Does a relationship exist between chest girth and body weight of Florida Black Bears?

Variables for Analysis & Assumptions

This analysis aims to explore the potential correlation between body weight and chest girth in Florida Black Bears. It will involve distinct examinations for both male and female Black Bears.

Head of male_bears_df:

	body_wt	chest_girth
3	32.21	62.0
4	36.29	63.0
6	35.38	63.5
7	43.09	66.0
11	44.45	67.0

Head of female_bears_df:

	body_wt	chest_girth
0	26.31	54.2
1	27.22	59.0
2	22.68	60.0
5	44.45	63.0
8	40.82	66.0

Assumptions:
For each value of the predictor (x), the probability distribution of the regression error has a mean of zero, has constant equal variance, is normal and are independent.

Summary Statistics and Variable Distributions:

Male Bear Summary Statistics		
Variable	**body_wt**	**chest_girth**
mean	98.88	98.10
std	40.67	18.83
min	32.21	62.00
Q1 (25%)	70.31	86.48
median	91.86	95.75
Q3 (75%)	117.93	108.00
max	212.28	152.00

Female Bear Summary Statistics		
Variable	**body_wt**	**chest_girth**
mean	74.33	88.08
std	24.56	13.32
min	22.68	54.20
Q1 (25%)	58.97	78.88
median	70.31	86.25
Q3 (75%)	86.75	95.13
max	162.39	126.00

Male Bear Body Weight:

The mean body weight of male bears is 98.88, with a median of 91.86, resulting in a difference of 7.02 units between them. This discrepancy suggests a potential rightward skew in the data distribution, as the mean surpasses the median value. The body weight data for male bears also exhibits a notable range, with a maximum of 212.28 and a minimum of 32.21, yielding a range of 180.07. Additionally, the standard deviation for this data set stands at 40.67.

Male Bear Chest Girth:

The mean chest girth for male bears is 98.10, with a median of 95.75, resulting in a modest 2.35-unit difference between the two measures. This slight disparity suggests a distribution that is nearly symmetrical or potentially slightly skewed to the right. The chest girth data for male bears exhibits a range spanning from a minimum of 62.00 to a maximum of 152.00, encompassing a total range of 90.00 units. It is noteworthy that the standard deviation, standing at 18.83, is relatively low, especially when considering that the sample data encompasses bears of all age groups.

Female Bear Body Weight:

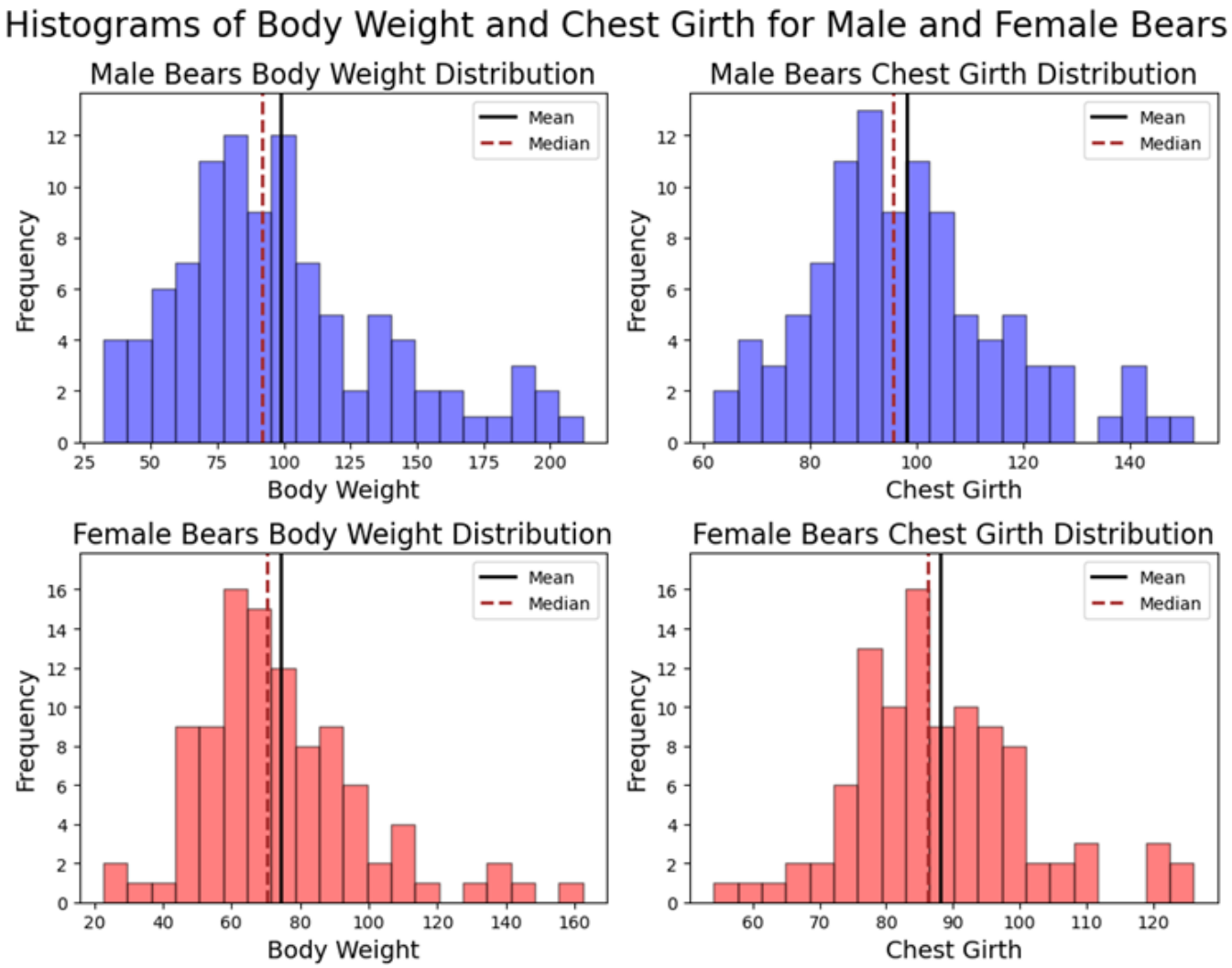
The mean body weight for female bears is 74.33, with a median of 70.31, resulting in a modest 4.02-unit difference between the two metrics. This relatively slight difference, coupled with the mean exceeding the median, suggests a potential, albeit very subtle, rightward skew in the distribution. Female bear body weight ranges from a minimum of 22.68 to a maximum of 162.39, encompassing a wide range of 139.71 units. Notably, the standard deviation for this dataset stands at 24.68.

