

## H7: Analysis of variance

Lieven Clement

Statistiek: 2<sup>de</sup> bach. in de Biochemie en Biotechnologie, Biologie,  
Biomedische Wetenschappen, en de Chemie

## Voorbeeld: Prostacycline studie

Onderzoekers bestuderen het effect van arachidonzuur op het prostacycline niveau in het bloedplasma. Ze gebruiken hierbij 3 verschillende concentraties van arachidonzuur: laag, gemiddeld en een hoge dosis en ze meten het prostacycline niveau in het bloed plasma a.d.h.v. een gecalibreerde elisa fluorescentie meting. 12 ratten worden at random toegekend aan elke behandelingsgroep.

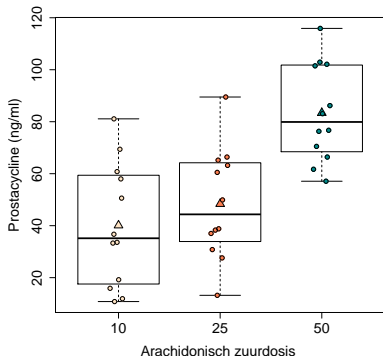


- Experimentele eenheid?
- Response variabele?
- Verklarende variabele(n)?
- Behandelingen?

**completely randomized design (CRD)**: experimentele eenheden at random toegekend aan de behandelingen.

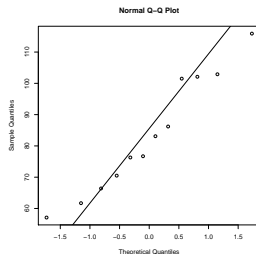
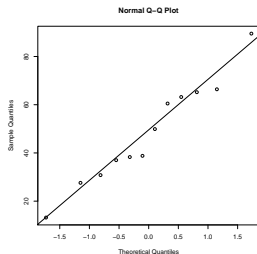
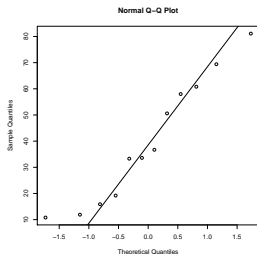
## Data exploratie

Onderzoekers bestuderen het effect van arachidonzuur op het prostacycline niveau in het bloedplasma. Ze gebruiken hierbij 3 verschillende concentraties van arachidonzuur: laag, gemiddeld en een hoge dosis en ze meten het prostacycline niveau in het bloed plasma a.d.h.v. een gecalibreerde elisa fluorescentie meting.



# Voorbeeld: Prostacycline studie

Hoe kunnen we data modelleren?



Data in de drie groepen lijkt normaal verdeeld en de variantie is ongeveer gelijk:

$$Y_i | \text{groep } j \sim N(\mu_j, \sigma^2),$$

met  $j = 1, 2, 3$

Vraagstelling kan vertaald worden in de hypotheses

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

en

$H_1$  : minstens twee gemiddelden zijn verschillend.

Vraagstelling kan vertaald worden in de hypotheses

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

en

$H_1$  : minstens twee gemiddelden zijn verschillend.

**naïeve benadering**: nulhypothese op splitsen in **partiële hypotheses**

$$H_{0jk} : \mu_j = \mu_k \text{ versus } H_{1jk} : \mu_j \neq \mu_k$$

- Elk van deze partiële hypotheses testen met two-sample  $t$ -testen
- Probleem van meervoudig toetsen + verlies van power.
- $H_0 : \mu_1 = \mu_2 = \mu_3$  testen met **1 enkele test**.

- Correcte oplossing voor het testprobleem: **variantie-analyse**, afgekort door ANOVA (ANalysis Of VAriance)
- We leiden de methode af voor de meest eenvoudige uitbreiding met 3 groepen (prostacycline voorbeeld)
- De veralgemening naar  $g$  groepen  $g > 3$  is triviaal

Stel dat  $Y_i$  de uitkomst voorstelt van observatie  $i$  ( $i = 1, \dots, n$ ), dan beschouwen we het lineaire regressiemodel

$$Y_i = g(x_{i1}, x_{i2}) + \varepsilon_i$$

met  $\varepsilon_i$  i.i.d.  $N(0, \sigma^2)$  en met de dummy-regressoren  $x_{i1}$  en  $x_{i2}$ .



Stel dat  $Y_i$  de uitkomst voorstelt van observatie  $i$  ( $i = 1, \dots, n$ ), dan beschouwen we het lineaire regressiemodel

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \quad (1)$$

met  $\varepsilon_i$  i.i.d.  $N(0, \sigma^2)$  en met de dummy-regressoren

$$x_{i1} = \begin{cases} 1 & \text{als observatie } i \text{ behoort tot middelste dosisgroep (M)} \\ 0 & \text{als observatie } i \text{ behoort tot een andere dosisgroep} \end{cases}$$

en

$$x_{i2} = \begin{cases} 1 & \text{als observatie } i \text{ behoort tot de hoogste dosisgroep (H)} \\ 0 & \text{als observatie } i \text{ behoort tot een andere dosisgroep} \end{cases}$$

De lage behandelingsgroep (L) met  $x_{i1} = x_{i2} = 0$  wordt in deze context de **referentiegroep** genoemd.

