HINF 6030: Principles of Biostatistics

Term Project

Can Adiponectin, Leptin, Resistin, Glucose, Age and Insulin predict Breast Cancer in Females?

Oluwaseun Tope Kiakubu

B00790204

RESEARCH QUESTION

"Can Adiponectin, Leptin, Resistin, Glucose, Age and Insulin predict breast cancer in females”?

ABSTRACT

BACKGROUND: Breast Cancer is one of the leading causes of death among women. Early diagnosis using predictors can help improve prognosis.

OBJECTIVE: This study is aimed at predicting the presence of breast cancer in females using predictors such as Adiponectin, Leptin, Resistin, Glucose, Age and Insulin.

METHOD: The data set consists of 6 independent variables [Age (years), Glucose (mg/ dL), Insulin (µU/mL), Leptin (ng/mL), Adiponectin (µg /mL), Resistin (ng/mL)] and 1 outcome variable called classification of subjects (1=Healthy controls; 2=Patients). Statistical analysis used for this study includes descriptive statistics, univariate logistic regression, multivariate logistic regression, and odds ratio. Odds ratio (OR) was used to measure the odds of experiencing breast cancer with independent variables while the receiver operating characteristics curve (ROC) was used to determine the trade-off between sensitivity and specificity.

RESULTS: Univariate logistics analysis showed that Glucose, Insulin, and Resistin were signific ant at $P<0.05$, while multivariate logistic analysis showed that Glucose [OR = 1.09, 95% (CI= 1.0 5 - 1.15)], Leptin [OR = 0.97, 95% (0.94-0.99)], Resistin [OR = 1.06, 95% (1.02-1.14)] were sign ificant at $P<0.05$. The Area under the Receiver Operator Characteristic (ROC) Curve is 0.835.

CONCLUSION: The combination of Glucose, Leptin, and Resistin may be considered to be good predictors of breast cancer in females.

KEYWORD: Adiponectin, Leptin, Resistin, Glucose, Age, Insulin, Breast Cancer

INTRODUCTION

Breast cancer is one of the leading causes of death among women worldwide. A report from the Canadian Cancer Society shows that 26,900 women will be diagnosed with breast cancer and 5,000 will die of it in 2019 (2). Early screening procedures have shown to be a crucial method to reduce mortality rates caused by this disease (3). At the moment, ultrasound and mammography are techniques used for screening breast cancer but these techniques present a drawback on computer aided diagnosis (CAD) systems pivoted on breast mammogram and ultrasound images (4,5,6). These techniques are computationally expensive and there are possibilities of inaccurate diagnosis and capture of unrelated features which can lead to vague prediction. (7). Therefore, an alternative method using biomarkers from routine consultation and blood analysis can be used to save cost, improve early detection and better prognosis. Hence, this study is aimed at developing and assessing a prediction model as well as determining the odds of the presence of breast cancer using predictors such as Age, Adiponectin, Leptin, Resistin, Glucose and Insulin.

DATA

Data used for this study was obtained from Breast Cancer Coimbra Data Set of the UCI Machine Learning Repository. The data set consist of 6 independent variables [Age (years), Glucose (mg/ dL), Insulin (μU/mL), Leptin (ng/mL), Adiponectin (μg /mL), Resistin (ng/mL)] and 1 outcome variable called classification of subjects (1=Healthy controls; 2=Patients).

Exploratory Analyses

Exploratory analyses were carried out using descriptive statistics, density plot of independent variables, box plot, scatter plot and correlation. The tables and figures present the graphical and tabular representation of results.

Table 1: Numeric Summaries of each predictor variable

	Mean	SD	Median	Q1	Q3	Min	Max	Range
Age	57.30	16.11	56.00	45.00	71.0	24.00	89.00	65.00
Glucose	97.79	22.53	92.00	85.75	102.00	60.00	201.00	141.00
Insulin	10.01	10.07	5.92	4.36	11.19	2.43	58.46	56.03
Leptin	26.2	19.18	20.27	12.31	37.38	4.31	90.28	85.97
Adiponectin	10.18	6.84	8.35	5.47	11.82	1.66	38.04	36.38
Resistin	14.73	12.39	10.83	6.88	17.55	3.21	82.10	78.89

