Authors	**Amy Good (agood4@elon.edu)** **621 Brandy Courts, Waxhaw North Carolina, 28173** **Phone: 704-681-1542** **Hannah Matthews (hmatthews4@elon.edu)**
Title	**The Prediction of PSA Level Based on Cancer Volume**
Institution	**Department of Statistics, Elon University**
Instructor Sponsor	**Laura Taylor (ltaylor18@elon.edu)**

THE PREDICTION OF PSA LEVEL IN CANCER PATIENTS BASED ON CANCER VOLUME

The Scenario

- A sample size of 97 men with advanced prostate cancer
- The objective of this project was to create a model that uses cancer volume to predict the level of PSA in men with advanced prostate cancer and assess how appropriate this model is based on several criteria
- Multiple prognostic clinical measurements were taken by the team of urologists that originally conducted this study. All variables are displayed in a table on the next slide. For the purpose of this assigned project, we only observed the relationship between Variables 2 and 3—PSA Level and Cancer Volume, respectively.
- SAS software was used to perform all calculations and generate all graphs

Complete List of Variables Used in Original Study

Variable Number	Variable Name	Description
1	Identification Number	1 – 97
2	PSA level	Serum prostate-specific antigen level (mg/ml)
3	Cancer volume	Estimate of prostate cancer volume (cubic centimeters, cc)
4	Weight	Prostate weight (gm)
5	Age	Age of patient (years)
6	Benign prostatic hyperplasia	Amount of benign prostatic hyperplasia (cm^2)
7	Seminal vesicle invasion	Presence or absence of seminal vesicle invasion: 1 if yes; 0 if otherwise
8	Capsular penetration	Degree of capsular penetration (cm)
9	Gleason score	Pathologically determined grade of disease using total score of two patters (summed scores were eight 6, 7 or 8 with higher scores indicating worse prognosis)

Original Model

Estimated (PSA level) = 1.1249 + 3.2299 (cancer volume)

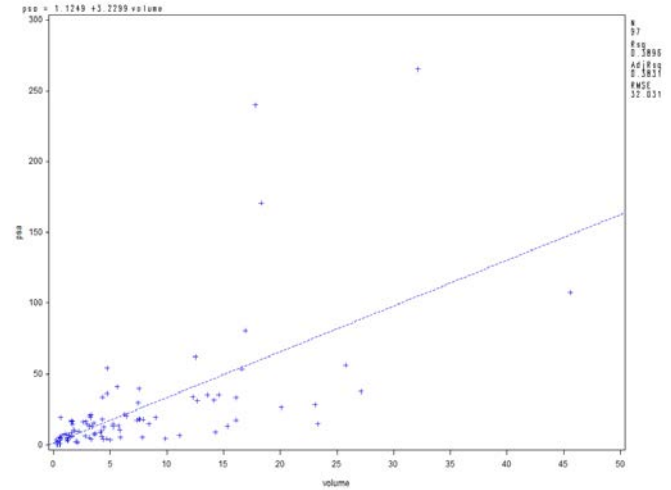
- Interpretation of Slope: For every 1 cubic centimeter increase in cancer volume, the mean PSA level is expected to increase by 3.2299 mg/ml.
- Interpretation of Y-Intercept: When the cancer volume of a patient is 0 cubic centimeters, the mean PSA level is expected to be 1.1249 mg/ml.
- Are these values significant?
  - Y-Intercept: P-Value = 0.7970 (Large P-Value=Not Significant)
  - Slope of Cancer Volume: P-Value = <0.0001 (Small P-Value=Significant)

