

## KAJIAN KOLINEARITAS FAKTOR NON-GENETIK PADA SAPI PERAH FH DI BBPTUHPT BATURRADEN

Agus Susanto\*, Dattadewi Purwantini, Setya Agus Santosa, Dewi Puspita Candrasari, dan Imbang Haryoko

Fakultas Peternakan, Universitas Jenderal Soedirman, Purwokerto, Indonesia

\*Email korespondensi: agus.susanto0508@unsoed.ac.id

**Abstrak.** Akurasi model pendugaan pada regresi berganda dapat dipengaruhi oleh faktor-faktor independen yang memiliki kolinearitas. Penelitian ini bertujuan untuk mengetahui hubungan faktor-faktor non-genetik terhadap produksi susu dan mengetahui tentang kolinearitas peubah-peubah tersebut pada sapi perah FH di BBPTUHPT Baturraden. Penelitian dilakukan menggunakan data sekunder dari Balai Pembibitan Ternak Unggul dan Hijauan Pakan Ternak (BBPTUHPT) Baturraden yang terdiri dari 1.942 catatan individu produksi susu per laktasi dari 1.015 ekor sapi perah FH yang lahir pada tahun 2000-2014. Variabel penelitian meliputi produksi susu per laktasi (laktasi 1-6) (myield) sebagai variabel dependent dan jumlah hari pemerahan (jhp), periode laktasi (lak), tahun saat lahir (yob), bulan saat lahir (mob), umur saat partus (agepar), tahun saat partus (yopar) dan bulan saat partus (mopar) sebagai variabel independent. Evaluasi terhadap kolinearitas antar peubah bebas ditentukan menggunakan matriks korelasi dan variance inflation factors (VIF) yang dilakukan menggunakan program antar muka Rstudio yang berbasis pada program R. Hasil analisis menunjukkan adanya tingkat keeratan hubungan antar peubah dari lemah hingga tinggi. Peubah yang memiliki tingkat keeratan hubungan sedang hingga tinggi adalah yob-agepar (-0,42), yob-laktasi (-0,4), laktasi-agepar (0,94) dan yib-yopar (0,89). Hasil analisis regresi myield menggunakan full model (menggunakan semua peubah bebas) menghasilkan nilai VIF berturut-turut sebesar 8,51, 1,08, 2641,09, 22,06, 2140,42, 20,13 dan 533,8 untuk peubah lak, jhp, yob, mob, yopar, mopar dan agepar. Dapat disimpulkan bahwa peubah bulan saat induk lahir, tahun saat induk lahir, bulan saat induk beranak, tahun saat induk beranak dan umur saat induk beranak memiliki kolinearitas. Peubah tersebut mengganggu hasil taksiran dalam analisis multipel regresi sehingga harus dikeluarkan dari model regresi.

**Kata kunci:** multikolinearitas, linear regresi berganda, produksi susu sapi, model analisis

**Abstract.** Independent variables with collinearity can affect the precision of the prediction model in multiple regression. This study aims to determine the relationship of non-genetic factors to milk production and to find out about the collinearity of these variables in FH dairy cows at BBPTUHPT Baturraden. The study was conducted using secondary data from the Baturraden Superior Cattle Breeding and Forage Cattle Center (BBPTUHPT) consisting of 1,942 individual records of milk production per lactation from 1,015 FH dairy cows born in 2000-2014. Research variables included milk production per lactation (lactation 1-6) (myield) as the dependent variable and number of milking days (jhp), lactation period (lac), year at birth (yob), month at birth (mob), age at parturition (agepar), year of parturition (yopar) and month of parturition (mopar) as independent variables. Evaluation of collinearity between independent variables was determined using the correlation matrix and variance inflation factors (VIF) which was carried out using the Rstudio interface program based on the R program. The results of myield regression analysis using the full model (using all independent variables) yielded VIF values of 8.51, 1.08, 2641.09, 22.06, 2140.42, 20.13 and 533.8 for the variables jhp, yob, mob, yopar, mopar and agepar, respectively. It can be concluded that the variable month when the parent was born, year when the parent was born, month when the parent gave birth, year when the parent gave birth and age when the parent gave birth have collinearity. These variables interfere with the estimated results in the multiple regression analysis, so they must be excluded from the regression model.

