

# **Blok EE**

## **Epidemiologie & Economie**

### **Werkcolleges en Zelfstudies**

# Inhoudsopgave

<b>Deel 1: STATISTIEK .....</b>	<b>3</b>
<i>Werkcollege 1 - Lineaire modellen I: Correlatie en Regressie.....</i>	<i>4</i>
<i>Werkcollege 2 - Lineaire modellen II: ANOVA.....</i>	<i>9</i>
<i>Werkcollege 3 - Lineaire modellen III: Uitbreidingen.....</i>	<i>13</i>
<i>Zelfstudie 01 - Survivalanalyse en toetsingsoverzicht .....</i>	<i>17</i>
<b>Deel 2: ECONOMIE .....</b>	<b>22</b>
<i>Werkcollege 4 - Waarde, Kosten en Baten.....</i>	<i>23</i>
<i>Werkcollege 5 - Rekenmethodes.....</i>	<i>30</i>
<b>Deel 3: EPIDEMIOLOGIE.....</b>	<b>37</b>
<i>Werkcollege 6 - Associatiematen en onderzoeksopzet.....</i>	<i>38</i>
<i>Werkcollege 7 - Test eigenschappen .....</i>	<i>43</i>
<i>Werkcollege 8 - Koppeldiagnostiek en serieel of parallel testen.....</i>	<i>47</i>
<i>Zelfstudie 2 - Populatie dynamica van infectieziekten.....</i>	<i>51</i>
<i>Zelfstudie 3 - Critical appraisal / publicaties lezen.....</i>	<i>55</i>

# **Deel 1: STATISTIEK**

# Werkcollege 1 - Lineaire modellen I: Correlatie en Regressie

## Inhoud

In dit werkcollege leer je de basis van lineaire modellen. Je gaat aan de slag met correlatie coëfficiënt en leert hoe regressiemodellen werken.

## Voorbereiding

Bekijk de videoclips 'WC 1 – Lineaire modellen I' en 'WC1 – R script'.

Petrie & Watson (3rd ed): 10.1 - 10.4.3(b), 10.4.6

Statistiekteksten: h. 1 en h. 2

Bestudeer het bestand 'TOETSINGSOVERZICHT STATISTIEK' onder 'course content' >> 'statistiek'. In dit bestand leer je welke toets je nodig hebt voor welk type onderzoeksvraag.

## Onderwerpen

- Scatterplot
- Correlatie coëfficiënt
- Voorkomen misbruik (prevention of misuse)
- De regressie lijn: intercept en regressiecoëfficiënt (the slope)
- $R^2$ : R-squared
- Toetsen (statistical testing) van de correlatie coëfficiënt en de richtingscoëfficiënt
- Residuele variantie
- Regression diagnostics: residuen analyse

## Script code met RStudio/R

```
#####  
# WC1  
# Environment - Import Dataset - From text (base)... kies Table10.1.txt  
#  
  
attach(Table10.1)  
names(Table10.1)  
summary(Table10.1)  
  
# de scatterplot, pch staat voor pointcharacter, als je het nummer verandert  
# krijg je andere symbolen  
  
plot(wBAP, PICP, pch=20, col="red")  
?cor # help informatie  
?cor.test # help informatie  
cor.test(wBAP, PICP)  
detach(Table10.1)
```

```
####
# Environment - Import Dataset - From text (base)... kies Table10. 2. txt
#

attach(Table10. 2)

# de scatterplot xlab en ylab geven de labels op de assen

plot(chestgth, livewt, xlab="Chest girth (cm)", ylab="Live weight (kg)",
pch=20, col="red")
?glm # help informatie
fit=glm(livewt~chestgth, family="gaussian")
summary(fit)

# In de kolom Estimate staan de intercept (Intercept) en de slope (chestgth).
# In de kolom Std. Error staan de standard errors.
# In de volgende kolom staan de t-toetsen daarna de p-waardes (Pr(>|t|)).
# Vergelijk met display 10.1 op blz 136 van het boek.
# Om de lijn in de scatterplot toe te voegen:

?abline # help informatie
abline(fit, col="blue")

# of

abline(a=- 46.041, b=1.043, col="blue")

# Om R^2 (kwadraat van de correlatie) uit te rekenen:
1- (fit$deviance/fit$null.deviance)

# of

(cor(livewt, predict(fit)))^2

# voor predict(fit) zie hieronder
# 95% Betrouwbaarheidsintervallen voor de intercept en de slope:

confint(fit)

#####
# Om de kwadraatsommen te krijgen gebruiken we drop1.

drop1(fit, test="F")

# Deze functie fit (schat) meerdere modellen door steeds 1 variabele uit het
model weg te laten en geeft residu kwadraatsommen van die modellen. In dit
geval:
# 1.) een model met het intercept en chestgth met livewt als afhankelijke
variabele. Dit is het volledige model. Er gaat dan geen variabele uit dus
drop is dan <none>. Dit is de eerste regel. De residu kwadraatsom is 452.3.
# 2.) een model waaruit chestgth is weggelaten, dus een model met alleen de
constante erin. De residu kwadraatsom van dit model is SS-totaal. Hier zitten
```

de effecten van chestgth nog in, omdat chestgth weggelaten is uit het model. Dus in de res. kwadraatsom uit de eerste regel zitten geen chestgth effecten meer, in de residu kwadraatsom uit de tweede regel nog wel. Het verschil geeft dus de regressie kwadraatsom.

# Nu is de anova tabel te maken. Zie ook Display 10.1 op blz 136 (P&W).

# De residuen en de predicted values (de punten op de lijn) krijg je door:

```
residuals(fit)
```

```
predict(fit)
```

# dus figuur 10.8a en c krijg je met

```
plot(chestgth, residuals(fit), xlab="Chest girth (cm)", ylab="Residuals (kg)", pch=20, col="red")
```

```
abline(h=0)
```

```
lw<-lowess(residuals(fit)~chestgth)
```

```
lines(lw$x, lw$y)
```

```
plot(predict(fit), residuals(fit), xlab="Predicted values (cm)", ylab="Residuals (kg)", pch=20, col="red")
```

#Om te zien of er afwijkingen zijn van de normaliteit gebruiken we een normal probability plot. Deze plot residuen tegen wat verwacht mag worden als de residuen normaal verdeeld zouden zijn. Als de residuen uit een normale verdeling komen zal dus deze plot punten laten zien, die op een rechte lijn liggen:

```
qqnorm(residuals(fit))
```

```
abline(0, sd(residuals(fit)))
```

#of voor de standardiz ed residuals:

```
qqnorm(rstandard(fit))
```

```
abline(0, 1)
```

```
detach(Table10.2)
```

## Opdrachten

### Opdracht 1.1

Men wil het verband vastleggen tussen het bloedvolume in geiten (n=30) en hun lichaamsgewicht. De afhankelijke variabele is bloedvolume.

- Zoek in het toetsingsoverzicht op welke analyse geschikt is voor deze vraagstelling.
- Lees de dataset geit.txt in en stel de hypothesen voor dit model op.
- Bepaal de correlatiecoëfficiënt.
- Geef aan waar de correlatiecoëfficiënt een maat voor is.
- Is de correlatie significant? Doorloop de volgende stappen:
  - stel de hypothesen op,
  - beoordeel het spreidingsdiagram,

- 3) geef de toetsingsgrootte en de bijbehorende kansverdeling,
  - 4) geef de uitkomst van de toetsingsgrootte,
  - 5) bepaal de p-waarde en trek de conclusie (zie blz. 129 en 130, P&W)
- f. Bereken de  $R^2$  en interpreteer deze.
  - g. Bepaal de regressielijn. Wat is de waarde van het intercept van de regressielijn. Heeft dit intercept een (biologische) betekenis in dit onderzoek?
  - h. Bepaal op basis van de computeruitvoer de waarde van de regressiecoëfficiënt en geef de betekenis hiervan.
  - i. Maak een scatterplot en zet de regressielijn daarin.
  - j. Geef de formule voor het berekenen van het 95% betrouwbaarheidsinterval voor de regressiecoëfficiënt van bovenstaande gegevens.
  - k. Bepaal het 95% betrouwbaarheidsinterval van de regressiecoëfficiënt en geef hiervan een interpretatie.
  - l. Bereken het verwachte bloedvolume bij een gewicht van 20, 40 en 85 kg lichaamsgewicht.
  - m. Bereken het residu voor de 3 verwachte waarden uit k). (Voor de werkelijke waarden: zie de gegevenstabel.)
  - n. Geef de onderliggende modelaannamen voor een lineair regressiemodel (paragraaf 10.4.3).
  - o. Geef per modelaanname welke grafiek je nodig hebt om deze aanname te kunnen controleren en hoe die grafiek er uit moet zien als de modelaanname klopt.
  - p. Beoordeel de grafieken en geef per modelaanname aan of deze wel of niet valide is.

## Opdracht 1.2

In een steekproef van 13 spreeuwen van verschillende leeftijd (in dagen) is de vleugellengte (in cm) gemeten. De onderzoeksvraag is of de leeftijd kan worden gebruikt om de vleugellengte lineair te voorspellen.

- a. Zoek in het toetsingsoverzicht op welke analyse geschikt is voor deze vraagstelling.
- b. Lees de dataset `spreeuw.txt` in.
- c. Bereken de correlatiecoëfficiënt en het 95% betrouwbaarheidsinterval.
- d. Geef de interpretatie van dit interval.
- e. Bepaal de regressielijn
- f. Maak een scatterplot en zet de regressielijn daarin.
- g. Stel de hypothesen op (denk goed na!) die bij de onderzoeksvraag horen.
- h. Toets deze hypothesen met stap 2-5 van het toetsingsprotocol.
- i. Bereken de  $R^2$  en interpreteer dit getal.

- j. Beschrijf in een kort verslag (enkele regels) uw bevindingen en uw antwoord op de onderzoeksvraag.

### **Opdracht 1.3**

Omdat de snuitlengte van in het wild levende alligators gemakkelijker te observeren valt dan het totale gewicht, wil men bekijken of dit laatste te voorspellen is uit de eerste. Men ging eerst na of er een correlatie bestaat tussen de snuitlengte ( $x$ ) en het gewicht ( $y$ ). Men vond een lineair verband nadat men de logaritme van beide variabelen had genomen. In `aligator.txt` staan de logaritme van het gewicht (pounds) en de logaritme van de snuitlengte (inches) van 15 alligators die gevangen werden in centraal Florida.

- a. Zoek in het toetsingsoverzicht op welke analyse geschikt is voor deze vraagstelling.
- b. Lees de dataset `aligator.txt` in en bepaal hoe snuitlengte samenhangt met lichaamsgewicht, in lijn met hoe dat in opdracht 1.2 is gedaan.



# Werkcollege 2 - Lineaire modellen II: ANOVA

## Inhoud

In dit werkcollege leer je hoe je kunt testen of gemiddelden van meerdere groepen verschillen. Dit doe je door middel van een variantie analyse (ANOVA). Ook leer je hoe je kunt corrigeren voor herhaald testen door middel van de Bonferroni-correctie.

## Vorbereiding

Bekijk de videoclips 'WC 2 – Lineaire modellen II' en 'WC2 – R script'.

Petrie & Watson (3rd ed): 4.5, 7.4, 8.1, 8.2, 8.5.1, 8.5.2, 8.6

Statistiekteksten: h. 3.