- Voor observaties in **dosisgroep L** wordt het model (1)

$$Y_i = \beta_0 + \varepsilon_i$$

met  $\varepsilon_i$  i.i.d.  $N(0, \sigma^2)$ .

- Voor observaties in **dosisgroep M** wordt het model (1)

$$Y_i = \beta_0 + \beta_1 + \varepsilon_i$$

met  $\varepsilon_i$  i.i.d.  $N(0, \sigma^2)$ .

- Voor observaties in **dosisgroep H** wordt het model (1)

$$Y_i = \beta_0 + \beta_2 + \varepsilon_i$$

met  $\varepsilon_i$  i.i.d.  $N(0, \sigma^2)$ .

De interpretaties van de  $\beta$ -parameters is dus eenvoudig:

$$\beta_0 = E\{Y_i \mid \text{behandeling met lage dosisgroep L}\}$$

$$\beta_1 = (\beta_0 + \beta_1) - \beta_0 = E\{Y_i \mid \text{behandeling M}\} - E\{Y_i \mid \text{behandeling L}\}$$

$$\beta_2 = (\beta_0 + \beta_2) - \beta_0 = E\{Y_i \mid \text{behandeling H}\} - E\{Y_i \mid \text{behandeling L}\}.$$

Anders geformuleerd:

- parameter  $\beta_0$  is de gemiddelde uitkomst in de lage dosis groep L.
- Parameter  $\beta_1$  is effect van groep M t.o.v. groep L.
- Parameter  $\beta_2$  is effect van hoge dosis groep H t.o.v. groep L.

We herformuleren de modellen gebruik makend van de  $\mu$ -notaties:

$$Y_{i|\text{dose=L}} = \beta_0 + \varepsilon_i = \mu_1 + \varepsilon_i$$

$$Y_{i|\text{dose=M}} = \beta_0 + \beta_1 + \varepsilon_i = \mu_2 + \varepsilon_i$$

$$Y_{i|\text{dose=H}} = \beta_0 + \beta_2 + \varepsilon_i = \mu_3 + \varepsilon_i.$$

met  $\varepsilon_i$  i.i.d.  $N(0, \sigma^2)$  en met

$$\mu_j = E\{Y_i \mid \text{behandelingsgroep } j\}.$$

De oorspronkelijk nulhypothese

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

kan equivalent geformuleerd worden als

$$H_0 : \beta_1 = \beta_2 = 0.$$

De oorspronkelijk nulhypothese

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

kan equivalent geformuleerd worden als

$$H_0 : \beta_1 = \beta_2 = 0.$$

Gezien model (1) een lineair regressiemodel is, kunnen de methoden van lineaire regressie gebruikt worden voor

- het schatten van de parameters en hun varianties
- het opstellen van hypothesetesten en betrouwbaarheidsintervallen
- Testen van  $H_0 : \beta_1 = \beta_2 = 0$  gebeurt dmv  $F$ -test.
- Hiermee is bijna de volledige oplossing bekomen.

Net zoals bij enkelvoudige regressie kunnen we opnieuw de kwadratensom van de regressie hiervoor gebruiken:

$$\begin{aligned}
 \text{SSR} &= \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \\
 &= \sum_{i=1}^n (\hat{g}(x_{i1}, x_{i2}) - \bar{Y})^2 \\
 &= \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}) - \bar{Y})^2 \\
 &= \sum_{i=1}^{n_1} (\hat{\beta}_0 - \bar{Y})^2 + \sum_{i=1}^{n_2} (\hat{\beta}_0 + \hat{\beta}_1 - \bar{Y})^2 + \sum_{i=1}^{n_3} (\hat{\beta}_0 + \hat{\beta}_2 - \bar{Y})^2 \\
 &= \sum_{i=1}^{n_1} (\bar{Y}_1 - \bar{Y})^2 + \sum_{i=1}^{n_2} (\bar{Y}_2 - \bar{Y})^2 + \sum_{i=1}^{n_3} (\bar{Y}_3 - \bar{Y})^2
 \end{aligned}$$

$$SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

- Kwadratensom opnieuw equivalent aan vergelijken van model (1) en een gereduceerd model met enkel een intercept.
- Voor gereduceerd model zal intercept worden geschat door steekproefgemiddelde.
- Deze kwadratensom heeft dus  $g-1=2$  vrijheidsgraden:  
g=3 model parameters - 1 parameter voor steekproefgemiddelde  
of  
g=3 par. van complexe model - 1 par. van gereduceerde model.



