第 29 卷 第 2 期　　西北农林科技大学学报(自然科学版)　　Vol. 29 No. 2
2001 年 4 月　　Jour. of Northwest Sci-Tech Univ. of Agri. and For. (Nat. Sci. Ed.)　　Apr. 2001

多元回归估计关中黑猪胴体瘦肉量的研究*

张慧林, 刘小林
(西北农林科技大学 畜牧兽医学院, 陕西 杨陵 712100)

[摘　要]　以活重 90 kg 左右的 30 头关中黑猪的宰前活重、体长、胴体重、胴体长、眼肌面积、腿臀重、背膘厚度为自变量, 胴体瘦肉量为依变量, 采用逐步回归的多元回归分析方法, 逐步引入作用较大的自变量又逐步剔除作用较小的自变量, 选择最优回归方程, 最后建立了 $R = 0.9515$ 的"最优"三元回归方程: $Y = -30.5194 + 0.5537X_3 + 0.2742X_4 + 0.3316X_5$, 经对 43 头关中黑猪的胴体瘦肉量进行估测, 与实测值差异不显著($P > 0.05$)。

[关键词]　关中黑猪; 胴体; 瘦肉量; 回归分析
[中图分类号]　S828.9+12　　[文献标识码]　A　　[文章编号]1000-2782(2001)02-047-04

关中黑猪是一个优良的培育品种, 具有繁殖力强, 生长快, 肉质好, 抗病力及适应性强的特点。为了进一步提高关中黑猪的种质特性, 加强关中黑猪的选育, 胴体瘦肉量的测定是一项重要工作。该指标也是现代猪育种中选择的主要经济性状之一。传统的测定方法仍然是依靠左半胴体的肉、脂、皮、骨的剥离称重, 工作浩繁, 准确性差且破坏了胴体的完整性, 难以给选择和生产提供及时准确的信息。国内外许多学者致力于对该性状进行间接估测和间接选择方法的研究[1~8], 但限于各自选择的猪品种和取材范围的不同, 尚无理想的可行方案。瘦肉量这一内部结构性状既受其他内部结构性状的影响, 又与外部形态性状有一定相关, 且与任何一个性状间的相关程度都不很高, 因此, 用简单的一元线性回归方程不可能很准确进行估计。为了提高估计的准确程度, 同时考虑到应用上的方便, 本研究以测量较为简便且与胴体瘦肉量相关程度较高的一些活体性状和胴体性状为自变量, 配置了估测胴体瘦肉量的"最优"多元回归方程, 以为关中黑猪的选育和生产提供理论依据。

1　材料和方法

本研究采用 1987～1997 年在咸阳市分 4 次屠宰的 30 头关中黑猪测定的结果。供试猪平均为 7.5 月龄, 平均体重为 91.46 kg, 空腹 24 h 称量宰前活重, 测量体尺指标。屠宰测定按全国统一方法进行, 宰后将左半胴体称测胴体重, 由第一肋骨与胸骨联合点起至耻骨联合前缘测定胴体长度, 在第 6～7 肋骨处测定背膘厚, 在腰荐结合处与背线成直角作垂直切割, 称测腿臀重, 在腰荐结合处用卡尺测定背最长肌的横截面积(宽度×厚度×0.70)为眼肌面积[9]。将左半胴体进行肉、脂、皮、骨分离并分别称重。

根据活体和胴体指标测定结果, 剔除了与瘦肉量相关程度不大的体尺指标和共线性的脂、皮、骨重指标, 瘦肉量估测采用表型相关分析和逐步回归的多元分析方法[10]。

2　结果

2.1　各性状的表型参数

所测各性状的平均数、标准差及变异系数列于表 1。

2.2　各性状间的表型相关系数

所测定的各性状间表型相关系数计算结果列于表 2。

* [收稿日期]　2000-06-12
[基金项目]　西北农业大学青年科学基金资助项目(005-Y08183)的部分内容
[作者简介]　张慧林(1961-), 女, 陕西户县人, 讲师, 主要从事家畜遗传育种与繁殖研究工作。

