 0.93$$

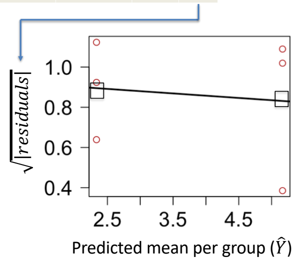
In ANOVAs, predicted values are the predicted mean values per group, and residuals (e) represent variation around the observed group mean not explained by the regression model (or ANOVA).

28

Diagnostic plot of the square root of the absolute ANOVA residuals (deviations from predicted group means) versus the predicted mean per group.

Response (Y)	Constant (β_0)	Predictor (β_1)	$\hat{Y}$	e
1.2	1	0	2.33	-1.13
2.7	1	0	2.33	0.37
3.1	1	0	2.33	0.77
4.1	1	1	5.17	-1.07
5.3	1	1	5.17	0.13
6.1	1	1	5.17	0.93

Residual variance appears reasonable, particularly considering the limited number of replicates per group.



Predicted means per group (squared values) and residuals (observed value – predicted mean; small circles)

29

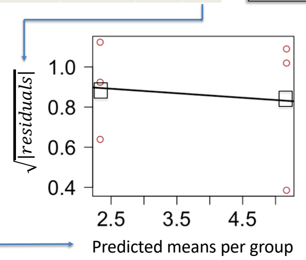
Plot of residuals on predicted values (ANOVA as a regression model) **versus** standard Levene's test for testing for homoscedasticity among groups

Response (Y)	Constant (β_0)	Predictor (β_1)	$\hat{Y}$	e
1.2	1	0	2.33	-1.13
2.7	1	0	2.33	0.37
3.1	1	0	2.33	0.77
4.1	1	1	5.17	-1.07
5.3	1	1	5.17	0.13
6.1	1	1	5.17	0.93

Levene's test

```
      Df F value Pr(>F)
group  1  0.0034 0.9562
       4
```

Variance of residuals are ok!



30

Coding for predictors with 3 groups (more groups and more factors, more predictors)

Response	Factor	Constant (β_0)	Predictor (β_1)	Predictor (β_2)
1.2	control	1	0	0
2.7	control	1	0	0
3.1	control	1	0	0
4.1	Group_1	1	1	0
5.3	Group_1	1	1	0
6.1	Group_1	1	1	0
8.1	Group_2	1	0	1
9.4	Group_2	1	0	1
10.1	Group_2	1	0	1

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$$

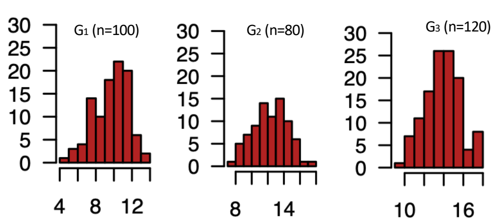
Multifactorial ANOVAs are a special case of multiple regression models

31

How does heteroscedasticity affect ANOVAs?

32

We now return to the simulated, normally distributed, homoscedastic data.

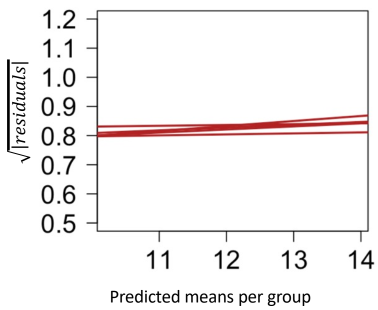


One-factorial design - 3 groups, normally distributed homoscedastic data ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 4$), varying in means ($\mu_1^2 = 10, \mu_2^2 = 12, \mu_3^2 = 14$)

33

One-factorial design - 3 groups, normally distributed homoscedastic data ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 4$), varying in means ($\mu_1^2 = 10, \mu_2^2 = 12, \mu_3^2 = 14$)

