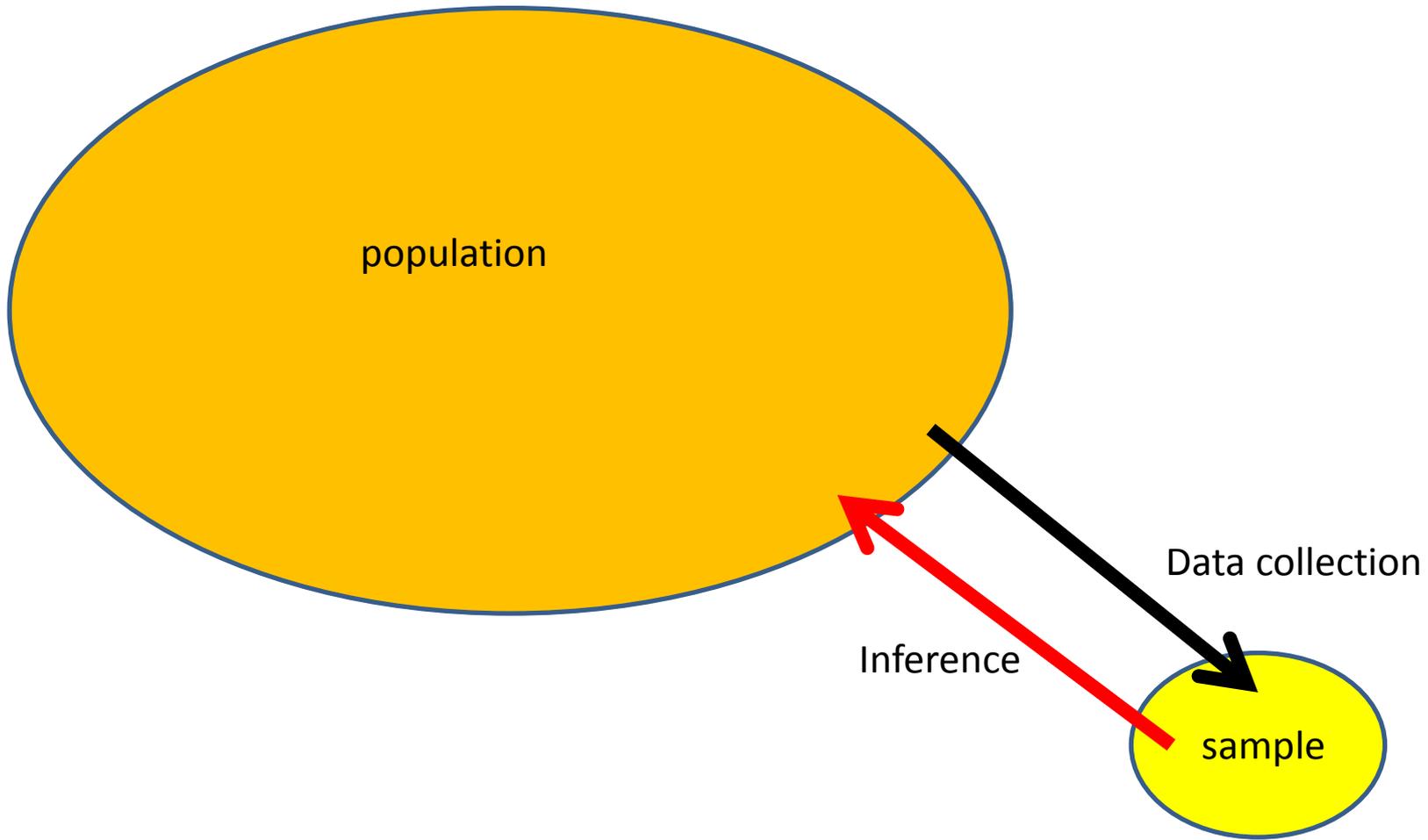
