

ANDMETE KOGUMINE

Eesmärk

- Andmete kogumise eesmärk on koguda vajalikku ja kehtivat informatsiooni tunnuste kohta, mis on seotud loomade haiguste, tootlikkuse ja neid määravate teguritega, loomadelt, proovidest ja loomade keskkonnast (looduslikud tegurid ja pidamistingimused).

Nõuded

- Informatsiooni kogutakse sellisel viisil, et juhuslik ja süstemaatiline viga on lubatavates piirides.
- Informatsiooni saab püsivalt salvestada ja seda saab uuesti välja otsida, töödelda ning edastada minimaalsete jõupingutustega.
- Informatsioon on jälgitav läbi kogumise, töötlemise, analüüsimise ja tõlgendamise protsessi.

Meetodid

- Andmete kogumise meetodid sõltuvad andmete kogumise esmasest protsessist ja informatsiooni allikatest.

Erinevad uuringud annavad erinevaid andmeid. Siiski, epidemioloogilise uuringu sellistele tüüpidele, nagu näiteks läbilõike-, juht-kontroll- või kohort- ja levimuse uuringud, on ühiselt omane, et huviobjektiks ei ole mitte ainult sihttunnus (näiteks mastiidi esinemine) ja üksik tegur (näiteks nisakahjustused). Teaduslik lähenemine loomade haiguste uurimisele näeb ette informatsiooni kogumist teguri mõju potentsiaalselt muutvate ja segavate faktorite kohta. Väga sageli on uuringu all olev nähtus (haigus, tootlikkuse vähenemine, viljatus jne) paljuteguriline ja seetõttu tuleb lisaks sihttunnusele registreerida paljusid tunnuseid. Arvesse tulevad järgmised andmete allikad.

BIOLOOGILISED JA KLIINILISED ANDMED

Allikas. Need andmed kogutakse laudas ja need puudutavad üksikut looma (mõnikord loomarühma). Näited: tõug, vanus, sugu, kliinilised tunnused, kehakaal ja konditsioon.

Meetod. Informatsiooni saadakse tunnuste, koemuutuste ja füüsiliste omaduste mõõtmise või subjektiivse (kuid teadliku) hindamise kaudu. Informatsioon registreeritakse tabeli vormis paber kandjale (joonised 33, 34). Tavaliselt kogutakse erinevate farmide andmed erinevatele lehtedele. Loomade ja farmi identifitseerimise andmed peavad olema dokumenteeritud.

DIAGNOSTILISED JA LABORATOORSED ANDMED

Allikas. Need andmed saadakse isenditelt või loomarühmadelt (koondproovid) võetud proovide edasiste uuringute tulemusena. Diagnostilised meetodid võivad olla lihtsad ja kohapeal teostatavad (näiteks California mastiiditest) või vajada laboritingimusi (näiteks seroloogilised testid või polümeraasi ahelreaktsioon).

Meetod. Laudas registreeritavad diagnostilised andmed kantakse andmelehtedele vastavalt eespool kirjeldatule. Laboratoorsete analüüside tulemused kantakse tabelisse laboratooriumi registritest (joonised 35 ja 36). Kasulikud on instrumendi- ja arvutiliidesed, kui neid on võimalik kasutada. Huvi pakuvad ainult lõpptulemused (pärast täpsustamist ja korrigeerivate faktorite rakendamist). Iga proov peab olema tuvastatav. See tähendab, et dokumenteeritud peavad olema looma ja farmi identifitseerimise andmed ja proovi võtmise kuupäev.

FARMITASANDI ANDMED

Allikas. Farmi (tootmisüksuse) lähem uurimine ja intervjuu farmeriga annavad väärtuslikku informatsiooni farmitasandil mõjuvate tegurite kohta. Intervjueeriya vaatlustele põhinedes saab registreerida lisainformatsiooni. Kõik ühe farmi kohta käivad andmed kehtivad selle farmi loomade suhtes.

Meetod. Andmeid kogutakse küsimustikke kasutades. Farmeritele või farmi personalile suunatud küsimustikud ei tohi sisaldada tehnilist ega meditsiinilist terminoloogiat. Küsimustikud ei tohi olla sõnastatud selliselt, et annavad mõista eeldatavaid vastuseid. Küsitletavate koostöövalmidust, küsimuste eesmärgipärasust ja arusaadavust tuleb enne kasutamist testida.

