



Analysis of Covariance of Gleason Score: A Case Study of 100 Prostate Cancer Patients Undergoing Treatment at the University of Port Harcourt Teaching Hospital, Rivers State, Nigeria

Sylva Ligeiaziba^{1*}, Maxwell A. Ijoma² and Emmanuel I. Biu²

¹Department of Mathematics, Bayelsa Medical University, Nigeria.

²Department of Mathematics and Statistics, University of Port Harcourt, Nigeria.

Authors' contributions

This work was carried out in collaboration among all authors. Authors SL, MAI and EIB designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Authors SL and MAI managed the analyses of the study. Author EIB managed the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2021/v11i330269

Editor(s):

(1) Dr. Oguntunde, Pelumi Emmanuel, Covenant University, Nigeria.

Reviewers:

(1) Umesh Chandra Dutta, India.

(2) Selcuk Erdem, Istanbul University, Turkey.

Complete Peer review History: <http://www.sdiarticle4.com/review-history/60860>

Original Research Article

Received: 20 July 2020

Accepted: 25 September 2020

Published: 27 March 2021

Abstract

Prostate cancer is the second most common cause of cancer related deaths in men. It is detected using many screening methods. Like every other cancer, there are risk factors associated with prostate cancer. This include but not limited to, Family History (FH) of the disease, smoking habit, alcohol intake, age and Body Mass Index (BMI). The survival of prostate cancer patients is dependent on many factors such as, early detection of the disease, age of patient and the aggressiveness of the cancer. Gleason score is used to measure the level of aggressiveness of a prostate cancer in a patient. the score ranges from 6 to 10. It is made up of two Gleason grades that ranges from 3 to 5. This study was carried out to determine whether there are significant differences in the mean of Gleason score by the various categories of BMI and FH of patients while controlling for the number of hospital visits. Gleason score was used as the dependent variable while FH and BMI and Number of hospital visits were used as the independent variables. Descriptive statistical measures were used to summarize the basic features of the data. Spearman correlation coefficient was used to measure if there is a significant statistical relationship

*Corresponding author: E-mail: sylva.ligeiaziba@bmu.edu.ng, gedaliahsylva@gmail.com;

between the Gleason score, age and BMI, while Analysis of Covariance (ANCOVA) was used to measure the differences in the mean of Gleason score by the categories of FH and BMI while controlling for number of hospital visits. The analysis was done using Statistical Programme for Social Science (SPSS 25.0) and Intellectus Statistics software. Results from the analyses were presented in tabular form. The results showed a significant effect of Body Mass Index (BMI) on Gleason score and that Gleason score increases, as age tends to increase.

Keywords: Prostate cancer; Gleason score; Body Mass Index; family history.

1 Introduction

A cell is the basic building block of a living things [1]. Trillions of cells make up the human body and these cells are responsible for the structure of the human body. When these cells begin to grow abnormally, it causes diseases. Cancer refers to a large number of diseases characterized by the abnormal cells growth that divide uncontrollably with the ability to spread, infiltrate and destroy neighboring normal body tissues. This disease is the among the leading causes of death globally. There are hundreds of different types of cancer and they are named after the primary organ or type of cell that is originally affected. Prostate cancer is among the most common type of cancers found in men. It is called prostate cancer because it begins developing in the prostate gland [2]. The prostate is located around the urethra of the male reproductive system and is responsible for the production of protective fluids for sperm cells. This fluid nourishes and protect sperm cells in the semen. The prostate becomes cancerous when it's cells begin to grow abnormally from the prostate gland to neighboring tissues, especially lymph nodes and bones the bones. Prostate cancer may cause many abnormalities in men, such as, painful and difficult urination, erectile dysfunction during sex and many others. Most patients are asymptomatic during the early stages of the disease. The disease has become a global health burden because of its prevalence and incidences in men worldwide. Prostate cancer accounts for over 7% of new cancer cases and over 14% of cases in men worldwide. Nigeria is not left out, as the disease is the most commonly reported cancer among men. it has hospital prevalence of about 182.5 per 100,000 men in 2010 in Osun state of Nigeria. It has been scientifically proven that the survival rate of the disease is high if reported or diagnosed early. Many countries have recorded significant decline in prostate cancer mortality due to early screening and detection. Prostate cancer is detected using many screening methods depending on the risk level of the patient. involved. However, Prostate-Specific Antigen (PSA) test, is the most used screening method which measures the PSA level in the blood. PSA levels:

1. <2.5 ng/ml
2. <3.5 ng/ml
3. <4.5 ng/ml
4. <6.5 ng/ml

Are considered abnormal. Other known screening methods are urine test and digital test. Several researchers have identified risk factors associated with prostate cancer [3,4]. Some of the identified risk factors include but not limited to age, tumor size, family history of prostate cancer level of alcohol intake and body mass index (BMI) [5], The Gleason score is used to measure the level of aggressiveness of a prostate cancer, it will be of great interest to study the association of the Gleason score to the risk factors associated with prostate cancer.

2 Materials and Methods

2.1 Data

The dataset used in this retrospective study was a secondary data collected from the data registry of the University of Port Harcourt Teaching Hospital (UPTH), Port Harcourt, River State, Nigeria. The dataset is the clinical record of 100 prostate cancer patients undergoing treatment and followup at the Hospital. The

variables include; age of patient, Gleason score, capra score, number of visits, family history, blood group and genotype. The data was presented in tabular form.

2.2 Statistical methods

2.2.1 Descriptive analysis

Descriptive statistics were used to summarize the basic features of dataset. Mean, standard deviation and standard errors was calculated for each variable. Frequencies and percentages were used to count occurrences of the nominal variables. Also maximum and minimum values was used to show the range.

2.3 Spearman correlation analysis

A Spearman correlation was used to analyze the relationship between (BMI) and age to Gleason score. Cohen's standard was used to evaluate the strength of the relationship Coefficients values between 0.10 and 0.29 represent a small effect size and between 0.30 and 0.49 represent a moderate effect size, while above .50 indicate a large effect size [6].

2.4 Analysis of covariance (ANCOVA)

ANCOVA is a statistical procedures used to analyze or study the differences between categories of the explanatory variable based the response variables while controlling the covariate. it is a systematic extension of analysis of variance (ANOVA). The statistical procedure has some assumption that must be met for the variable under study to conform to an ANCOVA procedures

2.4.1 Assumptions

- 1) **Normality:** The assumption of normality is examined by plotting the quantiles of the model residuals against the quantiles of a Chi-square distribution, also called a Q-Q scatterplot [7]. For the assumption of normality to be met, the quantiles of the residuals must not strongly deviate from the theoretical quantiles. Strong deviations could indicate that the parameter estimates are unreliable. Also a Shapiro-Wilk test can be used to determine whether the model residuals could have been produced by a normal distribution [8].

2.4.2 Homoscedasticity

The evaluation of homoscedasticity is checked by examining the plot of the residuals against the predicted values [9,10]. The assumption is met if the points appear randomly distributed with a mean of zero and no apparent curvature.

- 2) **Outliers:** To examine influential points, Studentized residuals were calculated and the absolute values were plotted against the observation numbers [11]. Studentized residuals are calculated by dividing the model residuals by the estimated residual standard deviation. An observation with a Studentized residual greater than 3.17 in absolute value, the 0.999 quartile of a t distribution with 99 degrees of freedom, was considered to have significant influence on the results of the model.