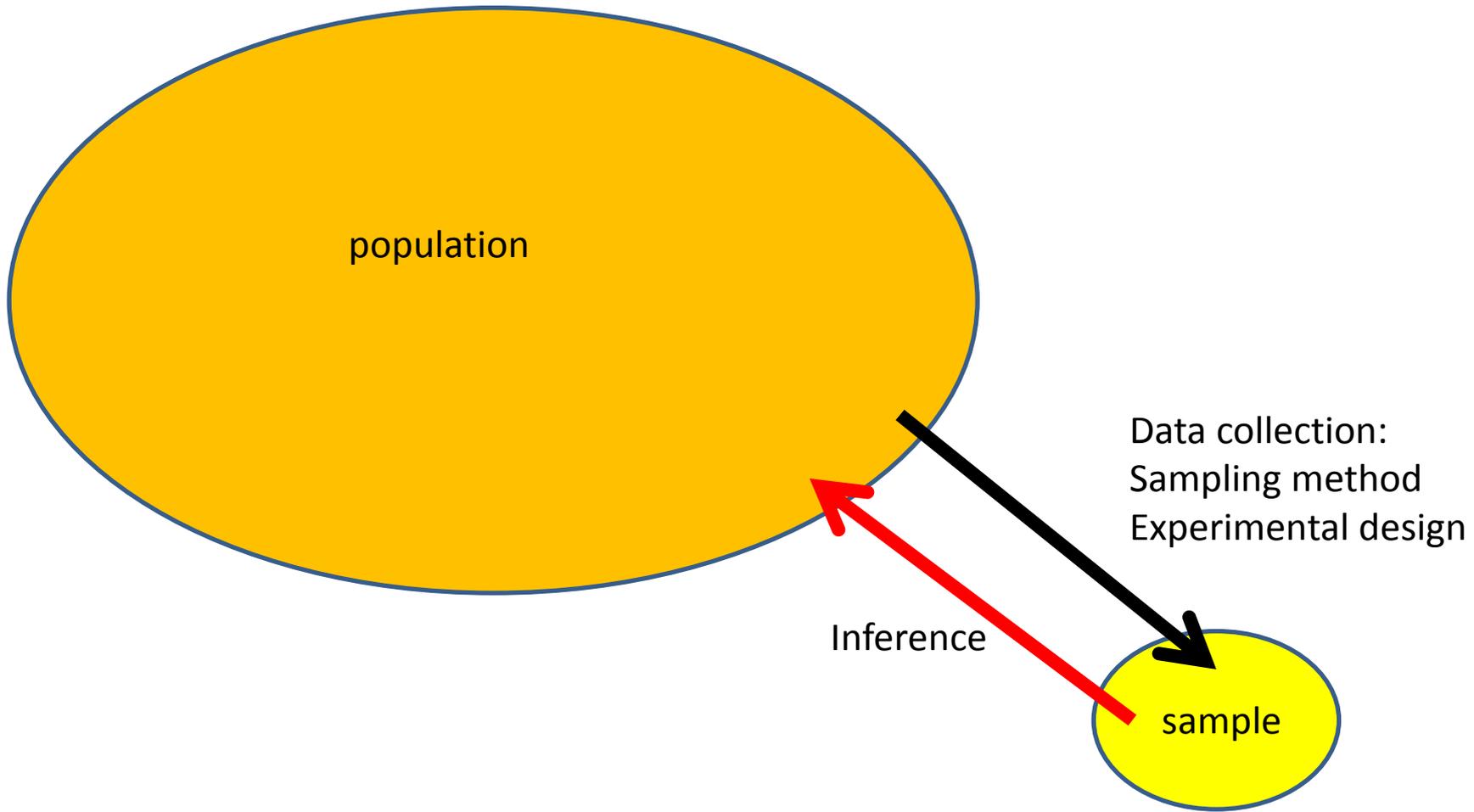


