Dealing with multicollinearity in predicting egg components from egg weight and egg dimension

Tarek M. Shafey,1 Ahmed H. Mahmoud,2 Mohamed A. Abuheif1
1Department of Animal Production, King Saud University, Riyadh, Saudi Arabia
2Department of Zoology, King Saud University, Riyadh, Saudi Arabia

Abstract

Measurements of 174 eggs from meat-type breeder flock (Ross) at 36 weeks of age were used to study the problem of multicollinearity (MC) instability in the estimation of egg components of yolk weight (YKWT), albumen weight (ALBWT) and eggshell weight (SHWT). Egg weight (EGWT), egg shape index (ESI) = egg width (EGWD) * 100/egg length (EGL) and their interaction (EGWTESI) were used in the context of un-centred vs centred data and principal components regression (PCR) models. The pairwise phenotypic correlations, variance inflation factor (VIF), eigenvalues, condition index (CI), and variance proportions were examined. Egg weight had positive correlations with EGWD and EGL (r = 0.56 and 0.50, respectively; P < 0.0001) and EGL had a negative correlation with ESI (r = -0.79; and 0.50, respectively; P < 0.0001). The components of eggs are of particular interest to producers and its acceptability to consumers (Stadelman, 1977). An average egg weights 58 g approximately. Of this weight, the yolk constitutes 16.8 g (29%), the albumen, 35.7 g (61.5%) and the eggshell 5.5 g (9.5%). The weight and dimensions of the chicken egg are very important characteristics that influence its components and grading and consequently the economic outcome of production (Pandey et al., 1986; Farooq et al., 2001, 2003; Wilson and Suarez, 1993). These egg characteristics are highly correlated and used to predict egg components (Choprakarn et al., 1998; Pandey et al., 1986; Wilson and Suarez, 1993; Farooq et al., 2003; Abanikandana et al., 2007). However, the idea of having many highly correlated variables of egg measurements entered into a regression model can lead to multicollinearity (MC) problems. Multicollinearity refers to a situation in which there is an exact (or nearly exact) linear relation among two or more of the input variables (Hawking and Pendleton, 1983). This has a potentially serious effect on the standard errors of the coefficients, which may mislead the interpretation of the results. Multicollinearity inflates the standard errors. Thus, it makes some variables statistically insignificant while they should be otherwise significant. This may create some difficulties to understand how the different external egg measurements impact on egg components. It also makes determining the contribution of each explanatory variable difficult because the effects of these variables are mixed. This relation among variables induces numerical instability into the estimates and alters the size of the coefficient of determination (Aziz and Sharaby, 1993). There are many statistical solutions to the MC problem, such as mean centring or using the principal component regression (PCR) model (Yu, 2008). Much of the published research seems to only predict egg components from a set of egg measurements, which is acceptable to disregard MC. The predictions will still be accurate and the overall R² quantifies how well the model predicts egg components. However, the model in this case does not help to understand how the various egg measurements impact the values of egg components. There is no available information on the problem of MC among egg weight and dimensions, the predictors of egg components. In addition, most studies examine main effects or linear effects, and interactions are only available in few publications. This is due to the fact that such interaction effects are difficult to interprept.

This study aimed at dealing with the problem of MC in the prediction of egg components of yolk weight (YKWT), albumen weight (ALBWT) and eggshell weight (SHWT) from egg weight (EGWT) and egg shape index (ESI). The effects of EGWT and ESI and their interaction (EGWTESI) in the regression models of un-centred data, centred data and principal components (PCs) on the variance inflation factor (VIF), eigenvalue and condition index (CI), variance proportions and prediction of egg components were examined.

Materials and methods

A total of 174 freshly laid eggs produced by a meat-type breeder flock (Ross-Awadi, Riyadh, Saudi Arabia) at 36 weeks of age were collected. Eggs were candled using light to detect cracks and other defects. Eggs were numbered and weighed individually by electronic scale. Measurements of length and width of the eggs (EGL and EGWD, respectively) were taken by steel vernier caliper graduated to 1/10 of 1 mm. Eggs were broken and albumen and yolk separated. The yolk was then carefully rolled on a...
employed as follows: the following formula given by Panda (1996):

\[ Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + \epsilon \]

where: \( Y = \) dependent variable (YKWT, ALBWT, SHWT); \( a = \) intercept; \( b_n = \) regression coefficients; \( X_n = \) independent variables (X1, X2 and X3 for EGWT, ESI and EGWTESI, respectively).