**Keywords:** multicollinearity, multiple linear regression, milk production, comparing model

### Pendahuluan

Sistem recording pada pusat pembibitan sapi perah termasuk Balai Besar Pembibitan Ternak Unggul dan Hijauan Pakan Ternak (BBPTUHPT) Baturraden meliputi pencatatan beberapa informasi tambahan selain data utama dari pembibitan ternak sapi perah. Data utama yang dicatat dalam sistem recording meliputi produksi susu harian dan komponen susu yaitu lemak dan protein susu. Informasi

tanggal lahir ternak, tanggal induk beranak, tanggal induk diperah, jumlah service per conception, tetua jantan/betina juga dimasukkan ke dalam sistem recording.

Informasi yang dicatat dalam sistem recording akan bermanfaat dalam program evaluasi mutu genetik ternak misalnya dalam mengelompokkan ternak ke dalam faktor non-genetik secara sistematis. Misalnya pengelompokan ternak berdasarkan asal kandang, bulan/tahun saat ternak lahir maupun saat induk beranak sebagai representasi musim, jumlah hari pemerahan, umur beranak dan lain-lain. Penaksiran nilai pemuliaan ternak menggunakan metode Best Linear Unbiased Prediction (BLUP) (Henderson, 1984; Mrode, 2014). Penaksiran nilai pemuliaan ternak pada umumnya menggunakan model ternak (animal model) yang kemudian efek faktor tetap (lingkungan) dan efek acak (nilai pemuliaan ternak) secara simultan ditaksir melalui serangkaian persamaan linear (Susanto et al., 2018).

Multikolinearitas peubah-peubah bebas pada linear berganda mendapat perhatian serius dari para peneliti dan perlu pemecahan masalah (Jinn et al., 2007; Garamszegi, 2006; Chen, 2014). Hasil taksiran parameter dalam regresi linier berganda dipengaruhi oleh ada atau tidaknya kolinearitas antar faktor-faktor independen tersebut. Pada regresi linear berganda, peubah-peubah yang memiliki korelasi tinggi satu sama lain akan menyebabkan kolinearitas. Kolinearitas menyebabkan hasil taksiran efek individual menjadi tidak akurat. Kesimpulan yang salah dapat terjadi akibat adanya kolinearitas antar peubah karena hasil estimasi koefisien variabel prediktor menjadi tidak stabil dan memiliki standar eror yang besar. Tujuan penelitian adalah untuk mengetahui adanya kolinearitas pada faktor-faktor non-genetik pada produksi susu sapi perah di Balai Besar Pembibitan Ternak Unggul dan Hijauan Pakan Ternak (BBPTUHPT) Baturaden.