變方分析診斷與補救

**ANOVA Diagnostics and remedial
measures**

黃家康





試驗設計重於統計分析

Assumptions of anova

- Normality
- Homogeneity
- Independence

Residual (殘差)

- $e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}_i.$
- $E(r)=0$
- e_i and \hat{y}_i are independence and normally distributed

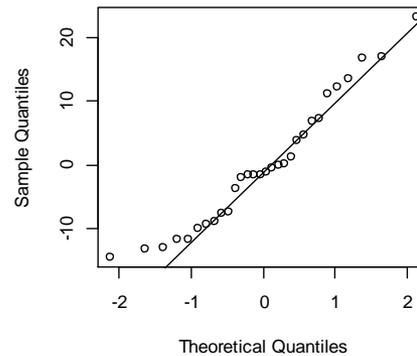
Normality

- Mean and variance are independent.
- Skewness=0
- Kurtosis=3
- The actual significance may be different from the specified one.

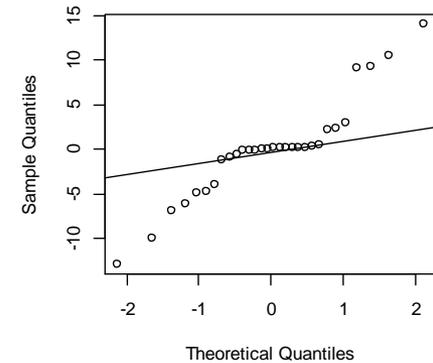
QQplot

```
data1=round(10+rnorm(10,0,10),1)
data2=round(15+rnorm(10,0,10),1)
data3=round(20+rnorm(10,0,10),1)
dataA=cbind(data1,data2,data3)
trt=rep(1:3,each=10)
FA=lm(as.numeric(dataA)~trt)
summary(FA)
qqnorm(FA$resid,main="residuals with equal variance")
qqline(FA$resid)
```

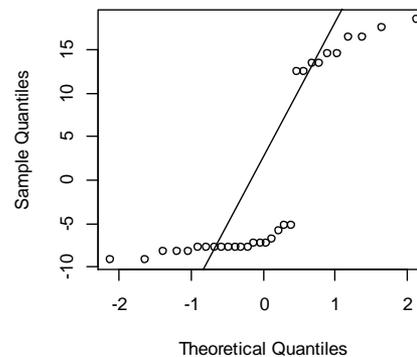
residuals with equal variance



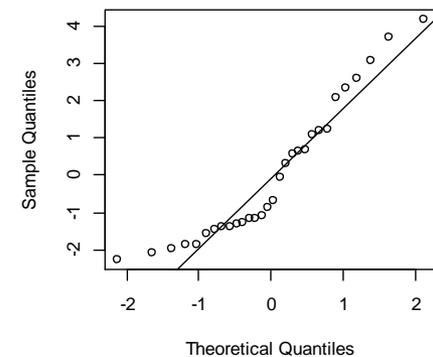
residuals with unequal variance



binomial



chisquare



Normality

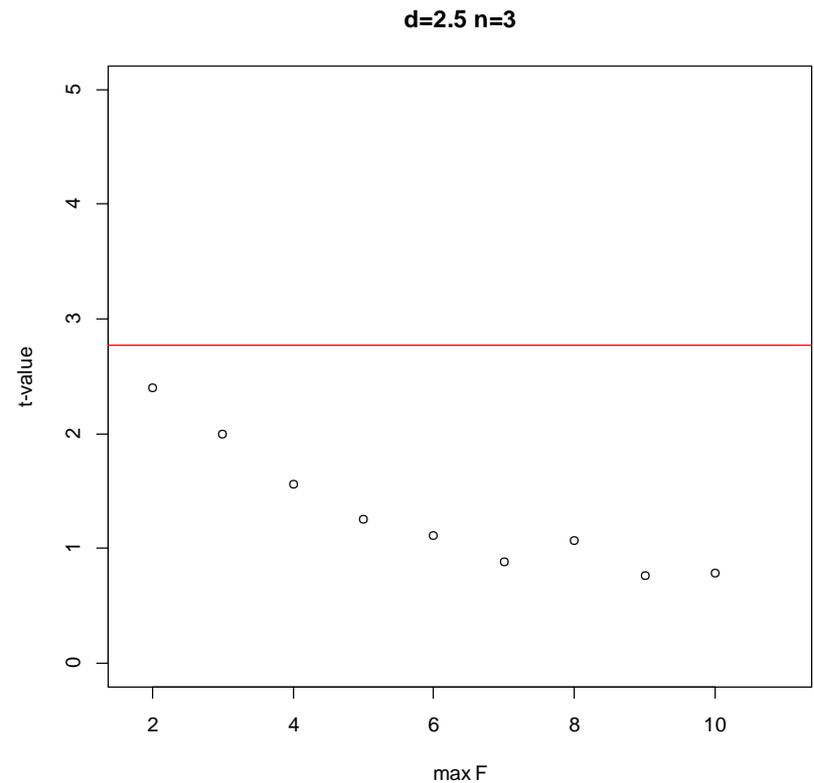
- Shapiro-Wilk
- Kolmogorov-Smirnov
- Cramer-von Mises
- Anderson-Darling