表1 各性状的平均数、标准差及变异系数

Table 1 The average value, standard deviation and variation coefficient of the traits

性状 Trait	活重/kg Live weight (X_1)	体长/cm Body length (X_2)	胴体重/kg Dressed weight (X_3)	胴体长/cm Carcass length (X_4)	眼肌面积/cm² Eye area (X_5)	腿臀重/kg Ham joint weight (X_6)	背膘厚度/cm Back-fat thickness (X_7)	瘦肉量/kg Lean meat weight (Y)
平均数 $\bar{X}$	91.457 7	119.653 8	65.263 1	76.307 9	25.986 2	18.282 3	3.088 7	35.160 8
标准差 S	2.933 3	10.459 0	1.645 1	3.152 6	2.499 8	1.012 4	0.035 2	2.026 5
变异系数 CV(%)	3.207 3	8.741 1	2.520 7	4.131 4	9.619 7	5.537 6	1.139 6	5.763 5

表2 性状间表型相关系数

Table 2 The phenotype correlation coefficient between the traits

性状 Trait	X_1	X_2	X_3	X_4	X_5	X_6	X_7	Y
X_1	1	0.187 2	0.527 7**	0.391 1**	0.285 3	0.485 6**	0.272 7	0.598 3**
X_2		1	0.368 9*	0.492 6**	0.407 1*	0.290 2	-0.364 5*	0.542 8**
X_3			1	0.157 4	0.376 9*	0.489 1**	-0.120 8	0.671 1**
X_4				1	0.444 2*	0.363 2*	-0.423 7*	0.679 5**
X_5					1	0.472 4**	-0.501 8**	0.768 0**
X_6						1	0.115 2	0.601 6**
X_7							1	-0.454 5*

注(Note): $r_{0.05,28}=0.361$, $r_{0.01,28}=0.463$。

2.3 胴体瘦肉量逐步回归的最优回归方程的建立

2.3.1 第一个自变量的引入 根据性状间的表型相关系数,分别计算各性状对瘦肉量的方差贡献,即标准回归平方和。由于各性状与瘦肉量的相关程度不同,所以对瘦肉量的方差贡献也各不相同。在7个自变量中,眼肌面积(X_5)的方差贡献($U_5^{(1)}=(r_{5y}^{(0)})^2=0.589\ 8$)最大,$F$检验时,$F_5^{(1)}=(N-2)U_5^{(1)}/(1-U_5^{(1)})=(30-2)\times0.589\ 8/(1-0.589\ 5)=40.259\ 4$。由于$F_5^{(1)}>F_{0.01,(1,28)}=2.89$, $P<0.01$,故眼肌面积(X_5)入选。

对相关系数矩阵$R^{(0)}$施行L_5变换,即$L_5R^{(0)}=R^{(1)}$,则$C_{55}^{(1)}=r_{55}^{(1)}=1$, $b_5^{(1)}=0.768\ 0$。

建立的标准化回归方程为$Y=0.768u_5$。

一般回归方程为$Y=18.982\ 0+0.622\ 6X_5$, $R^2=0.589\ 8$, $R=0.768\ 0$。

2.3.2 第二个自变量的引入 重新计算其余6个自变量对瘦肉量的方差贡献,得胴体重(X_3)的方差贡献$U_3^{(2)}=(r_{3y}^{(1)})^2/r_{33}^{(1)}=0.169\ 5$最大,其$F_3^{(2)}=(N-3)U_3^{(2)}/(r_{yy}^{(1)}-U_3^{(2)})=27\times0.169\ 5/(0.410\ 2-0.169\ 5)=19.013\ 3$, $F_3^{(2)}>F_{0.01,(1,28)}=2.90$, $P<0.01$故X_3入选。对$R^{(1)}$施行L_3变换,$L_3R^{(1)}=R^{(2)}$,则$b_3^{(2)}=r_{3y}^{(2)}=0.444\ 5$, $b_5^{(2)}=r_{5y}^{(2)}=0.601$。建立的二元标准化回归方程为$Y=0.444\ 5u_3+0.601u_5$。