VAREM KOGUTUD INFORMATSIOON

Allikas. Valitsuse aruanded, projektide aruanded, ametlik statistika, loenduste andmed, farmerite ühenduste andmed, publikatsioonid.

Meetod. Oluline informatsioon registreeritakse vastavalt konkreetsele eesmärgile.

MUUTUJATE LIIGID

Binaarsed (dihhotoomsed):

muutujal on ainult kaks võimalikku tulemust (0/1, jah/ei).

Nominaalsed (kategoriaalsed):

muutujal on piiratud arv võimalikke tulemusi (kuiv/märg hooaeg, farmi number).

Järjestusandmed (ordinaalsed, semikvantitatiivsed, järjestatud kategooriad):

muutujal on piiratud arv võimalikke tulemusi. Tulemusi saab reastada mingi loogilise korra alusel (-/+/++/+++), ei/harva/korduv ravi).

Mitmevalikulised muutujad (multiple-choice variable):

võib kasutada esmaseks andmete kogumiseks (kasutades paberkaartide). Vt näiteks muutuja "Nisakahjustused" joonistel 33, 34. Lubatud kanded sellesse muutujasse on EV, EP, TV ja TP udara nelja veerandi kohta. Siiski võib ühel lehmal olla kahjustusi rohkem kui ühel nisal. Tegelikult võivad ühel lehmal olla kõik nidad kahjustusega. Sellisel juhul oleks sissekanne lahtrisse "Nisakahjustused" "EV, EP, TV, TP". Muutuja on mitmevalikuline. Pange tähele, et see kodeering on aktsepteeritav üksnes esmaste andmete kogumise jaoks. Niipea, kui informatsioon sisestatakse andmebaasi, on vajalik luua neli binaarset muutujat – üks iga nisa kohta. Sellisel juhul kirjutatakse muutuja sulgudesse, kuna ülaltoodud definitsiooni (nimelt et muutuja sisaldab informatsiooni ainult ühe vaadeldava üksuse kohta) on rikutud. Seetõttu ei ole mitmevalikuline muutuja **vastuvõetav elektroonilistele andmebaasidele**.

Avatud nimekirjaga muutujad

Vaatame, näiteks, küsimustiku küsimust: "Milliseid ravimeid olete te kasutanud oma veiste ravimiseks nelja viimase nädala jooksul?" Muutuja on mitmevalikuline, kuna kasutada võidi rohkem kui üht ravimit. Võimalike ravimite nimekirja on uurijale teadmata, nii et fikseeritud kategooriaid ei saa andmete kogumise lehele kavandada. Mitmevalikuline muutuja kodeeritakse pärast andmete kogumise lõpetamist. Kasutatud ravimite arv on käesoleva näite puhul siis teada ja iga ravim (nüüd suletud nimekirjas) registreeritakse, kasutades ühte binaarset muutujat iga ravimi jaoks. Mitteolulised üksikasjad võib võtta kokku ühte muutujasse. Näiteks mõned ravimid, mida kasutati ainult aeg-ajalt, võib kokku võtta kategooriasse "muud ravimid".

Märkused

1. Binaarseid, nominaalseid ja järjestusmuutujaid saab registreerida kas **tekstiliste või numbriliste väärtuste** abil. Andmete sisestamise hõlbustamiseks tuleks valida sobiv vorming. Igal juhul on oluline dokumenteerida koodid ja kasutada neid järjekindlalt. Mõne statistikatarkvara puhul tuleb analüüsimiseks tekstilised väärtused muuta numbrilisteks.
2. Pidevad (kvantitatiivsed) muutujad võivad (teoreetiliselt) omada mis tahes väärtust võimalike väärtuste ulatuses. **Arvude vorming** (astendaja ja komakohad) sõltub väärtuste ulatusest ja vajalikust täpsusest. Minimaalne täpsus on määratud mõõtmisprotsessi tehnilise täpsusega (näiteks 3 komakohta ELISA optilise tiheduse lugemisel). Väärtusi võib ümardada, kajastades mõõtmisprotsessi korratavust (näiteks võib ELISA optilise tiheduse esitada ainult 2 komakohaga, kui analüütiline korratavus on väiksem kui OD 0,005).
3. Binaarsed, nominaalsed ja järjestusmuutujad eeldavad, et kategooriad on **teineteist välistavad**. Näiteks binaarne vaatlus võib olla kas "jah" või "ei", "haige" või "mittehaige". Nominaalsed vaatlused võivad olla kas "1", "2", "3", "4" või "5" viie farmi jaoks. Järjestusvaatlused võivad olla kas "-" või "+", või "++" või "+++" jne. Lubatud kategooriad on ka **ühiselt ammendavad**. See tähendab, et mingid muud vaatlused peale lubatud väärtuste ei ole võimalikud.