Homogeneity

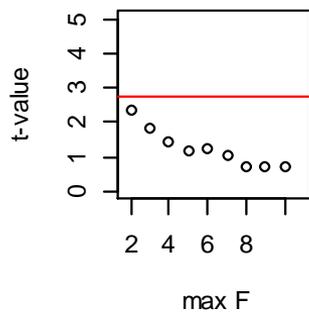
Homogeneity

- Box-plot
- Hartley
- Bartlett
- Box
- Levene

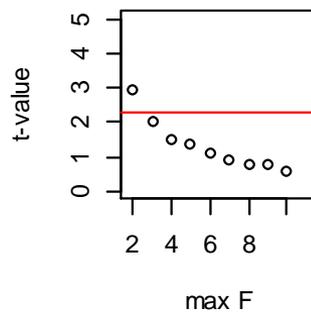
- `n1=3`
- `mu=0`
- `d=2.5`
- `sigma2=1`
- `data=0`
- `t_avg=0`
- `for(i in 1:10){`
- `c=i`
- `for(j in 1:300){`
- `set1=round(rnorm(n1,mu+d,c*sigma2),2)`
- `set2=round(rnorm(n1,mu,sigma2),2)`
- `aa=t.test(set1,set2,var.equal =T)`
- `data[j]=aa[1]`
- `}`
- `t_avg[i]=mean(as.numeric(data))`
- `}`
- `plot(as.numeric(t_avg),main="d=2.5
n=3",xlab="max F",ylab="t-
value",xlim=c(1.75,11),ylim=c(0,5))`
- `abline(h=qt(.975,as.numeric(aa[2])),col="red")`