一般的二元回归方程为

$Y=-13.182\ 3+0.547\ 6X_3+0.485\ 2X_5$

$R^2=0.760\ 7$, $R=0.872\ 2$

对二元回归方程进行方差分析(表3)表明,二元回归方程的回归关系极显著。

表3 二元回归方程的方差分析

Table 3 Analysis of variance of two-way regression equation

方差来源 Variance source	Df	SS	MS	F	临界F值 Critical F value
回归 Regress(X_3X_5)	2	$1-r_{yy}^{(2)}=0.759\ 3$	0.379 7	42.592**	$F_{0.05,(2,27)}=3.35$
剩余 Residual	27	$r_{yy}^{(2)}=0.240\ 7$	0.008 9		$F_{0.01,(2,27)}=5.49$
总和 Total	29	1			

2.3.3 判断是否应从回归方程中剔除变量 X_3为刚经检验引入,不能剔除。对于X_5,则$F_5^{(2)}=(N-3)U_5^{(2)}/Q^{(2)}=(30-3)\times0.3094/0.2407=34.7063>F_{0.01,(1,27)}=2.90$, $P<0.01$,故X_3和X_5均应保留。

2.3.4 第三个自变量的引入 在剩下的5个自变量中,胴体长(X_4)的方差贡献($U_4^{(3)}=(r_{4y}^{(3)})^2/r_{44}^{(2)}=0.146\ 1$)最大,$F_4^{(3)}=(N-4)\times U_4^{(3)}/(r_{yy}^{(2)}-U_4^{(3)})=26\times0.146\ 1/(0.240\ 7-0.146\ 1)=40.514\ 3>F_{0.01,(1,26)}=7.72$, $P<0.01$,故X_4入选。对$R^{(2)}$施行L_4变换,即$L_4R^{(2)}=R^{(3)}$,则$b_3^{(3)}=r_{3y}^{(3)}=0.449\ 5$, $b_4^{(3)}=r_{4y}^{(3)}=0.426\ 6$, $b_5^{(3)}=r_{5y}^{(3)}=0.409\ 1$,建立的三元标

准化回归方程为

$$Y = 0.4495u_3 + 0.4266u_4 + 0.4091u_5。$$

对三元回归方程进行方差分析(表 4), 表明三元回归方程的回归关系极显著。

表 4　三元回归方程的方差分析

Table 4　Analysis of variance of three-variate regression equation

方差来源 Variance source	Df	SS	MS	F	临界 F 值 Critical F value
回归(X_3, X_5, X_5) Regress	3	$1 - r_{yy}^{(3)} = 0.7593$	0.2531	27.34**	$F_{0.05,(3,26)} = 3.37$
剩余 Residual	26	$r_{yy}^{(3)} = 0.2407$	0.0093		$F_{0.01,(3,26)} = 5.53$
总和 Total	29	1			

2.3.5　判断是否应从回归方程中剔除变量

$F_3^{(3)} = (N - 4)U_3^{(3)}/Q^{(3)} = 26 \times 0.4495^2/(0.0946 \times 1.1658) = 47.6341^{**}$

$F_4^{(3)} = (N - 4)U_4^{(3)}/Q^{(3)} = 26 \times 0.4266^2/(1.2460 \times 0.0936) = 40.1543^{**}$

$F_5^{(3)} = (N - 4)U_5^{(3)}/Q^{(3)} = 26 \times 0.40912/(1.4163 \times 0.0946) = 32.4777^{**}$

$F_{0.05,(1,26)} = 4.23$, $F_{0.01,(1,26)} = 7.72$, 表明与 3 个变量均呈现极显著的回归关系, 即 $P < 0.01$, 故 X_3, X_4, X_5 均应保留。