ESMASTE ANDMETE KOGUMISE PÕHIMÕTTED

1. Määratlege informatsiooni liik (nominaalne, järjestus, pidev).
2. Määratlege ja dokumenteerige kodeering (lubatud väärtused ja ulatus).
3. Määratlege vaadeldav üksus (piirkond, farm, loom, elundid).
4. Koguge ainult antud uuringu jaoks vajalikke andmeid.
5. Kontrollige andmete sisestamist ja parandage vajadusel andmete kogumise vorme.
6. Püüdke viia paberitöö minimaalseks, vältige keerulisi koode.
7. Kindlustage, et kõik vaatlused oleks hõlpsasti looma ja farmi tasandil kergesti tuvastatavad.

8. Pidage meeles, et informatsiooni saab analüüsimiseks hõlpsasti vähendada (näiteks vanusegrupid) ja et informatsiooni sisu peab kajastama andmete kasutamise eesmärke.

Käsitletud põhimõtteid demonstreeritakse hüpoteetilise piimafarmi projekti "Milky Way 2003" näitel.

PIIMAFARMI PROJEKT "MILKY WAY 2003"

Loomade andmed

Farm 1 (Charly farm)

Külastuskuupäev 2003/5/13

Külastusaeg 6:45

Loom	Tõug	Vanus	BC tulemus	Poegimise kuupäev	CMT	Nisakahjustused	Muud andmed ...
1	C	3,5	++	03/2/23	-		
2	C	13	+	03/1/30	-		
3	C	7½	-	03/4/1	++	EV, TV	
4	C	4	-	02/12/24	-		
5	C	18 m	+	03/3/5	-		

Teised loomad

Kokku: 5

Loom: looma identifitseerimisnumber (unikaalne farmis)

Tõug: L = kohalik, C = ristatud, E = eksootiline

Vanus: looma vanus aastates, tühi = teadmata

KK hinne: ++ = hea, + = keskmine, - = halb

Poegimise kuupäev: viimase poegimise kuupäev (A/K/P) (aasta/kuu/päev), tühi = teadmata

CMT: California mastiiditest (piim kõigist udaraneljandikest): ++ = tugevalt positiivne, + = positiivne, +/- = märk, - = negatiivne, tühi = mitte tehtud

Nisakahjustused: EV = eesmine vasak, EP = eesmine parem, TV = tagumine vasak, TP = tagumine parem, tühi = kahjustus puudub

Joonis 33. Andmete registreerimise lehe näidis looma tasandil. Pange tähele, et mõned sissekanded ei ole järjekindlad!

Loomade andmed (järg)							
Farm	<u>2 (Mr. Bull)</u>			Külastuskuupäev	<u>2003/5/14</u>		
				Külastusaeg	<u>5:30</u>		
Loom	Tõug	Vanus	KK hinne	Poegimise kuupäev	CMT	Nisakahjustused	Muud andmed
1	E	4,5	++	03/3/25	-		
2	E	4	++	03/3/12	-		
3	E	3,5	++	03/3/28	-		
4	E	4,5	++	02/2/16	-		
5	E	7,5	++	03/1/20	+/-		
6	E	5,5	++	03/3/13	-		
7	E	5	++	03/3/1	-		
Teised loomad							
<u>Kokku: 7</u>							
Loom: looma identifitseerimisnumber (unikaalne farmis)							
Tõug: L = kohalik, C = ristatud, E = eksootiline							
Vanus: looma vanus aastates, tühi = teadmata							
KK hinne: ++ = hea, + = keskmine, - = halb							
Poegimise kuupäev: viimase poegimise kuupäev (A/K/P) (aasta/kuu/päev), tühi = teadmata							
CMT: California mastiiditest (piim kõigist udaraneljandikest): ++ = tugevalt positiivne, + = positiivne, +/- = märk, - = negatiivne, tühi = mitte tehtud							
Nisakahjustused: EV = eesmine vasak, EP = eesmine parem, TV = tagumine vasak, TP = tagumine parem, tühi = kahjustus puudub							