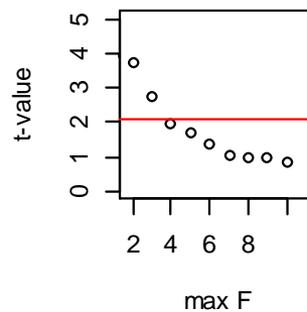
d=2.5 n=3



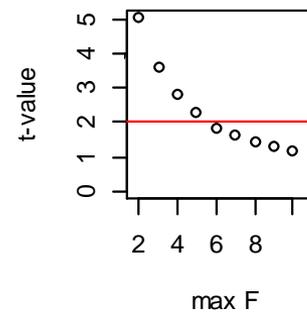
d=2.5 n=5



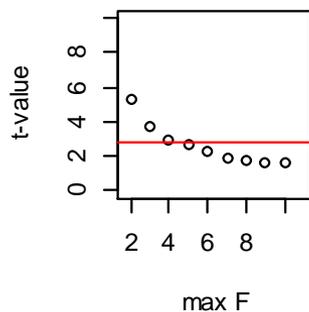
d=2.5 n=10



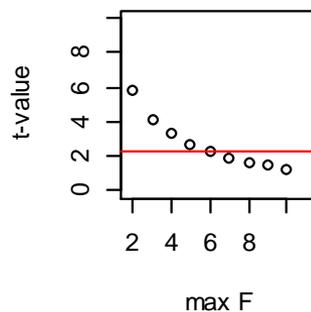
d=2.5 n=20



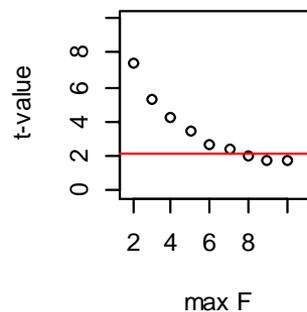
d=5 n=3



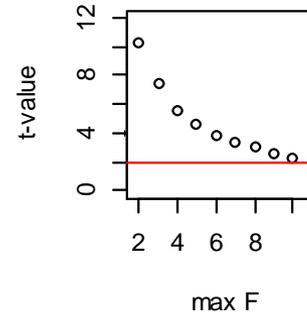
d=5 n=5



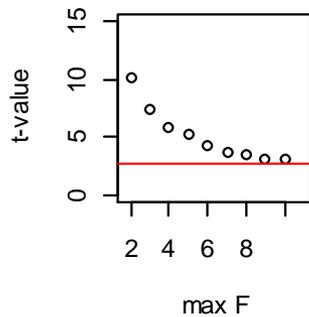
d=5 n=10



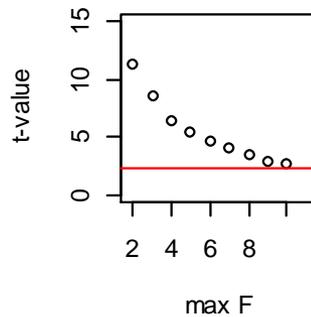
d=5 n=20



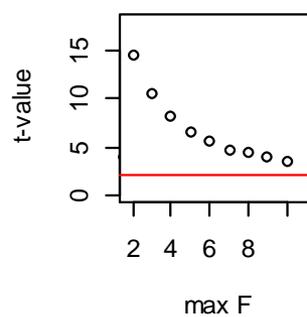
d=10 n=3



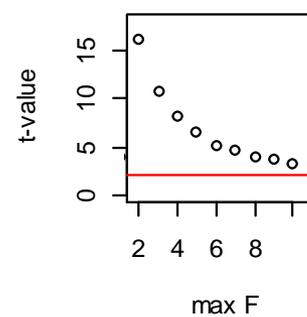
d=10 n=5



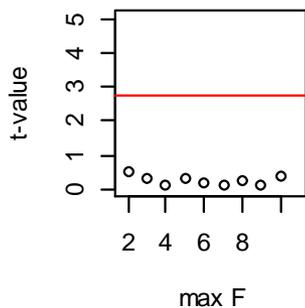
d=10 n=10



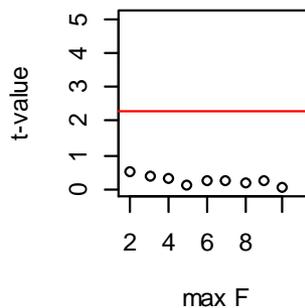
d=10 n=20



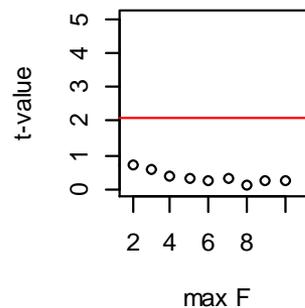
d=2.5 n=3 v=5



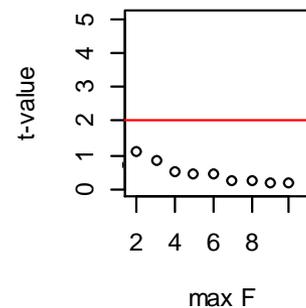
d=2.5 n=5 v=5



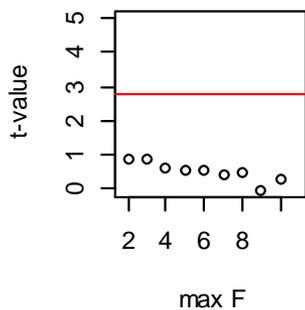
d=2.5 n=10 v=5



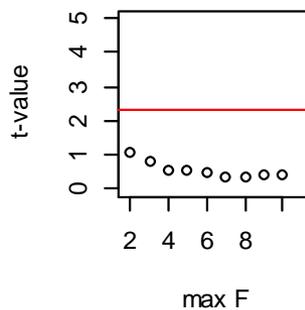
d=2.5 n=20 v=5



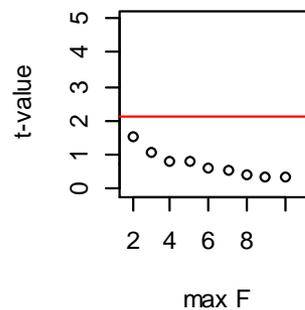
d=5 n=3 v=5



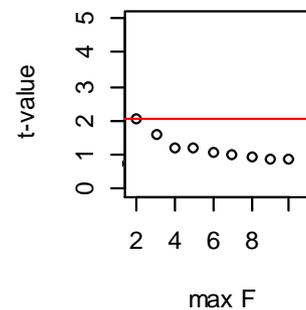
d=5 n=5 v=5



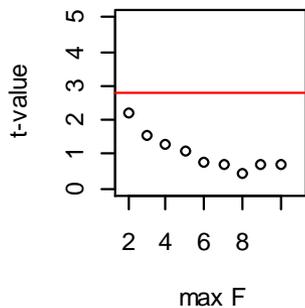
d=5 n=10 v=5



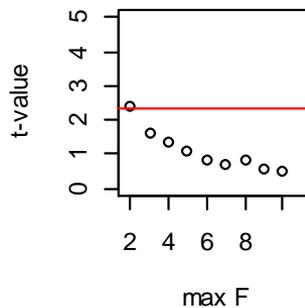
d=5 n=20 v=5



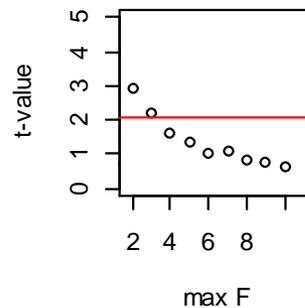
d=10 n=3 v=5



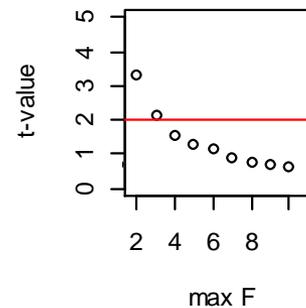
d=10 n=5 v=5



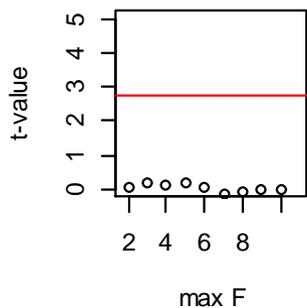
d=10 n=10 v=5



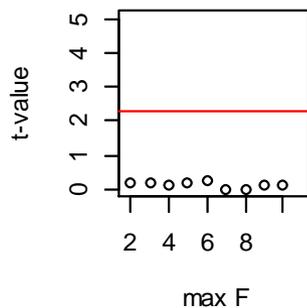
d=10 n=20 v=5



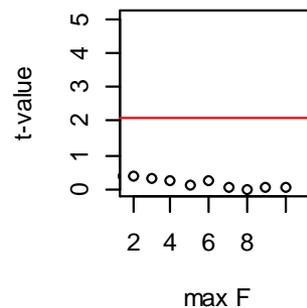
d=2.5 n=3 v=10



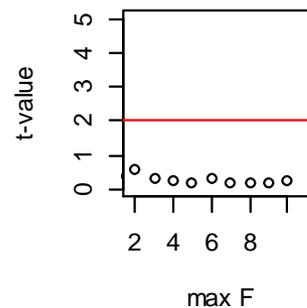
d=2.5 n=5 v=10



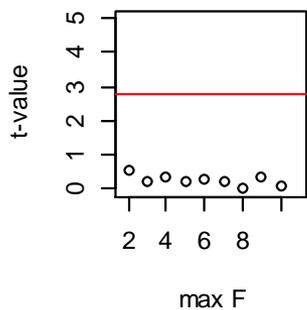
d=2.5 n=10 v=10



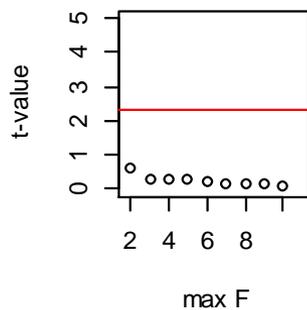
d=2.5 n=20 v=10



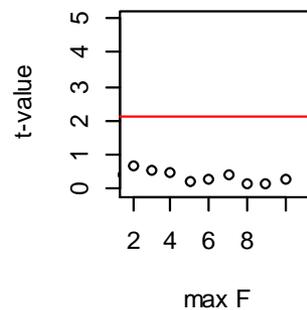