## Onderwerpen

- Variantie analyse (analysis of variance)
- Multiple comparison met Bonferroni-correctie

## Script code met RStudio/R

```
#####
```

```
#WC2
```

```
# Environment - Import Dataset - From text (base)... kies Table8.1.txt
```

```
attach(Table8.1)
```

```
names(Table8.1)
```

```
summary(Table8.1)
```

```
# Om duidelijk te maken dat de getallen in de kolom Group geen numerieke  
waarden voorstellen maar labels/groepen zijn, gebruiken we factor(Group).  
Wat R dan precies doet komt in een later werkcollege. Eerst de data plotten:
```

```
plot(factor(Group), Calc, xlab="Diet supplement", ylab="Calculus index")
```

```
# of
```

```
boxplot(Calc~Group, xlab="Diet supplement", ylab="Calculus index")
```

```
# gemiddelden en standaard afwijkingen zie display 8.1 blz 108
```

```
?tapply # helpinformatie
```

```
tapply(Calc, Group, mean) # berekent het gemiddelde van Calc per Group
```

```
tapply(Calc, Group, sd)
```

```
fit=glm(Calc~factor(Group), family=gaussian)
```

```
summary(fit)
```

```
#####
```

```
# Om de kwadraatsommen te krijgen gebruiken we drop1.
```

```
drop1(fit, test="F")
```

```
# Dit werkt net als bij de regressie:
# 1.) een model met een constante en Group. Dit is het volledige model. Er
gaat dan geen variabele uit dus drop is dan <none>. Dit is de eerste regel.
De residukwadraatsom is 3.1120
# 2.) een model waaruit Group is weggelaten, dus een model met alleen een
constante erin. De residukwadraatsom van dit model is SS-totaal. Hier zitten
de effecten van Group nog in. Omdat Group niet in het model zit, wordt hier
ook geen kwadraatsom voor uitgerekend. Dus in de res. kwadraatsom uit de
eerste regel zitten geen Group effecten meer, in de res. kwadraatsom uit de
tweede regel nog wel. Het verschil geeft dus de kwadraatsom voor group. Nu
is de anova table te maken. Zie ook Display 8.1 op blz 108.

# Ook van dit model kan je de residuen en de predicted values (de
groepsgemiddelden) krijgen door:

residuals(fit)
predict(fit)

# en de grafische relatie tussen predicted values en de residuen met

plot(predict(fit), residuals(fit), xlab="Predicted values
(cm)", ylab="Residuals (kg)", pch=20, col="red")

# Om te zien of er afwijkingen zijn van de normaliteit gebruiken we weer een
normal probability plot. Deze zet de geobserveerde residuen af tegen de
verwachte residuen als de residuen normaal verdeeld zouden zijn. Als de
residuen uit een normale verdeling komen zal dus deze plot punten laten zien
die op een rechte lijn liggen.

qqnorm(residuals(fit))
abline(0, sd(residuals(fit)))
detach(Table8.2)
```

## Opdrachten

### Opdracht 2.1

In een experiment is het glycogeengehalte bepaald van rattenlevers bij 3 verschillende behandelingen: een controlebehandeling, compound 217 en compound 217 plus suiker. De onderzoeksvraag was of er bij de twee laatste groepen een verlaging optreedt voor wat betreft het glycogeen gehalte t.o.v. de controle groep.

- Zoek in het toetsingsoverzicht op welke analyse geschikt is voor deze vraagstelling.
- Lees de dataset rattenlever.txt in.
- Bepaal per groep de gemiddelden en de standaarddeviaties.
- Maak een grafiek van het glycogeengehalte tegen de behandeling

- e. Gebruik de uitvoer van drop1 om een anova tabel te maken.
- f. Hoe luiden de hypothesen bij vraag c?
- g. Geef een antwoord op de onderzoeksvraag. Betrek daarin de groepsgemiddelden en geef commentaar op het aantal waarnemingen per groep.
- h. Geef de definitie van de onbetrouwbaarheid.

## Opdracht 2.2

Varkens blijken een verschillende mate van verzet (resistentie) te hebben. Om bij biggen dit verzet te meten is de zogenaamde rugtest gedaan 17 dagen na de geboorte. De uitkomst van deze test bepaalde of de big in een van de volgende 3 groepen terechtkwam:

Groep 1: low resistant

Groep 2: high resistant

Groep 3: middle resistant

Zo zijn in totaal 742 biggen ingedeeld in de groepen. Een tijd later is van deze varkens het vleespercentage bij slachten gemeten. De onderstaande tabel geeft een samenvatting van de resultaten:

Groep	Aantal biggen	Gemiddeld vleespercentage
1	348	53.51
2	227	54.67
3	167	53.94

De onderzoeksvraag was of biggen met een verschillende verzet, verschilden voor wat betreft het vleespercentage bij de slacht.

- a. Zoek in het toetsingsoverzicht op welke analyse geschikt is voor deze vraagstelling.
- b. Stel de hypothesen op die bij de onderzoeksvraag horen.

In onderstaande tabel zijn de 3 groepen met elkaar vergeleken.

	Sum of squares	df	Mean square	F	P-value
Between groups	...	...	...	...	0.0020
Within groups	10844.58	...	...		

Total	11028.94				
-------	----------	--	--	--	--

- c. Hoe noemt men deze analyse en aan welke voorwaarden dient te zijn voldaan wil deze analyse goed zijn?
- d. Op welke wijze zou je de aannamen kunnen controleren?
- e. Neem de tabel over en maak deze af.
- f. Geef de uitkomst van de toetsingsgrootheid uit de ANOVA tabel.
- g. Geef de p-waarde die bij deze uitkomst hoort.
- h. Wordt de nulhypothese verworpen? Geef een toelichting.
- i. Welke groep(en) wijken van elkaar af?

### Opdracht 2.3

Bij pluimvee werd een 100-dagen legtest uitgevoerd. Van twee rassen, A en B, werd een steekproef van 10 kippen genomen. Het aantal eieren is geteld in de 99 dagen volgend op het eerste ei.

- a. Zoek in het toetsingsoverzicht op welke analyse geschikt is voor deze vraagstelling.
- b. Lees de dataset pluimvee.txt in.
- c. Stel de hypothesen op.
- d. Bepaal het gemiddeld aantal eieren in de 100 dagentest voor beide rassen. Bepaal ook de standaarddeviatie.
- e. Druk de relatie tussen de 100 dagentest en de rassen grafisch uit.
- f. Bepaal m.b.v. de t-toets of er verschil is tussen beide rassen voor wat betreft hun gemiddelde (zie de R-commando's uit lijn 1, ?t.test).
- g. Stel voor de variantieanalyse (ANOVA) de hypothesen op.
- h. Gebruik drop1 om de anova tabel te maken. Geef de uitkomst van de toetsingsgrootheid en de bijbehorende p-waarde.
- i. Stel uw conclusie op voor de pluimveefokker.
- j. Bespreek de overeenkomsten tussen beide toetsen.
- k. Is hier gerandomiseerd? Welke consequenties kan dit hebben?
- l. Is de aanname van een normale verdeling redelijk?

# Werkcollege 3 - Lineaire modellen III : Uitbreidingen

## Inhoud

In dit werkcollege leer je hoe je de variantie analyse kunt uitbreiden met meerdere verklarende variabelen: lineaire regressie. Ook zul je ontdekken dat je deze regressie analyse kunt doen met verklarende variabelen die niet continue zijn, maar klasse variabelen.

## Voorbereiding

Bekijk de videoclips 'WC 3 – Lineaire modellen III' en 'WC3 – R script'.

Petrie & Watson (3rd ed): 11.1 - 11.3.

Statistiekteksten: h. 4

## Onderwerpen

- Lineaire regressie met meerdere onafhankelijke variabelen
- Variantieanalyse met meerdere factoren
- Indicator /klasse/groepsvariabele als onafhankelijk variabele

## Script code met RStudio

```
#####
```

```
#WC3
```

```
# Environment - Import Dataset - From text (base)... kies Display11.1
```

```
attach(Display11.1)
```

```
names(Display11.1)
```

```
summary(Display11.1)
```

```
# lees example 11.3.5
```

```
# de scatterplot xlab en ylab geven de labels op de assen
```

```
plot(Htirth, Bodywt, xlab="Heart girth (cm)", ylab="Body weight (kg)", pch=20, col="red")
```

```
fit=glm(Bodywt~Age+Htirth+Height+Length+Sex+Umbgth, family="gaussian")
```

```
summary(fit)
```

```
# in de kolom 'estimate' staan de intercept (constante) en de slopes/regressie coefficients;
```

```
# in de kolom 'Std. Error' staan de standard errors voor de schattingen.
```

```
# In de volgende kolom staan de t-toetsen daarna de p-waardes (Pr(>|t|)).
```

```
# Vergelijk met display 11.1 op blz 154 van het boek
```

```
# Om R^2 (kwadraat van de correlatie) uit te rekenen:
```

```
1- (fit$deviance/fit$null.deviance)
```

```

# 95% Betrouwbaarheidsintervallen voor de parameters:

confint(fi t)

#####
# Om de kwadraatsommen te krijgen gebruiken we drop1.

drop1(fi t, test="F")

# De residuen en de predicted values (de punten op de lijn) krijg je door

resi dual s(fi t)
predi ct(fi t)

# dus fi guur 11.1 krijg je

pl ot(Htgi rth, resi dual s(fi t), xlab="Chest gi rth (cm)", ylab="Residual s
(kg)", pch=20, col="red")
pl ot(predi ct(fi t), residuals(fi t), xlab="Predi cted values
(cm)", ylab="Residual s (kg)", pch=20, col="red")

# Om te zien of er afwijkingen zijn van de normaliteit gebruiken we een
normal probability plot. Deze plot residuen tegen wat verwacht mag worden
als de residuen normaal verdeeld zouden zijn. Als de residuen uit een
normale verdeling komen zal dus deze plot punten laten zien, die op een
rechte lijn liggen:

qqnorm(residual s(fi t))
abline(0, sd(residual s(fi t)))
detach(Displ ay11. 1)

# anova met indicator variabelen
attach(Table8. 1)
fi t=gl m(Calc~factor(Group))
summary(fi t)

# in de regel factor(group)2 staat in de kolom estimate, het verschil in
gemiddel de tussen groep 2 en 1
# in de regel factor(group)3 staat in de kolom estimate, het verschil in
gemiddel de tussen groep 3 en 1
# nu een regressie met 2 indicatoren. Eerst maken we de groep2 en groep 3
i ndi catoren:

?"==" # help over ==
gr2<-1*(Group==2) # indicator variabele voor groep 2
gr3<-1*(Group==3) # indicator variabele voor groep 3
fi t.i<-gl m(Calc~gr2+gr3)
summary(fi t.i)

#vergelijk met de anova uitvoer (fit)
detach(Table8. 1)

```

## Opdrachten

### Opdracht 3.1 (regressie met meerdere continue variabelen)

Er is een onderzoek verricht onder 39 paarden naar de hartslag (HR in slagen/minuut) bij zwemmende paarden. Naast de HR is het volgende gemeten:

speed:            snelheid in m/s  
sex:              0=ruin, 1=hengst  
goodbad:        een inschatting van hun prestaties, 0=bad, 1=good  
talent:           inschatting van de trainer 1=slecht, 2=gemiddeld, 3=goed  
trainer:         de trainer van het paard  
Age:             leeftijd van het paard  
swimrating:     indeling door jockey club; hoger is beter.

- Zoek in het toetsingsoverzicht op welke analyse geschikt is voor de vraag welke variabelen samenhangen met HR.
- Lees de dataset zwempaarden.csv in.
- Maak grafieken om een idee te krijgen van de samenhang tussen HR en de andere variabelen in de dataset.
- Maak een lineair regressiemodel met HR als afhankelijke variabele en met speed, age en swimrating als onafhankelijke variabele en interpreteer de estimates.
- Bepaal  $R^2$  voor het model onder c.
- Stel de hypothesen op die bij het regressiemodel horen.
- Welke richtingscoëfficiënten zijn significant?
- Geef de betrouwbaarheidsintervallen van de richtingscoëfficiënten en geef daarvan de interpretatie.
- Check de residuen op normaliteit (zie 10.4.3)

### Opdracht 3.2 (anova=regressie met indicatorvariabelen)

- Maak een anova model voor de variabele HR uit opgave 1 met Trainer en talent.
- Maak de anova tabel door gebruik te maken van drop1.
- Stel de hypothesen op die bij deze anova tabel horen.
- Geef de toetsingsgrootheden, hun uitkomst en de conclusies.
- Check de residuen op normaliteit en op homogene varianties (zie 10.4.3)
- Maak een summary van het regressiemodel.
- Interpreteer de estimates van talent.
- Geef de betrouwbaarheidsintervallen voor de estimates en interpreteer deze.

### Opdracht 3.3 (regressie met continue en groepsvariabelen)

In het eerste jaar diergeneeskunde hebben studenten een enquête onder katteneigenaren gehouden om informatie te verkrijgen over obesitas bij katten. De volgende variabelen zijn onder andere gemeten:

Gewicht: gewicht van de kat  
LftEig: leeftijd van de eigenaar  
GewEig: gewicht van de eigenaar  
BMI25: of de bodymass index van de eigenaar groter is dan 25 (0/1 variabele)  
LftKatDec: leeftijd van de kat in jaren (decimalen)  
Huishouden: grootte van het huishouden (1, 2 of >2 personen)

- a. Lees de dataset KattenLijn1.txt in.
- b. Maak een variabele die aangeeft of het huishouden uit meer dan twee personen bestaat: `groot_huish=1*(Huishouden=="Meer dan twee personen")`
- c. Maak een regressiemodel voor Gewicht met als onafhankelijke variabelen LftEig, GewEig, BMI25, groot\_huish en LftKatDec.
- d. Bekijk of de residuen normaal verdeeld zijn met een normal probability plot. Geef commentaar.
- e. Maak een plot van de residuen tegen de predicted values en bespreek deze.
- f. De katten met de hoogste residuen zijn nr 230, 186 en 79. Geef aan hoe dat komt.
- g. Geef een summary van het regressie model en bespreek de interpretatie van de resultaten.
- h. Geef een interpretatie van de estimate voor groot\_huish.
- i. Geef het betrouwbaarheidsinterval van het/de significante effecten.



