

X Simpósio Brasileiro de Melhoramento Animal Uberaba, MG – 18 a 23 de agosto de 2013

Distribuição Normal e Poisson para GWAS considerando características reprodutivas em suínos¹

Lucas Lima Verardo², Fabyano Fonseca e Silva³, Paulo Sávio Lopes⁴, Simone Eliza Facione Guimarães⁴

¹Trabalho financiado pela FAPEMIG

²Programa de Pós-Graduação em Zootecnia – UFV, Viçosa. Bolsista do CNPq. e-mail: lucas_verardo@yahoo.com.br

³Departamento de Estatística – UFV, Viçosa. e-mail: fabyanofonseca@ufv.br

⁴Departamento de Zootecnia – UFV, Viçosa. e-mail: plopes@ufv.br; sfacione@ufv.br

Resumo: As características reprodutivas estão diretamente relacionadas com o número de leitões desmamados por porca mostrando sua importância para a indústria de suínos. Para o melhoramento genômico destas características, detalhados estudos de associações genômica ampla são necessários. Mas, ao contrário da maioria dos estudos de associação, que consideram os fenótipos contínuos e assumem distribuição normal, características reprodutivas são caracterizadas como dados discretos, ou como uma variável de contagem, o que poderia seguir uma distribuição de Poisson. Assim, é apresentada uma comparação de distribuição Normal e Poisson a partir de estudos de associação de SNP em número total de leitões nascidos e número total de leitões desmamados. As análises comparativas permitiram confirmar a superioridade da distribuição Normal para os dados estudados em estudos de associação genômica.

Palavras-chave: Numero total de leitões desmamados, nascidos, SNP.

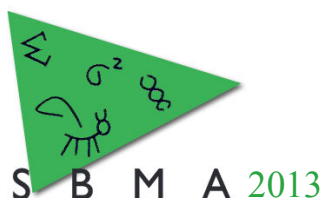
Normal and Poisson distribution for GWAS considering pig's reproductive traits

Abstract: Reproductive trait is directly related to the number of piglets weaned per sow showing its importance to pigs industry. For genomic improvement of these traits a detailed whole-genome associations studies in these are require. But as opposite to most association studies, which consider continuous phenotypes and normal assumption, reproductive traits are characterized as discrete traits, or as a counting variable, which could potentially, follow a Poisson distribution. Towards this, we present a comparison of Gaussian and Poisson distribution from SNP association studies in total number born and total number weaned piglets. Our comparative analyses allowed us to confirm the superiority of Gaussian distribution in association studies for the analyzed data.

Keywords: Total number born, weaned, SNP.

Introdução

Características reprodutivas estão diretamente relacionadas com o número de leitões desmamados por porca mostrando sua importância para a indústria de suínos. Estas têm sido estudadas amplamente em produção de suínos por meio de características como o número total de leitões nascidos, número de leitões nascidos vivos, número de leitões natimortos dentre outras (Schneider et. Al., 2012). Sendo estas características indicadas pela literatura como de baixa a média herdabilidade, e de grande importância econômica, a associação de painéis de marcadores de Polimorfismos de Nucleotídeos Únicos (SNP) para essas características é necessária. Atualmente, muitos estudos tem sido realizados usando associação genômica ampla (GWAS) (Uimari, et al, 2011; Onteru et al, 2011 e Schneider et al, 2012), sugerindo associação entre características reprodutivas e SNPs assumindo que os fenótipos do modelo são normalmente distribuídos. Porém, características reprodutivas como o número total de leitões nascidos e desmamados são caracterizados como discretos, mais especificamente como uma variável de contagem, o que poderia potencialmente seguir uma distribuição de Poisson (Silva et. al., 2011). Modelos de Poisson são alternativas estatísticas potencialmente mais apropriadas em detecção de Locus de características quantitativas (QTL) para estas características reprodutivas, sendo também aplicado em modelos mistos, mas ainda não em GWAS, sendo um ponto a ser estudado. A partir destes dados, o objetivo deste trabalho foi a comparação entre as distribuições Normal e Poisson em estudos de associação de SNP para número de leitões nascidos e desmamados.



Material e Métodos

Neste trabalho foi utilizada uma população de terceira geração. Resumidamente, dois avôs da raça Piau foram cruzados com dezoito avós, compostas das raças Large White, Landrace e Pietrain, produzindo a geração F1 a partir do qual foram selecionados 11 machos e 54 fêmeas. O cruzamento desta geração F1 produziu a população F2 sendo esta a terceira geração. O DNA foi extraído no Laboratório de Biotecnologia Animal do Departamento de Zootecnia da UFV. O DNA genômico foi extraído de células brancas de animais parentais, F1 e F2. O SNPChip personalizado com 384 SNPs baseou-se no BeadChip SNP60 de suíno da Illumina (San Diego, CA, EUA). Os SNPs usados foram selecionados de acordo com o seu espaçamento no interior dos cromossomos sendo utilizados 237 marcadores SNPs distribuídos como segue: ssc1 (56), SSC4 (54), SSC7 (59), SSC8 (30), SSC17 (25) e SSCX (13), sendo a distância entre eles (Mb), respectivamente: 5,17, 2,37, 2,25, 3,93, 2,68 e 11,0. Um modelo de regressão múltipla bayesiana considerando distribuições diferentes para os dados (Normal e Poisson) foi utilizado. Nestes modelos todos SNPs são ajustados simultaneamente (Bayes A) e foi usada a opção de especificar distribuições informativas a priori, assim, a fim de comparar o modelo Normal e Poisson, os valores de *deviance* a posteriori (D) foram usados, sendo os menores valores de D implicando em melhor ajuste do modelo. Para as análises, foi utilizado o software R por meio do pacote MCMCpack (Martin et al, 2011.) considerando, respectivamente, as funções MCMChregress e MCMChpoisson (Chib e Carlin, 1999) para ajustar os modelos Normal e de Poisson (maiores informações dos modelos consultar as literaturas dos pacotes utilizados).