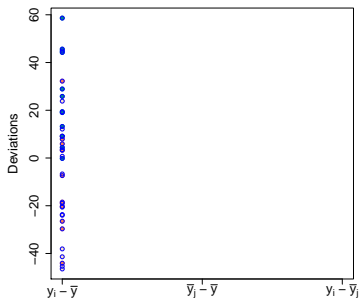
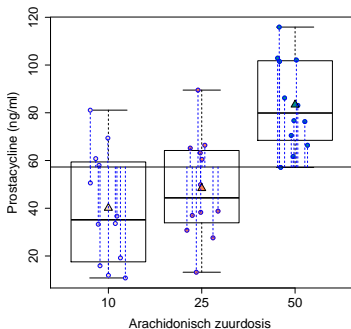
## Ontbinding van de Totale Kwadratensom

- De conventie in een Anova setting is om de kwadratensom te noteren als SST, de **kwadratensom van de behandeling (treatment)** of als SSBetween.
- De kwadratensom van de regressie geeft voor model (1) inderdaad de variabiliteit weer tussen de groepen. (zie volgende slide)
- De overeenkomstige gemiddelde kwadratensom wordt dan  $MST = SST/2$ .

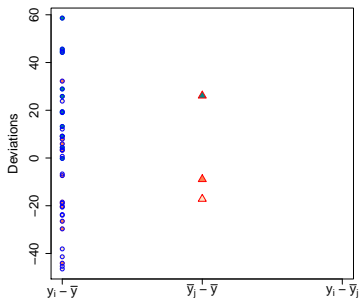
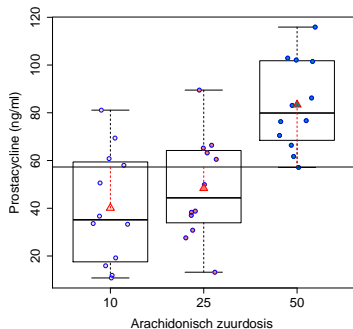
De decompositie van  $SST_{Tot}$  wordt dan geschreven als

$$SST_{Tot} = SST + SSE.$$

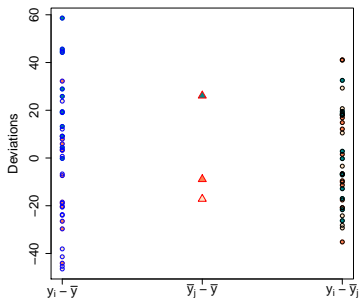
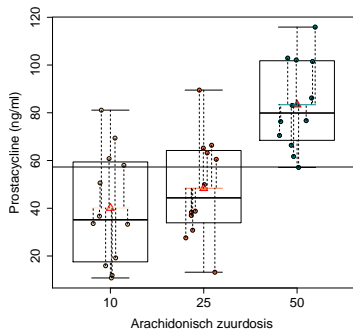
# Interpretatie ontbinding totale kwadratensom



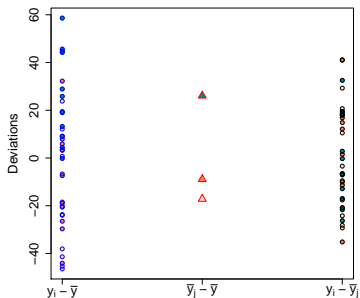
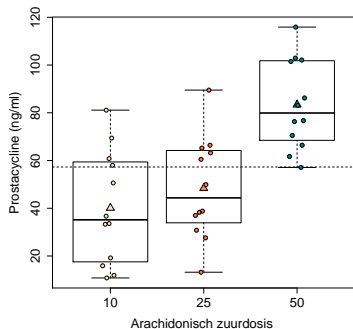
# Interpretatie ontbinding totale kwadratensom



# Interpretatie ontbinding totale kwadratensom



# Interpretatie ontbinding totale kwadratensom



## Anova test

- Testen van  $H_0 : \beta_1 = \beta_2 = 0$  gebeurt dmv  $F$ -test.

$$F = \frac{MST}{MSE}$$

met  $MST = SST/(g - 1)$  en  $MSE$  de gemiddelde residuele kwadratensom uit het niet-gereduceerde model (1); deze heeft  $n - g$  vrijheidsgraden.

- De teststatistiek vergelijk dus variabiliteit verklaard door het model (MST) met de residuele variabiliteit (MSE)

of

- Variabiliteit tussen groepen (MST) met variabiliteit binnen groepen (MSE)
- Onder de nulhypothese geldt  $F \sim F_{g-1, n-g}$ , met  $g=3$ .

