

AVLSDATA FRA FØLLFESTIVALEN – DEL 1

av Unn Reierstad, cand.scient (NLH/UMB), veterinær (NVH) / RR Reierstad Ridehest

MATERIALE & METODER :

AVLSLÆRE

For ethvert dyr er $P = GEN + ENV$, der P, GEN og ENV er et positivt eller negativt tall i forhold til gjennomsnittet i dyrets populasjon. Dersom alle dyr i en populasjon har nøyaktig samme miljø, er $ENV = 0$ for disse dyrene. Da vil forskjellene i PEN mellom ulike dyr i populasjonen uttrykke de genetiske forskjellene (GEN) alene, noe som kalles Arvbarhet (ARV), som da = 1. ARV er et tall mellom 0 og 1 dersom $ENV > 0$.

$$ARV (h^2) = \frac{\text{variansen av additive genetiske effekter}}{\text{variansen av additive genetiske effekter} + \text{variansen av genetiske interaksjoner/ korrelasjoner (INT)} + \text{variansen av dominans/epistasie (DOM)} + \text{variansen av miljøet (ENV)}}$$

Der variansen uttrykker variasjonen/spredningen av effekten i populasjonen. Arvbarhetsestimater krever mange registreringer pr egenskap ($n > 500$), strukturert slektskap og en god modell. **ARV er altså et statistisk beregnet mål på status vedr. det genetiske vs det miljømessige bidraget bak en egenskap i en populasjon, og gjelder bare for det datafangstoppellet (test- og målesystemet) den er beregnet på grunnlag av.** Kunnskap om genetiske korrelasjoner (KORR), deres størrelse og fortegn er også viktig fordi de sier hvordan det går med andre egenskaper når en selekterer for en spesiell egenskap. KORR mellom 0 og +1 = proporsjonalitet (den ene egenskapen øker om den andre øker), KORR mellom 0 og -1 = omvendt proporsjonalitet (den ene egenskapen minker om den andre øker). $K = +1$ eller -1 betyr fullstendig korrelasjon/at egenskapene er en og samme egenskap, $KORR=0$ betyr ingen korrelasjon/helt uavhengige egenskaper. ENV deles i tilfeldige og systematiske miljøfaktorer. De tilfeldige kan korrigeres for ved gjentatt registrering. De systematiske kan korrigeres for (statistisk) eller fjernes ved å skape likt miljø (stasjonstesting). Korrigerings for systematiske miljøfaktorer vil føre til høyere ARV. Endring av genetiske egenskaper i populasjonen vil også kunne endre på ARV. Derfor må ARV beregnes på nytt av og til i ethvert avlsarbeid. Likevel er nivået på ARV relativt konstante over tid og over populasjoner og man kan snakke om ARV for en egenskap som et relativt konstant tall. **Forutsatt at ARV er regnet ut på grunnlag av objektive egenskapsmål (målesikkerhet), kan ARV antyde hvor sterkt arvelig en egenskap er. Gjentatt beregnede ARV sier også noe om seleksjonen over tid, hvor sterkt selekterte egenskaper kan utvikle høyere ARV.** $ARV < 0,2$ regnes som lave, ARV mellom 0,2-0,5 regnes som middels, og $ARV > 0,5$ regnes som høye.

Statistiske beregninger - Utdrag pr. 18.02.2010:

Avlsmateriale og genetisk bidrag:

Føllenes foreldre og besteforeldredyr ble gruppert etter stambok de var registrert inn i som føll (mål på genetikk / hvilket avlsforbund de er avlet), jf. Tabell 5A, og morddyrene ble gruppert i kvalitetsklasser på grunnlag av genotypiske og fenotypiske egenprestasjoner, jf. Tabell 5B.

Foreldre og besteforeldredyrene i utvalget ble talt opp etter stambok og kvalitetsklasser (mordyr) og rangert. Ved kvantifisering av genetisk bidrag ble foreldredyr kun talt en gang (individnivå, hhv. $n=29$ og $n=87$), mens besteforeldredyr ble talt på stamboknivå (dvs. individer/forfedre som kom igjen som morfar eller mormor på et annet individ ble talt om igjen). Marginalt genetisk bidrag ble altså ikke beregnet etter standard prosedyre, da det kun var registrert individinformasjon i 2 generasjoner.

Tabell 5A: Gruppering av føllenes foreldredyr etter opprinnelsesstambok (basert på førstegangsregistrering), kategoriske uavhengige data). I 2.generasjon går noen individer igjen flere ganger (jf. marginalt genetisk bidrag).