The descriptive statistics table (above) shows good variability among the predictor variables

Table 2: Correlation analysis table between predictors

	Age	Glucose	Insulin	Leptin	Adiponectin	Resistin
Age	1.0000	0.2301	0.0324	0.1026	-0.2198	0.0027
Glucose	0.2301	1.0000	0.5046	0.3051	-0.1222	0.2913
Insulin	0.0324	0.5047	1.0000	0.30146	-0.0313	0.1467
Leptin	0.1026	0.3051	0.3015	1.0000	-0.0954	0.2523
Adiponectin	-0.2198	-0.1221	-0.0312	-0.0953	1.0000	-0.2524
Resistin	0.0027	0.2913	0.1467	0.2562	-0.2523	1.0000

The correlation analysis table shows that glucose has a moderate association with Insulin [ρ=0.5] (1), while Resistin and Age, Insulin and Age, Adiponectin and Insulin show no linear relationship. Also, a weak downhill (negative) linear relationship is seen in variables Adiponectin and Age, Glucose as well as Resistin respectively. A weak uphill (positive) linear relationship is seen between variables Glucose and Age, Glucose and Leptin, Glucose and Resistin, and Insulin and Leptin.

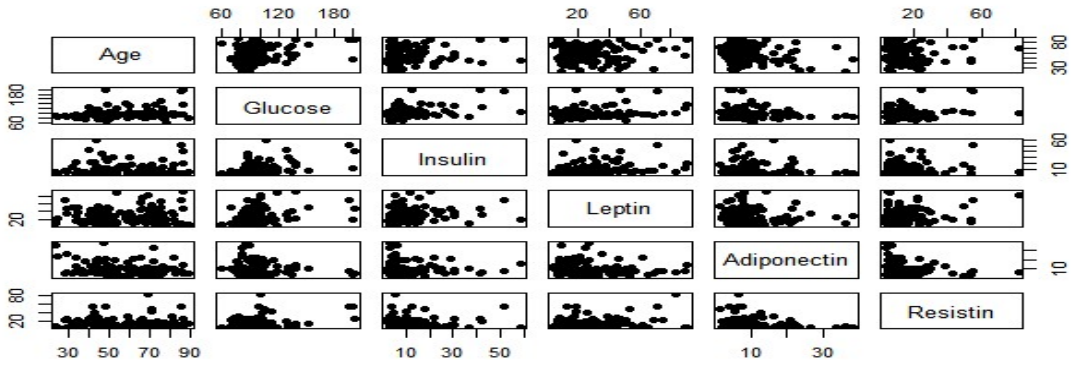

Figure 1: **Scatter plots showing the distribution of all predictors (Age, Glucose, Insulin, Leptin, Adiponectin, and Resistin)**

Table 3: Covariance Analysis table between Predictors

	Age	**Glucose**	**Insulin**	**Leptin**	**Adiponectin**	**Resistin**
Age	259.6212	83.5151	5.2714	31.7213	-24.2377	0.5474
Glucose	83.5151	507.3829	114.4443	131.8271	-18.8247	81.3099
Insulin	5.2714	114.4442	101.3599	58.22212	-2.1562	18.3041
Leptin	31.7213	131.8271	58.2222	367.9988	-12.5224	60.9050
Adiponectin	-24.2377	-18.8247	-2.1562	-12.5224	46.8313	-21.399
Resistin	0.5474	81.3099	18.3041	60.9050	-21.3988	153.5281

Table 3B: Multicollinearity Analyses

Age	**Glucose**	**Insulin**	**Leptin**	**Adiponectin**	**Resistin**
1.141442	1.291418	1.160439	1.313498	1.127829	1.14754

The table above shows the assumption check for the presence of multicollinearity among predictor variables.