d=5 n=3 v=10



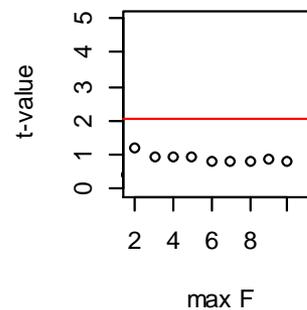
d=5 n=5 v=10



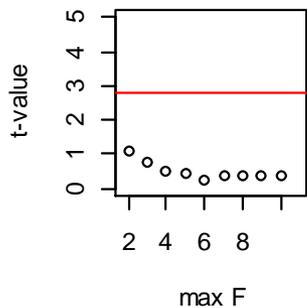
d=5 n=10 v=10



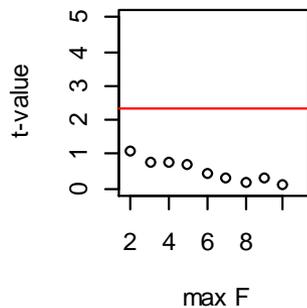
d=5 n=20 v=10



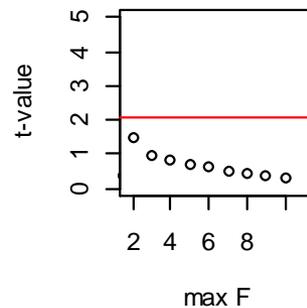
d=10 n=3 v=10



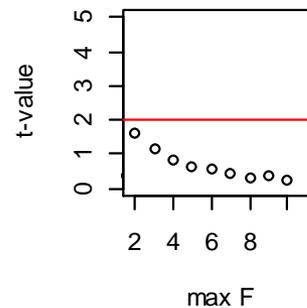
d=10 n=5 v=10



d=10 n=10 v=10



d=10 n=20 v=10



Independence

- Lack of independence is difficult to correct.
- Biased result
- Randomization in experimental study

Independence

- Run test
- Residual plot

Independence

- The good news is ...

Remedial measures (I) Homogeneity

- Weighted least squares
- Transformation

Remedial measures (II) Normality

- Nonparametric

Numerical case (I)Poisson

- `> data1=rpois(5,1)`
- `> data2=rpois(5,2)`
- `> data3=rpois(5,3)`
- `> data=cbind(data1,data2,data3)`
- `> datap=as.numeric(data)`
- `> trt=rep(1:3,each=5)`
- `> fm=aov(datap~trt)`
- `> summary(fm)`

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
trt	1	6.40	6.400	2.253	0.157
Residuals	13	36.93	2.841		
- `>`
- `> datanew=sqrt(datap+0.5)`
- `> trt=rep(1:3,each=5)`
- `> fmnew=aov(datanew~trt)`
- `> summary(fmnew)`

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
trt	1	0.493	0.4934	1.74	0.21
Residuals	13	3.687	0.2836		

```
> summary(fm)
      Df Sum Sq Mean Sq F value Pr(>F)
trt    1  12.1  12.100  4.333 0.0577 .
Residuals 13  36.3  2.792
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> datanew=sqrt(datap+0.5)
> trt=rep(1:3,each=5)
> fmnew=aov(datanew~trt)
> summary(fmnew)
      Df Sum Sq Mean Sq F value Pr(>F)
trt    1  1.383  1.3830  4.648 0.0504 .
Residuals 13  3.868  0.2975