Joonis 34. Andmete registreerimise lehe näidis looma tasandil

Seroloogiliste uurimiste andmedUurimise kuupäev 2003/5/15

Proov	Farm	Loom	ELISA	KSR	Muud andmed
1678	1	1	0,06	32	
1679	1	2	0,00	16	
1680	1	3	0,12	–	
1681	1	4	0,00	Nd	
1682	1	5	0,02	–	
1683	2	1	0,00	–	
1684	2	2	0,00	–	
1685	2	3	0,05	–	
1686	2	4	0,25	32	
1687	2	5	0,11	–	
1688	2	6	0,02	–	
1689	2	7	0,04	–	

Teised loomad

Kokku: 12

Proov: seroloogiliste uurimiste registri number

Farm: farmi number

Loom: looma number

ELISA: seerumi ELISA brutselloosile (tootja/partii number);

suhteline positiivsus (optiline tihedus jagatud kontrollproovi optilise tihedusega),

tühi = puudub seerum

KSR: komplemendi sidumise reaktsiooniga saadud lõpptiitri pöördväärtus,

nd = ei ole määratud antikomplementaarse reaktsiooni tõttu,

– = tiiter puudub,

tühi = puudub seerum

Joonis 35. Andmete registreerimise lehe näide looma tasandil

Mikrobioloogiliste uurimiste andmed					
Uurimise kuupäev <u>2003/5/16</u>					
Proov	Farm	Loom	Veerand	Kultuur	Muud andmed
2003	1	3	EV	–	
2004			EP	Sta	
2005			TV	Sta	
2006			TP	–	
2007	2	5	EV	–	
2008			EP	–	
2009			TV	–	
2010			TP	–	
2011			EV		
2012			EP		
2013			TV		
2014			TP		
Teised kirjed					
<u>Kokku: 8</u>					
Valim: mikrobioloogiliste uurimiste registri number					
Farm: farmi number					
Loom: looma number					
Neljandik: VE = eesmine vasak, EP = eesmine parem, TV = tagumine vasak, TP = tagumine parem					
Kultuur: Sta = <i>Staphylococcus spp.</i> , Str = <i>Streptococcus spp.</i> , Ent = <i>Enterobacteria</i> , Cor = <i>Corynebacterium</i> , Act = <i>Actinomyces</i> , Yea = pärm, muud on välja kirjutatud, – = kasvu puudumine					

Joonis 36. Looma organi tasandil andmete registreerimise lehe näide

Märkused

Joonised 33, 34. Kodeeringud "C" ja "L" võivad andmete käsitsi kirjutamisel segi minna. Farmist 1 pärit looma 5 vanus on sisse kantud kuudes. Väärtus tuleb muuta aastateks elektroonilises andmebaasis. CMT tuleb teha iga udaraveerandi kohta eraldi. Nelja nisa kahjustuste registreerimine ühe muutujana ei ole andmebaasi elektroonilise versiooni jaoks vastuvõetav. Lehe päises olevat olulist informatsiooni (farm, külastuskuupäev jne) tuleb andmebaasi elektroonilises versioonis vaadelda kui täiendavaid muutujaid.

Joonis 35. KSRi tulemused võivad puududa kahel põhjusel: seerumi antikomplementaarsus (kodeeritud "nd") ja seerumi puudumine (kodeeritud tühjana). Mõlemat tüüpi puuduvate väärtustega loomad jäävad välja KSRi andmete analüüsist. Siiski võib antikomplementaarsete seerumite osakaal olla huvipakkuv. Seepärast kogutakse selle kohta informatsiooni. Vajadusel võib kasutada mitut puuduva väärtuse koodi sõltuvalt väärtuse puudumise põhjusest.