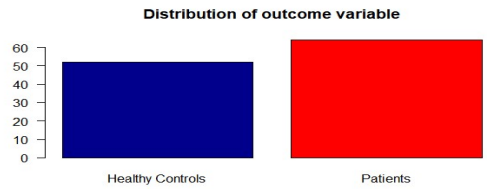

Figure 2: Bar plot showing the distribution of outcome variable

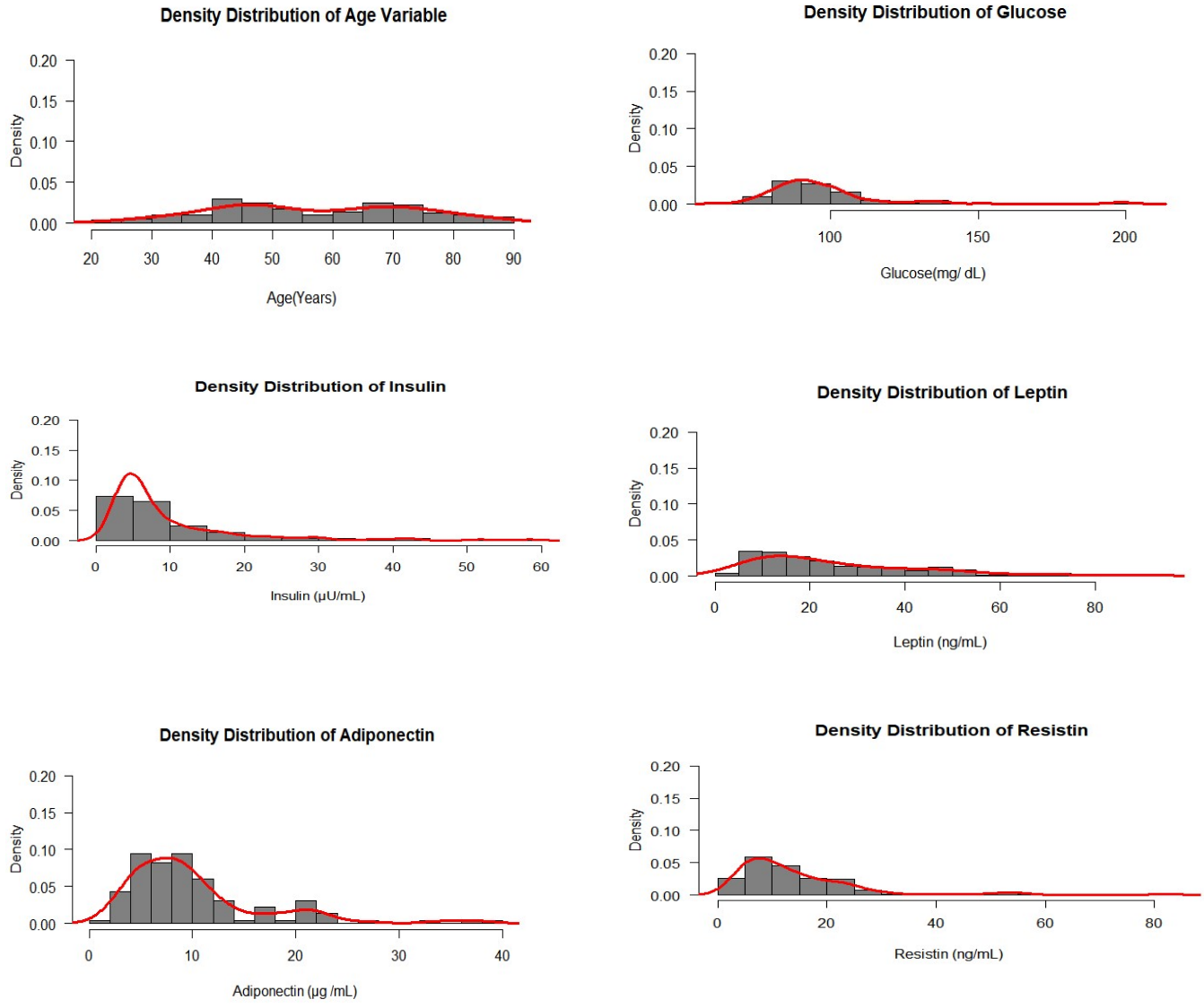

Figure 3: Shows the density distribution of each predictor variable

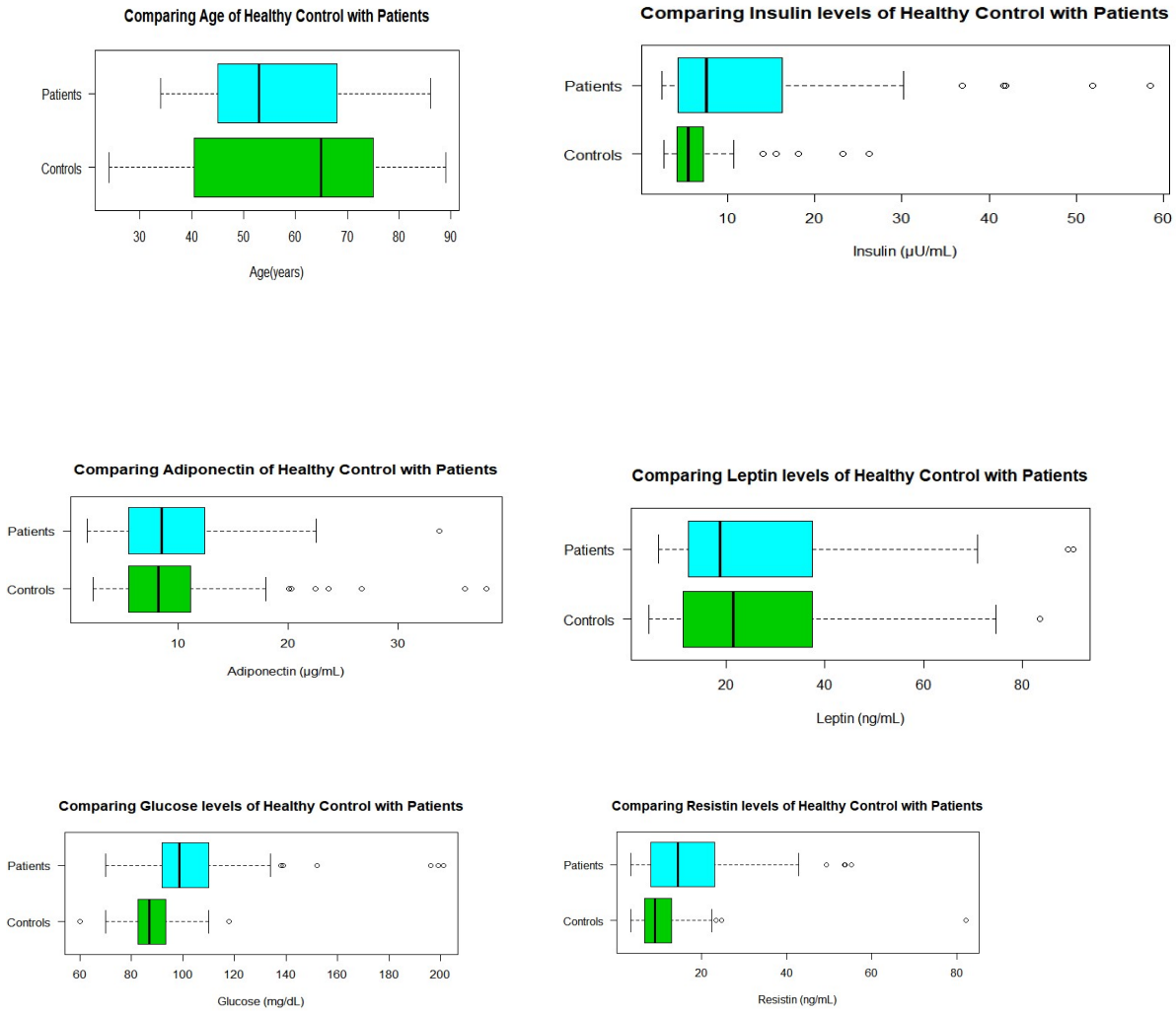

Figure 4: Boxplot showing the distribution of Age, Glucose, Insulin, Leptin, Adiponectin, Resistin with the outcome variable (Healthy Control/Patients)

Table 4: Univariate logistic regression table

	Estimate	Standard Error	z value	Pr(>|z|)
Age	-0.0055	0.0117	-0.469	0.639
Glucose	0.0787	0.0203	3.869	0.000109 ***
Insulin	0.0839	0.0315	2.665	0.00769 **
Leptin	-0.0001	0.0098	-0.012	-0.991
Adiponectin	-0.0057	0.0273	-0.210	0.834
Resistin	0.0489	0.0214	2.284	0.0224 *

The table above shows that Glucose, Insulin, and Resistin are significant at P<0.05. Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 (Dispersion parameter for the binomial family taken to be 1).

