Zagazig Veterinary Journal, ©Faculty of Veterinary Medicine,
Zagazig University, 44511, Egypt.
Volume 47, Number 2, p. 203-212, June 2019
DOI: 10.21608/zvjz.2019.11121.1034

RSEARCH ARTICLE

Application of Different Biostatistical Methods in Biological Data Analysis

Khairy M. El-Bayomi, Fatma D. Mohamed, Mahmoud S. El-Tarabany and Hagar F. Gouda*

Animal Wealth Development Department, Faculty of Veterinary Medicine, Zagazig University, 44511, Egypt

Article History: Received: 25/03/2019 Received in revised form: 04/04/2019 Accepted: 08/04/2019

Abstract

Logistic regression is one of the popular methods used in genetic data analysis. That is applied to predict a categorical binary dependent variable on basis of predictor variables, and to test the probability of getting a particular value of the dependent variable that is related to the explanatory variable. The objective of this study is to highlight the crucial role of biostatistical methods in increasing the accuracy of the results in veterinary and biological practices. Statistical analysis of previously published data in the National Research Center, Dokki, Giza, Cairo, Egypt was done using SPSS version, 24 to predict hepatocellular carcinoma metastasis by knowing the genotypes, age, and gender of the patients. The genotypes and gender displayed a significant effect on metastasis ($P < 0.05$) while age had no significant effect on metastasis ($P > 0.05$). There are other types of data (animal breeding and production) which were analyzed by repeated measures ANOVA and principal component analysis (PCA). The repeated measures ANOVA is equivalent to normalized ANOVA, but for related, not independent groups. Data of this test was obtained from a study aimed to measure body weight of three breeds of rabbits at 4 time points 4th, 6th, 8th and 10th weeks of the experiment. The main effect of breed types of rabbits was significant ($P < 0.05$), the time (weeks) was highly significant ($P < 0.001$) and their interaction was also highly significant ($P < 0.001$). Principal component analysis (PCA) is used to reduce a large set of variables to a small set that still contains most of the information in the large set. A reduced set is easier to analyze and interpret. Data with 6 variables reduced to only 2 variables where initial eigenvalues were > 1 for two variables and their values were (2.768 and 1.147).

Keywords: Principal component analysis (PCA), Repeated measures ANOVA, Sphericity assumption, Logistic regression, Odds ratio.

Introduction

The purpose of statistics has mainly two objectives: The first one is collecting and handling quantitative and qualitative information in the form of annual reports and listing the numbers or numerical details about animal or plant life. The second objective is organizing, summarizing and describing quantifiable data, and methods of drawing inferences and generalizing upon them [1]. Moreover, statistical methods, theories, techniques, and models play important roles in several stages of the scientific method to analyze and interpret data, so some knowledge of statistics is an important part of the purview of every biologist, statistics as a tool for distinguishing between random “noise” in the data and the real signal, then someone who incorrectly uses statistics may produce a result that is distorted or even artificial [2].

In order to reflect the structure of data and the possible correlations between variables, data can be analyzed by different statistical methods depending on study design and type of outcome variable, categorical variables often analyzed by contingency tables, logistic regression, or generalized estimation equation (GEE) models. Meanwhile, the continuous variables were analyzed by t-test, ANOVA, correlation and regression [3]. A well-designed study with a lower biased sampling method

***Corresponding author e-mail: (stathagarfathi@gmail.com), Animal Wealth Development Department, Faculty of Veterinary Medicine, Zagazig University, 44511, Egypt.**

and accurate choosing statistical model provides a well-founded, precise, valid and reliable results [4].

Principal component analysis (PCA) is the cornerstone method used in the dimension reduction data analysis. Even though widely used, it is poorly understood. Principal component analysis extracts the most benefit information in a large set of variables, but in fewer variables explaining most variation in data [5]. Principal component analysis investigates the relationship between a group of variables to select a subset of those are linearly correlated and explain most of the variance among all observed variables in order to derive the first summary component. The first component accounts for the largest amount of variation among all observed variables. This is expressed by eigenvalues or by the proportion or percentage of the total variance [6].