# Zelfstudie 01 - Survivalanalyse en toetsingsoverzicht

## Inhoud

In deze zelfstudie behandelen we de analyse van overlevingsdata. Bijvoorbeeld data over hoe lang een dier nog overleeft na een operatie. Maar overlevingsanalyse kan ook gebruikt worden voor andere data waarin je de tijd tot een gebeurtenis wilt weten, zoals de tijd tot genezing, de tijd tot drachtig worden of de tijd tot ontwikkelen van ziekte. Deze techniek heet survival analyse en in deze zelfstudie zal je met de basis daarvan aan de slag gaan.

Naast overlevingsanalyse gaan we oefenen welke test gekozen moet worden voor welk soort data, aan de hand van het toetsingsoverzicht.

## Voorbereiding

Bekijk de videoclips 'ZS01 – Survivalanalyse' en 'ZS01 – R script'.

Statistiek teksten H. 6

Bestudeer het toetsingsoverzicht stroomschema (staat op blackboard onder 'course content' >> 'Statistiek').

## Onderwerpen

- Survival
- Censurering
- Hazard
- Kaplan Meier

## Script code met RStudio

```
#####
```

```
#ZS01
```

```
# Environment - Import Dataset - From text (base)... kies katten.txt
```

```
attach(katten)
```

```
names(katten)
```

```
summary(katten)
```

```
# De commando's voor survival analyse zijn niet automatisch bekend in R.
```

```
# Eerst moet de survival library geladen worden
```

```
library(survival)
```

```
# eerst bekijken we hoe de events en de censurering verdeeld is over de  
tijd, om een indruk te krijgen van de dataset:
```

```
table(tijd, tumor)
```

# Nu moet R verteld worden dat de afhankelijke variabele een combinatie is van de variabele tijd en de variabele die aan geeft of er een event is geweest of een censurering op dat tijdstip. Dus we maken om in R-termen te spreken een survival object met `Surv(tijd, tumor)` en gebruiken dat als afhankelijke variabele in de `survfit` functie die dan de Kaplan Meier-tabel uitrekent:

```
fit=survfit(Surv(tijd, tumor)~1)
summary(fit)
```

#Om de KM-grafiek te krijgen en de onder- en bovengrens van de betrouwbaarheidsintervallen:

```
plot(fit)
detach(katten)
```

## Opdrachten

### Opdracht ZS01.1

In een artikel over de overleving van baby-chimpansees werden de volgende gegevens gehaald. Het betrof een onderzoek naar de overleving van in totaal 65 baby-aapjes. Deze dieren werden gedurende een periode van 7 maanden gevolgd (Folia primatology 52: 167-177, 1989).

Levensmaand	Gecensureerd	Overleden
1	4	7
2	3	3
3	2	1
4	5	0
5	3	1
6	2	0
7	0	1

- Zoek in het toetsingsoverzicht op welke analyse geschikt is voor deze vraagstelling.
- Bereken de Kaplan-Meier overlevingscurve met de hand.
- Lees de dataset `babychimps.txt` in.
- Controleer uw antwoord van a.
- Zet de overlevingscurve met de betrouwbaarheidsintervallen uit.

### Opdracht ZS01.2

Een onderzoek bij beren (verzonnen dataset):

Een onderzoek naar de populatie van nog voorkomende beren en berensoorten op aarde, schatte enige jaren geleden dat er nog ongeveer 1000 zwarte beren op aarde leefden. Men besloot de populatie zwarte beren zo goed mogelijk te volgen. In gebieden waar ze veel voorkwamen en waar ze gemakkelijk waren te vangen en te voorzien van een halsband met zender, werd gedurende 5 jaren bijgehouden hoe het de zwarte beren verging. In een steekproef van 240 zwarte beren werd bij ieder zwarte beer een halsband omgehangen. Gemiddeld waren deze beren 2 jaar oud op het moment dat ze de halsband om kregen. Ieder kwartaal werd vastgesteld of de beer nog in leven was. Sommige beren raakten gedurende het onderzoek hun halsband kwijt. De gegevens zijn in de onderstaande tabel weergegeven samen met de berekeningen voor de Kaplan- Meier survival. De tijd geeft aan welk kwartaal het betreft, waarbij  $r$  het aantal dieren at risk aan het begin van die periode vertegenwoordigt en  $d$  en  $c$  respectievelijk het aantal dode beren en het aantal gecensureerde beren gedurende die periode weergeven.

Tijd (kwartalen)	$r$	$d$	$c$	$h$	$S(i)$
1	240	8	12	.0333	$D_1$
2	220	12	$B$	.0545	.9139
3	198	8	10	$C$	.8770
4	$A_1$	4	6	.0222	.8575
5	$A_2$	8	2	.0471	$D_2$
6	160	8	2	.0500	.7763
7	150	0	4	.0000	.7763
8	146	0	4	.0000	.7763
9	142	10	2	.0704	.7216
10	130	0	0	.0000	.7216
11	130	3	2	.0231	.7050
12	125	3	2	.0240	.6881
13	120	2	3	.1670	.6766

Op 6 plaatsen aangegeven met  $A_1$ ,  $A_2$ ,  $B$ ,  $C$ ,  $D_1$  en  $D_2$  zijn de getallen weggelaten.

- Bepaal welke getallen op plaats  $A_1$  en  $A_2$  moeten staan. Geef toelichting.
- Bepaal welke getallen op plaats  $B$  en  $C$  moeten staan. Geef toelichting.
- Bepaal welke getallen op plaats  $D_1$  en  $D_2$  moeten staan. Geef toelichting.
- Teken in een grafiek de overlevingscurve.
- Geef een maat voor de overlevingsprognose en bepaal hiervan de uitkomst.

### Opdracht ZS01.3

Aan de kliniek voor gezelschapsdieren is onderzoek gedaan naar de prognose van honden met een operatie ter correctie van een maagtorsie. In totaal was informatie bekend van 47 honden. Van deze honden is over een periode van tenminste 2 jaar na operatie bekend dat bij 17 honden sterfte optrad ten gevolge van de maagtorsie in de maanden: 1, 1, 1, 1, 2, 2, 3, 4, 5, 8, 9, 10, 10, 12, 16, 17, 19, terwijl bij 12 honden censurering optrad in de maanden 8, 8, 10, 10, 10, 12, 14, 17, 21, 21, 21, 22 na de operatie op  $t=0$ .

- Zoek in het toetsingsoverzicht op welke analyse geschikt is voor deze vraagstelling.
- Lees de dataset maagtorsie.txt in.
- Maak een tabel met de aantallen en bepaal de survivalfracties.
- Maak de grafiek met de Kaplan-Meier curve.
- Geef de 10-maandse overleving met het 95% betrouwbaarheidsinterval
- Geef de mediane overlevingsduur.

### Opdracht ZS01.4

Bepaal met behulp van het toetsingsoverzicht (zie blackboard) hoe u de volgende datasets zou analyseren.

- Maagtorsie is een zeldzame maar ernstige afwijking bij honden. De maag van de hond draait in de buikholte en zorgt voor een afsluiting zowel naar craniaal, als naar caudaal. De honden zijn vaak zeer acuut ernstig ziek, en kunnen snel overlijden. Er wordt gesuggereerd dat er een relatie is tussen de frequentie van voeren (vaak of niet vaak) en het optreden van deze aandoening. Een grotere voerfrequentie geeft mogelijk minder kans op het ontstaan van maagtorsies. Ook het soort voer zou van invloed kunnen zijn op de frequentie van maagtorsies.
- In de jaren '70 is onder warmbloedpaarden een onderzoek gedaan naar kreupelheid (podotrochleose). Gedurende een aantal jaren is uit de populatie "alle kreupele warmbloedpaarden" een steekproef getrokken van 1491 paarden en uit de populatie "niet-kreupele warmbloedhengsten" een steekproef van 1008. Van deze paarden is röntgenologisch de hoefkatrol beoordeeld. Deze beoordeling kon de uitkomst goed of slecht opleveren. De beoordeling is slecht wanneer er o.a. sprake was van ernstige diffuse ontkalking of verdichting, of van een cysteuze laesie. Van de niet-kreupele paarden hadden er 161 een slechte beoordeling, van de kreupele paarden waren dit er 910.

- c. Men wil het verband vastleggen tussen het bloedvolume in geiten ( $n=30$ ) en hun lichaamsgewicht. De afhankelijke variabele is bloedvolume.
- d. Van koeien die een keizersnede hebben ondergaan wil men weten hoe lang het gemiddeld duurt voordat de koe weer drachtig is. Van 112 koeien is na de keizersnede elke week bepaald of ze drachtig waren of niet.
- e. In een experiment is het glycogeengehalte bepaald van rattenlevers bij 3 verschillende behandelingen: een controlebehandeling, compound 217 en compound 217 plus suiker. De onderzoeksvraag was of bij de twee laatste groepen het gemiddelde glycogeen gehalte lager was dan in de controle groep.
- f. In een experiment is het glycogeengehalte bepaald van rattenlevers bij 3 verschillende behandelingen: een controlebehandeling, compound 217 en compound 217 plus suiker. Als het glycogeengehalte met minstens 20% was afgenomen te opzichte van de beginsituatie, werd dit aangemerkt als een daling. Als het gelijk bleef of steeg, werd dit aangemerkt als geen daling. De onderzoeksvraag was of bij de compound 217 en compound 217 plus suiker groepen vaker een daling optrad dan in de controle groep.
- g. In het eerste jaar diergeneeskunde hebben studenten een enquête onder katteneigenaren gehouden om informatie te verkrijgen over obesitas bij katten. Ze willen weten welke factoren samenhangen met het gewicht van de kat. De volgende variabelen zijn onder andere gemeten:
  - Gewicht: gewicht van de kat
  - LftEig: leeftijd van de eigenaar
  - GewEig: gewicht van de eigenaar
  - BMI25: of de bodymass index van de eigenaar groter is dan 25
  - LftKatDec: leeftijd van de kat in jaren (decimalen)
  - Huishouden: grootte van het huishouden
- h. In de jaren 1995 tot 1998 is er een onderzoek gedaan onder 1243 dalmatiërs. Er is gemeten of ze doof waren aan tenminste 1 oor (deaf; 0=nee, 1=ja). De onderzoeksvraag was of er een relatie tussen doofheid en pigmentatie is. De mate van pigmentatie werd gemeten met behulp van de volgende variabelen:
  - spot: aantal vlekken op de huid; 1=weinig, 2=matig, 3=veel
  - blueeye: heeft de pup blauwe ogen ; 0=nee, 1=ja
  - headspot: heeft de pup een vlek op het hoofd; 0=nee, 1=ja

## **Deel 2: ECONOMIE**

# Werkcollege 4 - Waarde, Kosten en Baten

## Inhoud

Dit werkcollege bestaat uit 2 delen, namelijk een opdracht over waarde (deel A), en een opdracht over kosten en baten (deel B).

## Vorbereiding

- Lees de syllabus: h. 4
- Bekijk de kennisclips economie 1-7.
- Voor studenten met achternaam A-K vul de volgende enquête in voor 20 oktober.
  - [https://docs.google.com/forms/d/e/1FAIpQLScxr-2AAW2HX6GEfbGRgeVwjCISSF6YwBjrJO9gQ9-K0dIKBA/viewform?usp=sf\\_link](https://docs.google.com/forms/d/e/1FAIpQLScxr-2AAW2HX6GEfbGRgeVwjCISSF6YwBjrJO9gQ9-K0dIKBA/viewform?usp=sf_link)
  - Vul deze enquête in met de aanname dat je een student bent zonder een bijbaan (je hebt dus weinig geld beschikbaar)
- Voor studenten met achternaam L-Z vul de volgende enquête in voor 20 oktober.
  - [https://docs.google.com/forms/d/e/1FAIpQLSfxbFMbTqumhhyrEBPW7xQAGPm7t9WuCyINCJkPIq6CnOxRMA/viewform?usp=sf\\_link](https://docs.google.com/forms/d/e/1FAIpQLSfxbFMbTqumhhyrEBPW7xQAGPm7t9WuCyINCJkPIq6CnOxRMA/viewform?usp=sf_link)
  - Vul deze enquête in met de aanname dat je een student bent met een bijbaan (je hebt dus geld beschikbaar, maar weinig tijd)
- Als je geen EXCEL ervaring hebt, bekijk dan de volgende filmpjes:
  - <https://www.youtube.com/watch?v=nXf86ikzOjA>
  - <https://www.youtube.com/watch?v=en20BxrpIRc>
  - <https://www.youtube.com/watch?v=3XKq7KDem2o>

## Deel A van het werkcollege

Als toekomstig dierenarts moet je later beslissingen gaan nemen over bijvoorbeeld welk tarief je gaat vragen voor een consult en welk tarief je gaat vragen voor een uur bedrijfsbegeleiding. Voor optimale economische resultaten van je praktijk is het belangrijk dat je inzicht hebt in hoe je klanten kosten waarderen en hoeveel ze willen betalen. Dit inzicht kun je krijgen door direct aan je klanten te vragen hoeveel ze zouden willen betalen. Dit resulteert echter in wenselijke antwoorden, waarbij altijd voor de laagste kosten wordt gekozen. Een betere methode om inzicht te krijgen in het waarderen van kosten is het uitvoeren van een (adaptive) conjoint analyse. In een conjoint analyse moeten namelijk door de respondent afwegingen

worden gemaakt, en dit zorgt ervoor dat er meer inzicht wordt verschaft in voorkeuren van respondenten (bijv voorkeuren voor prijzen, behandelingen, etc).