Assumptions to Check for Original Model

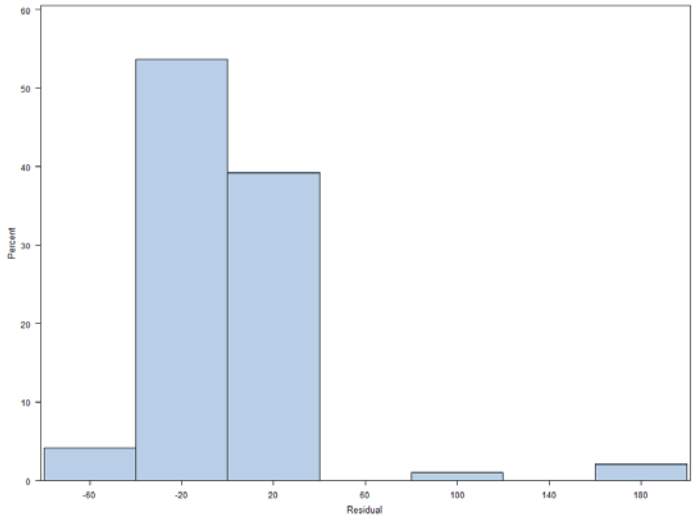
- **Linearity**
  - Checked through observing a linear trend on a scatter plot
- **Independence**
  - Independence of observations can only be checked with more information from the client. Independence could be compromised if any patients were related or came from the same hospital, etc. For the purpose of completing the assignment, we continued with analysis as though the assumption of independence was met, although we must note that independence cannot be officially assumed.
- **Normality**
  - Checked through a qq-plot of the residuals and a histogram
- **Equality of Variance**
  - Checked through a scatterplot of the error terms. We are looking for random scatter above and below the line of y=0.
- **Sample is random**
  - Randomness of the sample is another assumption we cannot assume to be true based on the information given, but we disregarded randomness for the purpose of completing this assignment.

Graphs Generated in SAS to Check Assumptions

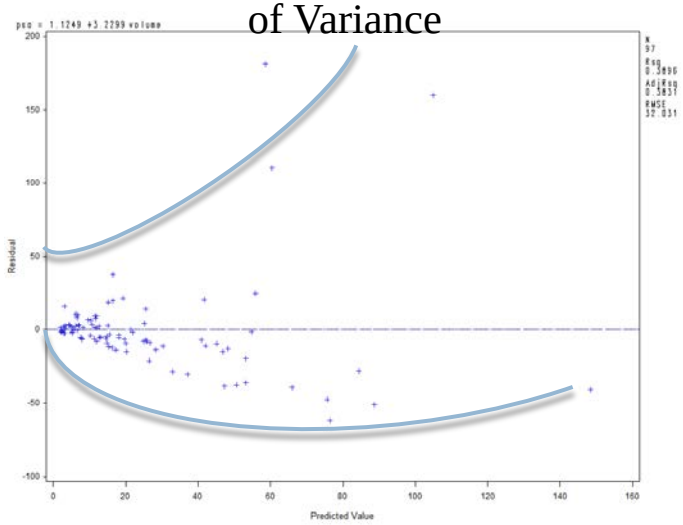
Scatterplot-Linearity



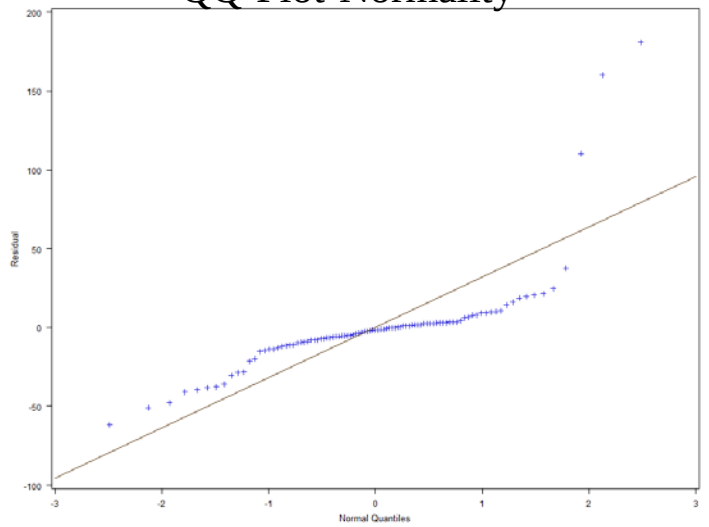
Histogram-Normality



Scatterplot of the Error Terms-Equality of Variance



QQ-Plot-Normality



Assumptions Met?