Table 5: ANOVA Table showing significant predictor variables

	LR Chisq	Df	Pr(>Chisq)
Glucose	26.833	1	2.219e-07 ***
Insulin	11.006	1	0.0009081 ***
Resistin	6.962	1	0.008326 **

Tables 4 and 5 shows that variables glucose (p value=<0.001, Pr(>Chisq) =<0.001), Insulin (P=0.008 Pr(>Chisq) =<0.001) and Resistin (p value=0.022, Pr(>Chisq) =0.008) are all significant at $p < 0.05$.

COMPLEX ANALYSIS

Table 6: Multivariate logistics analysis

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-7.68557	2.13874	-3.594	0.000326 ***
Age	-0.0198	0.01508	-1.313	0.189189
Glucose	0.0894	0.02293	3.898	9.68e-05 ***
Insulin	0.05819	0.03781	1.539	0.123801
Leptin	-0.03521	0.01382	-2.548	0.010842 *
Adiponectin	0.01645	0.03439	0.478	0.63252
Resistin	0.06267	0.02851	2.198	0.027923 *

Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 (Dispersion parameter for binomial family taken to be 1). Multivariate analyse shows that Glucose (p=9.68e-05), Resistin (p=0.028) and Leptin (p=0.0108) were significant at p value < 0.05.

Table 6b Anova Table showing significant predictor variables

	LR Chisq	Df	Pr(>Chisq)
Age	1.7557	1	0.185158
Glucose	23.1537	1	1.496e-06 ***
Insulin	2.9081	1	0.088138.
Leptin	6.9787	1	0.008249 **
Adiponectin	0.2265	1	0.634112
Resistin	6.9888	1	0.008202 **

Glucose, Insulin, Leptin, and Resistin are significant at P<0.05.

Table7: Odds Ratio and Confidence interval table of Classification (Health Control/ Patients) and Predictors (Age, Glucose, Insulin, Leptin, Adiponectin, and Resistin)

	ODDS RATIO	2.50%	97.50%
(Intercept)	0.000459411	4.71E-06	0.02214118
Age	0.98039279	9.51E-01	1.00948268
Glucose	1.093514674	1.05E+00	1.14910088
Insulin	1.059915353	9.92E-01	1.15371399
Leptin	0.965404151	9.38E-01	0.99109029
Adiponectin	1.016582538	9.48E-01	1.08729895
Resistin	1.064678095	1.01E+00	1.13546191

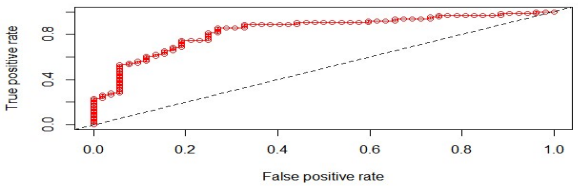

Figure 5: Receiver Operator Characteristic (ROC) Curve

The figure above presents the Receiver Operator Characteristic (ROC) Curve to determine the tra de-off between sensitivity and specificity (AUC= 0.835).

DISCUSSION

Selection of reliable biomarkers for breast cancer early detection and treatment play a crucial role in reducing morbidity and mortality rate. Biomarkers have shown to present an alternative to computer aided diagnosis (CAD) with the added advantage of the non-invasive detection method. This study presents a model that can predict breast cancer early detection using Age, Glucose, Resistin, Leptin, Adiponectin, and Insulin as predictors. Descriptive statistics show variability among the predictors and pair wise correlation present not very high association among the predictors. The Variance Inflation Factor (VIF) for detecting multicollinearity and measure of imprecision was less than 2 which indicates a low correlation among the predictor variables. Univariate analysis showed that Glucose, Insulin, and Resistin were significant at $P<0.05$. This agrees with previous studies that Glucose, Resistin and Insulin are predictors of breast cancer. Analysis of variance (ANOVA) also showed the same significant pattern. Complex analysis using multivariate logistic analysis showed that Glucose [OR = 1.09, 95% (CI= 1.05 - 1.15)], Leptin [OR = 0.97, 95% (0.94-0.99)], Resistin [OR = 1.06, 95% (1.02-1.14)] were all significant at $P<0.05$ presence. Age [OR = 0.98, 95% (CI= 4.71e-06 – 0.022)], Insulin [OR = 1.06, 95% (CI= 0.99 – 1.15)]and Adiponectin [OR = 1.02, 95% (CI= 0.95 – 1.08)] were not significant at $P<0.05$.Interestingly, Insulin and Leptin were significant for the analysis of variance (ANOVA).