Female Bear Chest Girth:

The mean chest girth of female bears measures 88.08, while the median is 86.25, yielding a minimal difference of just 1.83 units between the two values. This slight disparity suggests a distribution that is relatively symmetrical in nature. The range of female bear chest girth extends from a minimum of 54.20 to a maximum of 126.00, spanning a total range of 71.80 units. It is noteworthy that the standard deviation for this dataset is 13.32.

NOTE: The substantial range and deviation observed in the body weights of both male and female black bears can be attributed to the comprehensive nature of the sample data, which encompasses black bears of all age groups. To gain further insights, a more nuanced analysis might be considered, involving a segmentation of the data by age groups.

Visual Distributions:



Male Bear Body Weight:

The histogram depicting the body weight distribution of male bears aligns with the insights derived from the summary statistics analysis. The calculated skewness value for the male bear body weight sample data registers at 0.86. A visual examination of the histogram corroborates this observation, revealing a subtle rightward skew in the distribution.

Male Bear Chest Girth:

The histogram representing the chest girth distribution of male bears reinforces the findings from the summary statistics analysis. The calculated skewness value for the chest girth data from the male bear sample is 0.59. Upon visual inspection of the histogram, a very subtle rightward skew is discernible, but the distribution is predominantly symmetrical in nature.

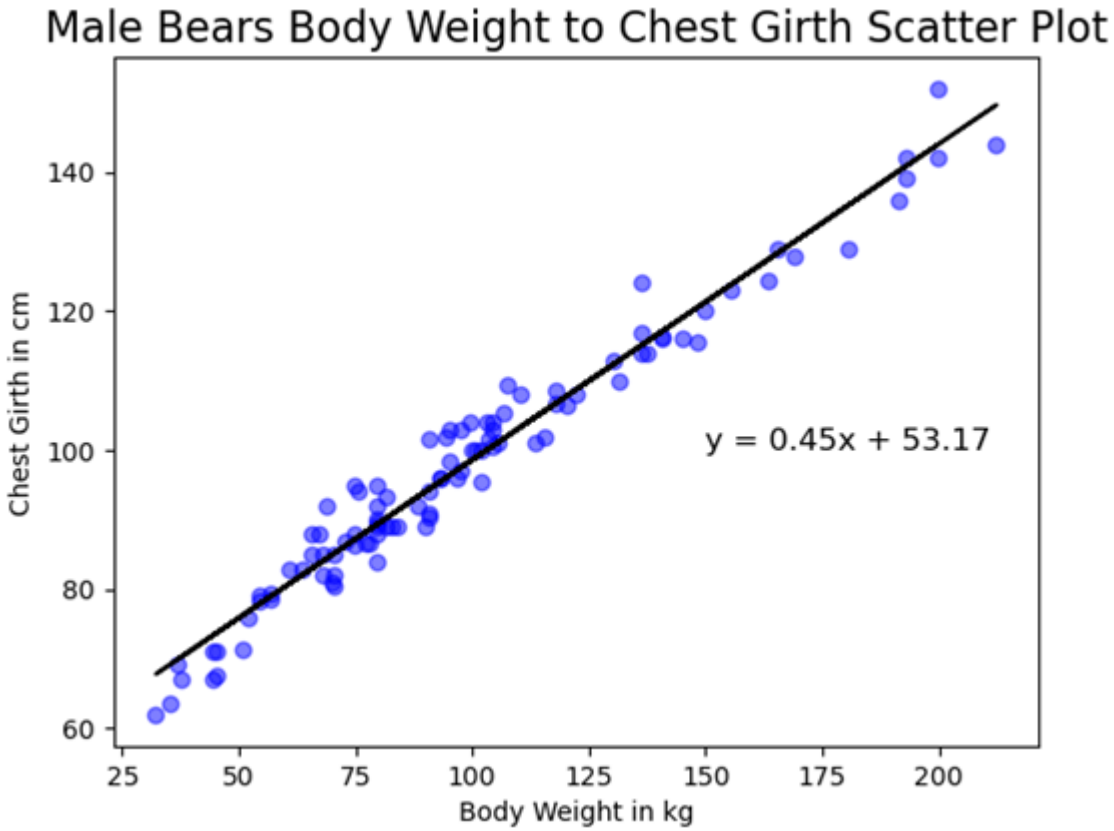
Female Bear Body Weight:

The histogram depicting the body weight distribution of female bears aligns with the insights derived from the summary statistics analysis. Notably, the calculated skewness value for the female bear body weight sample data stands at 1.03. A visual examination of the histogram indicates a rightward skew in the distribution.

Female Bear Chest Girth:

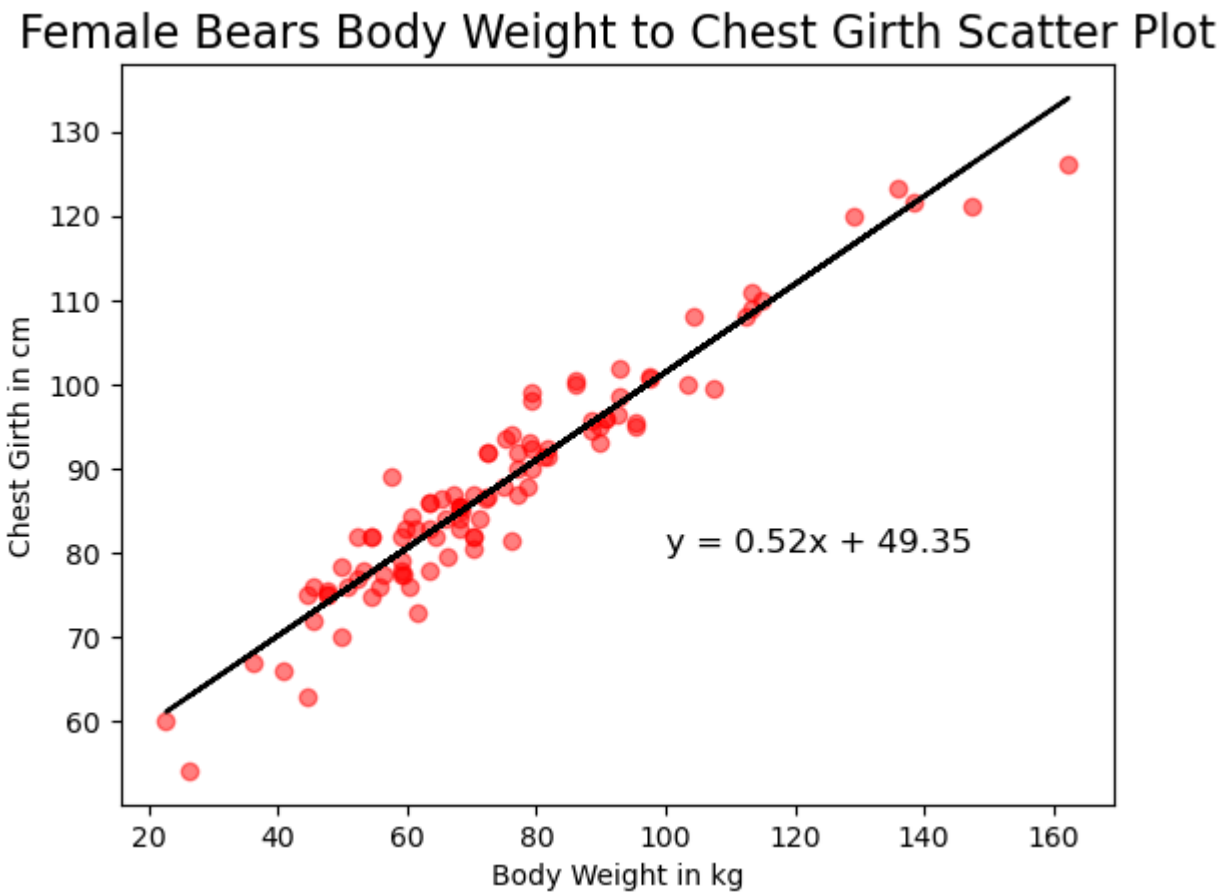
The histogram representing the chest girth distribution of female bears is consistent with the findings from the summary statistics analysis. Notably, the calculated skewness value for the female bear chest girth sample data is 0.56. Upon visual examination of the histogram, a very subtle rightward skew is discernible, but the distribution is predominantly symmetrical in nature.