SAS (2006) was used in the analysis.

**Results and discussion**

The mean and coefficient of variation of egg characteristics and correlations among egg characteristics are shown in Tables 1 and 2, respectively. Egg weight, EGWD, EGI, ESI, YKWT, ALBWT and SHWT averaged 58.83 g, 43.54 mm, 55.28 mm, 78.84%, 17.85 g, 35.74 g and 5.22 g, respectively. There were significant positive correlations (P<0.0001) between EGWT and egg dimensions (EGWD and EGL, \( r = 0.56 \) and 0.50, respectively), and between ESI and EGWD (\( r = 0.48 \)). There was a significant negative correlation (P<0.0001) between EGL and ESI (\( r = 0.79 \)). Mean centring did not alter the correlations between independent and dependent variables. Similar findings on the relationship between ESI with either EGWD or EGI were reported in chicken and Japanese quail eggs (Ozcelik, 2002; Kul and Seker, 2004; Abanikannda et al., 2007). In contrast, Olawumi and Ogunlade (2008) reported non-significant negative correlation value (-0.09) between the EGWT and ESI. The relationships between ESI and EGWD and between ESI and EGL (positive and negative relationships, respectively) in chicken and partridge eggs suggest that these relationships are more likely due to the way of ESI is calculated in which the EGWD is the numerator factor and EGL is the denominator factor (Panda, 1996; Gunlu et al., 2003). The weight and dimensions of egg (EGWT, EGWD and EGL) have positive significant (P<0.0001) correlations with YKWT (0.34 to 0.74), ALBWT (0.45 to 0.94) and SHWT (0.33 to 0.82). Thus, EGWT, EGWD and EGL have direct relations with the weight of egg components. Similar finding was reported by Olawumi and Ogunlade (2008) who found significant positive correlation between EGWT and SHWT. Egg shape index was negatively (P<0.05) correlated to ALBWT and SHWT. The strong relationship between EGWT or egg

### Table 1. Characteristics of meat-type breeder eggs (n=174).

| Egg parameter | Mean   | Minimum  | Maximum  | CV   |
|---------------|--------|----------|----------|------|
| Weight        | EGWT, g| 58.82    | 51.71    | 67.11| 5.23 |
| Dimension     | EGWD, mm| 43.54    | 34.60    | 45.79| 2.58 |
| ESI, %        | 78.84  | 60.91    | 94.08    | 4.11 |
| Component, g  | YKWT, g| 17.85    | 15.44    | 20.76| 7.56 |
| SHWT          | 5.22   | 4.63     | 5.82     | 5.10 |

CV, coefficient of variation; EGWT, egg weight; EGWD, egg width; EGL, egg length; ESI, egg shape index=EGWD*100/EGL; YKWT, yolk weight; ALBWT, albumen weight; SHWT, eggshell weight.

### Table 2. Correlation coefficients of weight, dimensions and components of meat-type breeder eggs.

| EGWT, g | EGWD, mm | EGI, mm | ESI, %  | EGWTESI |
|---------|----------|---------|---------|---------|
| 0.561***| 1        | 0.499***| -0.094  | EGWTESI |
| 0.561***| 1        | 0.499***| -0.094  | 1       |
| 0.763***| 0.781*** | -0.100  | 0.571***| 1       |
| 0.739***| 0.575*** | 0.341***| 0.049***| 0.643***|
| 0.942***| 0.485*** | 0.485***| -0.149  | 0.678***|
| 0.816***| 0.329*** | 0.419***| -0.163  | 0.564***|

EGWT, egg weight; EGWD, egg width; EGL, egg length; ESI, egg shape index=EGWD*100/EGL; EGWTESI=EGWT*ESI; YKWT, yolk weight; ALBWT, albumen weight; SHWT, eggshell weight. *P<0.05, ***P<0.0001. Mean centring did not alter the correlations among variables.
dimensions (EGWD and EGL) and egg components (YKWT, ALBWT and SHWT) suggests that the combination of these characteristics could be used to estimate egg components.