Een onderdeel van een conjoint analyse is het invullen van een enquête. Als voorbereiding op dit werkcollege heb je een enquête ingevuld over het behandelen van een hond. In dit werkcollege ga je zelf aan de slag met de uitkomsten van de enquête. Dit werkcollege geeft je inzicht in onder andere hoe kosten gewaardeerd worden door hondeneigenaren. Deel A van het werkcollege bestaat uit 6 opdrachten.

#### **Opdracht 4.1**

Uit welke 3 attributen bestond de enquête over het behandelen van een hond? Welke levels hoorden daarbij?

#### **Opdracht 4.2**

Welk attribuut had nog meer meegenomen kunnen worden in de enquête? Verzin minimaal 1 attribuut, en geef ook aan welke levels je bij dit attribuut zou kunnen definiëren.

Door middel van een lineaire regressie zijn de antwoorden van de enquête over het behandelen van de hond geanalyseerd. De afhankelijke variabele was de voorkeur score voor een behandeling en de onafhankelijke variabelen waren de attributen, die als klasse variabelen in het model zijn gedaan. De uitkomsten van deze analyse staan in Figuur 1 (zie "uitkomsten enquêtes voor werkcollege 4.pdf").

#### **Opdracht 4.3**

Wat zijn de schattingen ('estimates' of 'part-worth utilities') van alle levels van de attributen? Wat kun je concluderen op basis van deze schattingen?

#### **Opdracht 4.4**

Bereken de 'preference score' voor de 3 attributen? Welk attribuut is op basis van de 'preference score' het belangrijkste bij het behandelen van een hond, en welke het minst belangrijk?



Voor het invullen van de enquête had je de opdracht gekregen om de enquête in te vullen als een student met of zonder bijbaan. De schattingen van deze 2 situaties zijn weergegeven in Figuur 2 (zie "uitkomsten enquêtes voor werkcollege 4.pdf").

#### **Opdracht 4.5**

- a) Welke groep vindt lage kosten voor een consult het belangrijkste (studenten met of zonder bijbaan)?
- b) Welke groep vindt weinig behandeltime het belangrijkste (studenten met of zonder bijbaan)?
- c) Welke groep vindt het niet hebben van chronische pijn het belangrijkste (studenten met of zonder bijbaan)?

Door middel van de utilities heb je nu inzicht gekregen in hoe verschillende groepen mensen (studenten met of zonder bijbaan) kosten van een behandeling waarderen.

#### **Opdracht 4.6**

Is het (op basis van deze resultaten) aan te raden om voor studenten met of zonder bijbaan een ander tarief te hanteren? Motiveer je antwoord.

#### **Deel B van het werkcollege**

In het tweede deel van het werkcollege ga je zelf aan de slag om de kosten en baten van een melkveebedrijf uit te rekenen bij verschillende hoeveelheden overbezetting in de melkveestal. Dit werkcollege geeft je inzicht in de kosten en baten van een melkveebedrijf, maar ook in de marginale opbrengsten, marginale kosten en marginale winst. Deel B van het werkcollege bestaat uit 10 opdrachten.

Om dit deel van het werkcollege vlot te kunnen uitvoeren is het raadzaam om:

- Het EXCEL bestand werkcollege4B.xlsx te openen (staat op Blackboard) en te gebruiken
- Gebruik EXCEL voor het beantwoorden van de vragen, dus gebruik geen rekenmachine. Op het examen heb je echter geen EXCEL (maar wel een rekenmachine), en het is daarom het belangrijkste dat je snapt hoe berekeningen worden uitgevoerd.

### Overbezetting in de melkveestal

Een melkveehouder heeft een stal gebouwd met 140 ligplaatsen voor melkkoeien. De melkveehouder wil weten hoe veel koeien hij/zij zou moeten melken om de winst te maximaliseren. De melkveehouder vraagt jou om advies. Jij hebt literatuur doorzocht en hebt gevonden dat overbezetting in de stal leidt tot lagere melk productie per koe en een hogere voer behoefte per kg melk. Je schat dat de productiviteit veranderd als weer gegeven in tabel 1.

Tabel 1.

Aantal melkkoeien in de stal	Kg melk per koe per jaar	Kg droge stof voer per liter
100	9100	0,78
110	9050	0,78
120	9000	0,79
130	8750	0,79
140	8400	0,80
150	8000	0,82
160	7600	0,84

Voerprijs per kg ds	: <b>0.12 euro</b>
Melk prijs per kg melk	: <b>0.34 euro</b>
Andere variabele kosten per koe	: <b>450 euro</b>
Totale vaste kosten	: <b>150.000 euro</b>

#### Opdracht 4.7

Bereken de kosten, opbrengsten en winst bij elke bezettingsgraad van de stal en vul tabel 2 in (zie Excel file: werkcollege4B.xlsx).

#### Opdracht 4.8

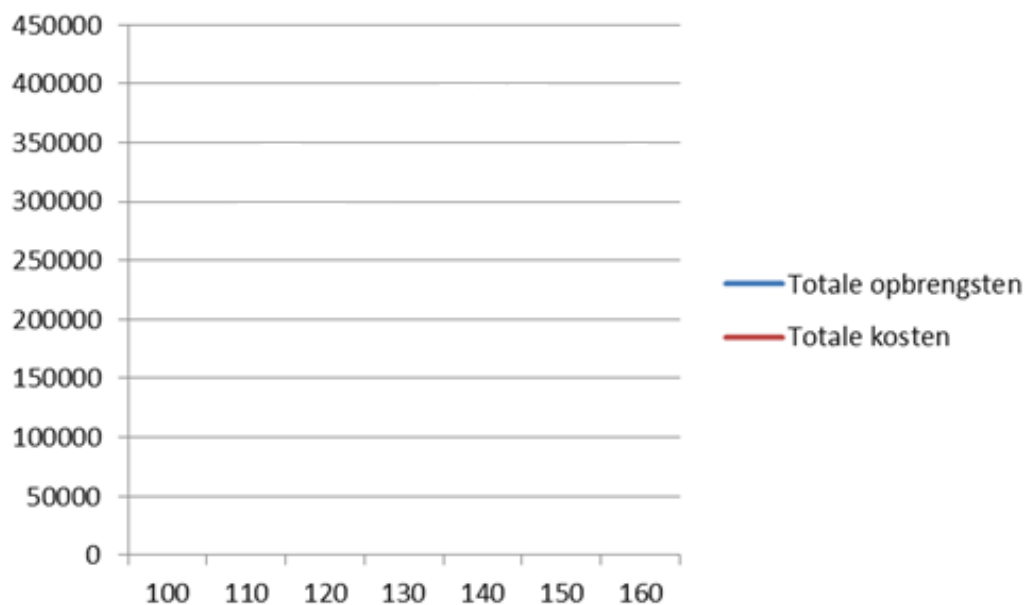
Wat zijn vaste kosten? Welke kostenposten op een melkveebedrijf zijn vaste kosten? Verzin minimaal 2 kostenposten.

#### Opdracht 4.9

Bereken ook de marginale opbrengsten, marginale kosten en marginale winst van elke bezettingsgraad en vul tabel 3 in (zie Excel file: werkcollege4B.xlsx).

#### Opdracht 4.10

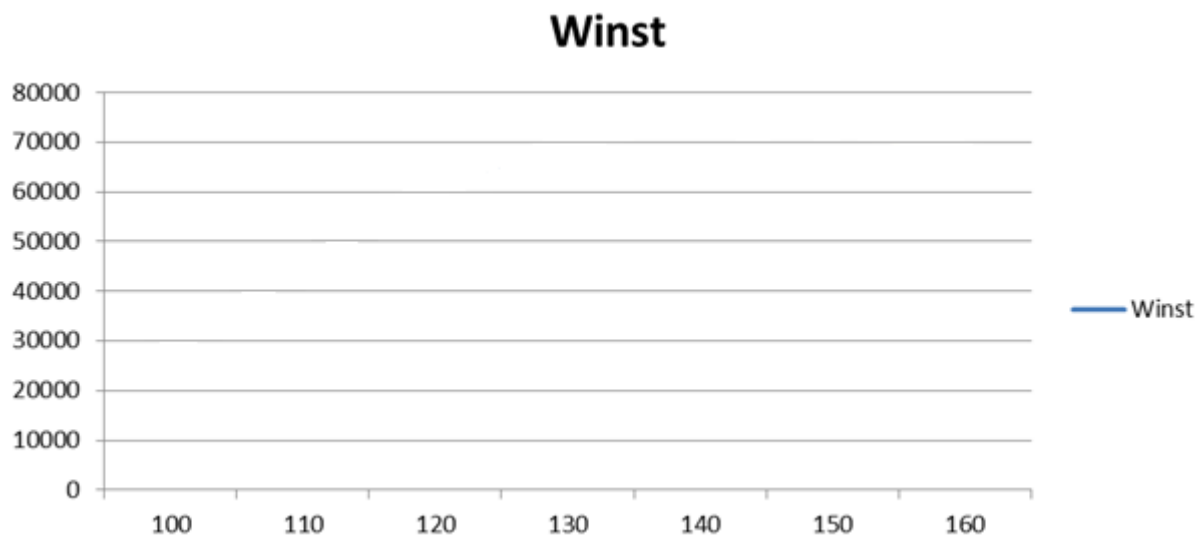
Plot de grafieken met totale kosten en totale opbrengsten met aantal melkkoeien op de x-as in figuur 1



*Figuur 1.*

### Opdracht 4.11

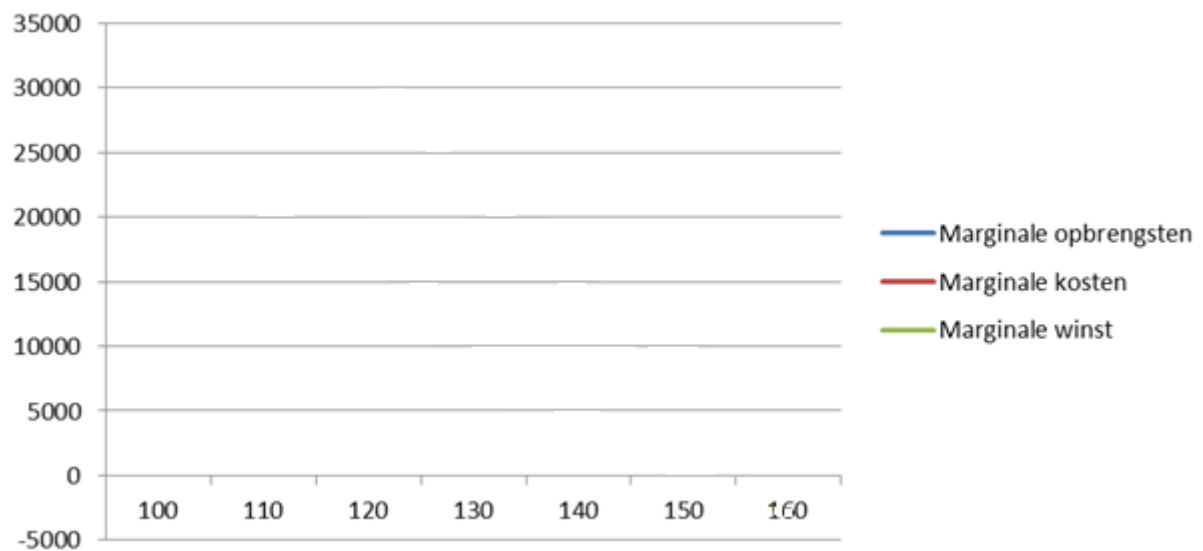
Plot nu de winst in figuur 2



*Figuur 2.*

### Opdracht 4.12

Plot de marginale opbrengsten, marginale kosten en marginale winst in één figuur



*Figuur 3.*

### **Opdracht 4.13**

Welke bezettingsgraad is de optimale? Waarom?

### **Opdracht 4.14**

Bekijk de kosten en de opbrengsten. Wanneer wordt de marginale winst curve negatief? Wat is het verband met de winst?