2.3.6　检验是否接受新变量

在剩下的 4 个变量中, 活重(X_1)的方差贡献(0.00102)最大, $F_1^{(4)} = (N - 5)U_1^{(4)}/(r_{yy}^{(3)} - U_1^{(4)}) = 25 \times 0.0102/(0.0946 - 0.0102)) = 3.0213 <$ $F_{0.05,(1,25)} = 4.24$, $P > 0.05$, 因此, 再没有变量可以引入方程, 从而逐步回归结束。所得最优回归方程为

$Y = -30.5194 + 0.5537X_3 + 0.2742X_4 + 0.3316X_5$, $R^2 = 0.9054$, $R = [1 - r_{yy}^{(4)}]1/2 = 0.9515^{**}$。

2.4　最优回归方程的回归估测检验

用逐步回归程序建立的估测胴体瘦肉量的最优回归方程对实际符合程度如何, 需对方程式进行验证。对原样本的 30 头关中黑猪和后来测定的 13 头关中黑猪的胴体瘦肉量进行估测, 计算了估测值与实测值的平均数和标准差, 平均离差和平均误差 $1/n \sum[(Y_e - Y_o)/Y_o \times 100\%]$, 并进行了均数差异显著性检验($t$ 检验), 其结果如表 5。

表 5　回归方程验证结果

Table 5　The test results of regression equation

样本 Sample	头数 Number	实验测值 Observed value Y_0	实验测值 Observed value Sy_0	估测值 Estimated value Y_e	估测值 Estimated value Sy_e	平均离差 Average deviation	平均误差值/% Average error	t 检验 t test	
Ⅰ	30	35.1608	2.0265	34.9521	2.4978	0.2087	0.19	$t = 0.355$	$P > 0.05$
Ⅱ	13	35.0069	2.1241	35.2415	2.0984	0.2346	0.50	$t = 0.283$	$P > 0.05$

表 5 检验结果表明, 2 个样本平均离差仅为 0.2087 和 0.2346 kg, 平均误差值为 0.19% 和 0.50%, 群体实测平均值与估测平均值差异在 2 个样本均不显著($P > 0.05$), 证明本研究的最优回归方程准确性好, 与实际符合程度高, 精确度也高。

3　分析与讨论

3.1　回归分析的自变量与依变量要显著相关

本研究所列的 7 个性状与胴体瘦肉量之间的表型相关系数均达极显著或显著水平, 是影响瘦肉量的主要因素, 被剔除的体高、胸围、胴体宽、皮厚等指标对瘦肉量的相关程度都未达到显著水平。

3.2　多性状回归估计准确性高

猪的胴体瘦肉量是一个受多因素影响的经济性状, 用简单相关的方法寻求个别相关性状予以估测, 准确性较差, 必须利用与瘦肉量相关程度高的几个性状, 建立多元回归方程, 才会使准确度有大的提高。本研究中, $R_{y \cdot (x_3, x_4, x_5)} > R_{y \cdot (x_3, x_5)} > R_{y \cdot (x_5)} > r_{5y} > r_{4y} > r_{3y} > r_{6y} > r_{1y} > r_{2y} > r_{7y}$, 完全证明了此结论。

3.3　高度相关的自变量不能在回归方程中共存

在所研究的 7 个自变量中, 活重(X_1)、体长(X_2)、腿臀重(X_6)和膘厚(X_7)虽然与瘦肉量的相关程度达到极显著或显著水平, 但均未进入最优多元回归方程。这主要是因为它们分别与入选的 3 个自变量相关性很强, 不能在最优回归方程中共存。

3.4　自变量的度量应简单、经济、准确、可靠、适用

建立估测胴体瘦肉量的最优多元回归方程, 关键问题是选择合适的自变量, 所选自变量既要与瘦肉量的相关程度高, 度量简便、经济, 估测准确可靠, 又要便于在科研和生产中推广应用。本研究提出的