Leptin which was significant during multivariate analysis was not a surprise because several kinds of literature have reported that Leptin is a good predictor of breast cancer. Furthermore, Analysis of variance (ANOVA) showed that glucose, leptin, and Resistin were still significant. The odds ratio analysis showed that the odds of experiencing breast cancer is 1.09 times higher for every unit increase of glucose. Also, for every 1 unit increase in Leptin, Resistin, Adiponectin and Insulin, the odds of experiencing breast cancer are 0.97, 1.06, 1.02 and 1.06 times higher respectively. Receiver Operator Characteristic (ROC) curve shows the area under the curve is 0. 835. This suggests that the model is a good fit and we are able to predict the presence of breast cancer in females using the independent variables. Also, AUC showed that the regression model was a better fit compared to the null model.

CONCLUSION

The results showed that Adiponectin, Leptin, Resistin, Glucose, Age and Insulin are a good model for the prediction of breast cancer in females with AUC 0. 84. Also, Glucose, Leptin, and Resistin could be good predictors for the presence of breast cancer. This suggests that the combination of Glucose, Leptin and Resistin may be considered to be good predictors of breast cancer in females. To further evaluate the roles of these predictors, I suggest that a well-designed study with therapeutic implications beyond odd ratio be considered. Also, a robust statistical technique should be considered for better decisions on outliers. Since the origin of data, demographic and other relevant information about the patient is not known, there is no assurance that there are no other factors affecting the patients used for the study. The main advantage of using biomarkers is that they can provide diagnostic and therapeutic predictive information as well as prognosis.

References

1. Adel M. A. Assiri, Hala F. M. Kamel, and Mohamed F. R. Hassanien Resistin, Visfatin, Adiponectin, and Leptin: Risk of Breast Cancer in Pre- and Postmenopausal Saudi Females and Their Possible Diagnostic and Predictive Implications as Novel Biomarkers (2015)
2. Ralph D'Agostino Sr., Lisa Sullivan, Alexa Beiser. Introductory Applied Biostatistics. Brooks Cole (2005).
3. Soler, M.; Estevez, M.C.; Villar-Vazquez, R.; Casal, J.I.; Lechuga, L.M. Label-free nanoplasmonic sensing of tumor-associate autoantibodies for early diagnosis of colorectal cancer. *Anal. Chim. Acta* 2016, *930*, 31–38.
4. Zhu Q, et al. Pathologic response prediction to neoadjuvant chemotherapy utilizing pretreatment near-infrared imaging parameters and tumor pathologic criteria. Breast Cancer Res. 2014;16(5):456
5. Cole KD, He HJ, Wang L. Breast cancer biomarker measurements and standards. Proteomics Clin Appl. 2013;7(1–2):17–29. 3.
6. Hwa HL, et al. Prediction of breast cancer and lymph node metastatic status with tumour markers using logistic regression models. J Eval Clin Pract. 2008;14(2):275–80. 4.
7. Santillan-Benitez JG, et al. The tetrad BMI, Leptin, Leptin/Adiponectin (L/a) ratio and CA 15-3 are reliable biomarkers of breast cancer. J Clin Lab Anal. 2013;27(1):12–20. 5.
8. Dalamaga M, et al. Serum resistin: a biomarker of breast cancer in postmenopausal women? Association with clinicopathological characteristics, tumor markers, inflammatory and metabolic parameters. Clin Biochem. 2013;46(7–8):584–90