When measuring certain values on same individuals or subjects at different consecutive time points, the appropriate test is repeated measures ANOVA not traditional between subjects ANOVA, as it does not consider dependencies between observations within-subject in the analysis [7]. Repeated measures ANOVA can be applied in another form when results were obtained by handling two explanatory variables, one represents the groups or the treatment (between subject) applied and the other variable is the variable repeatedly occurred (within subject). This design called mixed model design or split-plot or within-between subjects design [8]. Regression models increasingly used in the data analysis for describing the relationship between a response variable and other/s predictor variable/s. In addition, logistic regression is the most appropriate when the response variable is binary coded or even the response is dummy variable [9].

The aim of this study is to display the most important statistical methods that can be applied in analyzing genetic data as well as data of animal breeding and production, such type of data are commonly in the veterinary field. Also the objectives of this work extend further to present the correct application to obtain valid and accurate results by testing conditions of each statistical test.

Materials and Methods

Three different types of statistical tests were applied according to different types of data under study.

Genetic data

Source of data and variable definition

Data were obtained from a previous study [10].The study was performed at the National Research Center, Dokki, Giza, Cairo, Egypt between January and December 2016. The experimental units were divided into three groups: group I consisted of 90 (Hepatitis C virus) HCV patients with (Hepatitis C Carcinoma) HCC (including 45 patients with metastasis and 45 patients without metastasis). Group II included 99 HCV patients without HCC, and group III 90 unrelated healthy controls. Genotype of individuals was determined by allele-specific polymerase chain reaction (AS-PCR).

Metastasis is the word describes the tumor that has spread into foreign sites out of its primary site [11]. The categorical response variable was hepatic metastasis that coded 1 (presence of metastasis) and 0 (no metastasis), and the independent variables were, one continuous variable that was the age and two categorical independent variables, one was genotype with three categories (AA, CA and CC) and the other categorical was gender with two levels (male and female). The reference genotype was CC and the reference gender was female.

Statistical analysis

Logistic regression is a widely used flexible technique neither obliges normality nor homogeneity of variance for all variables like many regression models [12].

Mathematical model of binary logistic regression:

$$\boldsymbol{log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_m x_m}$$

$\boldsymbol{\pi}$ indicates the probability of an event (metastasis).

$\boldsymbol{\beta_1,\ldots,\beta_m}$ are the regression coefficients associated with the reference group.

$x_1,\ldots,x_m$ are explanatory variables (genotype, age and gender).

The reference group was represented by **β_0**.

In logistic regression, the response or dependent variable y is the log odds (log (p/1-p), which is called the logit:

$$\log\left(\frac{p}{1-p}\right) = a + b \times x$$

a: is the intercept (constant).

b :is the regression coefficient of x, and **x** is the categorical predictor [13].

The odds ratio of explanatory variables is the value that explains how likely specific value of independent variable occurs in the response variable [14].

Animal breeding and production data

Repeated measures ANOVA

Source of data and variables definition

Data were obtained from experimental study conducted at the Department of Animal Wealth Development, Faculty of Veterinary Medicine, Zagazig University, Egypt within the period from April to August 2014 [15]. A total of 129 newly weaned rabbits over the course of four weeks from three different commercial breeds (49 NewZeland (NZW), 34 California (CAL) and 46 Rex (RX)) of both sexes. This study was conducted to investigate the effect of ambient heat stress (30± 2°C) on three different rabbit breeds of both sexes. The body weight was measured at four time points; 4th, 6th, 8th and 10th weeks. All procedures concerning animals were conducted according to the Zagazig University Animal Ethics Committee guidelines.