## Materi dan Metode Penelitian

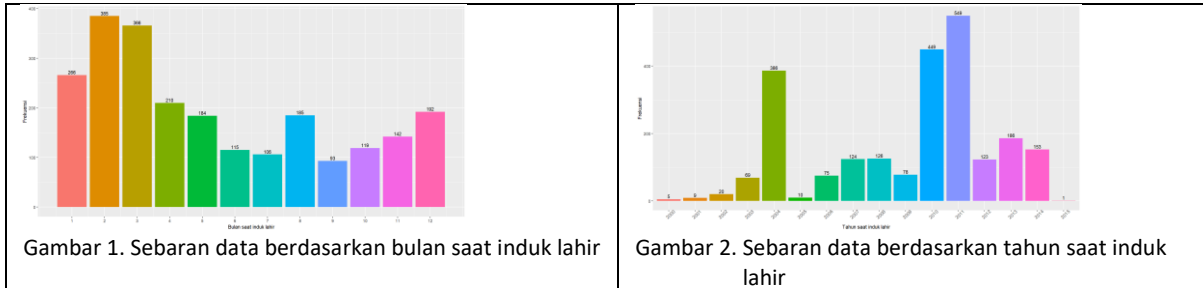
Data penelitian diekstrak dari *data base* catatan individu sapi perah FH yang terdiri dari 728.212 baris data produksi susu harian yang dilengkapi dengan informasi tanggal/bulan/tahun lahir induk, tanggal/bulan/tahun induk saat beranak, hari saat induk diperah, dan periode laktasi. Sebanyak 1.942 data catatan harian produksi susu dari 1.015 ekor induk sapi perah sapi perah FH yang dicatat dalam *data base* BBPTUHPT Baturaden digunakan dalam penelitian ini. Data base yang tersimpan dalam format MS EXCEL tersebut kemudian diekstrak dan di-*filter* menggunakan program R (R Core Team, 2022) pada antar muka Rstudio (Team, 2023) untuk beberapa kriteria. Kriteria yang digunakan dalam melakukan *filtering* data adalah periode laktasi 1-7, hari induk diperah (*Days in Milk*) 7-600, produksi susu harian 10-60 kg, umur saat induk beranak 500-3000 hari dan jumlah hari pemerahan 40-750 hari.

Bulan saat induk lahir (*mob*), tahun saat induk lahir (*yob*), bulan saat induk partus (*mopar*), tahun saat induk partus (*yopar*), umur saat induk beranak (*agepar*), jumlah hari pemerahan (*jhp*) dan periode laktasi (*laktasi*) dijadikan sebagai peubah-peubah bebas dalam penelitian sedangkan peubah tak bebas dalam penelitian adalah produksi susu per laktasi. Umur saat induk beranak dihitung menggunakan tanggal lahir dan tanggal saat beranak.

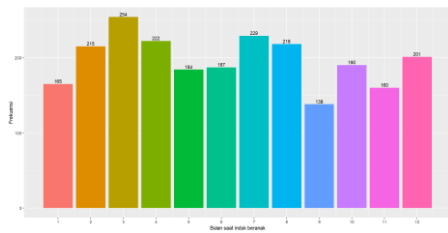
Analisis linear regresi berganda, matriks korelasi, penghitungan VIF dan visualisasi grafik dalam penelitian dilakukan menggunakan program R (R Core Team, 2022) dalam program antar muka Rstudio (Team, 2023). Kriteria VIF wajar yang digunakan untuk mengevaluasi kolinearitas peubah bebas adalah <10 mengikuti Dormann et al. (2012).

## Hasil dan Pembahasan

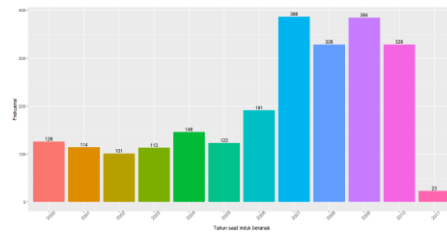
Distribusi data untuk masing-masing peubah bebas (bulan saat induk lahir, tahun saat induk lahir, bulan saat induk beranak, tahun saat induk beranak, umur saat induk beranak, periode laktasi dan jumlah hari pemerahan) disajikan pada Gambar 1, 2, 3, 4, 5, 6, dan 7.