Joonis 36. Analüüsiüksuseks on üks udaraveerand, kuna mikrobioloogilise analüüsi üksuseks on üks proov. Eeldatakse, et igalt CMT-positiivselt lehmalt võetakse proov igast udaraveerandist mikrobioloogiliseks uuringuks. Kui ühel lehmalt on vähem kui 4 lüpsvat udaraveerandit, siis tuleb puuduv veerand märgistada puuduva väärtuse koodiga (või jätta lahter tühjaks).

ANDMEBAASI STRUKTUUR

Andmebaas on andmete süsteemne kogu. Tänapäev on toonud kaasa andmete registreerimise elektroonilisel kujul. Andmebaas koosneb ridadest ja veergudest.

1 rida esindab vaatluse 1 üksust (valimielementi)

Iga rida esindab vastavalt kas ühte looma või ühte farmi vastavalt looma tasandil andmebaasis või farmi tasandil andmebaasis. Kõik lahtrid antud reas sisaldavad vaatlusi vastava üksuse (loom või farm) kohta. Läbilõikeuuringutes esineb iga element ainult üks kord. Pikaajalistes uuringutes võivad elemendid (loomad, farmid) esineda mitu korda. Korduvad vaatlused pikaajalises uuringus on eristatavad üksuse identifikaatori (looma, farmi number) ja vaatluskuupäeva järgi. Vaatlusperioodi vältel jäävad mõned tunnused püsivaks, teised muutuvad.

1 veerg esindab 1 vaadeldud tunnust (muutuja)

Iga veerg esindab ühte vaadeldud tunnust (muutujat). Kõik samasse veergu (muutujasse) kuuluvad lahtrid kirjeldavad sama tunnust andmebaasi erinevate elementide kohta. Ka skaala, mille järgi tunnus on mõõdetud, on sama kõikide elementide jaoks (näiteks aastates või kuudes või päevades mõõdetud vanus kõikide loomade kohta).

ELEKTROONILINE REGISTREERIMINE

Esmane andmete kogu on sageli vasturääkiv ja järgib pigem ergonoomilisi kui analüütilisi vajadusi. Näiteks loomade vanus on mõnikord registreeritud aastates või kuudes, sõltuvalt tegelikust east. Korratud vaatlused on mõnikord kirjutatud lihtsalt külgnesse lahtrisse. Kõik see on lubatav, kuni informatsioon on täpne ja arusaadav. Siiski on elektroonilisel andmekogul teised nõuded kui paberil esitatud andmestikul.

Andmebaaside loogika on see, et **iga lahter on määratletud kui ühe muutuja vaatlus ühe elemendi kohta ja kõrvalolevate lahtrite sisu ei saa määrata ega muuta selle lahtri sisu.**

Järgnevalt on toodud näiteid näidisandmete elektrooniliste versioonide kohta (joonis 37). Andmed saab sisestada arvutustabelisse (Microsoft EXCEL) või andmebaasi (näiteks Microsoft Access).

Andmed										
Farm	Loom	Tõug	Vanus	KK hinne	Poegimise kuupäev	CMT	EV	EP	TV	TP
1	1	1	3,5	2	99/02/23	0				
1	2	1	13	1	99/01/30	0				
1	3	1	7,5	0	99/04/01	2	1		1	
1	4	1	4	0	98/12/24	0				
1	5	1	1,5	1	99/03/05	0				
2	1	2	4,5	2	99/03/25	0				
2	2	2	4	2	99/03/12	0				
2	3	2	3,5	2	99/03/28	0				
2	4	2	4,5	2	99/02/16	0				
2	5	2	7,5	2	99/01/20	1				
2	6	2	5,5	2	99/03/13	0				
2	7	2	5	2	99/03/01	0				

Koodid eraldi lehel					
Farm	Loom	Tõug	Vanus	KK hinne	Poegimise kuupäev
Farmi ID	Looma ID	Tõug: 0 = kohalik, 1 = ristand, 2 = eksootiline	Vanus aastates	Koha konditsioon: 0 = halb, 1 = keskmine, 2 = hea	Poegimise kuupäev: AA/KK/PP (aasta/kuu/päev)