The comparison of treatments by performing separate statistical tests at each time point is an inappropriate, as it often fails to address relevant research questions and is subject to statistical deficiencies such as ignoring those observations on a given subject are likely to be correlated and multiple testing [16]. The repeated measures ANOVA was used to measure live body weight repeatedly for same animals of three breeds (New Zealand White, Californian and Rex) at different times 4th, 6th, 8th and 10th weeks age.

Statistical analysis

The mathematical model of repeated measures ANOVA:

$$y_{ij} = \mu + \alpha_t + \beta_j + (\alpha\beta)_{tj} + b_i + \varepsilon_{ij}$$

y_{ij} is the live body weight of rabbits.

μ is the overall mean.

α_t is the breed effect.

β_j is the time of experiment effect.

$(\alpha\beta)_{tj}$ is the interaction between breed **t** and time **j**.

b_i is the random subject effect.

ε_{ij} is the random error

The model measures the main effects of breed type and time and the interaction effects between breed type and time (weeks of the experiment) on live body weight.

Principal component analysis:

Source of data and variable definition

Data were obtained from a designed experimental study on rabbits at the experimental unit belonging to Animal Wealth Development Department, Faculty of Veterinary Medicine, Zagazig University, Egypt, during the period from January to May 2014. The rabbits were injected intramuscularly at the 40th and 47th days of age with three doses (control dose: 0.25 mL sesame oil/kg body weight (BW), normal dose of boldenone undecylenate (BUL): 4.4 mg/kg BW and a double dose of BUL: 8.8 mg/kg BW), then the effect on growth traits and body dimensions were investigated at 44, 58 and 72 days of age [17]. The variables in the model represent growth traits and body dimensions body length (BL), chest circumference (CHC), abdominal circumference (ABDC), thigh circumference (THC), ear length (EL), and ear width (EW). All variables are quantitative continuous and measured by ratio level. Experimental procedures were conducted in accordance with the Zagazig University Animal Ethics Committee guidelines.

Statistical analysis

The PCA transformed the variables in a multivariate dataset x1, x2, ---, xp, into new variables, y1, y2,---, yp which are uncorrelated with each other and account for decreasing proportions of the total variance of the original variables [18]. Principal component analysis was performed. The correlation coefficients were determined for all six variables that represent body measurements at 58th day of age with normal dose of BUL.

Results

Genetic data

Logistic regression

The continuous predictor variable (the age) was described by mean and standard deviation (59.17± 6.9). The logistic regression model Nagelkerke R square value equals 0.30 so model explained 30% of variation in metastasis based on genotype and gender of patients. The P value of Hosmer and Lemeshow (0.65) was non-significant (P > 0.05) so it showed a good fit of the model to the data. The genotype (AA) and male gender were statistically significant (P < 0.01) so significantly used to predict metastasis, while genotype CA and age removed from model as non-significant result (P > 0.05). The effect of genotype AA was significant and odds ratio = (0.064) with 95% confidence level (0.012 to 0.35) indicating that individuals with genotype AA were 0.94% less probably susceptible to metastasis compared to those with genotype CC. The effect of gender (male) was statistically significant and odds ratio = (7.29) with 95% confidence level as this value falls within the interval (2.23 to 23.8), so the males were 29% more likely susceptible to metastasis compared to females (Table 1).

Table 1: Logistic regression results for prediction of hepatocellular carcinoma metastasis in concerned patients

	β	S.E.	Wald statistic	*P*-value	Odds ratio	95% C.I.	
						Lower	Upper
Genotype (CC)							
Genotype (AA)	-2.744	0.856	10.268**	0.001	٠.064	٠.012	0.35
Genotype (CA)	-1.330	0.811	2.685NS	0.101	٠.27	0.054	1.29
Gender (male)	1.988	0.605	10.803**	0.001	7.29	2.23	23.88
Age	0.010	0.035	0.079NS	0.779	1.01	0.943	1.08
Constant	-0.116	2.241	0.003NS	0.959	٠.89		

** Highly significant difference P < 0.01, NS non-significant difference P > 0.05, β is regression coefficient, S.E. is standard error, C.I is confidence interval, Wald statistic is a chi-square value test whether a predictor variable is significant in the model or not.