### **Opdracht 4.15**

De melkprijs fluctueert de laatste jaren flink. Neem aan dat de melkprijs fluctueert tussen de 30 en de 38 eurocent per kg melk. Hoe beïnvloedt dit je keuze bij vraag 4.13? (Tabel 4 en 5 (zie Excel file: werkcollege4B.xlsx) kunnen gebruikt worden om de marginale opbrengsten, marginale kosten en marginale winst opnieuw te berekenen)

### **Opdracht 4.16**

De dierenarts verwacht dat door overbezetting de gezondheidskosten zullen stijgen (door meer ziekteproblemen). Hij verwacht dat de variabele kosten zullen stijgen van €450 per koe naar €480 per koe vanaf 140 koeien in de stal.

Bereken opnieuw de overige variabele kosten, totale kosten en winst. Gebruik hiervoor Tabel 6 (zie Excel file: werkcollege4B.xlsx).

Bereken ook de marginale opbrengsten, marginale kosten en marginale winst bij de stijgende variabele kosten. Gebruik hiervoor Tabel 7 (zie Excel file: werkcollege4B.xlsx).

Welke bezettingsgraad is optimaal als inderdaad de gezondheidskosten stijgen vanaf 140 koeien? Hebben de gezondheidskosten invloed op de optimale bezettingsgraad?

# Werkcollege 5 - Rekenmethodes

## Inhoud

In dit werkcollege ga je zelf een aantal rekenmethodes uitproberen zodat je hiermee oefent en meer begrip voor deze methodes krijgt. Deze methodes zijn partial budgetting, kosten/baten analyse en beslisbomen. Deze methodes worden veel gebruikt en vormen de basis van het denken rondom veterinaire economie.

## Voorbereiding

- Lees de Syllabus: h. 4.
- Bekijk de kennisclips Economie 8-10.

Dit werkcollege bestaat uit vier delen:

1. Partial Budget van overbezetting op een melkveebedrijf
2. Kosten/baten analyse van de bestrijding van leucose
3. Een beslisboom over de behandelen van een hond met ziekte van Cushing
4. Een beslisboom over de behandeling van hernia bij de hond

De eerste twee opdrachten maken gebruik van EXCEL-spreadsheets. Opdracht 3 is een zwaar vereenvoudigd probleem, voornamelijk bedoeld om te oefenen met de software. Voor beslisbomen wordt namelijk een speciaal programma gebruikt, genaamd DATA (versie 3.0). Opdracht 4 is een iets ingewikkelder beslissingsprobleem, dat met behulp van DATA 3.0 opgelost wordt.

Let op: de formules van Economie komen niet op formuleblad dat je bij het tentamen krijgt!

## Opdrachten

### Opdracht 5.1 Partial Budget van verminderen overbezetting op een melkveebedrijf

Voor deze opdracht heb je de EXCEL file werkcollege5.xls nodig, deze staat op Blackboard. De gegevens die nodig zijn voor de berekeningen zijn opgenomen in de spreadsheet. Op basis van de gegevens wordt een partial budget voor de economische gevolgen van het verminderen van overbezetting op een melkveebedrijf bepaald.

In de spreadsheet voor dit werkcollege zijn een groot aantal gegevens ingevoerd. Aan de hand van deze gegevens wordt u gevraagd om zelf berekeningen uit te voeren. U kunt deze berekeningen met de hand uitvoeren en intypen in Excel, of u kunt de berekeningen door Excel zelf laten doen door middel van het invoeren van

formules in de rood omljnde cellen. Door het gebruik van formules, rekent Excel een verandering in invoer direct door. Als het antwoord juist is, verschijnt er een "Goed!" boodschap op het scherm.

### **Opdracht 5.2 Kosten/baten analyse van de bestrijding van leucose**

Ook voor deze opdracht heb je de file werkcollege5.xls nodig. In deze file staat het begin van een kosten/baten analyse voor een bestrijdingsprogramma. Vul deze file verder in en bepaal welk bestrijdingsprogramma economisch het aantrekkelijkst is voor de veehouder.

In de spreadsheet voor dit werkcollege zijn een groot aantal gegevens ingevoerd. Aan de hand van deze gegevens wordt u gevraagd om zelf berekeningen uit te voeren. U kunt deze berekeningen met de hand uitvoeren en intypen in Excel, of u kunt de berekeningen door Excel zelf laten doen door middel van het invoeren van formules in de rood omljnde cellen. Door het gebruik van formules, rekent Excel een verandering in invoer direct door. Als het antwoord juist is, verschijnt er een "Goed!" boodschap op het scherm.

### **Opdracht 5.3 Een beslisboom over de behandeling van een hond met ziekte van Cushing: Oefenen met DATA 3.0.**

Voordat het probleem en de opdracht gegeven wordt, worden eerst de basisbeginselen van DATA 3.0 uitgelegd. Dit gebeurt aan de hand van een zeer eenvoudig beslissingsprobleem:

Een hondeneigenaar wil weten welke van de 3 mogelijke behandelingen het goedkoopst is voor zijn hond met de ziekte van Cushing. De eigenaar heeft 3 opties:

1. Een operatie
2. Bestralen
3. Medicatie

Na de behandeling kunnen er twee dingen gebeuren, namelijk:

1. De hond geneest
2. De hond geneest niet

Als de hond bij bestralen niet geneest wordt de bestraling herhaald. Voor de eenvoud wordt aangenomen dat de eigenaar de hond laat inslapen als de hond niet geneest.

Gegeven is verder:

- Een hersenoperatie kost €500, en de kans op genezing is 90%.
- Het bestralen van de hond kost €400, en de kans op genezing is 60%. Als er na bestralen geen genezing optreedt wordt de hond opnieuw bestraald. Deze nieuwe behandeling heeft dezelfde kans op genezing en dezelfde kosten.
- Onderdrukken met medicatie kost €600 en de kans op genezing is 30%.
- Verder wordt er aangenomen dat als er geen genezing optreedt de dierenarts de hond laat inslapen (kosten €100).

Bereken met behulp van DATA 3.0 welke behandeling economisch gezien verstandig is. Het kan handig zijn, om de beslisboom eerst op papier te tekenen voordat deze in DATA 3.0 ingevoerd wordt.

#### **Het gebruik van DATA 3.0**

- Het programma, data32.exe, staat op Blackboard in de folder *Economie*. Je kunt het programma kopiëren naar je eigen gebied. Start het programma door dubbel te klikken op de filenaam.
- De muis is een belangrijk hulpmiddel in DATA 3.0. Het verdient aanbeveling deze te gebruiken.
- Een scherm verschijnt met daarin een rechthoek en rechts onder in de rechthoek een blauw blokje. Dit is het startpunt van het maken van een nieuwe boom. Het blauwe blokje staat voor een beslisknoop (decision node).
- Typ in de rechthoek de naam van het probleem dat je wilt analyseren, bijv. behandeling hond

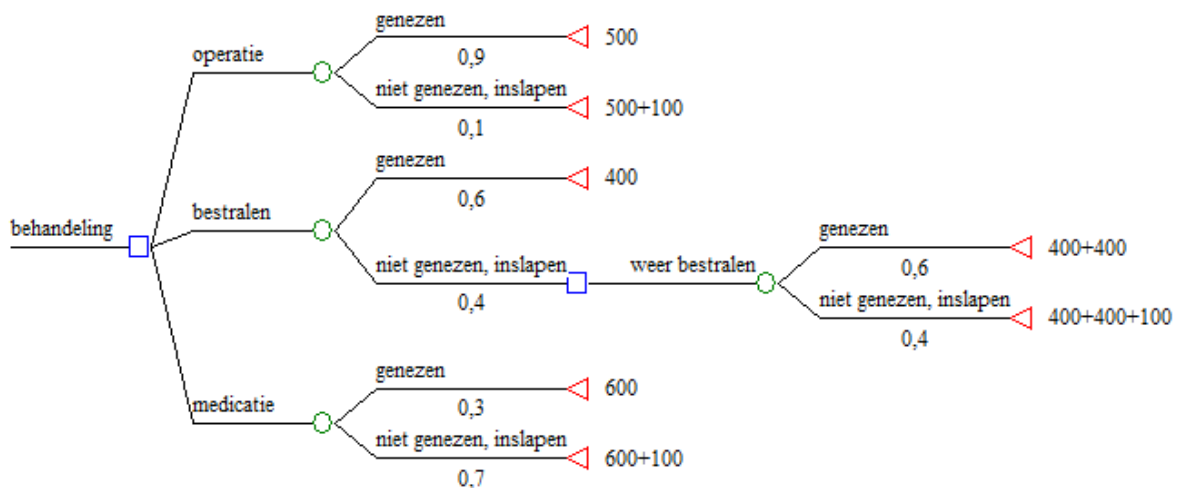


- Ga met de muis op het blauwe blokje staan; een symbool van een opsplitsende tak verschijnt. Dubbelklikken met links zorgt voor 2 takken vanuit het blokje, beide uitkomend op een groen rondje, de kansknoop (probability node). Nog een keer dubbelklikken levert een derde tak op.
- Voorzie nu elke tak van bijbehorende tekst. Doe dit door eenmaal te klikken met de linker muisknop op het eind van de tak (op het rondje). Een rechthoek verschijnt waarin de beschrijving van het alternatief kan worden ingevuld, bijv. operatie.
- Door op de kansknoop dubbel te klikken met de linker muisknop ontstaan 2 nieuwe takken. Doe dit voor beide alternatieven, immers beiden hebben twee uitkomsten.
- Voorzie elke tak van een omschrijving, t.w. de uitkomst van de kansknoop, dus genezing of geen genezing. Doe dit door de tak te selecteren door eenmaal klikken op de knoop (het rondje wordt groen en een rechthoek voor tekst verschijnt). Tekst toevoegen kan ook door boven de tak te klikken (ook dan verschijnt de rechthoek).
- Verder is het noodzakelijk om takken vanuit kansknopen te voorzien van kansen. De kansen dienen onder de tak te staan (de omschrijving erboven). Door onder de tak te klikken verschijnt er een rechthoek waarin de kans behorende bij de tak wordt genoteerd, bijv. 0.90. (Ook mogelijk om m.b.v. de TAB-toets te springen van omschrijving naar kans en weer terug). Belangrijk: Afhankelijk van de instelling van de computer moet je een komma of een punt gebruiken (dus 0,90 of 0.90), gebruik je het verkeerde decimale teken dan krijg je een foutmelding.
- Het aantal takken is nu compleet. Wat nog rest is totale kosten of opbrengsten per tak per tak. Eerst moet aangegeven worden dat de takken compleet zijn. Daartoe moet het type van de laatste kansknopen worden gewijzigd. Doe dit door met de rechter muisknop eenmaal op de knoop te klikken. Kies uit het menu "Change Node Type" (aanklikken). Kies vervolgens voor het type "Terminal" en OK. Vervolgens wordt gevraagd de financiële uitkomsten van deze tak te geven (scherm met ruimte voor payoffs). Voor operatie en genezing is dat 600 (de kosten van een operatie). Hier mag ook een formule ingevuld worden (bijv.  $600+100$ ). Het is voldoende om alleen payoff 1 van een waarde te voorzien.
- Voorzie op deze wijze elk eind van de takken van de financiële uitkomsten.
- Alvorens te rekenen dient aangegeven te worden hoe de oplossing moet worden berekend. Er kan geminimaliseerd (voor zo laag mogelijke kosten) of gemaximaliseerd (voor zo hoog mogelijke opbrengsten) worden. In dit geval wordt er met kosten gewerkt en wordt de VFW geminimaliseerd. Doe dit als volgt.
- Klik met de rechter muisknop ergens willekeurig op de ruimte om de boom heen. Er wordt een menu geopend. Selecteer op dit rolmenu "Preferences".

Onder edit (in de menubalk bovenin het scherm) zijn de “preferences” overigens ook te vinden.

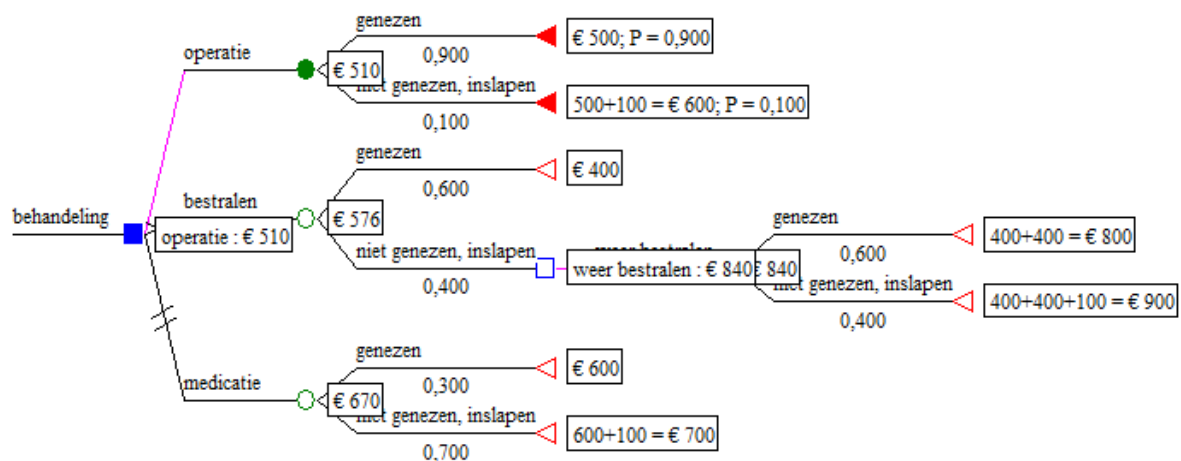
- Selecteer bij Category: Calculation method. En geef invulling aan de parameters van Calculation method: Method: simple; Use payoff: 1; Optimal path: low (immers minimaliseren van de kosten). Sluit af met OK.