## Prostacycline Voorbeeld

```
> model1=lm(prostac~dose,data=prostacyclin)
> anova(model1)
```

Analysis of Variance Table

Response: prostac

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dose	2	12658	6329.0	13.944	4.081e-05 ***
Residuals	33	14979	453.9		

---

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

## Prostacycline Voorbeeld

```
> model1=lm(prostac~dose,data=prostacyclin)
> anova(model1)
Analysis of Variance Table
```

Response: prostac

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dose	2	12658	6329.0	13.944	4.081e-05 ***
Residuals	33	14979	453.9		

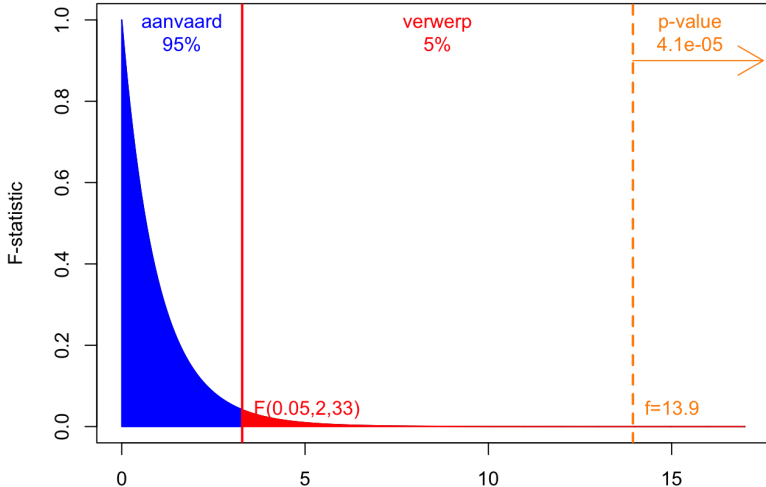
---

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

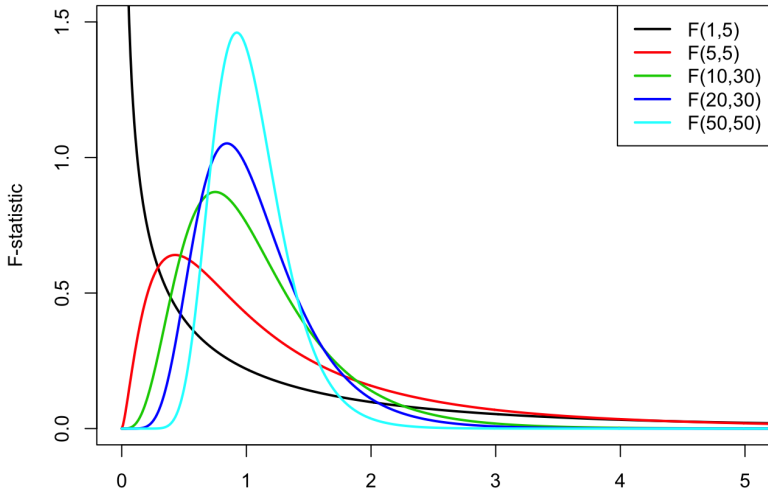
Er is een extreem significant effect van de arachidonzuurdosering op de gemiddelde prostacycline concentratie in het bloed bij ratten ( $p \ll 0.001$ ).



# F-verdeling



# F-verdeling



## Prostacycline Voorbeeld: welke groepen zijn verschillend?

```
> summary(model1)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	40.108	6.150	6.521	2.10e-07	***
dose25	8.258	8.698	0.949	0.349	
dose50	43.258	8.698	4.974	1.99e-05	***

---

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

Residual standard error: 21.3 on 33 degrees of freedom

Multiple R-squared: 0.458, Adjusted R-squared: 0.4252

F-statistic: 13.94 on 2 and 33 DF, p-value: 4.081e-05

## Prostacycline Voorbeeld: welke groepen zijn verschillend?

```
> summary(model1)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	40.108	6.150	6.521	2.10e-07	***
dose25	8.258	8.698	0.949	0.349	
dose50	43.258	8.698	4.974	1.99e-05	***

---

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

Residual standard error: 21.3 on 33 degrees of freedom

Multiple R-squared: 0.458, Adjusted R-squared: 0.4252

F-statistic: 13.94 on 2 and 33 DF, p-value: 4.081e-05

Met output van model kunnen we ook testen of de gemiddelde prostacycline concentratie verschillend is tussen de matige en lage dosis groep ( $\beta_1$ : dose25) en tussen de hoge en lage dosis groep ( $\beta_2$ : dose50).

# Post hoc Analyse: Meervoudig Vergelijken van Gemiddelden

In het eerste deel van dit hoofdstuk hebben we een  $F$ -test gezien die gebruikt kan worden voor het testen van

$$H_0 : \mu_1 = \dots = \mu_g \text{ versus } H_1 : \text{niet } H_0.$$

Dus als de nulhypothese verworpen wordt, dan wordt besloten dat er minstens twee gemiddelden verschillen van elkaar, maar de methode stelt ons niet in staat om te identificeren welke gemiddelden van elkaar verschillen.