Stambok	Far (1.gen)	Far-far	Far-mor	Mor (1.gen)	Mor-far	Mor-mor	TOTAL 1.gen	% Genetisk bidrag	TOTAL 2.gen	% Genetisk bidrag
Norsk Varmblod	2	0	0	17	3 (0*)	4 (1*)	19	16,4	7 (1*)	3,0
Importert:**	27	29	29	70	83	83 (86)	97	83,6	224(227)	97,0
Dansk Varmblod	5	0	6	18	8	23	23	19,8	37	16,0
Svensk Varmblod	2	2	3	11	13	13	13	11,2	31	13,4
Hannoveraner	4	3	5	6	12	7	10	8,6	27	11,7
Nederlandsk Varmblod (KWPN****)	0	0	0	12	9	15	12	10,3	24	10,4
Holsteiner	5	8	5	3	8	3	8	6,9	24	10,4
Eng.fullblod***	1	2	1	7	12	7	8	6,9	22	9,5
Oldenburger	4	9	4	0	5	0	4	3,4	18	7,8
Fransk Varmblod (S.F.= Selle Francais)	1	2	1	3	5	3	4	3,4	11	4,8
Belgisk Varmblod (BWP****)	3	1	3	1	0	3	4	3,4	7	3,0
Trakehner	0	0	0	1	4	1	1	0,9	5	2,2
Mecklenburger	0	1	0	1	0	1	1	0,9	2	0,9
Zangersheide (Z)	0	0	0	2	1	0	2	1,7	1	0,4
Westhpaler	1	0	1	0	0	0	1	0,9	1	0,4
Reinlender	0	1	0	0	1	0	0	0,0	2	0,9
Irsk Sportshest (ISH)	0	0	0	0	0	2	0	0,0	2	0,9
Brandenburger	1	0	0	0	0	0	1	0,9	0	0,0
Andre (Østeuropeisk, britisk, ponni, ukjent)	0	0	0	5	5	5	5	4,3	10	4,3
TOTAL	N=29	N=29	N=29	N=87	N=87	N=87	N=116	100	N=231	100

*Antall hvor en eller begge foreldredyr er Norsk Varmblod..

**Importgruppen underoppdelt etter stambok, rangert på grunnlag av % genetisk bidrag i 2.generasjon av foreldredyr.

***XX = Engelsk fullblod regnes som importert uavhengig av 1.gangsregistrering da produksjonen av rasen / det genetisk materialet har sitt opphav i utlandet og populasjonen er fullstendig internasjonalsert..

**** Koninklijk Warmbloed Paardenstamboek Nederland/Royal Dutch Sport Horse.

***** Belgisch warmbloed Paardenstamboek.

Tabell 5B: Gruppering av føllenes mordyr basert på avkomsresultater, kåringsgrad og sportsresultater slik at lavest tall gir best geno- eller fenotype for egenprestasjoner (kategoriske uavhengige data, alle individer kun talt en gang):

Gruppe	Forfremmelse på grunnlag av avkoms resultater i et europeisk forbund (1)*	Ant. Kåret i et europeisk forbunds Hovedstambok (2)**	Sportsresultater LA eller høyere (3)**	Ant. Kåret i et europeisk forbunds Stambok (4)**	Ant. Kåret i et europeisk forbunds Register eller Forregister (5)**	Ant. Uten kåring eller sportsresultater (eller ukjent status) (6)***	Totalt antall
TOTAL	3	19	26	16	7	16	N=87
%	3,4	21,8	29,9	18,4	8,0	18,4	100

*Genotypiske egenprestasjoner.

**Fenotypiske egenprestasjoner. Mordyr som hadde både kåringsgrad og sportsresultater ble plassert i den gruppen som ga det beste resultatet. En hoppe er ponni.

*** Fenotypiske egenprestasjoner. Unge hopper bedekt som 3-års/hvor resultater ikke ennå er oppnådd (n=3) ble kategorisert på grunnlag av mors prestasjoner (føllets mormor) i de tilfeller det var kjent, ut i fra en betraktning om kompensasjon for redusert generasjonsintervall. En hoppe er ponni.

Faste miljøeffekter, avhengighet og summeringsvariable:

Test på faste miljøeffekter og støyeffekt (tilfeldig variasjon fra ulike faktorer) ble utført med prosedyre GLM i statistikkprogrammet SPSS for å få fram en passende modell for variansanalysene, og med prosedyre MEANS (ANOVA) for hver av egenskapsmålene og hver av avkomsgruppene d (dressurbetonede føll) og s (sprangbetonede føll) for effektene kjønn (hingsteføll vs. hoppeføll), alder (født før 1.juni vs. født etter 1.juni), mordyrkvalitet (genetisk effekt av mor, gruppe 1 til 6 jf. Tabell 5B), startrekkefølge (startgruppe 1 vs. 2 (mellom kl 08-10 og 10-12) og 3 vs. 4 (mellom kl 12-14 og 14-16) for hhv sprangbetonte føll (alle før lunsj) og dressurbetonte føll (alle etter lunsj)) og år (2007 vs. 2009) jf. Tabell 4.