Opdracht 5.3 moet geresulteerd hebben in de volgende beslisboom:



- Voer de berekeningen uit door met de rechter muisknop in de ruimte om de boom te klikken. Kies op het rolmenu RollBack (of klik op Analysis boven in de menubalk, en kies dan RollBack).
- RollBack werkt de formules uit en gebruikt de kansen om voor elk alternatief de VFW van de kosten te bepalen. Daarbij wordt aangegeven welk alternatief het optimale alternatief is.

Door RollBack te kiezen verschijnt de volgende boom met de bijbehorende VFW's



- Uit deze boom is af te lezen welke behandeling economisch gezien het aantrekkelijkst is.
- Door opnieuw voor RollBack te kiezen verdwijnen de VFW's e.d. en kunnen kansen en/of financiële uitkomsten worden gewijzigd.
- Test dit uit door de berekeningen te herhalen voor 60% genezing na een operatie.

#### **Opdracht 5.4 Welke behandeling voor een hond met een hernia**

Dit probleem betreft een economische afweging van de verschillende behandelmethoden nadat een dierenarts bij een hond hernia geconstateerd heeft. De volgende mogelijkheden tot behandeling werden gegeven:

- Volledige rust gedurende 4 weken, waarbij de hond door middel van kalmeringsmiddelen tot rust gedwongen wordt. Kosten van de kalmeringsmiddelen bedragen € 150,-. De kans op genezing (succes) is 40 %. Na afloop van de behandeling wordt tijdens een consult de genezing (of niet) vastgesteld. Kosten van een dergelijk consult zijn € 25,-.
- Opereren. Een hernia operatie is een moeilijke operatie en de kosten bedragen € 500,-. De kans op genezing is 90%.

Direct euthanaseren is natuurlijk ook een optie maar de eigenaar van de hond wil alles proberen om de hond te genezen.

Wanneer de hond na 4 weken rust niet genezen is, kan de rustperiode verlengt worden met 3 weken, kosten middelen bedragen € 100,- en ook na deze 3 weken is een consult nodig om te bepalen of het dier daadwerkelijk genezen is of niet. De kans op succes van deze tweede periode van rust is 30 %. Als na deze tweede periode van rust de hond nog steeds niet genezen is, zal langer rust niet meer helpen. Er kan dan alsnog geopereerd worden. De kans op succes bij de dieren die door rust niet genezen is 75 %.

De revalidatie na een operatie of langdurige rust is gelijk. Als opereren niet helpt zal het dier geëuthanaseerd moeten worden. Euthanaseren kost € 50,-.

Maak met behulp van bovenstaande gegevens een beslisboom in DATA 3.0 voor een hond waarbij je een hernia hebt geconstateerd. Bereken wat economisch gezien de beste behandelingsstrategie is

Wanneer een beslisboom eenmaal gemaakt is, kunnen factoren gewijzigd worden. Door dit wijzigen wordt de gevoeligheid van een beslissing voor onderliggende parameters (vaak zijn dit aannames) bepaald. Reken ter illustratie de volgende zaken uit:

- Wat is de optimale beslissing als een consult van de dierenarts € 35,- is in plaats van € 25,-?

- Wat is de optimale beslissing als de kans op genezing na 4 weken rust niet zo hoog is: 25 % in plaats van 40 % en dat pas na 7 weken rust meer kans op genezing is (40 % in plaats van 30 %).

## **Deel 3: EPIDEMI OLOGIE**

## Werkcollege 6 - Associatiematen en onderzoeksopzet

### Inhoud

In dit werkcollege maak je kennis met de odds ratio (OR), het relatieve risico (RR) en onderzoeksopzetten waarbij je deze associatiematen kunt gebruiken. Je gaat zelf een aantal van die associatiematen uitrekenen.

### Voorbereiding

Bekijk de volgende videoclips op Blackboard:

- 'Oorzaken van ziekte 1 tm 4'
- 'WC6 – Associatiematen en onderzoeksopzet'

Lees:

- Petrie & Watson (3rd ed): 4.7, 9.4
- Syllabus: H. 1
- Statistiek teksten EE, H. 5

De volgende formules zijn de basis voor dit werkcollege:

Disease status	Exposure status		Total
	Exposed	Non-exposed	
Diseased	$a_1$	$a_0$	$m_1$
Non-diseased	$b_1$	$b_0$	$m_0$
Total	$n_1$	$n_0$	$n$

$$p_1 = \frac{a_1}{n_1}$$

$$p_0 = \frac{a_0}{n_0}$$

$$OR = \frac{a_1/a_0}{b_1/b_0} = \frac{a_1/b_1}{a_0/b_0} = \frac{a_1 \cdot b_0}{a_0 \cdot b_1}$$

$$RR = \frac{p_1}{p_0} = \frac{a_1/n_1}{a_0/n_0}$$

### Opdrachten

Opdrachten 1 en 2 zijn voor tijdens het werkcollege, waarvoor korte eindantwoorden beschikbaar zijn op Blackboard. Opdrachten 3 en 4 zijn extra oefenopdrachten, waarvoor een uitgebreidere uitwerking beschikbaar is op Blackboard.

### Opdracht 6.1

Maagtorsie is een zeldzame maar ernstige afwijking bij honden. De maag van de hond draait in de buikholte en zorgt voor een afsluiting zowel naar craniaal, als naar caudaal. De honden zijn vaak zeer acuut ernstig ziek, en kunnen snel overlijden. Er wordt gesuggereerd dat er een relatie is tussen de frequentie van voeren en het optreden van deze aandoening. Een lagere voerfrequentie geeft mogelijk meer kans op het ontstaan van maagtorsies.

- a. Welke studiemethode is het meest geschikt voor de bestudering van de associatie tussen voerfrequentie en maagtorsie? Beargumenteer uw antwoord.

Van de 45 honden die aan de kliniek voor gezelschapsdieren werden aangeboden met een maagtorsie bleken er 8 te zijn die meer dan drie maal daags gevoerd werden. Bij een groep van 90 patiënten die om een andere reden bij de kliniek werden aangeboden bleken er 17 te zijn die meer dan drie maal daags werden gevoerd.

- b. Bereken met de hand of in Excel de meest correcte maat voor associatie tussen voerfrequentie en het optreden van maagtorsies. Maak hiervoor eerst de 2x2 tabel op basis van de gegeven data.

Bij verdere beoordeling van de gegevens bleek dat van de maagtorsie patiënten er 40 tot de grote rassen gerekend mochten worden, 7 van deze dieren bleek frequent gevoerd te worden. Bij de groep van 90 'overige' patiënten waren er 30 honden die tot de grote rassen gerekend mochten worden, van deze 30 werden er 15 frequent gevoerd.

- c. Bereken de meest correcte maat voor associatie tussen voerfrequentie en het optreden van maagtorsies bij de grote rassen. Idem voor kleine rassen.
- d. Interpreteer de resultaten van vraag (b) en (c). Betrek bij uw interpretatie ook de onderzoeksopzet (representativiteit).

## **Opdracht 6.2**

Bij katten in Nederland vond een uitbraak plaats van een zenuwaandoening (polyneuropathie) waarbij een partiële verlamming optrad. Van de in totaal 2 miljoen katten in Nederland bleken er 500 de aandoening gekregen te hebben.

- a. Bereken de ziektemaat die past bij het optreden van deze aandoening en bereken voor deze maat het 95% betrouwbaarheidsinterval (zie Petrie & Watson

paragraaf 9.3.1).

- b. Geef de interpretatie van een 95% betrouwbaarheidsinterval. Is dat in dit geval een logische interpretatie?
- c. Welke studieopzet zou u kiezen om erachter te komen wat de mogelijke oorzaak is voor de geobserveerde verschijnselen? Geef 2 argumenten.

Bij verder onderzoek bleek dat bij 53 van de 60 "ondervraagde dieren" met de zenuwaandoening een bepaald soort voer gebruikt ('type x') was. Bij dieren waarbij geen afwijking geconstateerd was bleek dat 17 van de 60 dieren dit voer kregen.

- d. Zet deze gegevens in een tabel, bepaal de relatieve frequenties en bereken een associatiemaat voor de relatie tussen het eten van het voer en het optreden van de aandoening. Interpreteer de resultaten.
- e. Is de associatie significant? Kun je met dit onderzoek zeggen of het voer werkelijk de oorzaak is van polyneuropathie, m.a.w. is er sprake van een causaal verband? Welke mededeling zou je aan de voerfabrikant geven?

Het is mogelijk om een dataset als deze, met een 0/1-variabele als afhankelijke variabele (dependent variable, in dit geval de zenuwaandoening) te analyseren met *logistische regressie*. Dit is een vorm van regressie analyse waarover je kunt lezen in de statistiektteksten EE, hoofdstuk 5. In dit geval is de onafhankelijke variabele ook een 0/1-variabele, maar je zou ook continue onafhankelijke variabelen kunnen gebruiken of zelfs meerdere typen onafhankelijke variabelen tegelijk, net als in lineaire regressie. Hieronder zie je het resultaat van de logistische regressie analyse met polyneuropathie (0/1) als afhankelijke variabele en voertype ('type x' versus ander voer) als onafhankelijke variabele:

```
> fit=glm(polyneuropathie~factor(voertype), family=binomial)
> summary(fit)
```

Call:

```
glm(formula = polyneuropathie ~ factor(voertype), family = binomial)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.68243	-0.54922	0.09835	0.74593	1.98298

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.8153	0.4076	-4.454	8.43e-06 ***
factor(voertype)type x	2.9524	0.4938	5.979	2.24e-09 ***

---

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



Onder 'Estimate' zie je de regressie coefficient van het effect van voertype x op het vóórkomen van neuropathie. Het leuke van logistische regressie is dat vanuit deze estimate (beta) een odds ratio berekend kan worden. Het blijkt namelijk zo te zijn dat de beta gelijk is aan de log-odds ratio van de samenhang tussen voertype en polyneuropathie. De inverse van het natuurlijke logaritmme ( $\ln$ ) heet de exponent ( $e^\wedge$ ). Door  $e^\wedge\text{beta}$  te berekenen, kan de odds ratio bepaald worden uit de output van de logistische regressie.

- f. Bereken de odds ratio van de associatie tussen voertype en polyneuropathie op basis van de gegeven logistische regressie.
- g. Bereken het 95% betrouwbaarheidsinterval rondom deze odds-ratio, uitgaande van de standaard normaal verdeling.

### **Extra Oefenmateriaal**

#### **Opdracht 6.3**

Op een melkveebedrijf heerst er een tekort aan het sporenelement selenium in het serum van rundvee. Selenium is een belangrijk element, dat bij een deficiëntie verschijnselen geeft van een slechtere immuunfunctie. Deze aandoening komt in Nederland voor, maar is nog wat onbekend bij veel practici. Er is op een bedrijf een volledig gerandomiseerd dubbel blind veld experiment uitgevoerd. Op dit bedrijf is er sprake van een slechte vruchtbaarheid. Het drachtpercentage na inseminatie voor het uitvoeren van het experiment bedraagt 30%. Het is te verwachten dat het drachtpercentage in de met selenium behandelde groep met 15 tot 20 procent zal stijgen.

Randomisatie vindt in deze studie dubbelblind plaats. De flesjes met selenium en de flesjes met placebo worden genummerd en alleen de fabrikant weet de code. Zowel de dierenarts als de veehouder is hiervan niet op de hoogte.

- a. Geef de belangrijkste reden waarom randomisatie een essentieel onderdeel is van klinische effectiviteit studies.

Bij de gegevensverzameling wordt ook bepaald of een koe mastitis heeft. De resultaten zijn: van de 53 dieren die met selenium behandeld werden kregen er 8 mastitis; van de 53 dieren die met een placebo behandeld werden kregen er 17 mastitis.

- b. Bereken de associatiemaat en toets of de gevonden relatie statistisch significant is. Kan je in dit geval spreken van een causaal verband?

Klinische effectiviteitstudies zijn een belangrijk onderdeel van het dossier wat ter beoordeling aangeboden wordt om een medicijn in Nederland geregistreerd te krijgen.

- c. Geef een aantal argumenten waarom klinische effectiviteitstudies meer of andere informatie leveren dan laboratoriumexperimenten.
- d. Waarom is het vaak moeilijk om een goede klinische effectiviteitstudie te doen?

#### **Opdracht 6.4**

In de jaren '70 is onder warmbloedpaarden een onderzoek gedaan naar kreupelheid (podotrochleose). Gedurende een aantal jaren is uit de populatie "alle kreupele warmbloedpaarden" een steekproef getrokken van 1491 paarden en uit de populatie "niet-kreupele warmbloedhengsten" een steekproef van 1008. Van deze paarden is röntgenologisch de hoefkatrol beoordeeld. Deze beoordeling kon de uitkomst goed of slecht opleveren. De beoordeling is slecht wanneer er o.a. sprake was van ernstige diffuse ontkalking of verdichting, of van een cysteuze laesie. Van de niet-kreupele paarden hadden er 161 een slechte beoordeling, van de kreupele paarden waren dit er 910.

- a. Welk type studie is hier gedaan? Geef toelichting.
- b. Zet de gegevens in een tabel.
- c. Geef een schatting voor de bij deze proefopzet behorende associatiemaat.
- d. Geef een interpretatie van de uitkomst.

De onderzoeksvraag was of de beoordeling van het hoefkatrol en kreupelheid met elkaar geassocieerd waren.

- e. Stel de hypothesen op die bij deze onderzoeksvraag horen.
- f. Toets de hypothesen uit vraag (e).
- g. Geef beargumenteerde kritiek op de onderzoeksopzet.

# Werkcollege 7 - Test eigenschappen

## Inhoud

In dit werkcollege ga je aan de slag met begrippen als sensitiviteit, specificiteit en positieve en negatieve voorspellende waarde. Je leert hoe deze test eigenschappen tot stand komen, en door welke factoren ze worden beïnvloed. Daarnaast leer je hoe je in de praktijk test eigenschappen moet interpreteren.

## Voorbereiding

Petrie & Watson (3rd ed): 14.2.1 - 14.2.3, 14.2.7.

Kuiper en Van Nieuwstadt: h. 2.

Syllabus h. 2.1 – h. 2.3

## Script code met RStudio