Animal breeding and production data

Repeated measures ANOVA

Live body weight was measured at different time points, mean and standard error of mean for body weight were measured at the four time points. The results for test of normality (Shapiro-Wilk) were not significant (P value > 0.05), so the data are normally distributed. Mauchly's test of sphericity was significant P < 0.05 so the assumption was violated, and epsilon correction required Greenhouse-Geisser was used. The sphericity is the condition where the variances of the differences between all combinations of related groups (levels) are equal. The repeated measures ANOVA results with Greenhouse-Geisser correction revealed that the effect of breed type at different time points (interaction) was statistically significant P < 0.05 (the most important results) (Table 2).

Table 2: Main effect of time points (4th, 6th, 8th and 10th weeks), breed and sex, and their interaction on body weight of NewZeland, California and Rex rabbit breeds

Source of variance		SS	df	MS	F	*P*-value
Within-subjects effects						
Weeks	Greenhouse-Geisser	81590059.13	2.65	30848288.10	3799.3**	0.000
Weeks*breed	Greenhouse-Geisser	305382.74	5.29	57730.89	7.110**	0.000
Weeks*sex	Greenhouse-Geisser	4087.64	2.65	1545.49	0.19^{NS}	0.88
Weeks * Breed * Sex	Greenhouse-Geisser	96794.98	5.29	18298.55	2.254^{NS}	0.065
Error(weeks)	Greenhouse-Geisser	2641408.45	325.3	8119.41		
Between-subjects effects						
Breed		589355.63	2	294677.82	3.802*	0.025
Sex		195058.42	1	195058.42	2.517^{NS}	0.115
Breed * Sex		69021.033	2	34510.52	0.445^{NS}	0.642
Error		9532479.3	123	77499.83		

SS is sum of squares; df is degrees of freedom; MS is mean of squares; F is F-value

Table 3: Duncan's multiple range post hock test for comparing means of body weights at different time points (4th, 6th, 8th and 10th weeks), for NewZeland, California and Rex rabbit breeds.

				95% Confidence Interval	
Breed	**Time(weeks)**	**Mean**	**Std. Error**	**Lower Bound**	**Upper Bound**
NewZeland	4th	444.265^{h}	19.726	404.131	484.398
	6th	887.500^{f}	29.120	828.255	946.745
	8th	1258.088^{d}	26.721	1203.723	1312.454
	10th	1541.618^{b}	22.329	1496.188	1587.047
California	4th	586.471^{g}	18.521	548.789	624.152
	6th	1021.471^{e}	30.492	959.433	1083.508
	8th	1334.412^{c}	36.297	1260.564	1408.259
	10th	1584.265^{b}	31.135	1520.920	1647.609
Rex	4th	547.794^{g}	17.658	511.868	583.720
	6th	970.294^{e}	27.006	915.350	1025.238
	8th	1339.235^{c}	27.160	1283.977	1394.494
	10th	**1680.0^{a}**	23.681	1631.821	1728.179

Means with different superscript are statistically significant at $P < 0.05$ according to Duncan's multiple range test.

To determine which means present significant effect, a post hoc test of Duncan's multiple range was used. The post hoc test revealed that the body weight of the Rex breed at the 10th week was the highest body weight of the all three breeds at different times (1680 ± 23.68). There was no statistically significant difference in body weight between NewZeland (1541.62 ± 22.33) and California (1584.27 ± 31.14) at the 10th week of experiment. Furthermore the body weight of California and Rex did not significantly differ at the 6th week (1021.47 ± 30.49 and 970.29 ± 27.01, respectively) and 8th week (1334.41 ± 36.29 and 1339.24 ± 27.16, respectively). Besides, NewZeland breed exhibited the smallest body weight at the 4th week (444.27 ±19.73), while California and Rex did not significantly differ at 4th week (586.47 ±18.52 and 547.79 ± 17.66, respectively). Body weight of NewZeland breed was the smallest at the 4th week (444.27 ±19.73), while California and Rex weren’t significantly differ at 4th week (586.47 ± 18.52 and 547.79 ± 17.66, respectively) (Table 3).