The VIFs, eigenvalues, CIs and variance proportion for the relationships among EGWT, ESI and EGWTESI are shown in Table 3. The regression analysis indicated that there were collinearity problems in the two variables (EGWT and ESI) and EGWTESI as showed by VIFs. The VIFs were higher than 10.00 (VIF=835.5, 519 and 1228.5 for EGWT, ESI and EGWTESI, respectively). According to Gill (1986), no absolute standard exists for judging the magnitude of the VIF. However, a crude rule of thumb is to be suspicious of collinearity if VIF is greater than 10.00. This is consistent with the report of Belsley et al. (1980), Rook et al. (1990) and Belsley (1991). Collinearity problems were further confirmed from the computations of the eigenvalues of the correlation matrix, CIs and variance proportions (Table 3). The eigenvalues of EGWT, ESI and EGWTESI were near zero (0.003, 0.002 and 8.1 E-7, respectively) indicating that the correlation matrix approached singularity. Condition indexes were higher than 30 (CIs=38, 49 and 2220 for EGWT, ESI and EGWTESI, respectively). Belsley (1991) suggested that moderate to strong relations are associated with CI numbers of 30 to 100. Results in Table 3 from un-centred regression model indicated that EGWT, ESI and EGWTESI have high proportions of variance (0.999) with respect to EGWTESI. Obviously, EGET and ESI are highly correlated to EGWTESI (r=0.76 and 0.57 for EGWT and ESI with EGWTESI, respectively) with an eigenvalue of 8.11 E-7 and CI of 2220 causing MC in the regression model. This made the variances of estimates become inflated and consequently overfitting the regression model. A similar finding was reported by Greene (2000). This finding suggests that using variable selection methods in the presence of MC may be an inappropriate way to find the correct relationships between variables. Results from the regression of centred means model indicated that EGWT and ESI had high variance proportions in relations to EGWT and EGWTESI (0.44 and 0.45, 0.32 and 0.32, respectively) with eigenvalue and CI values of 1.09 and 0.79, and 1.03 and 1.21, respectively (Table 3).

The un-centred, centred means and PCR models of egg measurements predicting egg components are shown in Table 4. The un-centred model predicted YKWT, ALBWT and SHWT from egg weight between 50.71 and 67.11 g and ESI between 60.9 and 94.1 (Table 1). The fitted regression models to these data were:

\[
YKWT=53.577-0.664 \text{EGWT}-0.676 \text{ESI}+0.012 \text{EGWTESI}
\]

\[
ALBWT=42.674+1.381 \text{EGWT}+0.525 \text{ESI}-0.009 \text{EGWTESI}
\]

\[
SHWT=10.903+0.283 \text{EGWT}+0.152 \text{ESI}-0.003 \text{EGWTESI}
\]

The centred model provides a more meaningful interpretation when compared with that of the un-centred model, where the intercept values of 17.85, 35.74 and 5.22 g represent values of YKWT, ALBWT and SHWT when the mean values of EGWT and ESI were 58.82 g, and 78.84%, respectively (Table 1). The slopes of the centred means represent the linear relationships between egg measurements (EGWT and ESI) and egg components (YKWT, ALBWT and SHWT) when characterising the egg at the mean of the collected data. This is advantageous because the interpretations of the intercept and slope are now in the range of the observed data. Mean centring does not alter the correlations between independent and dependent variables. It seems that means centring does not affect the interpretation of regression results, but it improves the meaning of the intercept values.

The PCR model was used to reduce MC. The reduction is accomplished by using less than the full set of PCs to explain the variation in the response variable. When all three PCs are used, the ordinary least squares solution can be reproduced. The model can be as follows:

\[
\text{Egg component}=\alpha_1 Z_1+\alpha_2 Z_2+\alpha_3 Z_3+\epsilon
\]

The PC scores were calculated with eigenvectors as weights as follows:

\[
Z_1=0.566 \text{EGWT}+0.396 \text{ESI}+0.723 \text{EGWTESI}
\]

\[
Z_2=-0.597 \text{EGWT}+0.802 \text{ESI}+0.028 \text{EGWTESI}
\]

\[
Z_3=0.569 \text{EGWT}+0.448 \text{ESI}-0.690 \text{EGWTESI}
\]

The \(Z_1, Z_2\) and \(Z_3\) had eigenvalues of 1.909484, 1.090131 and 0.00038759, and variances of \(\lambda_1=1.909, \lambda_2=1.090, \lambda_3=0.0039\) for \(Z_1, Z_2\) and \(Z_3\), respectively. The regression of egg

---

Table 3. The regression models of un-centred and mean centred data and principal component of egg measurements predicting egg components form egg weight, egg shape index and their interaction.