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Numerical case (I)unequal variance

```
> data1=round(10+rnorm(6,0,2),1)
> data2=round(15+rnorm(6,0,10),1)
> data3=round(20+rnorm(6,0,20),1)
> data=cbind(data1,data2,data3)
> datap=as.numeric(data)
> trt=rep(1:3,each=6)
> fm1=aov(datap~trt)
> summary(fm1)
      Df Sum Sq Mean Sq F value Pr(>F)
trt    1  38.9  38.88  0.386 0.543
Residuals 16 1610.5 100.66
>
> data1a=data1/var(data1)
> data2a=data2/var(data2)
> data3a=data3/var(data3)
> dataa=cbind(data1a,data2a,data3a)
> dataap=as.numeric(dataa)
> trt=rep(1:3,each=6)
> fm1a=aov(dataap~trt)
> summary(fm1a)
      Df Sum Sq Mean Sq F value Pr(>F)
trt    1  616.4  616.4  47.3 3.72e-06 ***
Residuals 16  208.5   13.0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
data1=round(10+rnorm(6,0,2),1)
data2=round(15+rnorm(6,0,10),1)
data3=round(20+rnorm(6,0,20),1)
data=cbind(data1,data2,data3)
datap=as.numeric(data)
trt=rep(1:3,each=6)
fm1=aov(datap~trt)
summary(fm1)

data1a=data1/var(data1)
data2a=data2/var(data2)
data3a=data3/var(data3)
dataa=cbind(data1a,data2a,data3a)
dataap=as.numeric(dataa)
trt=rep(1:3,each=6)
fm1a=aov(dataap~trt)
summary(fm1a)
```

Residual analysis

- Before making inferences using hypothesis testing and confidence intervals, it is important to assess the model assumptions:
- (i) Have all important effects been captured?
- (ii) Are the errors independent and normally distributed?
- (iii) Do the errors have constant variance?

- The main purpose of examination of the appropriateness of the model is therefore to detect **serious departures** from the conditions assumed by the model.

- Thank you