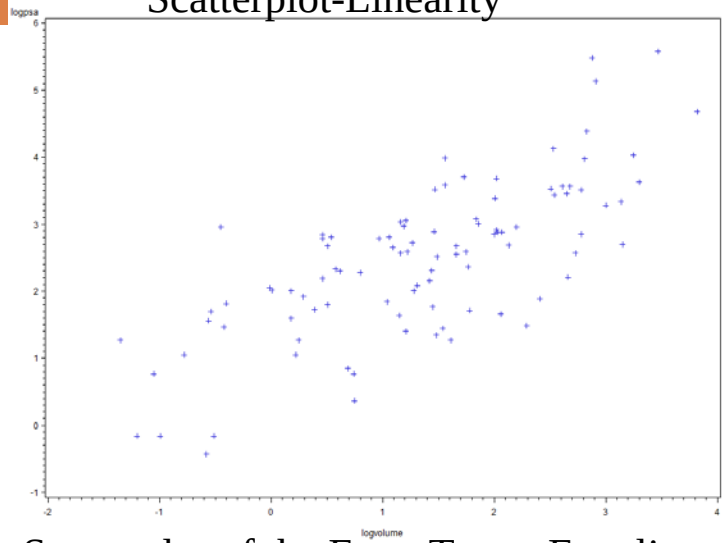
- The assumptions of Linearity, Normality, and Equality of Variance do not appear to have been met by the original model
  - No clear linear trend in scatterplot
  - Histogram is skewed right
  - Cornucopia shape in scatterplot of residuals
  - QQ-Plot does not follow the 45-degree reference line

Improvements on Assumptions

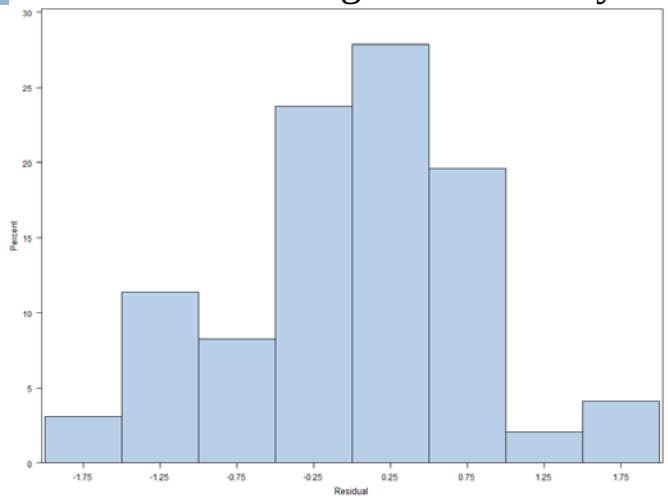
- Remedial measures were taken to meet the assumptions of Linearity, Normality, and Equality of Variance. Both the dependent and independent variables were log-transformed using SAS.
- Log-Transformed Model (Final Model):
  - $\ln(\text{PSA level}) = 1.5092 + 0.7183 \ln(\text{cancer volume})$