Een eerste, maar naïeve benadering van het probleem bestaat erin om de nulhypothese op te splitsen in partiële hypothesen

$$H_{0jk} : \mu_j = \mu_k \text{ versus } H_{1jk} : \mu_j \neq \mu_k$$

en deze partiële hypothesen te testen met  $m = g(g - 1)/2$  two-sample  $t$ -testen. Voor het vergelijken van groep  $j$  met groep  $k$  wordt de klassieke two-sample  $t$ -test onder de veronderstelling van homoscedasticiteit gegeven door

$$T_{jk} = \frac{\bar{Y}_j - \bar{Y}_k}{S_p \sqrt{\frac{1}{n_j} + \frac{1}{n_k}}} \stackrel{H_0}{\sim} t_{n-2}$$

waarin  $S_p^2$  de gepoolde variantieschatter is,

$$S_p^2 = \frac{(n_j - 1)S_j^2 + (n_k - 1)S_k^2}{n_j + n_k - 2}$$

met  $S_j^2$  en  $S_k^2$  de steekproefvarianties van respectievelijk de uitkomsten uit groep  $j$  en  $k$ .

In een ANOVA setting wordt echter verondersteld dat in alle  $t$  groepen de variantie van de uitkomsten dezelfde is (de residuele variantie  $\sigma^2$ ). Indien we dus  $S_p^2$  gebruiken, dan is dit niet de meest efficiënte schatter omdat het niet van alle data gebruik maakt. We kunnen dus efficiëntie winnen door MSE te gebruiken. Ter herinnering, MSE kan geschreven worden als

$$\text{MSE} = \sum_{j=1}^g \frac{(n_j - 1)S_j^2}{n - g}.$$

De  $t$ -testen voor het twee-aan-twee vergelijken van alle gemiddelden worden dus best gebaseerd op

$$T_{jk} = \frac{\bar{Y}_j - \bar{Y}_k}{\text{MSE} \sqrt{\frac{1}{n_j} + \frac{1}{n_k}}} \stackrel{H_0}{\sim} t_{n-g}.$$

We zullen hier eerst demonstreren dat het werken met  $t$ -testen op het  $\alpha$  significantieniveau een foute aanpak is die de kans op een type I fout niet onder controle kan houden. Dit zal aanleiding geven tot een meer algemene definitie van de type I fout.



## Prostacycline Voorbeeld

Alvorens de denkfout in de naïeve aanpak te demonsteren via simulaties, tonen we hoe de naïeve benadering in zijn werk zou gaan voor het prostacycline voorbeeld.

```
> pairwise.t.test(prostac,dose, data = prostacyclin,"none")
```

Pairwise comparisons using t tests with pooled SD

```
data:  prostac and dose
```

```
      10      25
25 0.34927 -
50 2e-05   0.00031
```

```
P value adjustment method: none
```

Deze output toont de tweezijdige  $p$ -waarden voor het testen van alle partiële hypothesen. We zouden hier kunnen besluiten dat :

- Het gemiddelde prostacycline niveau extreem significant verschillend is tussen de hoge en de lage dosis groep
- Het gemiddelde prostacycline niveau extreem significant verschillend is tussen de hoge en de matige dosis groep
- Het gemiddelde prostacycline niveau niet significant verschillend is tussen de matige en de lage dosis groep

- In onderstaande R code wordt simulatiestudie opgezet (herhaalde steekproefname)
- Simualtie uit een ANOVA model met  $t = 3$  groepen
- De gemiddelden in het ANOVA model zijn gelijk aan elkaar, zodat de nulhypothese

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

opgaat.

- Voor iedere gesimuleerde dataset drie paarsgewijze two-sample  $t$ -testen
- Zodra minstens één van de  $p$ -waarden kleiner is dan het significantieniveau  $\alpha = 5\%$ , wordt de nulhypothese  $H_0 : \mu_1 = \mu_2 = \mu_3$  verworpen omdat er minstens twee gemiddelden verschillend zijn volgens de  $t$ -testen.
- Relatieve frequentie van het verwerpen van de globale nulhypothese (i.e. de kans op een type I fout van de test voor  $H_0 : \mu_1 = \mu_2 = \mu_3$ )

```
> g<-3 # aantal behandelingen (g=3)
> ni<-12 # aantal herhalingen in iedere groep
> n<-g*ni # totaal aantal observaties
> alpha<-0.05 # significantieniveau van een individuele test
> N=10000 #aantal simulaties
> set.seed(302) #seed zodat resultaten exact geproduceerd kunnen
> trt=factor(rep(1:g,ni)) #factor
> cnt<-0 #teller voor aantal foutieve verwerpingen
> for(i in 1:N) {
+ if (i%%1000==0) cat(i,"/",N,"\n")
+ y <- rnorm(n)
+ tests<-pairwise.t.test(y,trt,"none")
+ verwerp<-min(tests$p.value,na.rm=T)<alpha
+ if(verwerp) cnt<-cnt+1
+ }
1000 / 10000
...
> cnt/N
[1] 0.1209
```