Principal component analysis (PCA)

There was enough correlation between variables to conduct the analysis as the correlation coefficients were highly significant ($P < 0.01$) and their values predominantly were above 0.3. As the PCA depends on reducing the large number of correlated variables to fewer uncorrelated others. Value of Kaiser-Meyer-Olkin was 0.6. Bartlett’s test of sphericity with an associated (chi-square value = 241.3; P value < 0.001) which provided enough support for the validity of the principal component analysis of the data set.

Table 4: Total variance explained by extracted components

Total Variance Explained									
	Initial Eigenvalues			**Extraction Sums of squared Loadings**			**Rotation Sums of Squared Loadings**		
Component	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	**2.768**	46.136	46.136	2.768	46.136	46.136	2.361	39.354	39.354
2	**1.147**	19.115	**65.251**	1.147	19.115	65.251	1.554	25.897	65.251
3	0.826	13.765	79.016						
4	0.665	11.080	90.095						
5	0.420	7.003	97.098						
6	0.174	2.902	100						

Table 5: Component matrix and rotated component matrix for body dimensions in experimentally inoculated rabbits with a normal dose of BUL.

Component Matrix			**Rotated Component Matrix**		
	Component			Component	
	1	**2**		**1**	**2**
ABDC	0.866		**THC**	0.818	
CHC	0.844		**CHC**	0.802	0.300
THC	0.776		**BL**	0.709	
BL	0.540	-0.483	**ABDC**	0.691	0.535
EW	0.540	0.436	**EL**		0.847
EL	0.347	0.777	**EW**		0.648

ABDC= abdominal circumference; CHC= chest circumference; THC= thigh circumference; BL= body length; EW= ear width; EL=ear length

Table 4 showed that the data matrix was best described by two principal components as their eigenvalues were above 1 and their values were 2.768 and 1.147, and both accounted for 65.25% of the total variability. Table 5 revealed that the factor pattern coefficients that were used to assess the relative contributions of the various body dimensions in determining the most important factors that explained and summarized the data. There were two extracted principal components. The most important variables that found to be highly associated with the first component were abdominal circumference, chest circumference and thigh circumference with correlation coefficients of 0.866, 0.844, and 0.776, respectively, while ear length was highly associated with the second principal component (0.77). After rotation abdominal circumference, chest circumference and thoracic circumference remained highly associated with the first principal component (0.818, 0.802, and 0.709, respectively), and ear length became higher associated with the second principal component (0.847).

Discussion

The current research demonstrated different applied statistical methods in the field of genetics and animal production and breeding data, and showed the appropriate tests according to the type of the data. One of the most common problem in statistics is the wrong application of statistical methods. In this study logistic regression was applied to predict the probability of metastasis that is a categorical binary variable obtained by counting, so application of ANOVA or simple regression was not correct as these tests require continuous data type but the accurate decision here was the application of logistic regression in such type of data and this agreed with Agresti [19] and Kleinbaum and Klein [20] but disagreed with Rao [21] who reported that analysis of variance is an appropriate test can be used in count data type. Our result revealed the significant role of genetics (genotype) which was previously supported by a study [22] applied logistic regression to determine the relationship between genetic effect and metastasis of HCC, which had showed a significant association between metastasis and genotypes. The effect of gender had a significant effect on metastasis of HCC and males revealed more susceptibility than females, this result is in contrast with a previous-study [23] which used logistic regression to study the effect of gender on metastasis of HCC, and the results showed no significant difference between males and females in metastasis. The age factor showed non-significant effect on metastasis and this result is in contrast with previous studies [24, 25], which revealed that hepatocellular carcinoma is prominent in older ages.