Mordyrkvalitet kan dels sees som en fast miljøeffekt, og dels som en variabel genetisk effekt. For føll antas den siste å ha størst effekt. Mordyrkvalitet ble likevel inkludert i testkjøringer for å evaluere metodikk.

Pga. design med gruppevis bedømmelse samlet etter avkomsgrupper og alle sprangbetonte føll på morgenen/alle dressurbetonte føll på ettermiddagen samt få observasjonsår (kun fra ett år), vil startrekkefølgen også inneholde en systematisk effekt av gruppering på bedømmelsesresultatene. Startrekkefølge ble likevel tatt med som faktor i testkjøringer for å evaluere metodikk.

Tidsfaktorene startrekkefølge og år inneholder ellers effekt på bedømmelsesresultatene av tidsaspekt / skalakalibrering, ulike arrangørsted / personal / rutiner / værforhold, dommer / skriver, ulikt genetisk grunnlag, genetisk endring osv.

Inndeling av alder, startrekkefølge og mordyrkvalitet i grupper ble gjort for å få et større antall sammenlignbare n. For mordyrkvalitet vil det pga. inndeling i mange grupper (6 innen hver av d og s) likevel bli et veldig lite antall observasjoner i noen av gruppene pga. lite data.

Tabell 4. Gruppering av avkom etter år, alder, kjønn og startrekkefølge (kategoriske data, uavhengighet)

	Dressurbetonet	Sprangbetonet	Total
Hingsteføll (h)	N=30	N=26	N=56
Hoppeføll (hp)	N=25	N=20	N=45
Født før 1. juni (alder1)	N=30	N=31	N=61
Født etter 1. juni (alder2)	N=25	N=15	N=40
Bedømt mellom kl 08-10 (start1)*	N=3**	N=14	N=17
Bedømt mellom kl 10-12 (start2)*	-	N=18	N=18
Bedømt mellom kl 12-14 (start3)*	N=18	-	N=18
Bedømt mellom kl 14-16 (start4)*	N=17	-	N=17
År 2007	N=17	N=14	N=31
År 2009	N=38	N=32	N=70

*Mangler data for 2007. **Avhengige data (hingst med dobbeltalent).

GLM-Modell (hvor hver variabel (x) i modellen + modellen selv testes for evnen til å bidra med variasjonen i den avhengige (målte) variabelen (Y=A, B, C, D, E, F, G)):

$$Y_{ijklmn} = \mu + \text{År}_i + \text{Kjønn}_j + \text{Alder}_k + \text{Hoppe}_l + \text{Startrekkefølge}_m + (\text{Individ}_n) + e_{ijklmn}$$

der:

Y_{ijklmn} = observert verdi for hest n for den målte egenskapen (A, B, C, D, E, F, G).

μ = populasjonsgjennomsnittet for den målte egenskapen (A, B, C, D, E, F, G).

År_i = fast effekt av år, i = 1 eller 2, 1=2007 og 2=2009, for bedømmelsene.

Kjøn_n_j = fast effekt av kjønn, j = 1 eller 2, 1= hoppe, 2 = hingst, for bedømmelsene.

Alder_k = fast effekt av alder, k = 1 eller 2, 1=født før 1. juni, 2=født etter første juni, for bedømmelsene.

Hoppe_l = fast effekt av mordyrets kvalitet, l = 1, ...,6, jf. Tabell 5B, for bedømmelsene.

Startrekkefølge_m = fast effekt av bedømmelsestidspunkt, m = 1/3 eller 2/4, 1=kl 08-10, 2=kl 10-12, 3=kl 12-14, 4=kl 14-16, for bedømmelsene.

Individ_n = additiv genetisk effekt av hest n for den målte egenskapen $\sim N(0, \sigma_a^2)$

E_{ijklmn} = tilfeldig residualeffekt $\sim N(0, \sigma_e^2)$

Korrigert modell:

$Y_{ijkl} = \mu + \text{År}_i + \text{Kjøn}_j + \text{Alder}_k + \text{Individ}_l + e_{ijklmn}$

der:

Y_{ijkl} = observert verdi for hest l for den målte egenskapen (A, B, C, D, E, F, G).

μ = populasjonsgjennomsnittet for den målte egenskapen (A, B, C, D, E, F, G).

År_i = fast effekt av år, i = 1 eller 2, 1=2007 og 2=2009, for bedømmelsene.

Kjøn_n_j = fast effekt av kjønn, j = 1 eller 2, 1= hoppe, 2 = hingst, for bedømmelsene.

Alder_k = fast effekt av alder, k = 1 eller 2, 1=født før 1. juni, 2=født etter første juni, for bedømmelsene.