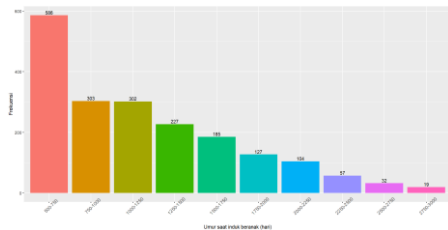
Balai Besar Pembibitan Ternak Unggul dan Hijauan Pakan Ternak (BBPTUHPT) Baturaden merupakan balai pembibitan yang terbuka (open nucleus) yang memungkinkan masuknya individu baru ke dalam populasi misalnya melalui ternak impor. Seperti pusat pembibitan di negara tropis lainnya, waktu kelahiran pedet tidak hanya terbatas pada bulan-bulan tertentu meskipun adanya perbedaan frekuensi induk yang beranak (Gambar 1). Sebaran jumlah induk yang beranak berdasarkan tahun juga tidak seragam. Hal ini diduga terkait dengan struktur populasi dan bukan berhubungan dengan bulan dan tahun itu sendiri.



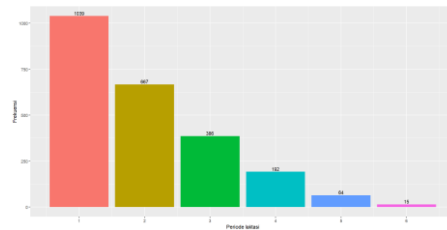
Gambar 3. Sebaran data berdasarkan bulan saat induk beranak



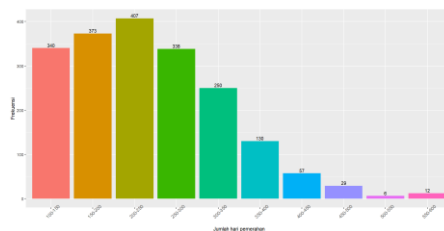
Gambar 4. Sebaran data berdasarkan tahun saat induk beranak



Gambar 5. Sebaran data berdasarkan umur induk



Gambar 6. Sebaran data berdasarkan periode laktasi



Gambar 7. Sebaran data berdasarkan jumlah hari pemerahan

Distribusi data berdasarkan bulan maupun tahun saat induk beranak juga tidak seragam namun data berdasarkan bulan saat induk beranak cenderung seragam (Gambar 3). Jumlah data produksi

susu berdasarkan tahun saat induk beranak menunjukkan adanya perbedaan utamanya pada tahun kelahiran 2007-2010 (Gambar 4). Distribusi data yang demikian juga diduga terkait dengan struktur populasi dari data.

Distribusi data berdasarkan umur induk dan periode laktasi cenderung memiliki pola yang mirip yaitu sebagian besar data produksi susu berasal dari induk umur muda dengan periode laktasi awal kemudian bergeser ke kelompok induk dengan umur tua dengan periode laktasi akhir (Gambar 5 dan 6). Distribusi data produksi susu berdasarkan jumlah hari pemerahan (JHP) cenderung mendekati normal yaitu sebagian besar data berasal dari produksi susu dengan JHP di sekitar nilai tengah. Hasil analisis korelasi antara peubah bebas divisualisasikan pada Gambar 8 yang menunjukkan matriks korelasi semua peubah bebas yang dianalisis. Peubah periode laktasi dan umur pada saat induk beranak dan peubah tahun lahir induk dan tahun saat induk beranak memiliki tingkat hubungan yang sangat tinggi. Koefisien korelasi yang sangat tinggi bisa dijadikan indikator adanya kolinearitas antar peubah tersebut (Alin , 2010). Namun demikian, koefisien korelasi yang tinggi tidak cukup akurat digunakan untuk menentukan adanya kolinearitas antar peubah bebas (Mihola and Bílková , 2014).



Gambar 8. Matriks korelasi antar peubah bebas

Mihola and Bílková (2014) menyatakan bahwa untuk mengukur tingkat multikolinearitas dalam model regresi, metrik tambahan seperti faktor inflasi varians (VIF) dan nomor kondisi juga harus digunakan. Deteksi adanya multikolinearitas pada linear regresi berganda yang paling umum adalah menggunakan Variance Inflation Factor (VIF) atau Tolerance dan Condition Number (CN) (Salmerón-Gómez et al. , 2020). Selain itu deteksi awal kolinearitas dapat dilakukan secara deskriptif membandingkan hubungan antar peubah-peubah melalui matriks korelasinya.