Repeated measures ANOVA was run in this study in which body weight of three breeds of rabbits was measured at different consecutive time points, this is the correct statistical models for this type of data, as it takes into consideration the relation between repeated measurements and the serial effect of time. A previous study [26] stated that application of separate ANOVA tests for each time point is inappropriate to be used, as this method doesn't consider certain important items as the effect of the consecutive time points on the measured parameters and the relation between measurements that cannot be obtained by traditional ANOVA. The results in Table 2 show within-subject effect with Greenhous-Gisser correction that was used due to violation of sphericity assumption as the P value of Mauchly's test of sphericity < 0.05. The use of Greenhous-Gisser correction is in consistence with previously published articles [7, 27 and 28]. The results in Table 3 show that at young age California breed is heavier than NewZeland and Rex breed in the 4th weeks of age. These results agree with Hassanien and Baiomy [29]. The final body weight of California is superior to NewZeland this is in contrary to the results reported by Marai *et.al* [30].

Principal component analysis method requires an adequate correlation between variables to be suitable in process of reduction. Our results showed that many correlation coefficients were greater than 0.3 and this is in consistence with Tabachnick and Fidell [31], but for Kaiser-Meyer-Olkin it was 0.6 that is acceptable value indicating suitability of PCA analysis of this data. Several authors have suggested that the minimum accepted value for the test is 0.5as Hair [32]. Our data with six variables reduced to only two which had eigenvalue > 1, this extraction criterion was used previously [6,33]. This study is similar to the previous study applied on rabbits and used principal component analysis to determine the important factors from all body measurements [18]. Body length was moderately correlated (0.54) to the first component and ear length was weakly correlated to the first component (0.35), this is in contrary with the results published by Udeh [34] who revealed a high correlation with 0.901 and 0.884 for body length and ear length respectively.

Conclusion

Choosing the correct statistical method for analysis of data produces valid and accurate results that can help in assessing and developing of biological fields based data. Logistic regression is the suitable model for predicting the binary response variables. Data with repeated measures are common in animal field studies, so caution must be paid for correct application of repeated measures with verification of required assumptions. Indeed, principal component analysis is very helpful and satisfactory method to be used in recognition of more important characters representing animal breeds.

Conflict of interest

The authors have no conflict of interest to declare

References

[1] Fowler, J.; Cohen, L. and Jarvis, P.(2009): Practical statistics for field biology, 2d edition reprint. John Wiley & Sons. 15 p.

[2] Gardenier, J. and Resnik , D. (2002): The misuse of statistics: concepts, tools, and a research agenda. Acc Res, 9(2): 65-74.

[3] Fu, W.J.; Stromberg, A.J.;Viele, K.; Carroll, R.J. and Wu, G. (2010): Statistics and bioinformatics in nutritional sciences: analysis of complex data in the era of systems biology. J Nutr Biochem, 21 (7): 561-572.

[4] Binu, V.; Mayya, S.S. and Dhar, M.(2014): Some basic aspects of statistical methods and sample size determination in health science research. Ayu J, 35 (2): 119-123.

[5]Shlens, J.(2014): A tutorial on principal component analysis. e-prints arXiv:1404.1100, 1-13.

[6] Schürks, M.; Buring, J.E. and Kurth, T. (2011): Migraine features, associated symptoms and triggers: a principal component analysis in the Women's Health Study. Cephalalgia, 31 (7): 861-869.

[7] Park, E.; Cho, M. and Ki, C.S. (2009): Correct use of repeated measures analysis of variance. Korean J Lab Med, 29(1): 1-9.

[8] Chartier, S. and Cousineau, D. (2011): Computing mixed-design (split-plot) ANOVA. Mathematica J, 13: 13-17.

[9] Hosmer, D.; Lemeshow, S. and Sturdivant, R.X. (2013): Applied logistic regression,3rd edition.volum (398). John Wiley & Sons, New York 1p.