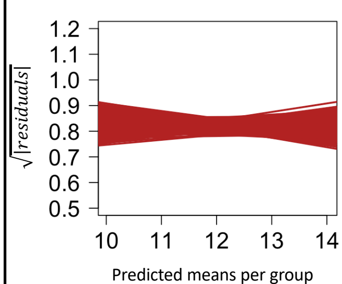
5 simulated data sets



34

One-factorial design - 3 groups, normally distributed homoscedastic data ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 4$), varying in means ($\mu_1^2 = 10, \mu_2^2 = 12, \mu_3^2 = 14$)

1000 simulated data sets



Sample variation in homoscedastic populations -

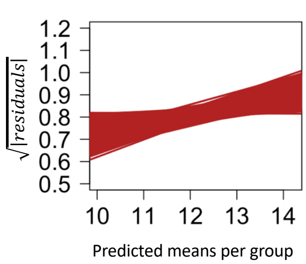
H_0 for population mean (μ) differences is set to **FALSE**.

H_0 for population variance differences (σ) is set to **TRUE**.

35

One-factorial design - 3 groups, normally distributed heteroscedastic data ($\sigma_1^2 = 4, \sigma_2^2 = 6.25, \sigma_3^2 = 9$), varying in means ($\mu_1^2 = 10, \mu_2^2 = 12, \mu_3^2 = 14$)

1000 simulated data sets



Sample variation in heteroscedastic populations -

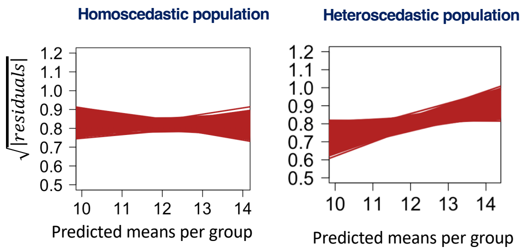
H_0 for population mean (μ) differences is set to **FALSE**.

H_0 for population variance differences (σ) is set to **FALSE**.

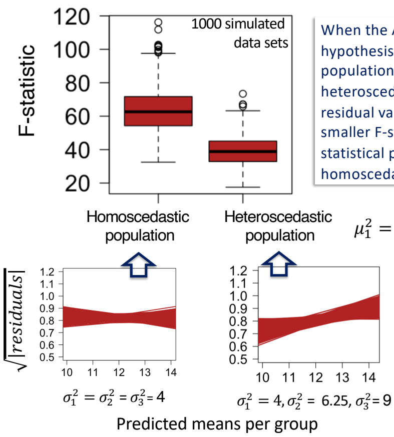
36

One-factorial design - 3 groups, normally distributed homoscedastic data ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 4$) and heteroscedastic data ($\sigma_1^2 = 4, \sigma_2^2 = 6.25, \sigma_3^2 = 9$),
varying in means ($\mu_1^2 = 10, \mu_2^2 = 12, \mu_3^2 = 14$)

1000 simulated data sets



37



When the ANOVA null hypothesis is false (i.e., population means truly differ), heteroscedasticity inflates residual variation, leading to a smaller F-statistic and reduced statistical power compared with homoscedastic populations.

$\mu_1^2 = 10, \mu_2^2 = 12, \mu_3^2 = 14$

38

Under a true ANOVA null hypothesis, heteroscedasticity can artificially increase the F-statistic, causing the test to reject the null hypothesis too often (inflated Type I error). This will be demonstrated in TUTORIAL 5.

AUSTRIAN JOURNAL OF STATISTICS
Volume 36 (2007), Number 3, 179–188

How to keep the Type I Error Rate in ANOVA if Variances are Heteroscedastic

Karl Moder

Institute of Applied Statistics and Computing,
University of Natural Resources and Applied Life Sciences, Vienna

Abstract: One essential prerequisite to ANOVA is homogeneity of variances in underlying populations. Violating this assumption may lead to an increased type I error rate. The reason for this undesirable effect is due to the calculation of the corresponding F-value. A slightly different test statistic keeps the level α. The underlying distribution of this alternative method is Hotelling's T^2. As Hotelling's T^2 can be approximated by a Fisher's F-distribution, this alternative test is very similar to an ordinary analysis of variance.