- Kans op een type I fout gelijk is aan 12.1%, wat meer dan dubbel zo groot is dan de vooropgestelde  $\alpha_E = 5\%$ .
- Als we simulatiestudie herhalen met  $g = 5$  groepen (i.e. 10 paarsgewijze  $t$ -testen) dan vinden we 28.0% in plaats van de gewenste 5%.
- Deze simulaties illustreren het probleem van **multipliciteit** (Engels: *multiplicity*): de klassieke  $p$ -waarden mogen enkel met het significantieniveau  $\alpha$  vergeleken worden, indien het besluit op exact één  $p$ -waarde gebaseerd is. Hier wordt het finale besluit (aldanniet verwerpen van  $H_0 : \mu_1 = \dots = \mu_g$ ) gebaseerd op  $m = g \times (g - 1)/2$   $p$ -waarden.
- We bespreken eerst een uitbreiding van het begrip van type I fout en vervolgens introduceren we enkele oplossingen.

## Family-wise error rate (FWER)

- Wanneer  $m > 1$  toetsen worden aangewend om 1 beslissing te vormen, is het noodzakelijk te corrigeren voor het risico op valse vondsten.

## Family-wise error rate (FWER)

- Wanneer  $m > 1$  toetsen worden aangewend om 1 beslissing te vormen, is het noodzakelijk te corrigeren voor het risico op valse vondsten.
- Meeste procedures voor meervoudig toetsen gaan ervan uit dat **alle  $m$  nulhypotheses waar** zijn.
- Ze proberen dan het **risico op minstens 1 valse vondst** te controleren op niveau  $\alpha_E$ ; typisch,  $\alpha_E = 0.05$ .  
(experimentgewijs significantie niveau)
- We noemen dit de **family-wise error rate (FWER)**.

## Bonferroni correctie (1)

Bij het uitvoeren van  $m$  onafhankelijke toetsen met elk significantieniveau  $\alpha$ , is

$$\begin{aligned}\alpha_E &= P(\text{minstens 1 Type I fout}) \\ &= 1 - (1 - \alpha)^m \approx m\alpha\end{aligned}$$



## Bonferroni correctie (1)

Bij het uitvoeren van  $m$  onafhankelijke toetsen met elk significantieniveau  $\alpha$ , is

$$\begin{aligned}\alpha_E &= P(\text{minstens 1 Type I fout}) \\ &= 1 - (1 - \alpha)^m \approx m\alpha\end{aligned}$$

### Bonferroni correctie

- Als we 5 toetsen uitvoeren op het 5% significantieniveau is FWER  $\approx 25\%$ .
- Door ze op het 1% significantieniveau uit te voeren, bekomen we FWER  $\approx 5\%$ .

## Bonferroni correctie (2)

**Bonferroni correctie** houdt de FWER begrensd op  $\alpha_E$  door

$$\alpha = \alpha_E / m$$

te kiezen:

- Toetsen op het  $\alpha_E / m$  significantieniveau
- of
- aangepaste p-waarden zodat we deze met het experimentgewijze  $\alpha_E$  niveau kunnen vergelijken:

$$\tilde{p} = \min(m \times p, 1)$$

- $(1 - \alpha_E / m)100\%$  betrouwbaarheidsintervallen rapporteren.

## Prostacycline Voorbeeld

Onderstaande R code geeft de resultaten (gecorrigeerde  $p$ -waarden) na correctie met de methode van Bonferroni.

```
> pairwise.t.test(prostac,dose, data = prostacyclin,  
+ p.adjust.method="bonferroni")
```

Pairwise comparisons using t tests with pooled SD

data: prostac and dose

	10	25
25	1.00000	-
50	6e-05	0.00094

P value adjustment method: bonferroni

De conclusies zijn hetzelfde behalve dat de FWER gecontroleerd is  $\alpha = 5\%$  en de  $\tilde{p}$ -waarden zijn een factor 3 groter.

Dezelfde analyse kan uitgevoerd worden met de multcomp R package.

```
> model1.mcp<-glht(model1,linfct=mcp(dose="Tukey"))
> summary(model1.mcp,test=adjusted("bonferroni"))
```

### Simultaneous Tests for General Linear Hypotheses

#### Multiple Comparisons of Means: Tukey Contrasts

Fit: `lm(formula = prostac ~ dose, data = prostacyclin)`

#### Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t )	
25 - 10 == 0	8.258	8.698	0.949	1.000000	
50 - 10 == 0	43.258	8.698	4.974	5.98e-05	***
50 - 25 == 0	35.000	8.698	4.024	0.000943	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
(Adjusted p values reported -- bonferroni method)

## Bonferroni aangepaste betrouwbaarheidsintervallen

We moeten eerst zelf functie definiëren om Bonferroni kritische waarde te bepalen.

```
> alpha_bon_t<-function(object,level) abs(qt((1-level)/2/nrow(object$linfct), object$df))
> confint(modell.mcp,calpha=alpha_bon_t)
```

Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts

Fit: lm(formula = prostac ~ dose, data = prostacyclin)

Quantile = 2.5222  
95% confidence level

Linear Hypotheses:

	Estimate	lwr	upr
25 - 10 == 0	8.2583	-13.6790	30.1957
50 - 10 == 0	43.2583	21.3210	65.1957
50 - 25 == 0	35.0000	13.0626	56.9374

## Effect van Bonferroni correctie op FWER

```
> g<-3 # aantal behandelingen (g=3)
> ni<-12 # aantal herhalingen in iedere groep
> n<-g*ni # totaal aantal observaties
> alpha<-0.05 # significantieniveau van een individuele test
> N=10000 #aantal simulaties
> set.seed(302) #seed zodat resultaten exact geproduceerd kunnen
> trt=factor(rep(1:g,ni)) #factor
> cnt<-0 #teller voor aantal foutieve verwerpingen
> for(i in 1:N) {
+   if (i%%1000==0) cat(i,"/",N,"\n")
+   y <- rnorm(n)
+   tests<-pairwise.t.test(y,trt,"bonferroni")
+   verwerp<-min(tests$p.value,na.rm=T)<alpha
+   if(verwerp) cnt<-cnt+1
+ }
...
> cnt/N
[1] 0.0457
```

We vinden dus een FWER van 4.6% (een beetje conservatief).  
Wanneer we de simulaties doen voor  $g = 5$  groepen, vinden we een FWER van 4.1% (meer conservatief).

## Tukey Method

De methode van Tukey is een minder conservatieve methode voor het uitvoeren van post hoc testen

```
> model1.mcp<-glht(model1,linfct=mcp(dose="Tukey"))
> summary(model1.mcp)
```

Multiple Comparisons of Means: Tukey Contrasts

Fit: `lm(formula = prostac ~ dose, data = prostacyclin)`

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t )
25 - 10 == 0	8.258	8.698	0.949	0.613373
50 - 10 == 0	43.258	8.698	4.974	< 1e-04 ***
50 - 25 == 0	35.000	8.698	4.024	0.000871 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
(Adjusted p values reported -- single-step method)



```
> confint(model1.mcp)
```

Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts

```
Fit: lm(formula = prostac ~ dose, data = prostacyclin)
```

```
Quantile = 2.4552
```

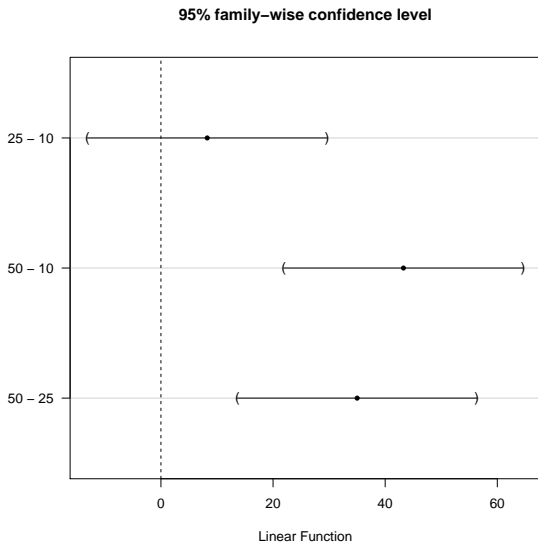
```
95% family-wise confidence level
```

Linear Hypotheses:

	Estimate	lwr	upr
25 - 10 == 0	8.2583	-13.0964	29.6130
50 - 10 == 0	43.2583	21.9036	64.6130
50 - 25 == 0	35.0000	13.6453	56.3547

Smallere BIs en kleinere p-waarden dan Bonferroni.

```
> plot(confint(model1.mcp))
```



# Effect van Tukey correctie op FWER

```
> g<-3 # aantal behandelingen (g=3)
> ni<-12 # aantal herhalingen in iedere groep
> n<-pgni # totaal aantal observaties
> alpha<-0.05 # significantieniveau van een individuele test
> N=10000 #aantal simulaties
> set.seed(302) #seed zodat resultaten exact geproduceerd kunnen worden
> trt=factor(rep(1:g,ni)) #factor
> cnt<-0 #teller voor aantal foutieve verwerpingen
> for(i in 1:N) {
+ if (i%%100==0) cat(i,"/",N,"\n")
+ y <- rnorm(n)
+ m<-lm(y~trt)
+ m.mcp<-glht(m,linfct=mcp(trt="Tukey"))
+ tests<-summary(m.mcp)$test
+ verwerp<-min(as.numeric(tests$pvalues),na.rm=T)<alpha
+ if(verwerp) cnt<-cnt+1
+ }
...
> cnt/N
[1] 0.0503
```

- We vinden dus een FWER van 5.03% wat heel dicht bij het nominale FWER= 5% ligt  
(theoretisch moet dit 5% zijn, maar we tonen “slechts” het resultaat gebaseerd op 10000 simulaties).
- Voor  $g = 5$  groepen, vinden we een FWER van 5.2%, wat ook vrij goed is.