3 Results

3.1 Frequencies and percentages

The observations for Visit had an average of 12.94 ($SD = 8.74$, $SE_M = 0.87$, $Min = 1.00$, $Max = 77.00$, $Skewness = 3.90$, $Kurtosis = 26.90$). The observations for Gleason score had an average of 7.37 ($SD = 0.77$,

$SE_M = 0.08$, Min = 6.00, Max = 9.00, Skewness = 0.18, Kurtosis = -0.32). The observations for age had an average of 72.25 ($SD = 5.98$, $SE_M = 0.60$, Min = 59.00, Max = 84.00, Skewness = -0.28, Kurtosis = -0.75). The observations for BMI had an average of 25.96 ($SD = 2.67$, $SE_M = 0.27$, Min = 20.00, Max = 32.90, Skewness = 0.04, Kurtosis = -0.46). When the skewness is above 2 in absolute value, the variable is considered to be asymmetrical about its mean. When the kurtosis is above or equal to 3, the variable's distribution is markedly different than a normal distribution in its tendency to produce outliers [12]. The summary statistics can be found in Table 1.

Table 1. Summary statistics table for interval and ratio variables

Variable	<i>M</i>	<i>SD</i>	<i>n</i>	<i>SE_M</i>	Min	Max	Skewness	Kurtosis
Visit	12.94	8.74	100	0.87	1.00	77.00	3.90	26.90
Gleason score	7.37	0.77	100	0.08	6.00	9.00	0.18	-0.32
Age	72.25	5.98	100	0.60	59.00	84.00	-0.28	-0.75
BMI	25.96	2.67	100	0.27	20.00	32.90	0.04	-0.46

Note. '-' denotes the sample size is too small to calculate statistic

Table 2. Frequency table for nominal variables

Variable	<i>n</i>	%
FH		
No	73	73.00
Yes	27	27.00
Missing	0	0.00
Blood Group		
O	34	34.00
A	38	38.00
AB	10	10.00
B	13	13.00
0	5	5.00
Missing	0	0.00
Age		
70-79	57	57.00
60-69	29	29.00
80-89	12	12.00
8	1	1.00
50-59	1	1.00
Missing	0	0.00
BMI		
Over weight	53	53.00
Normal weight	41	41.00
Obese	6	6.00
Missing	0	0.00
Genotype		
AA	54	54.00
AS	46	46.00
Missing	0	0.00
Gleason score		
High (8.0-10.0)	42	42.00
Moderate (7.0)	57	57.00
Low (6.0)	1	1.00
Missing	0	0.00

Note: Due to rounding errors, percentages may not equal 100%

3.2 Result of spearman correlation

The result of the correlation was examined based on an alpha value of 0.05. A significant positive correlation was observed between BMI and Gleason score ($r_s = 0.29, p = .003, 95\% \text{ CI } [0.10, 0.46]$). The correlation coefficient between BMI and Gleason score was 0.29, indicating a small effect size. This correlation indicates that as BMI increases, Gleason score tends to increase. Table 3 presents the results of the correlation the result of the correlation was examined based on an alpha value of 0.05. A significant positive correlation was observed between Gleason score and age ($r_s = 0.34, p < .001, 95\% \text{ CI } [0.15, 0.50]$). The correlation coefficient between Gleason score and age was 0.34, indicating a moderate effect size. This correlation indicates that as Gleason score increases, age tends to increase. Table 3 presents the results of the correlation. The result of the correlation was examined based on an alpha value of 0.05. There were no significant correlations between any pairs of variables. Table 3 presents the results of the correlation.

Table 3. Spearman correlation results

Combination	r_s	95% CI	p
Gleason score-Age	0.34	[0.15, 0.50]	< .001
BMI-Gleason score	0.29	[0.10, 0.46]	.003
Gleason score-Visit	-0.01	[-0.21, 0.18]	.902

Note: $n = 100$

3.3 Result of preliminary analysis

The results of the Shapiro-Wilk test were not significant based on an alpha value of 0.05, $W = 0.98, p = .069$. This result suggests the possibility that the residuals of the model were produced by a normal distribution cannot be ruled out, indicating the normality assumption was met. The assumption of homoscedasticity was met since points appear randomly distributed with a mean of zero and no apparent curvature (see Fig. 2). Studentized residuals are calculated by dividing the model residuals by the estimated residual standard deviation. An observation with a Studentized residual greater than 3.18 in absolute value, the 0.999 quartile of a t distribution with 97 degrees of freedom, was considered to have significant influence on the results of the model.