| Regression model | VIF | E   | CI                          | Variance proportions° |
|------------------|-----|-----|---------------------------|-----------------------|
|                  |     |     | Intercept  | EGWT   | ESI   | EGWTESI  |
| Un-centred       |     |     | Intercept  | 0.00   | 1.00  | 0.012 E-7 | 2.01 E-7 | 2.02 E-7 | 2.04 E-7 |
|                  |     |     | EGWT       | 835    | 0.003 | 38.1    | 0.0002  | 0.0004  | 0.0004  | 0.0002  |
|                  |     |     | ESI        | 518    | 0.002 | 49.2    | 0.0006  | 0.0003  | 0.0003  | 0.0007  |
|                  |     |     | EGWTESI    | 1228.5 | 0.011 | 2220.1  | 0.0099  | 0.0009  | 0.0009  | 0.0009  |
| Centred          |     |     | Intercept  | 0.00   | 1.15  | 1.00    | 0.237   | 0.073   | 0.073   | 0.463   |
|                  |     |     | EGWT       | 1.02   | 1.09  | 1.03    | 0.00008 | 0.445   | 0.450   | 0.00002 |
|                  |     |     | ESI        | 1.02   | 0.966 | 1.09    | 0.637   | 0.166   | 0.155   | 0.068   |
|                  |     |     | EGWTESI    | 1.02   | 0.786 | 1.21    | 0.126   | 0.316   | 0.322   | 0.477   |
| PCR              |     |     | Intercept  | 0.00   | 1.00  | 1.00    | 1.00    | 0.00    | 0.00    | 0.00    |
|                  |     |     | Z1         | 1.00   | 1.00  | 1.00    | 1.00    | 0.00    | 0.00    | 0.00    |
|                  |     |     | Z2         | 1.00   | 1.00  | 1.00    | 0.00    | 1.00    | 0.00    | 0.00    |
|                  |     |     | Z3         | 1.00   | 1.00  | 1.00    | 0.00    | 0.00    | 1.00    | 0.00    |

VIF, variation of inflation; E, eigenvalue; CI, condition index; EGWT, egg components form egg weight; ESI, egg shape index; EGWTESI, EGWT*ESI; PCR, principal component regression. °Proportion of variance for each regression coefficient attributable to each CI. MC exists when VIF>10 (Belsley et al., 1980) and CI>30 (10<CI<30 would indicate possible problems of MC) (Belsley et al., 1980; Belsley, 1991).
Conclusions

It is concluded that mean centring and PCR techniques may be used to overcome MC problem and develop a better model for the estimation of egg components (YKWT, ALBWT and SHWT) from EGWT and ESI.

References

Abanikanda, O.T.F., Olutogun, O., Leigh, A.O., Ajayi, L.A., 2007. Statistical modeling of egg weight and egg dimensions in commercial layers. Int. J. Poult. Sci. 6:59-63.

Aziz, M.A., Sharaby, M.A., 1993. Collinearity as a problem in predicting body weight from body dimensions of Najdi sheep in Saudi Arabia. Small Ruminant Res. 12:117-124.

Belsley, D.A., 1991. Conditioning diagnostics, collinearity and weak data in regression. John Wiley and Sons, New York, NY, USA.

Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression diagnostics: identifying influential data and sources of collinearity. John Wiley and Sons, New York, NY, USA.

Chopprakarn, K., Salangam, I., Janaka, K., 1998. Laying performance, egg characteristics and egg composition in Thai indigenous hens. J. Nat. Res. Council Thailand 30:1-17.

Farooq, K.A.M., Durran, F.R., Sarbiliand, K., Chaud, N., 2003. Predicting egg weight, shell weight, shell thickness and hatching chick weight of Japanese quails using various egg traits as regressors. Int. J. Poult. Sci. 2:164-167.

Farooq, M., Mian, M.A., Ali, M., Durrani, F.R., Asquar, A., Muqarrab, A.K., 2001. Egg traits of Fayomi bird under subtropical conditions. Sarad J. Agri. 17:141-145.

Gill, J.L., 1986. Outliers and influence in multiple regression. J. Anim. Breed. Genet. 103:161-175.

Greene, W.H., 2000. Econometric analysis. Prentice-Hall Publ., Upper Saddle River, NJ, USA.

Gunlu, A., Kiriki, K., Cetin, O., Carip, M., 2003. Some external and internal quality characteristics of partridge (A. graeca) eggs. Food Agrici. Environ. 1:197-199.

Hawking, R.R., Pendleton, O.J., 1983. The regression dilemma. Commun. Stat. A-Theor. 12:497-527.