... koodide järg				
CMT	EV	EP	TV	TP
California mastiididest: 0 = negatiivne, 1 = jälg, 2 = positiivne, 3 = tuge- valt positiivne	Nisakahjustused eesmisel vasakul neljandikul: 1 = jah, tühi = ei	Nisakahjustused eesmisel paremal neljandikul: 1 = jah, tühi = ei	Nisakahjustused tagumisel vasakul neljandikul: 1 = jah, tühi = ei	Nisakahjustused tagumisel paremal neljandikul: 1 = jah, tühi = ei

Joonis 37. Jooniste 33, 34 andmed ja koodid arvutustabelisse sisestatuna

Tavad, reeglid ja soovitused elektrooniliste andmebaaside jaoks

1. Kasutage lihtsaid muutujate nimetusi, mis poleks pikemad kui 8 märki. Sümboloid (! ? . ; -) mitte kasutada. Tühikuid mitte jätta. Kasutage tähendusega nimetusi, mida saab alles hoida kogu analüüsi ja dokumenteerimise vältel.
2. Ärge kasutage andmete sisestamise lehtedel valemeid.
3. Ärge sisestage andmebaasi lehele mingit muud informatsiooni kui ainult vaadeldud muutuja väärtused.
4. Kopeerige muutujate nimed eraldi lehele ning sisestage kodeeringuskeem ja lisainformatsioon sinna.
5. Kehtestage kodeerimisele reeglid (numbrilised või tähelised koodid, puuduvad väärtused), mis hoiavad ära andmesisestusvead.
6. Vigade ennetamiseks kasutage andmete sisestamisel andmete hindamise reegleid.
7. Kasutage looma- ja farmitasandi andmete jaoks eraldi faile.
8. Säilitage informatsiooni uuringu kohta, andmebaasi versiooni ja viimaseid muudatusi koos andmetega (näiteks kasutage eraldi töölehti EXCELis).
9. Kindlustage, et ainult ühte (ajakohast) andmebaasi versiooni kasutatakse kui peamist koopiat.
10. Salvestage andmebaasi ajakohane ja iganenud versioon vähemalt kahte sõltumatusse andmete säilitamise keskkonda, et ennetada andmete kaotsiminekut. Muudatused (näiteks muutujate ümberkodeerimine) on siis jälgitavad.

SEOSTATUD ANDMEBAASID

Analüüsi seisukohast on oluline, et loomatasandi informatsiooni oleks võimalik ühendada farmitasandi informatsiooniga. Vaatame meie näites ristklassifitseerimist CMT tulemuse ja tootmissüsteemi alusel. Andmed asuvad kahes eraldi tabelis.

Kuid **tabelite vahel on seos, mis on antud farmi numbri kaudu**. Skemaatiliselt, on meil nn 1 : n seos, sest 1 farm võib anda n loomatasandi vaatlust.

Seoseid saab tekitada ACCESSi jt andmebaasi programmidega.

Farm	Tootmissüsteem	Teised muutujad	Farm	Loom	CMT	Teised muutujad
1	0		1	1	0	
1	1		1	2	0	
			1	3	2	
			1	4	0	
			1	5	0	
			2	1	0	
			2	2	0	
			2	3	0	
			2	4	0	
			2	5	1	
			2	6	0	
			2	7	0	

Joonis 38. Seostatud andmebaas seosega 1 (farm) : n (looma)

Protseduur

1. Importige EXCELi tabelid andmebaasi faili.
2. Looge uus päring, lisage kaks tabelit.
3. Looge tabelite vahel seos, kasutades ühendavat muutujat (meie puhul on selleks farm).
4. Valige ühendtabelisse soovitud muutujad.
5. Valige andmete vaatamine ja te saate midagi joonise 39 sarnast.

Farm	Tootmissüsteem	Loom	CMT
1	0	1	0
1	0	2	0
1	0	3	2
1	0	4	0
1	0	5	0
2	1	1	0
2	1	2	0
2	1	3	0
2	1	4	0
2	1	5	1
2	1	6	0
2	1	7	0

Joonis 39. Farmitasandi ja loomatasandi andmete riskklassifitseerimine

Seostatud andmebaaside muud kasutused

Sarnaselt võib lisatabeleid (nagu joonis 35) ühendada loomatasandi tabeliga. See oleks 1:1 seos. Loomatasandi andmed kombineerituna organitasandi andmetega (vt joonis 36), meie puhul 4 udaraveerandit, esindavad seost 1:4-le.