Resultados e Discussão

Uma vez que os números totais de leitões nascidos e desmamados são considerados variáveis de natureza contáveis, tentou-se provar a eficiência da distribuição Normal em relação a Poisson em estudos de associação de SNPs. A distribuição de Poisson era esperada ter um melhor papel para esse fim, mas não foi o observado. A *deviance* da distribuição de Poisson foi significativamente maior em comparação com a distribuição Normal para as características analisadas como segue na Tabela 1 e Figura 1. Isto pode ser explicado pelo fato de que a distribuição de Poisson é assimétrica à direita e distorcida, e mesmo que seja contínua, a simetria da distribuição Normal assegurou o melhor ajuste, provavelmente, pois ela foi a mais consistente com a distribuição de dados das amostras observadas.

Tabela 1. Média da *deviance* e a desvio padrão (DP) de cada distribuição e o contraste entre elas (Poisson/Normal) para cada característica: Numero Total de Nascidos (NTN) e Numero Total de Desmamados (NTD).

Característica	Distribuição	Média	DP
NTN	Normal	1304,2155	35,51975
	Poisson	1697,2907	18,24981
	Poisson/ Normal	393,0752	39,05684
NTD	Normal	1264,3002	35,79486
	Poisson	1643,8523	17,39598
	Poisson/ Normal	379,5521	38,71987

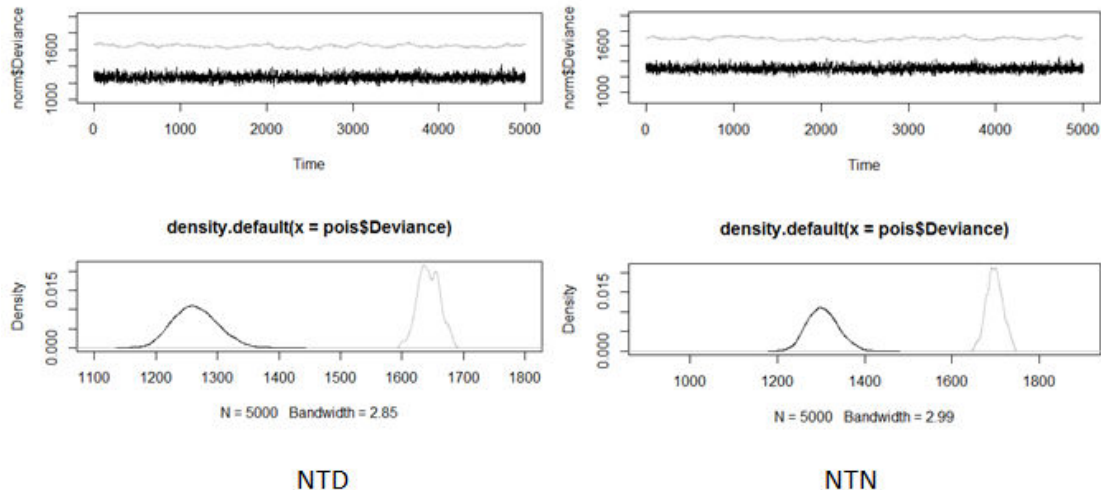


Figura 1. Curvas da *deviance* a posteriori e densidade para as distribuições Normal (linha preta) e Poisson (linha cinza) para cada característica. NTD: Número total desmamado e NTN: Número total de nascidos.

Conclusões

Com os resultados obtidos, possibilitou confirmar a superioridade da distribuição Normal para as características reprodutivas estudadas. Assim, é possível considerar que esta distribuição é a que melhor se ajusta em análises de GWAS para as características de número total de leitões nascidos e desmamados. Porém, outras distribuições discretas mais poderosas, como a Binomial Negativa, também devem ser comparadas com a Normal a fim de identificar aquela que melhor descreve características de contagens, como as reprodutivas, contempladas no presente trabalho.

Agradecimentos

FAPEMIG, CAPES e CNPq

Literatura citada

- MARTIN, A.D.; QUINN, K.M.; PARK, J.H. MCMCpack: Markov chain Monte Carlo (MCMC) Package. *Journal Of Statistical Software*, v.42(9), p.1-21, 2011.
- ONTERU, S.K.; FAN, B.; DU, Z-Q. et al. A whole-genome association study for pig reproductive traits. *Animal Genetics*, v.43, p.18–26, 2011.
- SCHNEIDER, J. F.; REMPEL, L.A.; ROHRER, G.A. Genome-wide association study of swine farrowing traits. Part I: Genetic and genomic parameter estimates. *Journal of Animal Science*, v.90, p.3353-3359, 2012.
- SILVA, F.F.; TUNIN, K.P.; ROSA, G.J.M. et al. Zero-inflated Poisson regression models for QTL mapping applied to tick-resistance in a Gyr x Holstein F2 population. *Genetics and Molecular Biology*, v.34(4), p.575-581, 2011.
- UIMARI, P.; SIRONEN, A.; SEVÓN-AIMONEN, M-L. Whole-genome SNP association analysis of reproduction traits in the Finnish Landrace pig breed. *Genetics Selection Evolution*, v.43, p.42, 2011.
- CHIB, S. & CARLIN, B.P. On MCMC Sampling in Hierarchical Longitudinal Models. *Statistics and Computing*, v.9, p.17-26, 1999.