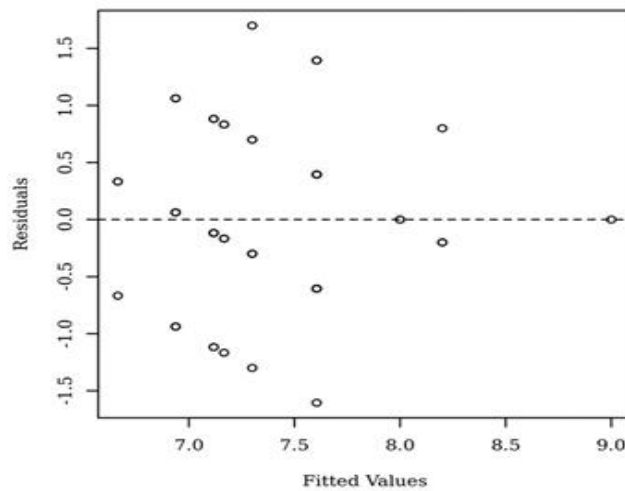


Fig. 1. Residuals scatterplot testing homoscedasticity

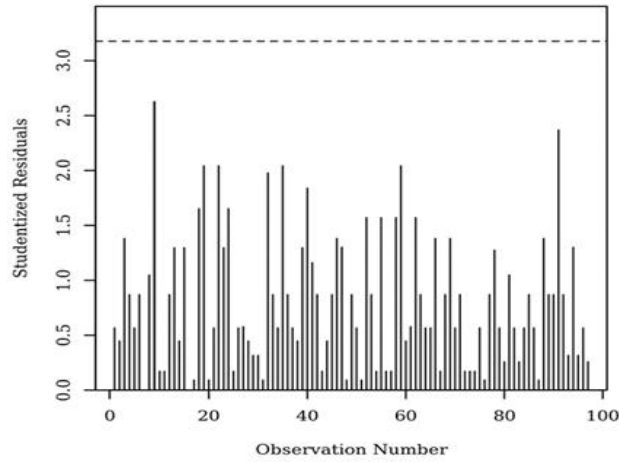


Fig. 2. Studentized residuals plot for outlier detection

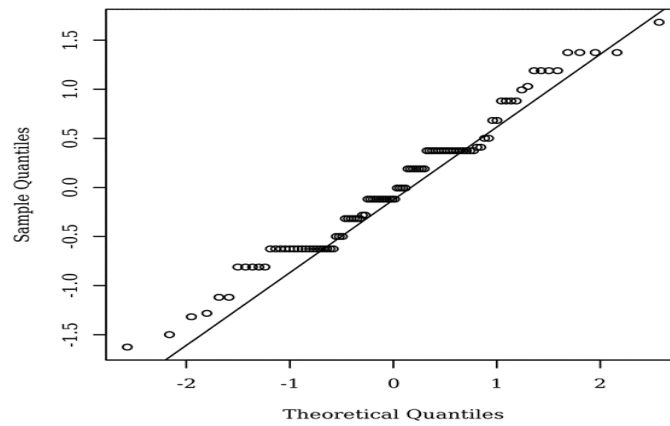


Fig. 3. Q-Q scatterplot for normality of the residuals for the regression model

3.4 Result of ANCOVA

The results of the ANCOVA were significant, $F(4, 95) = 3.65, p = .008$, indicating significant differences among the values of BMI and FH (Table 4). The main effect, BMI was significant, $F(2, 95) = 7.24, p = .001, \eta_p^2 = 0.13$, indicating there were significant differences in Glsc by BMI levels. The main effect, FH was not significant, $F(1, 95) = 0.88, p = .351$, indicating there were no significant differences of Glsc by FH levels. The means and standard deviations are presented in Table 5.