[10] Bakr, N.M.; Awad, A. and Moustafa, E.A. (2018): Association of genetic variants in the interleukin-18 gene promoter with risk of hepatocellular carcinoma and metastasis in patients with hepatitis C virus infection. IUBMB Life,70 (2): 165-174.

[11]Marx, J.(2001): Cance research. New insights into metastasis. Science, 294: 281-282.

[12] Josephat, P.K. and Ame, A. (2018): Effect of testing logistic regression assumptions on the improvement of the propensity scores. Int J Statist Appl, 8: 9-17.

[13] Pandis, N. (2017): Logistic regression: Part 1. Am J Orthod Dentofacial Orthop, 151: 824-825.

[14] Garcia Garmendia, J.L and Maroto Monserrat, F. (2018): Interpretation of statistical results. Med Intensiva, 42 (6): 370-379.

[15] Abdel-Hamid, T. and Dawod, A. (2014): Impacts of ambient heat stress on growing rabbit performance and carcass traits. J Vet Sci Technol, 4(2), 7-13.

[16] Matthews, J.N; Altman, D.G.; Campbell, M.N and Royston, P. (1990): Analysis of serial measurements in medical research. Bmj, 300 (6719): 230-235.

[17] Abdel-Hamid, T.M. and Farahat, M.H. (2015): Carcass traits and some blood stress parameters of summer stressed growing male rabbits of different breeds in response to boldenone undecylenate. J Adv Vet Anim Res, 2(3): 263-270.

[18] Yakubu, A. and Ayoade, J. (2009): Application of principal component factor analysis in quantifying size and morphological indices of domestic rabbits. Int J Morphol, 27(4):1013-1017.

[19] Agresti, A. (2013): Categorical data analysis, 3 rd edition. John Wiley & Sons. 163-195 p.

[20] Kleinbaum, D.G. and Klein, M. (2010): Logistic regression: A self learning text. 3rd edition, Springer, New York. 5 p.

[21] Rao, M.M. (1960): Some asymptotic results on transformations in the analysis of variance. ARL technical Note, Aerospace Research Laboratory, Wright-Patterson Air Force Base, Dayton, 60-126.

[22] Ren, N.; Wu, J.C.; Dong, Q.Z.; Sun, H.J.; Jia, H.L.; Li, G.C.; Sun, B.S.; Dai, C.; Shi, J. and Wei, J.W. (2011): Association of specific genotypes in metastatic suppressor HTPAP with tumor metastasis and clinical prognosis in hepatocellular carcinoma. Cancer Res, 71(9); 3278–3286.

[23] Sobotka, L.; Hinton, A. and Conteh, L. (2017): Women receive more inpatient resections and ablations for hepatocellular carcinoma than men. World J Hepatol, 28; 9(36): 1346-1351.

[24] Nzeako, U. C.; Goodman, Z. D. and Ishak, K. G. (1994) :Association of hepatocellular carcinoma in North American patients with extrahepatic primary malignancies. Cancer, 74 (10), 2765–2771.

[25] Wu, W.C.; Chen, Y.T.; Hwang, C.Y.; Su, C.W.; Li, S.Y.; Chen, T.J.; Liu, C.J.; Kao, W.Y.; Chao, Y.; Lin, H.C. and Wu, J.C. (2013): Second primary cancers in patients with hepatocellular carcinoma: a nationwide cohort study in Taiwan. Liver INT J, 33 (4), 616-623.

[26] Hickey, G.L.; Mokhles, M.M.; Chambers, D.J. and Kolamunnage-Dona, R. (2018): Statistical primer: performing repeated-measures analysis. Interact Cardiovasc Thorac Surg J, 26, 539-544.

[27] Wang, LA. and Goonewardene, Z. (2004): The use of mixed models in the analysis of animal experiments with repeated measures data. Can J Anim Sci, 84, 1-11.