Hasil analisis multiple regresi menggunakan semua peubah bebas (full model) yaitu menggunakan peubah bebas periode laktasi, jumlah hari pemerahan, bulan/tahun saat induk lahir, bulan/tahun saat induk beranak, dan umur saat induk beranak, diperoleh koefisien determinasi sebesar 90.42 %. Hasil analisis tersebut menunjukkan bahwa secara terpisah semua peubah bebas memiliki pengaruh yang signifikan terhadap produksi susu kecuali periode laktasi. Hasil analisis regresi menggunakan full model tersebut menghasilkan variance inflation factor (VIF) seperti yang disajikan pada Tabel 1.

Tabel 1. Nilai Probabilitas t hitung dan VIF dari analisis multiple regresi full model

Peubah bebas	Prob. ( $> t $ )	VIF	Keterangan♦
lakt	$>0,05$	8,5135	Normal
JHP	$<0,001^{***}$	1,0767	Normal
mob	$<0,05^*$	22,0619	Kolinear
yob	$<0,05^*$	2641,0943	Kolinear
mopar	$<0,001^{***}$	20,1346	Kolinear
yopar	$<0,05^*$	2140,4154	Kolinear
AgePar	$<0,01^{**}$	533,8004	Kolinear

Catatan:

lakt= periode laktasi; JHP= jumlah hari pemerahan; mob= bulan saat induk lahir; yob= tahun saat induk lahir; mopar= bulan saat induk beranak; yopar= tahun saat induk beranak; AgePar= kelas umur saat induk beranak.

♦) Status kolinearitas berdasarkan nilai VIF

Tabel 1 menunjukkan adanya VIF yang tidak wajar. Kriteria VIF untuk menentukan adanya kolinearitas dilaporkan berbeda-beda misalnya (Mihola and Bílková , 2014) menyatakan bahwa nilai VIF yang wajar adalah  $<5$  sedangkan (Dormann et al. , 2012) dalam artikel review-nya merangkum bahwa nilai VIF yang wajar adalah  $<10$ . Hasil penelitian menunjukkan bahwa beberapa peubah bebas memiliki nilai VIF yang tak wajar sehingga patut diduga peubah-peubah tersebut memiliki kolinearitas. Pada penelitian kriteria VIF wajar mengikuti Dormann et al. (2012) yaitu  $VIF < 10$ .

Hasil analisis multiple regresi menggunakan peubah bebas yang lebih sedikit (sub model) yaitu periode laktasi, jumlah hari pemerahan, bulan dan tahun saat induk beranak menghasilkan koefisien determinasi sebesar 90.3 % dan VIF seperti yang disajikan pada Tabel 2.

Ada beberapa metode untuk mengatasi kolinearitas dalam regresi linier berganda, antara lain: (1) Menghapus satu atau lebih variabel prediktor yang berkorelasi tinggi dari model, (2) Menggabungkan variabel prediktor yang sangat berkorelasi menjadi satu variabel, (3) Menggunakan teknik regularisasi, seperti regresi ridge atau regresi laso (Adnan et al. , 2006), yang dapat membantu mengurangi dampak kolinearitas pada estimasi parameter. Secara keseluruhan, penting untuk menyadari potensi kolinearitas dalam regresi linier berganda dan mengambil langkah-langkah untuk mengatasinya agar mendapatkan hasil yang akurat dan dapat diandalkan (Samsa et al. , 2005).

Tabel 1. Nilai Probabilitas t hitung dan VIF dari analisis multiple regresi sub model

Peubah bebas	Prob. ( $> t $ )	VIF	Keterangan
laktasi	$<0,001^{***}$	1,0045	Normal
JHP	$<0,001^{***}$	1,0741	Normal
mopar	$<0,001^{***}$	1,0548	Normal
yopar	$<0,001^{***}$	1,0350	Normal

Keterangan: sama dengan Error! Reference source not found.