Linearity, Normality, and Equality of Variance for Final Model

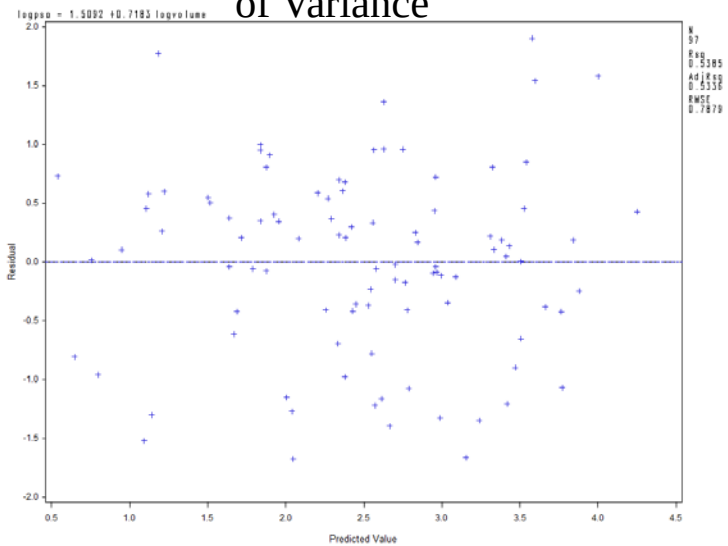
Scatterplot-Linearity



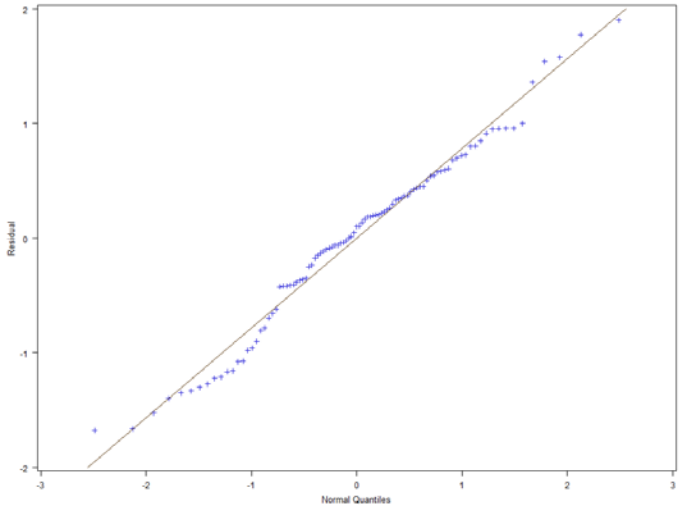
Histogram-Normality



Scatterplot of the Error Terms-Equality of Variance



QQ-Plot-Normality



Assumptions Met for Transformed Model?

- Linearity: There now appears to be a strong linear correlation between PSA level and Cancer volume in men with advanced prostate cancer (Correlation Coefficient (R)=.7338)
- Normality: We see a bell-shaped curve in our histogram of the residuals, indicating normality, and our qqplot shows that the residuals closely follow the 45 degree reference line for normal quartiles
- Equality of Variance: Our residual plot shows random and even scatter above and below the y=0 line
- Our conclusions regarding the assumptions of independence of observations and random sample are the same as before.

Prediction Intervals

- With 95% confidence, we predict that the median PSA level at 15 cc is between 6. 498 and 154.038 mg/ml for observation 98.
- With 95% confidence, we predict that the median PSA level at 1 cc is between 0.929 and 22.021 mg/ml for observation 99.
- With 95% confidence, we predict that the median PSA level at 31 cc is between 10.789 and 263.223 mg/ml for observation 100.

Final Model

$$\ln(\text{PSA level}) = 1.5092 + 0.7183 \ln(\text{cancer volume})$$

- An increase by a multiplicative factor of two in cancer volume in men with advanced prostate cancer is associated with a 1.645 multiplicative effect in the population median PSA level.
- When the cancer volume of a patient is 0 cubic centimeters, the median PSA level will be 4.523 mg/ml.
- Are these values significant?
  - Y-Intercept: p-value = <0.0001 (Yes, Significant)
  - Slope of ln(cancer volume) = <0.0001 (Yes, Significant)

Conclusion

- Our Coefficient of Determination is 0.5385.
  - 53.85% of the variation in PSA level is explained by the regression on cancer volume.
- Our transformed model appears to meet all assumptions and each coefficient is significant. Therefore, we believe that this model is appropriate to use when estimating PSA level based on cancer volume in a male patient with advanced prostate cancer.