```
#####
```

```
# Hieronder vind je de code die je nodig hebt in werkcollege 7
```

```
#
```

```
# Environment - Import Dataset - From text (base)... kies cushing.long.txt
```

```
attach(cushing.long)
```

```
names(cushing.long)
```

```
summary(cushing.long)
```

```
table(cort_conc, cushing)
```

```
cort4<-ifelse(cort_conc>4, 1, 0)
```

```
tab<-table(cort4, cushing)
```

```
tab
```

```
prop.table(tab, 2)
```

```
prop.table(tab, 1)
```

```
binom.test(x,n) #voor het uitrekenen van een betrouwbaarheidsinterval
```

```
install.packages("pROC") #het installeren hoeft alleen de eerste keer
```

```
library(pROC)
```

```
help(pROC)
```

```
my_roc=roc(cushing, cort_conc)
```

```
summary(my_roc)
```

```
cbind(my_roc$thresholds, my_roc$sensitivities, my_roc$specificities)
```

```
plot.roc(my_roc)
```

```
detach(cushing.long.txt)
```

## Opdrachten

### Opdracht 7.1

U hebt een hond op uw spreekuur die u verdenkt van het syndroom van Cushing en u hebt recent van een collega gehoord dat je die diagnose heel goed kunt bevestigen door cortisol te bepalen in een haar van de hond. Dat lijkt u gemakkelijk, dus u verzamelt enkele haren van de hond voor verzending naar het laboratorium. Voor u het monster opstuurt zoekt u echter naar wetenschappelijke onderbouwing van deze nadere diagnostiek en vindt een artikel waarin met deze test 22 honden met de ziekte van Cushing worden vergeleken met 40 gezonde honden. Het cortisolgehalte is uitgedrukt in picogram/milligram haar.

Concentratie	Aantal	Aantal
Cortisol	Cushing	Gezond
0	1	3
1	1	11
2	2	11
3	2	5
4	5	6
5	2	2
6	0	2
7	0	0
8	2	0
9	2	0
$\geq 10$	5	0
Totaal	22	40

Op grond van deze tabel denkt u dat 4 pg/mg een goede afkapwaarde is, d.w.z. bij concentraties van 5 pg/mg en hoger spreekt u van een verhoogde cortisolwaarde, wat uw vermoeden voor het syndroom van Cushing bevestigt.

- Lees de dataset `cushing.long.txt` in en reproduceer bovenstaande tabel.
- Maak een tabel van de ziektestatus versus de test-status ( $\text{cortisol} > 4$ ).
- Wat is de sensitiviteit en specificiteit van de cortisolbepaling in hondenhaar voor het vaststellen van het syndroom van Cushing bij deze afkapwaarde? Geef ook de bijbehorende 95% betrouwbaarheidsintervallen.
- Interpreteer de uitkomsten onder c.

U vindt de sensitiviteit nogal laag, maar vanuit blok 19 herinnert u zich dat voor de individuele patiënt de voorspellende waarde van de test belangrijker is dan de sensitiviteit en specificiteit van de test.

- e. Bereken de voorspellende waarde van een positieve test (positive predictive value = PPV) en de voorspellende waarde van een negatieve test (negative predictive value = NPV).
- f. Interpreteer de uitkomst onder e.
- g. Gelden de door u onder f berekende voorspellende waarden onverkort voor de uitslag die u krijgt van de hond op uw spreekuur?
- h. Teken een ROC curve met de gegevens uit de tabel. Dit kun je doen in excel, of met behulp van R.
- i. Bij welke afkapwaarde is de sensitiviteit van de test groter dan 90%? Wat is de bijbehorende specificiteit?
- j. Zijn gezonde honden eigenlijk een goede controlegroep voor honden met het Syndroom van Cushing?
- k. Beschrijf hoe je een studie zou opzetten waarbij je de bruikbaarheid van deze test in de praktijk vast kunt stellen.
- l. Wat is uw conclusie over de bruikbaarheid van deze test?

## **Opdracht 7.2**

De eigenaar van een bedrijf met 500 schapen wil vrij worden van zwoegerziekte. Met behulp van serologie kunnen besmette dieren worden opgespoord ruim voor ze klinisch worden.

U besluit om eerst maar eens van een aantal dieren een serumbloedmonster te nemen om een indruk te krijgen van de prevalentie van de ziekte op het bedrijf.

- a. Welke dieren selecteert u hiervoor?

U kunt de monsters laten testen met een ELISA met een sensitiviteit van 95% en een specificiteit van 99% welke 5 euro per bepaling kost, of een agar gel immuno diffusietest (AGID) met een sensitiviteit van 99% en een specificiteit van 99,9% welke 20 euro per bepaling kost.

b. Welke test kiest u voor deze screening?

De gevonden prevalentie is hoog, 60% van de dieren scoort positief in de test.

c. Hoe hoog is de werkelijke prevalentie in de koppel?

d. Bereken de voorspellende waarde van een positieve test in deze koppel. Doe dit ook voor een negatieve test. Bespreek de interpretatie van deze waarden.

U laat de eigenaar vanaf nu een gesloten bedrijfsvoering voeren met strikte hygiënemaatregelen en nog een aantal andere maatregelen. Bovendien adviseert u hem om dieren met klinische verschijnselen in verhoogd tempo af te voeren. U verwacht dat daarmee de prevalentie zal afnemen.

Na een aantal jaar heeft u uit een nieuwe steekproef geleerd dat de werkelijke prevalentie ongeveer 2% is geworden.

e. Bereken de positieve en negatieve voorspellende waarden bij deze prevalentie. Hoe zijn deze veranderd?

U adviseert de veehouder de nog aanwezige positieve dieren af te voeren.

f. Hoeveel dieren bemonstert u nu om de laatste positieve dieren te vinden?

g. Uw veehouder is risico mijdend en wil zeker zijn dat elk besmet dier ook werkelijk wordt verwijderd. Hoe kunt u die zekerheid vergroten?

# Werkcollege 8 - Koppeldiagnostiek en serieel of parallel testen

## Inhoud

In dit werkcollege komen aspecten van koppeldiagnostiek aan de orde zoals redenen voor koppeldiagnostiek, selectie van dieren voor (aanvullende) diagnostiek en steekproefgrootte. Er worden twee voorbeelden van koppeldiagnostiek uitgewerkt. Daarnaast komt het serieel of parallel combineren van verschillende testen aan de orde.

## Voorbereiding

Petrie & Watson (3rd ed): 3.2.3 - 3.4.2.

Syllabus h. 2.4 - h. 2.5

## Formule

Binomiale kansverdeling

Kans op  $x$  successen bij  $n$  trekkingen, als de kans op succes per trekking gelijk is aan  $p$

$$\Pr = \binom{n}{x} p^x (1-p)^{n-x}$$

Bijvoorbeeld: kans op 8 positieve monsters in een totaal van 10 monsters van geïnfecteerde dieren, als de sensitiviteit 90% is:

$$\Pr = \binom{10}{8} 0.9^8 0.1^2 = \frac{10!}{8!(10-8)!} 0.9^8 0.1^2 = 0.194$$

Dit kan ook eenvoudig in R gedaan worden:

```
help(dbinom)
dbinom(8, 10, 0.9)
```

Voor het maken van verschillende opdrachten kun je als je wilt ook gebruik maken van het bestand 'Formules Werkcollege 8.xlsx' op blackboard.

## Opdrachten

### Opdracht 8.1

U wordt in het voorjaar om hulp gevraagd door de eigenaar van een stoeterij. Op deze stoeterij zijn 120 merries. Daarvan hebben 50 er al geveulend; de 70 andere merries zijn nog drachtig en zullen in de komende weken veulen. De eigenaar vertelt u dat twee veulens malaise, gebrek aan eetlust en diarree vertonen. U loopt langs alle veulens en ziet dat er in totaal 20 veulens zijn met symptomen. U denkt aan een bacteriële oorzaak (*Salmonella*). Om dat te bevestigen wilt u BO doen op mestmonsters.

- a. Welke dieren kiest u uit voor individuele diagnostiek en waarom?

U besluit om van drie van de 20 veulens onderzoek van het digestiestelsel te doen en aanvullend onderzoek op de mest.

- b. Waarom neemt u een steekproef en laat u niet mestonderzoek doen op monsters van alle 20 zieke veulens?

U wilt BO doen naar *Salmonella* spp. Aangezien u een geroutineerd dierenarts bent kunt u er van uitgaan dat u de juiste selectie hebt gemaakt voor dit ziektebeeld (op bedrijven met een *Salmonella* probleem selecteert u alleen dieren die de infectie ook werkelijk hebben).

- c. Beredeneer of u de koppel besmet zou noemen bij tenminste 1 van de 3 testpositief, tenminste 2 van de 3 testpositief, of tenminste 3 van de 3 testpositief.

In de praktijk wordt meestal geconcludeerd dat een koppel positief is bij tenminste 1 dier positief, dus dat zullen wij hier nu ook doen. Van de bacteriologische test weet u dat die een specificiteit heeft van 90%. De sensitiviteit van BO is bij klinisch zieke dieren 60%, bij subklinisch geïnfecteerde dieren 45%.

- d. Bereken de koppelsensitiviteit en de koppelspecificiteit, gebaseerd op de steekproef van 3 veulens.

Stel dat u de uitslag van het lab terugkrijgt: in 1 van de 3 monsters is *Salmonella* aangetoond. Volgens de gekozen drempel (positief koppel indien tenminste 1 positief monster) zouden we nu concluderen dat het koppel positief is.

- e. Hoe zeker weet u dat deze conclusie correct is?

(Reken daarvoor uit hoe groot de kans is op één BO-positief mestmonster wanneer de koppel werkelijk besmet zou zijn met *Salmonella* (en uw selectie correct) en de kans op een dergelijke uitslag als *Salmonella* niet de oorzaak is van de problemen)

U vertelt de eigenaar dat merries symptoomloze dragers kunnen zijn van *Salmonella*. Aangezien de komende weken nog meer veulens worden geboren, wil de eigenaar dat van alle nog drachtige merries mest wordt onderzocht op



aanwezigheid van *Salmonella*. Hij zou ze dan kunnen behandelen om zo het risico van infectie bij de nieuwe veulens te verkleinen.

Stel dat 20% van de merries besmet is met *Salmonella*.

- f. Als nu bij alle nog drachtige dieren BO op mest wordt gedaan, hoeveel van de geteste dieren zullen dan positief zijn in de test?
- g. Bereken de voorspellende waarde van een positieve test in de symptoomloze dragers.
- h. Welk plan van aanpak zou u voorstellen op grond van de uitslag?

### Opdracht 8.2

De schapenhouder van het schapenbedrijf uit opdracht 7.2 wil graag een inschatting van de kosten van het vrijmaken van de kudde, aangezien de prevalentie al tot 2% gedaald was en hij nu helemaal vrij wil worden. Ter herinnering, de gegevens van de 2 testen:

	AGID	ELISA
Sensitiviteit	0.99	0.95
Specificiteit	0.999	0.99
Prijs	€20	€5

- a. Bereken het verschil in kosten voor de schapenhouder bij gebruikmaken van de ELISA en gebruik makend van de parallelle test ELISA - AGID. Ga er daarbij vanuit dat een onterecht afgevoerd schaap 50 euro kost.
- b. Bereken de kans dat er nog een of meer besmette schapen aanwezig zijn bij gebruik van de ELISA. En hoe groot is die kans bij de parallelle combinatie-test?

Het bedrijf wordt uiteindelijk vrijverklaard en krijgt het zwoegerziekte-vrij certificaat. U wilt i.v.m. de certificering jaarlijks een steekproef nemen om te laten zien dat het bedrijf nog steeds vrij is.

- c. Welke zaken spelen een rol bij de omvang van de steekproef?
- d. Welke eisen wat betreft sensitiviteit en specificiteit stelt u nu aan de test?

In het testschema voor zwoegerziekte moet uw veehouder nu eenmaal per twee jaar 129 schapen laten onderzoeken.

- e. Bereken de koppelspecificiteit bij gebruik van de onder 7.2. beschreven ELISA. Wat vindt u hiervan?
- f. U besluit serieel te gaan testen met de AGID om de koppel specificiteit te verhogen, wat houdt dit in?
- g. Bereken de koppelspecificiteit bij gebruik van de seriële test ELISA – AGID.
- h. Bereken de koppelsensitiviteit van de hierboven gevolgde testprocedure onder de aanname van het testschema dat de prevalentie in de koppel 2% is.

De veehouder vindt de kosten van de heronderzoeken nogal hoog.

- i. Wat kunt u doen om de kosten te reduceren bij een gelijk aantal bemonsterde dieren? Wat zijn de nadelen?

# Zelfstudie 02 - Populatiedynamica van infectieziekten

## Inhoud

In dit werkcollege komen begrippen uit de populatiedynamica aan de orde zoals reproductie ratio, infectieuze periode, en transmissie van kiemen. Er worden wat eenvoudige berekeningen gedaan met modellen die veel worden gebruikt in de epidemiologie. Toepassingen van deze modellen worden geïllustreerd aan de hand van voorbeelden.

## Voorbereiding

Lees uit de Syllabus hoofdstuk 3.

Kijk de kennisclips over Populatie Dynamica, behorend bij ZS02.

In Teams, in het channel VU02 Vragenuurtje Populatiedynamica kun je vragen stellen over ZS02. Stel je vragen uiterlijk de dag voor het vragenuurtje, dan worden ze op het vragenuurtje behandeld door de docent.

## Opdrachten

### Opdracht ZS02.1

Voor twee virusinfecties bij varkens is een schatting gedaan voor de reproductieratio. Voor het Aujeszkyvirus (ADV) is een  $R_0$  geschat van 10, voor klassieke varkenspestvirus (KVP) een  $R_0$  van 5. De generatietijd van ADV is ongeveer 7 dagen, van KVP ongeveer 10 dagen. Stel er komt virus binnen op een vleesvarkensbedrijf met 1000 varkens. Geen van de varkens is gevaccineerd.

- Welk agens verspreidt zich het snelst op het bedrijf, en waarom?
- Welke verklaringen zou u kunnen geven voor dit verschil in  $R_0$ ? Denk aan eigenschappen van de ziekte die tot uitdrukking komen in het SIR-model.

De KVP-epidemie van 1997 – 1998 heeft een enorme impact gehad. Na de epidemie werden tal van maatregelen en monitoringsprogramma's voorgesteld om de kans op weer zo'n epidemie of de gevolgen van een eventuele introductie te verkleinen. Monitoringsprogramma's hebben als doel de infectie zo snel mogelijk op te sporen, zodat de bestrijding kan aanvangen als de epidemie nog klein is. Veel programma's hangen af van de oplettendheid van de boer, die zijn dierenarts zal waarschuwen als zijn dieren ziek zijn.

- Hoe snel na introductie van KVP op een vleesvarkensbedrijf van 1000 varkens zal een boer gemiddeld zijn dierenarts waarschuwen? Doe hiervoor een aanname over het aantal zieke dieren waarbij een boer belt, en bereken hoe snel na introductie van het virus dit aantal bereikt zal zijn.

De KVP uitbraak in 1997 werd nu niet direct snel opgespoord. Op het moment dat het bekend werd waren er al zo'n 30 bedrijven besmet, al was dat op dat moment natuurlijk nog niet bekend. Stel nu dat de maatregelen die toen werden ingesteld wel zo effectief waren geweest dat de  $R_h$  (reproductieratio tussen bedrijven) zou zijn gedaald tot 0,6.

- d. Bereken hoeveel bedrijven er in totaal besmet zouden zijn geweest aan het eind van de epidemie?

### **Opdracht ZS02.2**

Tijdens de phocine distemper virus (PDV) epidemie onder zeehonden (met 90% mortaliteit) in de Waddenzee in 1988 is een  $R_0$  geschat op 2,5 (De Koeijer et al. 1998). Het virus was in 1989 uit de zeehondenpopulatie verdwenen.

- a. Waardoor kan een infectieziekte als PDV vanzelf uit de populatie verdwijnen?

Een nieuwe PDV epidemie vond plaats in 2002. Het virus was niet heel anders in vergelijking tot 1988, dus er werd al snel beweerd dat ook nu het virus wel zou verdwijnen uit de populatie.

- b. Wat zou er veranderd kunnen zijn t.o.v. 1988 waardoor deze redenering niet hoeft te kloppen?

### **Opdracht ZS02.3**

Vossen staan bekend als een reservoir voor het rabiesvirus. Om het risico op besmetting voor honden en mensen te beperken, wordt er aas met daarin een oraal vaccin verspreid in vossenpopulaties. Het uiteindelijke doel hiervan is om rabies uit de vossenpopulatie uit te roeien.  $R_0$  van rabies in vossen is 3,0 (Brochier et al, 1991), de infectieuze periode is een half jaar.

- a. Welk percentage van de populatie moet worden beschermd om dat doel (uitroeien) te bereiken?

In het verleden werd rabies bestreden door vossen te jagen.

- b. Hoe verwacht u dat dit de transmissie van rabies (en dus  $R_0$ ) zal beïnvloeden? Bedenk hiervoor welke factoren die  $R_0$  bepalen veranderen (paragraaf 4.4).

- c. Hoe verwacht u dat de prevalentie van rabies door het jagen zal worden beïnvloed? Motiveer uw antwoord.

#### **Opdracht ZS02.4**

IBR, veroorzaakt door BHV1 virus, is een endemisch voorkomende aandoening in de Nederlandse rundveepopulatie, met een geschatte  $R_0$  van 2.5. Besmette dieren kunnen levenslang drager blijven van de infectie. Begin jaren 90 is overgegaan tot een verplichte vaccinatie tegen het BHV1 virus en men hoopte op die manier het virus te kunnen uitroeien. Een vaccin moest gE-negatief zijn (markervaccin) en zou worden toegelaten tot de markt nadat het was getest op werkzaamheid. De bijbehorende diagnostische test was een gE-ELISA.

Een vaccin wordt getest in een vaccinatie-challenge experiment: dieren worden eerst gevaccineerd en dan na bijvoorbeeld één maand opzettelijk besmet (geïnoculeerd) met een wildtype BHV1 stam. Na inoculatie zijn de gevaccineerde runderen klinisch beschermd (d.w.z. ze worden niet meer ziek), maar ze scheiden nog wel virus uit. De onderzoeker concludeert dat met dit vaccin het virus niet kan worden uitgeroeid bij runderen.

- a. Bent u het eens met deze conclusie? Waarom wel/niet?

Mars et al. hebben een gE-BHV1 markervaccin in het veld getest. De geschatte  $R$ -waarde in gevaccineerde koppels was 1,2, dus ongeveer een halvering t.o.v. 2,5 zonder vaccinatie.

- b. Beredeneer hoe de waarde van  $R_h$  tussen bedrijven zal veranderen door vaccinatie. Bedenk hoe de virusuitstoot tijdens uitbraken van IBR binnen bedrijven verandert als gevolg van vaccinatie, en hoe dat  $R_h$  verandert. Zal  $R_h$  ook ongeveer halveren, of is de daling minder groot of juist groter?

#### **Opdracht ZS02.5**

De ziekte van Aujeszky kwam in de jaren 70 en 80 van de vorige eeuw endemisch voor in Nederland. In 1993 is een bestrijdingsprogramma gestart om Nederland vrij te krijgen van het oorzakelijk agens, t.w. het Aujeszky virus (ADV). Het middel daartoe was verplichte vaccinatie met een gE-negatief (marker) vaccin.

Op het grootste deel van de varkensbedrijven in NL werd gevaccineerd. Enkele vermeerderingsbedrijven konden een vaccinatie-ontheffing krijgen, maar daar moest worden gemonsterd om afwezigheid van infectie aan te tonen. Bedrijven die

niet exporteerden moesten 12 monsters per kalendermaand laten onderzoeken op antistoffen tegen ADV. Dit is voldoende om 25% prevalentie aan te tonen<sup>1</sup>, zowel op grote als op kleine bedrijven.

- a. Beredeneer het verschil tussen een klein en een groot bedrijf in de snelheid van opsporen van een beginnende infectie met ADV, als je bij bemonsteren 25% prevalentie kan aantonen.

De eigenaar van zo'n bedrijf met vaccinatie-ontheffing liet u uitslagen zien uit een maand in 2006. Een van de 12 monsters bleek seropositief te zijn.

- b. Welke verklaring(en) heeft u voor deze uitslag: 1 positief monster van het totale vereiste aantal?
- c. Wat had een vervolgstap kunnen zijn om erachter te komen of het bedrijf wel of niet werkelijk besmet is?

---

<sup>1</sup> met 95% koppelsensitiviteit

## Zelfstudie 03 - Critical appraisal / publicaties lezen


### Inhoud

In de zelfstudie oefen je kritisch lezen van wetenschappelijk artikelen, met het doel artikelen te kunnen plaatsen op een schaal van zwak naar sterk bewijs ('critical appraisal' stap in Evidence-based Veterinary Medicine (EBVM)).

Tijdens het vragenuurtje VU03 kun je onduidelijkheden bespreken met elkaar en de docent, en kun je discussie voeren over vraag 4 en 5. In Teams, in het channel VU03 Vragenuurtje Critical appraisal kun je vragen stellen over ZS03. Stel je vragen uiterlijk de dag voor het vragenuurtje.

### Zelfstudie


De materialen zijn te vinden op BlackBoard bij ZS03

- 5 kennisclips behorende bij ZS03
- FeedbackFruits  P&W 18.1-18.3
- Zelfstudie Bille (pdf 18.1-18.3 & pdf artikel met antwoordmodel)
- Zelfstudie Lori (pdf 18.1-18.3 & pdf artikel met antwoordmodel)

### Opdracht ZS03.1

Bekijk de 5 kennisclips van ZS03

### Opdracht ZS03.2

Bestudeer Petrie & Watson (3<sup>rd</sup> ed.): 18.1 – 18.3 in FeedbackFruits 

### Opdracht ZS03.3

Lees het 'Abstract' en 'Material and Methods' deel van de paper van Lori et al met behulp van de vragen uit 18.3 (3) voor 'M&M' en controleer je eigen antwoorden met het antwoordmodel (achter artikel in zelfde pdf).

### Opdracht ZS03.4

Lees het 'Abstract' en 'Results' deel van de paper van Bille et al met behulp van de vragen uit 18.3 (4) voor 'Results' en controleer je eigen antwoorden met het antwoordmodel (achter het artikel in zelfde pdf).

### Opdracht ZS03.5

Formuleer voor jezelf de kwaliteit van beide studies. Let daarbij op de volgende aspecten:

- Waar bevinden ze zich in de pyramide van bewijs
- Is er 'selection bias'

- Is er 'information bias'
- Is er 'confounding bias', al/niet met analytische correctie
- Is een passende statistische techniek gebruikt
- Zijn het binnen hun studietype sterke of zwakke studies

### **Opdracht ZS03.6**

Als deze 2 studies het enige wetenschappelijke bewijs zouden zijn, wat zou je dan adviseren aan de eigenaar over:

- De keus voor soort chemotherapie bij een hond met multicenter lymfoom
- Operatie of geen operatie voor een 12-jarige kat met een botbreuk in verband met risico op overlijden tijdens narcose