Hubungan yang sangat erat antara peubah-peubah bebas dalam linear regresi berganda akan menyebabkan terjadinya multikolinearitas yang dapat menyebabkan hasil koefisien taksiran peubah-peubah bebasnya menjadi bias. Pada kondisi kolinearitas *exact*, koefisien taksiran koefisien peubah bebasnya tak mungkin bisa diperoleh (Gómez et al. , 2012) (Salmerón-Gómez et al. , 2020).

## Kesimpulan

Dapat disimpulkan bahwa peubah bulan saat induk lahir, tahun saat induk lahir, bulan saat induk beranak, tahun saat induk beranak dan umur saat induk beranak memiliki kolinearitas. Peubah



tersebut mengganggu hasil taksiran dalam analisis multipel regresi sehingga harus dikeluarkan dari model regresi.

## Daftar Pustaka

- Adnan, N, MH Ahmad, and R Adnan. 2006. A comparative study on some methods for handling multicollinearity problems. *Matematika*. 22(2):109–119. Retrieved from <http://eprints.utm.my/3662/%5Cnpapers3://publication/uuid/EAAA50DB-3CD8-4D34-BF50-221ED0C31832>
- Alin, A. 2010. Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*. 2(3):370–374.
- Chen, Y. 2014. Covariance among independent variables determines the overfitting and underfitting problems in variation partitioning methods: with a special focus on the mixed co-variation. *ArXiv Preprint ArXiv:1402.3324*.
- Dormann, CF, J Elith, S Bacher, C Buchmann, G Carl, G Carré, JRC Marquéz, B Gruber, B Lafourcade, PJ Leitão, T Münkemüller, C McClean, PE Osborne, B Reineking, B Schröder, AK Skidmore, D Zurell, and S Lautenbach. 2012. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*. 35:001–020. <http://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Garamszegi, LZ. 2006. Comparing effect sizes across variables: generalization without the need for Bonferroni correction. *Behavioral Ecology*. 17:682–687.
- Gómez, MD, PJ Azor, ME Alonso, J Jordana, and M Valera. 2012. Morphological and genetic characterization of Spanish heavy horse breeds: Implications for their conservation. *Livestock Science*. 144(1–2):57–66. <http://doi.org/10.1016/j.livsci.2011.10.013>
- Henderson, CR. 1984. *Applications of Linear Models in Animal Breeding*. Guelph, ON, Canada: University of Guelph Press.
- Jinn, JH, 宋傳欽, and WB Mou. 2007. *An Application of the Generalized Shrunken Least Squares Estimator on Principal Component Regression*.
- Mihola, J, and D Bílková. 2014. *Measurement of Multicollinearity Using Determinants of Correlation Matrix*.
- Mrode, RA. 2014. *Linear Models for the Prediction of Animal Breeding Values: 3rd Edition*. CABI.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from <https://www.r-project.org/>
- Salmerón-Gómez, R, A Rodríguez-Sánchez, and C García-García. 2020. Diagnosis and quantification of the non-essential collinearity. *Computational Statistics*. 35(2):647–666.
- Samsa, G, G Hu, and M Root. 2005. Combining information from multiple data sources to create multivariable risk models: illustration and preliminary assessment of a new method. *Journal of Biomedicine & Biotechnology*. 2005(2):113–123. <http://doi.org/10.1155/JBB.2005.113>
- Susanto, A, Suyadi, VMA Nurgiartiningih, and L Hakim. 2018. (Co)variance components and genetics parameter estimation for linear traits in Holstein cattle in Indonesia: Traits related to foot/leg and udder. *Archives Animal Breeding*. 61(4):491–496. <http://doi.org/10.5194/aab-61-491-2018>
- Team, Rs. 2023. *RStudio: Integrated Development Environment for R*. Boston, MA. Retrieved from <http://www.posit.co/>