Individ_l = additiv genetisk effekt av hest l for den målte egenskapen $\sim N(0, \sigma_a^2)$

E_{ijkl} = tilfeldig residualeffekt $\sim N(0, \sigma_e^2)$

Test for effekt på bedømmelses-resultatene av dommer (iboende målevariasjon utover effekt av startrekkefølge og år) ble gjort gjennom test for normalitet i datasettet, og deretter gjennom test av hypotesene 2-6.

Chi-square test (χ^2) mellom A, B, C, D, E, F, G (jf. TABELL 1) ble kjørt for å teste for uavhengighet mellom egenskapsmålene.

TOTAL1 = A+B+C+D+E+F og **TOTAL2** = A+B+C+D+E+F+G ble beregnet for alle føllene (n=101) samt for farshingstene (n=10).

Korrigeringer:

Avhengighet (parede data)

Siden to av hingstene (Aurelio I og Tobajo Picez) hadde avkom i begge avstammingskategorier d og s (dvs. avhengige variable mht. genetisk påvirkning), ble avkommene for gruppen som var kategorisert på grunnlag av mors avstamning (hhv. d n=4 og d n=4) for disse to hingstene fjernet fra de kjøringene som sammenlignet gruppene d og s på genetisk nivå (AVLSDATA FRA FØLLFESTIVALEN - Del 2), slik at n=47 (d) og n=46 (s). Observasjonene ble ikke fjernet for undersøkelser på målesikkerhet (dommeren som måleenhet – Del 1), slik at n=55 (d) og n=46 (s).. Tabell 2.

Test for normalitet i datasettet:

Normal sannsynlighets-/Q-Q plot med residualplot, og Shapiro-Wilk test ble kjørt i statistikkprogrammet SPSS for egenskapsmålene A, B, C, D, E, F, G, TOTAL1 og TOTAL2 på datasettet totalt sett (n=101) og på avstammingsgruppene (d, n=55/47 og s, n=46).

Deskriptiv statistikk:

Middeltall (normalfordelingskurvens topp), median (midterste karakter), konfidensintervall og frekvensfordeling (varians, standardavvik (normalfordelingskurvens spredning), min, max, range, skjevhet og kurtosis) ble beregnet i statistikkprogrammet SPSS for egenskapsmålene A, B, C, D, E, F, G, TOTAL1 og TOTAL2 på datasettet totalt sett (n=101) og på avstammingsgruppene (d, n= 55/47 og s, n=46).

Hypotesetester:

Statistisk signifikans for

- **Hypotese 1** at gjennomsnitt av hhv. A, B, C, D, E, F, G for hhv. d (n=46) og s (n=47/55) er like for de 2 alders-, kjønns-, starts- og årsgruppene (jf. Tabell 4) ble testet med GLM-regresjon og enveis variansanalyse modell F-test prosedyre MEANS (ANOVA) i statistikkprogrammet SPSS.
- **Hypotese 1 og 7** at gjennomsnitt av hhv. A, B, C, D, E, F, G for hhv. s (n=46) og d (n=47/55) ikke er like for de 6 mordyrkvalitetgruppene (jf. Tabell 5B) ble testet med GLM-regresjon og enveis variansanalyse modell F-test prosedyre MEANS (ANOVA) i statistikkprogrammet SPSS.
- **Hypotese 2** at gjennomsnitt av TOT1 hhv. TOT2 (s) (n=46) ikke er lik gjennomsnitt av TOT1 hhv. TOT2 (d) (n=47/55) ble testet i enveis variansanalyse modell F-test prosedyre MEANS (ANOVA) i statistikkprogrammet SPSS.
- og at gjennomsnitt av A hhv. D og E (s) (n=46) ikke er lik gjennomsnitt av A hhv. D og E (d) (n=47/55) ble testet i enveis variansanalyse modell F-test prosedyre MENS (ANOVA) i statistikkprogrammet SPSS.
- **Hypotese 4** at gjennomsnitt av G (s) (n=46) er lik gjennomsnitt av G (d) (n=47/55) ble testet i enveis variansanalyse modell F-test prosedyre MEANS (ANOVA) i statistikkprogrammet SPSS.
- **Hypotese 3, 5 og 6** at gjennomsnittene for A, B, C, D, E, F, G (avhengige data) ikke er like i populasjonen som helhet (n=101) og for hhv. s (n=46) og d (n=47/55) ble testet i enveis variansanalyse modell paret t-test i statistikkprogrammet SPSS.
- **Hypotese 5** at gjennomsnittet av G korrelerer ulikt med A, B, C, D, E, F i populasjonen som helhet (n=101) og for hhv d (n=46) og s (n=47/55) ble testet i enveis variansanalyse modell paret t-test og med GLM-regresjon i statistikkprogrammet SPSS.