Table 4. Analysis of variance table for Gleason-score by BMI and FH

Term	SS	df	F	p	η_p^2
BMI	7.83	2	7.24	.001	0.13
FH	0.48	1	0.88	.351	0.01
Visit	0.13	1	0.24	.625	0.00
Residuals	51.40	95			

Table 5. Marginal means, standard error, and sample size for Glsc by BMI and FH controlling for visit

Combination	Marginal Means	SE	n
Over weight : yes	7.58	0.11	43
Obese : yes	6.99	0.13	26
3 : yes	7.43	0.31	4
Over weight : no	7.74	0.17	10
Obese : no	7.15	0.16	15
3 : no	7.59	0.32	2

3.5 Post-hoc

Paired *t*-tests were calculated between each pair of measurements to further examine the differences among the variables. Tukey pairwise comparisons were conducted for all significant effects based on an alpha of 0.05. For the main effect of BMI, the mean of Gleason score for over weight ($M = 7.60$, $SD = 0.74$) was significantly larger than for normal weight ($M = 7.05$, $SD = 0.69$), $p = .001$. No other significant effects were found.

4 Discussion and Conclusion

The results from the study showed a significant effect of Body Mass Index (BMI) on Gleason score. It showed that the mean of Gleason score for overweight patients was significantly higher than normal weight. We also noticed a significant effect of age of patients on the Gleason score, as 80-89yrs of age had a higher mean Gleason score. Body mass index was seen to have more effect on the Gleason score than the age of patients, however, further study of these effects on Gleason score should be investigated with a different sample before a definite and reliable conclusion can be made. No other significant effects were found.

Ethical Approval

After ethical approval from the University of Port Harcourt Teaching Hospital, clinical records of 100 prostate cancer patients were collected.

Competing Interests

Authors have declared that no competing interests exist.

References

- [1] Biologydictionary.net Editors. (2016, November 15). Cell. Available:<https://biologydictionary.net/cell/>
- [2] Summary of the epidemiology. American Journal of Epidemiology. 1996;144(11):1041-1047.
- [3] Chan JM, Stampfer MJ, Giovannucci EL. What causes prostate cancer? A brief summary of the epidemiology; 1998.
- [4] Lesko SM, Rosenberg L, Shapiro S. Family history and prostate cancer risk. Seminars in Cancer Biology. 1996;8:263-273.

- [5] Andreas Josefsson, Pernilla Wikström, Lars Egevad, Torvald Granfors, Lars Karlberg, Pär Stattin, Anders Bergh. Low endoglin vascular density and Ki67 index in Gleason score 6 tumours may identify prostate cancer patients suitable for surveillance. *Scandinavian Journal of Urology*. 2012; 46(4):247-257.
- [6] Cohen J. *Statistical power analysis for the behavior sciences* (2nd ed.). West Publishing Company; 1988.
- [7] De-Carlo LT. On the meaning and use of kurtosis. *Psychological Methods*. 1997;2(3):292-307.
- [8] Razali NM, Wah YB. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*. 2011;2(1):21-33.
- [9] Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4: arXiv preprint arXiv. *Journal of Statistical Software*; 2014.
- [10] Osborne J, Waters E. Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation*. 2002;8(2):1-9.
- [11] Field A. *Discovering statistics using SPSS* (4th ed.). Sage Publications; 2013.
- [12] Westfall PH, Henning KSS. *Texts in statistical science: Understanding advanced statistical methods*. Taylor & Francis Westfall, P. H., & Henning, K. S. S. (2013). *Texts in statistical science: Understanding advanced statistical methods*. Taylor & Francis; 2013.

© 2021 Ligeiaziba et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)

<http://www.sdiarticle4.com/review-history/60860>