最优多元回归方程估测瘦肉量的复回归关系极显著,准确性高($R=0.9515$),估测值与实际值的平均离差小,尤其是方程中的3个自变量简单易测,不破坏胴体完整性,是一个准确、简便、合理、适用的回归方程,但其是否适用于各品种猪的瘦肉量估测,有待进一步研究。

3.5 克服自变量间的多重共线性

多元回归分析是统计分析和预测的一种重要方法,在畜牧业生产及科研中得到了广泛应用,但自变量存在程度很高的多重共线性,将使最小二乘法失效,使多元回归分析失去重要信息而难以寻找最优回归方程。本研究中胴体重剖分为肉、皮、脂、骨重4部分,它们达到高度的共线性,通过删除皮、脂、骨重后,建立了理想的最优回归方程。

[参考文献]

[1] Aunan W J, Winters L M. A study of the variations of muscle, fat and bone of swine carcasses[J]. J Anim Sci, 1949, 8: 182-189.

[2] Cordray J C, Huffman D L, McGuite J A. Predictive equations for estimating protein and fat in the pork carcass[J]. J Anim Sci, 1978, 46: 666-673.

[3] Smith G C, Carpenter Z L. Evaluation of factors associated with the composition of pork carcasses[J]. J Anim Sci, 1973, 36: 473-478.

[4] 陈润生,王性善,汪嘉燮,等. 应用猪腿臀部个别肌肉估测猪胴体组成的研究[J]. 东北农学院学报, 1983, 39(3): 23-29.

[5] 李国豪,扈有蓉. 瘦肉量的不同估测方法对提高个体估测准确程度的探讨[J]. 中国畜牧杂志, 1984, (2): 12-17.

[6] 徐士清,翁经强,陈忠伟. 瘦肉量估测方法的研究[J]. 畜牧兽医学报, 1987, 18(4): 231-236.

[7] 张书松,杨国勋,任广志. 杜洛克猪胴体瘦肉量估测方法的研究[J]. 河南农业大学学报, 1995, 29(3): 282-284.

[8] 张慧林,刘小林. 关中黑猪活体及胴体性状与瘦肉量的相关分析[J]. 西北农林科技大学学报(自然科学版), 2001, 29(1): 106-109.

[9] 路兴中,郭传甲,吕志强,等. 现代猪肉生产理论与实践[M]. 北京: 中国农业科技出版社, 1994. 226-232.

[10] 西北农学院主编. 概率基础与数理统计[M]. 北京: 农业出版社, 1988. 253-262.

Prediction of the lean meat content of carcass with multivariate regression method in Guanzhong Black Pig

ZHANG Hu-lin, LIU Xiao-lin

(*College of Animal Sciences and Veterinary Medicine, Northwest Science and Technology University of Agricultural and Forestry, Yangling, Shaanxi* 712100, *China*)

Abstract: Thirty Guanzhong Black pigs with a live weight about 90 kg each were selected in this study. The live weight (X_1), final body length (X_2), dressed weight (X_3), carcass length (X_4), eye area (X_5), ham joint weight (X_6), back-fat thickness (X_7) were used as independent variables, and the lean meat weight of carcass (Y) was used as a dependent variable, using the method of stepwise regression analysis. After statistical examination and omitting the small insignificant independent variables, an optimum three-way regression equation for estimating the lean meat weight was established with $R=0.9575$: $Y=-30.5194+0.5537X_3+0.2742X_4+0.3316X_5$, in which X_3 refers to the dressed weight (kg), X_4 refers to the carcass length (cm), X_5 refers to the eye area (cm^2). The average value of prediction of the lean meat weight using the equation and the observed value were proved to have no significant difference through the t-test ($P>0.05$) in 43 Guanzhong Black Pigs. It shows this method is a simple and easy way to predict the lean meat content in pork carcass accurately, and can be used in both research and production.

Key words: Guanzhong black pig; carcass; lean meat content; regression analysis