39

The weighted least squares (WLS) approach

Special case $\beta = (X^{\mathrm{T}}X)^{-1}\, X^{\mathrm{T}}Y\ (OLS)$

General case $\beta = (X^{\mathrm{T}}WX)^{-1}\, X^{\mathrm{T}}WY\ (WLS)$

$W =$

1	0	0	0	0	0
0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0
0	0	0	0	1	0
0	0	0	0	0	1

Ordinary Least Squares (OLS) and Weighted Least Squares (WLS) are equivalent when the weight matrix is the identity matrix (i.e., all diagonal elements equal 1), meaning that all observations receive equal weight in the regression.

40

The weighted least squares (WLS) approach

Let's see how weighting observations changes statistical estimates, beginning with the weighted mean.

$$\frac{1\times2+2\times3+3\times4+4\times5}{14} = 2.86$$

Weighted mean
Weights = 2,3,4,5

$$\frac{1+2+3+4}{4} = 2.5$$

regular mean

$$\frac{1\times5+2\times4+3\times3+4\times2}{14} = 2.14$$

Weighted mean
Weights = 5,4,3,2

41

The weighted least squares (WLS) approach

Let's see how weighting observations changes statistical estimates, beginning with the weighted mean.

$$\frac{1\times2+2\times3+3\times4+4\times5}{14} = 2.86$$

Weighted mean
Weights = 2,3,4,5

$$\frac{1+1+2+2+2+3+3+3+3+4+4+4+4+4}{14} = \frac{40}{14} = 2.86$$

42

The weighted least squares (WLS) approach

Let's see how weighting observations changes statistical estimates, beginning with the weighted mean.

$$\beta = (X^{\mathrm{T}}WX)^{-1}\, X^{\mathrm{T}}WY\ (WLS)$$

Response (Y)	Constant (β_0)	Predictor (β_1)	$\hat{Y}$	e
1.2	1	0	2.33	-1.13
2.7	1	0	2.33	0.37
3.1	1	0	2.33	0.77
4.1	1	1	5.17	-1.07
5.3	1	1	5.17	0.13
6.1	1	1	5.17	0.93

Variance of residuals per group

1.003333

1.013333

43

Weighted Least Squares (WLS) - variability determines influence: more variance means less weight in the estimation.

$$\beta = (X^{\mathrm{T}}X)^{-1}\, X^{\mathrm{T}}Y\ (OLS)$$

$$\beta = (X^{\mathrm{T}}WX)^{-1}\, X^{\mathrm{T}}WY\ (WLS)$$

$$W = 1/s^2_{group}$$

In **OLS**, all observations receive the same weight and therefore contribute equally to the model. In **WLS**, observations are weighted according to their residual variance: observations from groups with smaller residual variance are treated as more *informative* about the underlying relationship between *X* and *Y*, whereas those with larger residual variance contribute less.

44

Weighted Least Squares (WLS) - variability determines influence: more variance means less weight in the estimation.

$$\beta = (X^{\mathrm{T}}WX)^{-1}\, X^{\mathrm{T}}WY\ (WLS)$$

$W = 1/$

0.997	0	0	0	0	0
0	0.997	0	0	0	0
0	0	0.997	0	0	0
0	0	0	0.990	0	0
0	0	0	0	0.990	0
0	0	0	0	0	0.990

1 / Variance of residuals per group

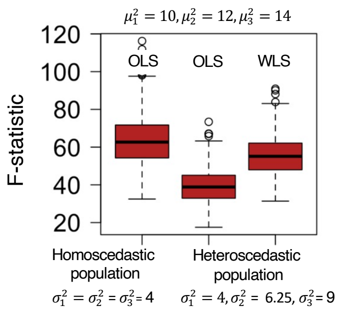
1 / 1.003333

1 / 1.013333

The influence of each observation is the inverse of its group residual variance (i.e., reciprocal, 1/variance)

45

When the ANOVA null hypothesis is false (i.e., population means truly differ), heteroscedasticity inflates residual variation, leading to a smaller F-statistic and reduced statistical power compared with homoscedastic populations. **WLS applied to ANOVA corrects** for unequal variances, leading to larger F-statistics and inference that closely resembles what we would expect if populations were homoscedastic.



46