# Prostacycline Voorbeeld: Conclusies

```
> anova(model1)
Analysis of Variance Table

Response: prostac
      Df Sum Sq Mean Sq F value    Pr(>F)
dose    2  12658   6329.0   13.944 4.081e-05 ***
Residuals 33  14979    453.9
---
```

## Prostacycline Voorbeeld: Conclusies

```
> anova(model1)
Analysis of Variance Table

Response: prostac
      Df Sum Sq Mean Sq F value    Pr(>F)
dose    2  12658   6329.0   13.944 4.081e-05 ***
Residuals 33  14979    453.9
---
```

Er is een extreem significant effect van de arachidonzuurdosering op de gemiddelde prostacycline concentratie in het bloed bij ratten ( $p < 0.001$ ).

## Prostacycline Voorbeeld: Conclusies

```
> summary(model1.mcp)
Multiple Comparisons of Means: Tukey Contrasts

Linear Hypotheses:
      Estimate Std. Error t value Pr(>|t|)
25 - 10 == 0    8.258     8.698  0.949 0.613373
50 - 10 == 0   43.258     8.698  4.974 < 1e-04 ***
50 - 25 == 0   35.000     8.698  4.024 0.000871 ***
```

Er is een extreem significant effect van de arachidonzuurdosering op de gemiddelde prostacycline concentratie in het bloed bij ratten ( $p < 0.001$ ).

## Prostacycline Voorbeeld: Conclusies

```
> summary(model1.mcp)
Multiple Comparisons of Means: Tukey Contrasts

Linear Hypotheses:
      Estimate Std. Error t value Pr(>|t|)
25 - 10 == 0    8.258      8.698  0.949 0.613373
50 - 10 == 0   43.258      8.698  4.974 < 1e-04 ***
50 - 25 == 0   35.000      8.698  4.024 0.000871 ***
```

Er is een extreem significant effect van de arachidonzuurdosering op de gemiddelde prostacycline concentratie in het bloed bij ratten ( $p < 0.001$ ). De gemiddelde prostacycline concentratie is hoger bij de hoge arachidonzuur dosisgroep dan bij de lage en matige dosisgroep (beide  $p < 0.001$ ).



## Prostacycline Voorbeeld: Conclusies

```
> confint(model1.mcp)
```

Linear Hypotheses:

	Estimate	lwr	upr
25 - 10 == 0	8.2583	-13.0964	29.6130
50 - 10 == 0	43.2583	21.9036	64.6130
50 - 25 == 0	35.0000	13.6453	56.3547

Er is een extreem significant effect van de arachidonzuurdosering op de gemiddelde prostacycline concentratie in het bloed bij ratten ( $p < 0.001$ ). De gemiddelde prostacycline concentratie is hoger bij de hoge arachidonzuur dosisgroep dan bij de lage en matige dosisgroep (beide  $p < 0.001$ ).

## Prostacycline Voorbeeld: Conclusies

```
> confint(model1.mcp)
```

Linear Hypotheses:

	Estimate	lwr	upr
25 - 10 == 0	8.2583	-13.0964	29.6130
50 - 10 == 0	43.2583	21.9036	64.6130
50 - 25 == 0	35.0000	13.6453	56.3547

Er is een extreem significant effect van de arachidonzuurdosering op de gemiddelde prostacycline concentratie in het bloed bij ratten ( $p < 0.001$ ). De gemiddelde prostacycline concentratie is hoger bij de hoge arachidonzuur dosisgroep dan bij de lage en matige dosisgroep (beide  $p < 0.001$ ). De gemiddelde prostacycline concentratie in de hoge dosis groep is respectievelijk 43.3 ng/ml (95% BI [21.9,64.6]) en 35.0 ng/ml (95% BI [13.7,56.3]) hoger dan in de lage en matige dosis groep.

## Prostacycline Voorbeeld: Conclusies

```
> confint(model1.mcp)
```

Linear Hypotheses:

	Estimate	lwr	upr
25 - 10 == 0	8.2583	-13.0964	29.6130
50 - 10 == 0	43.2583	21.9036	64.6130
50 - 25 == 0	35.0000	13.6453	56.3547

Er is een extreem significant effect van de arachidonzuurdosering op de gemiddelde prostacycline concentratie in het bloed bij ratten ( $p < 0.001$ ). De gemiddelde prostacycline concentratie is hoger bij de hoge arachidonzuur dosisgroep dan bij de lage en matige dosisgroep (beide  $p < 0.001$ ). De gemiddelde prostacycline concentratie in de hoge dosis groep is respectievelijk 43.3 ng/ml (95% BI [21.9,64.6]) en 35.0 ng/ml (95% BI [13.7,56.3]) hoger dan in de lage en matige dosis groep. Het verschil in gemiddelde prostacycline concentratie tussen de matige en lage dosisgroep is niet significant ( $p=0.61$ , 95% BI op gemiddelde verschil [-13.1,29.6] ng/ml).