Hu, F.B., Manson, J.E., Willett, W.C., 2001. Types of dietary fat and risk of coronary heart disease: a critical review. J. Am. Coll. Nutr. 20:5-19.

Kul, S., Seker, I., 2004. Phenotypic correlation between some external and internal egg quality traits in the Japanese quail (Coturnix japonica). Int. J. Poult. Sci. 3:400-405.

Malau-Aduli, A.E.O., Aziz, M.A., Kojina, T., Nibayashi, T., Oshima, K., Komatsu, M., 1991. Conditioning diagnostics, collinearity and weak data in regression. J. Anim. Breed. Genet. 103:161-175.

Nishimura, A., Shibata, M., Kojima, T., 1991. Conditioning diagnostics, collinearity and weak data in regression. J. Anim. Breed. Genet. 103:161-175.

Shafey, et al., 2018. Component was considered against Z1, Z2 and Z3 (egg component=α1Z1+α2Z2+α3Z3+E; Table 4). However, the proportions of variance explained by Z1 and Z2 were 63.6 and 36.3%, respectively. The cumulative variance explained by Z1 and Z2 were 63.6 and 36.3%, respectively. Approximately equal to zero and was the source of MC in the data. Regression of reduced model of egg component was considered against Z1 and Z2 (egg component=α1Z1+α2Z2+α3Z3+E; Table 4). Consequently, the R² and adjusted-R² of the reduced PCR model were slightly reduced when compared with those of the PCR model (56.16 vs 56.70 and 55.64 vs 55.93; 89.24 vs 89.00 and 89.05 vs 88.87; 66.74 vs 67.13 and 67.43 vs 66.74 for the R² and adjusted-R² of yolk weight, albumen weight and eggshell weight, respectively).
2004. Fixing collinearity instability using principal component and ridge regression analyses in the relationship between body measurements and body weight in Japanese Black cattle. J. Anim. Vet. Adv. 3:856-863.

Olawumi, S.O., Ogunlade, J.T., 2008. Phenotypic correlations between some external and internal egg quality traits in the exotic Isa Brown layer breeders. Int. J. Poul. Sci. 2:30-35.

Ozcelik, M., 2002. The phenotypic correlation among some external and internal quality characteristics in Japanese quail eggs. Vet. J. Ankara Univ. 49:67-72.

Panda, P.C., 1996. Shape and Texture. In: P.C. Panda (ed.) Egg and poultry technology. Vikas Publ., New Delhi, India, p 57.

Pandey, N.K., Mahapatra, C.M., Verma, S.S., Johari, D.C., 1986. Effect of strain on physical egg quality characteristics in white Leghorn chickens. J. Poul. Sci. 21:304-307.

Pimentel, E.C.G., Queiroz, S.A., Carvalheiro, R., Fries, L.A., 2007. Use of ridge regression for prediction of early growth performance in crossbred calves. Genet. Mol. Biol. 30:536-544.

Robinson, C., Schumacker, R.E., 2009. Interaction effects: centering, variance inflation factor, and interpretation issues. Multiple Linear Regression Viewpoints 35:6-11.

Rook, A.J., Dhanoa, M.S., Gill, M., 1990. Prediction of the voluntary intake of grass silages by beef cattle. 2. Principal component and ridge regression analyses. Anim. Prod. 50:439-454.

SAS, 2006. Guide for personal computers, version 9.1.3. SAS Inst. Inc., Cary, NC, USA.

Shrimpton, D.H., 1987. The nutritive value of eggs and their dietary significance. In: R.G. Wells and C.G. Beljavin (eds.). Egg quality. Current problems and recent advances. Butterworth Publ., London, UK, pp 11-25.

Stadelman, W.J., 1977. Quality identification of shell eggs. In: W.J. Stadelman and D.J. Cotterill (eds.) Egg science and technology. AVI Publ., Westport, CT, USA, p 33.

Wilson, H.R., Suarez, M.E., 1993. The use of egg weight and chick weight coefficient of variation as quality indicators in hatchery management. J. Appl. Poultry Res. 2:227-231.

Yu, C.H., 2008. Multi-collinearity, variance inflation, and orthogonalization in regression. Available from: http://www.creative-wisdom.com/computer/sas/collinear.html

Yu, C.H., 2011. Principal component regression as a counter measure against collinearity. Arizona State University ed., Tempe, AZ, USA. Available from: http://www.lexjansen.com/ wuss/2011/analy/Papers_Yu_C_73333.pdf