ANDMEPUHASTUS

Kui andmed on sisestatud elektroonilisse andmebaasi, tuleb neid enne edasist töötlemist ja analüüsi hoolikalt kontrollida.

Usutavuse ja ühtlikkuse kontrollimine

Kõlbmatute väärtuste või ebaühtlaste koodide avastamiseks kasutage EXCELi filterfunktsioone ja kontrollige, kas vaadeldud tulemused on lubatud ulatuste ja väärtuste piires. Muutke ebaühtlased andmed ja kontrollige kõlbmatuid väärtusi esmase andmekogumislehe andmetest. Kontrollige ka andmete loogilist usutavust (näiteks piimaanni andmed ainult lehmade kohta).

Ümberkodeerimine

Soovitud koodi (näiteks numbriline väärtus) saamiseks teatud vaatluste kohta (näiteks tekstilised muutujad) kasutage automaatset leia-asenda funktsiooni. Teisendage puuduva proovi koodid või puuduva väärtuse koodid andmesalvestuse ja andmeanalüüsi jaoks nõutavas vormis.

Ümberkodeerimist ei soovitata teha originaaltöölhel (mõeldud andmete sisestamiseks). Kasutage eraldi töölehte (näiteks andke sellele nimeks "Ümberkodeeritud andmed"). Ümberkodeerimine on paindlikum, kui leia-asenda funktsiooni asemel kasutada valemid. Siis on hiljem kergem ümberkodeerimisskeemi toimetada. Ümberkodeerimine tehakse antud muutuja kõigis lahtrites. Seepärast tuleks valem sisestada pigem reavalemina kui üksiku lahtri valemina (vt EXCELi kasutusjuhiseid). Muidugi peavad toorandmetes puuduvad väärtused olema ka puuduvad väärtused ümberkodeeritud andmetes. Võtame vaatluse alla kaks tüüpilist näidet. Sammsammulised juhtnöörid on antud EXCELi süntaksis.

Näide 1

Aastatel (vanus aastates) põhinevate vanusegruppide (Vanusegrupid) moodustamine kategooriatena: 0 = noorem kui 2aastane; 1 = 2aastane või vanem, noorem kui 5aastane; 2 = 5aastane või vanem.

Toorandmed	Vanus	Ümberkodeeritud andmed	Vanusegrupp
	3,5		=if(Age="";"";(Age>=2)*1+(Age>=5)*2+*0)
	13		Tõlge: = kui(Vanus ""; ""; (Vanus> = 2)+(Vanus> = 5))
	...		

1. Andke algandmete töölehel nimetus "Vanus" lahtritele, mis sisaldavad vanust aastates.
2. Märkige ümberkodeeritud andmete töölehel vanusegrupi väärtustele mõeldud lahtrid.
3. Kirjutage esimesse lahtrisse valem, mida ülalpool näete (või muutke vanusegruppide intervalle). Vajutage CONTROL+SHIFT+ENTER. Valem on nüüd loogelistes sulgudes {} ja kirjutatud kõikidesse lahtritesse.
4. Kontrollige, kas valem töötab korralikult. Pidage meeles, et "" tähendab "tühi lahter". Kui vaja, kasutage teisi puuduva väärtuse koode.

Näide 2

Muutuja „tõug” algsete kategooriatega – L, C, E, ümberkodeerimine, kasutades numbreid 0, 1, 2.

1. Järgige ülaltoodud samme 1 ja 2.
2. Kirjutage: =if(Breed="";"";(Breed="L")*0+(Breed="C")*1+(Breed="E")*2), jätkake ülalkirjeldatud viisil.
/Tõlge trükitavale valemile: = kui(Tõug = ""; "";(Tõug = "L") * 0 + (Tõug = "C") * 1 + (Tõug = "E") * 2)./

Viide.

Greiner, M. 2002. Data processing and interpretation. In: Course Reader: International Training Course: Globalisation of Trade and International Quality Standards: New Technical Regulations and Voluntary Standards in Infectious Animal Disease Control. Deutsche Stiftung für Internationale Entwicklung (DSE), Feldafing.