[28] Singh, V.; Rana, R.K. and Singhal, R. (2013): Analysis of repeated measurement data in the clinical trials. J Ayurveda Integr Med, 4, 77-81.

[29] Hassanien, H. and Baiomy, A. (2011): Effect of breed and parity on growth performance, litter size, litter weight, conception rate and semen characteristics of medium size rabbits in hot climates. Egypt Poult Sci J, 31-45.

[30] Marai, I.F.; HABEEB, A.S.i.A. and Gad, A.E. (2008): Performance of New Zealand White and Californian male weaned rabbits in the subtropical environment of Egypt. Anim Sci J, 79, 472-480.

[31] Tabachnick, B.G. and Fidell, L.S. (2018): Using multivariate statistics, 7th edition. Pearson Education.

[32] Hair, J.F.; Black, W.C.; Babin, B.J. and Anderson, R.E. (2010): Multivariate data analysis, 7 th edition. Pearson Education, Upper Saddle River

[33] Abdi, H. and Williams, L.J. (2010): Principal component analysis. WIREs: Comp Stat, 2: 433-459.

[34] Udeh, I. (2013): Prediction of body weight in rabbits using principal component factor scores in multiple linear regression model. Rabbit Genetics, 3: 1-6.

الملخص العربي

تطبيق طرق الاحصاء الحيوي المختلفة في تحليل البيانات الحيوية

خيري محمد البيومي، فاطمة دسوقي محمد، محمود محمد صلاح الطرباني و هاجر فتحي جودة

قسم تنمية الثروة الحيوانية - كلية الطب البيطري- جامعة الزقازيق

تهدف هذه الدراسة لتسليط الضوء علي تطبيقات بعض الطرق الاحصائية لتحليل البيانات الحيوية. وهي نموذج الانحدار اللوجستي الذي يستخدم للتنبؤ باحتمالية حدوث او عدم حدوث متغير تابع بناء علي متغير او اكثر من المتغيرات المستقلة. حيث تم تحليل البيانات ببرنامج SPSS اصدار٢٤ واستخدمت البيانات من دراسة ميدانية أجريت علي مجموعة من المرضي من الجنسين من اعمار مختلفة بتركيبات جينية مختلفة بالمركز القومي للبحوث بالدقي حيث تم تطبيق نموذج الانحدار اللوجستي لاكتشاف العلاقة بين انتشار الورم الكبدي الي خلايا اخري من عدمه بدلالة عمر المرضي والتركيب الجيني لهم وجنسهم اظهرت النتائج تاثيرا معنويا للتركيب الجيني والجنس بينما لم يظهر تاثير معنوي للعمر في امكانية التنبؤ بالقابلية لانتشار الاورام الكبدية. تم تطبيق تحليل المكون الرئيسي علي مجموعة بيانات تتكون من ستة متغيرات تمثل قياسات ابعاد الجسم لمجموعة من الارانب .اوضح تحليل المكون الرئيسي ارتباطا معنويا بين المتغيرات الستة وكانت البيانات تتبع التوزيع الطبيعي .تم اختزال الجزء الاكبر من الاختلافات في مكونين رئيسين فقط يقوما بتوضيح ٦٥.٢٥% من الاختلافات الكلية الحاصلة في مجموعة البيانات وكانت القيمة الذاتية لكل منهما اكبر من ١ .بينما تم تطبيق تحليل التباين للقياسات المتكررة علي مجموعة من الارانب قامت التجربة بقياس تاثيرعامل الحرارة علي وزن الارانب من الثلاث سلالات من الجنسين من خلال متابعة قياس الوزن علي فترات زمنية كالتالي كان اول قياس للوزن عند الاسبوع الرابع ثم السادس ثم الثامن واختتمت التجربة باخر قياس عند الاسبوع العاشر.اظهرت نتائج التحليل الاحصائي اختلافا معنويا عاليا -(القيمة الاحتمالية اقل من٠.٠١)-للاوزان في الاسابيع المختلفة للتجربة في الثلاث سلالات.