Scatter Plots



The scatter plot illustrating the connection between body weight and chest girth in male bears implies a moderately strong positive correlation. This correlation's strength is discernible from the proximity of data points to the regression line, and the upward inclination of the regression line reinforces the positive character of this relationship.

While the observed data points are closely packed around the regression line, indicating a substantial degree of consistency, it is essential to emphasize the need for a more comprehensive analysis to fully ascertain the potential relationship and its strength.



The scatter plot representing the correlation between body weight and chest girth in female bears also demonstrates a relatively strong positive association. This becomes evident as we observe the data points, which exhibit a tight clustering pattern around the ascending regression line. The upward trajectory of the regression line further underscores the positive character of this relationship.

It's noteworthy that a significant proportion of data points appear to be concentrated near the central region of the scatter plot. This suggests a considerable number of female bears with similar body weight and chest girth measurements within this dataset. However, to provide a more

comprehensive understanding of the nature and strength of this relationship, a more thorough analysis would be required.

Correlation Coefficient

Male Bears Correlation Coefficient: r = 0.9812

The correlation coefficient, denoted as "r," consistently falls within the range of -1 to 1. The proximity of the "r" value to either -1 or 1 indicates the strength of the relationship. Moreover, the sign of "r" conveys whether the relationship is negative or positive.

In the case of male bear body weight and chest girth, where "r" is calculated to be 0.98, it signifies a strong positive relationship between these two variables.

The scatter plot depicting the relationship between body weight and chest girth of male bears aligns with the calculated correlation coefficient (r) value. The plot reveals a distinct pattern, characterized by a closely clustered distribution of data points that closely follows an ascending regression line.

This observation indicates a positive correlation between body weight and chest girth among male bears. In other words, as body weight increases, chest girth tends to increase as well. The tight grouping of data points around the regression line suggests that this relationship is relatively strong and consistent.

It's important to note that a high correlation coefficient, such as a positive value of r in this case, signifies a strong linear association between the two variables. In this context, the scatter plot and the r value concur, supporting the notion that there is a noticeable and consistent connection between body weight and chest girth in male bears.

Female Bears Pearson Correlation Coefficient: r = 0.9611

An "r" value of 0.96, denoting the correlation between female bear body weight and chest girth, strongly signifies a positive relationship between these two variables.

The scatter plot substantiates the implications of this "r" value. The plot visually manifests a cohesive pattern with data points closely grouped around the ascending regression line. This pattern accentuates that as body weight increases, chest girth tends to increase similarly. In essence, the scatter plot corroborates the notion of a strong positive correlation between female bear body weight and chest girth.

Linear Regression Equation ($y = b_0 + b_1x$) and Coefficient of Determination (R^2)

In a regression equation, "y" denotes the dependent variable under investigation, which we aim to predict or understand. "x," on the other hand, signifies the independent variable, employed to make predictions regarding the dependent variable "y." In this context, "b_0" serves as the y-intercept, representing the value of "y" when "x" equals zero. It essentially defines the baseline value of the dependent variable when the independent variable exerts no influence. "b_1," meanwhile, denotes the slope of the regression coefficient and quantifies the change in the dependent variable "y" for each one-unit alteration in the independent variable "x."

So, the equation "$y = b_0 + b_1x$" essentially defines a linear relationship between the dependent variable "y" and the independent variable "x." The "b0" term is the baseline value, and "b_1" represents the rate of change in "y" associated with changes in "x."

The coefficient of determination, denoted as R-squared (R^2), is a statistical measure used in regression analysis to assess the goodness of fit of a regression model. R^2 indicates the proportion of the variance in the dependent variable that can be accounted for by the independent variable(s) in the regression model. The value of R^2 is always between 0 and 1 and the closer the value of R^2 is to 1 the better the data fits the model. In other words, the closer R^2 indicates the quality and suitability of the regression model, indicating how well the model aligns with the observed data.

Male Bear Liner Regression Equation: $y = 53.17 + 0.45x$

Male Bear Coefficient of Determination: $R^2 = 0.96$

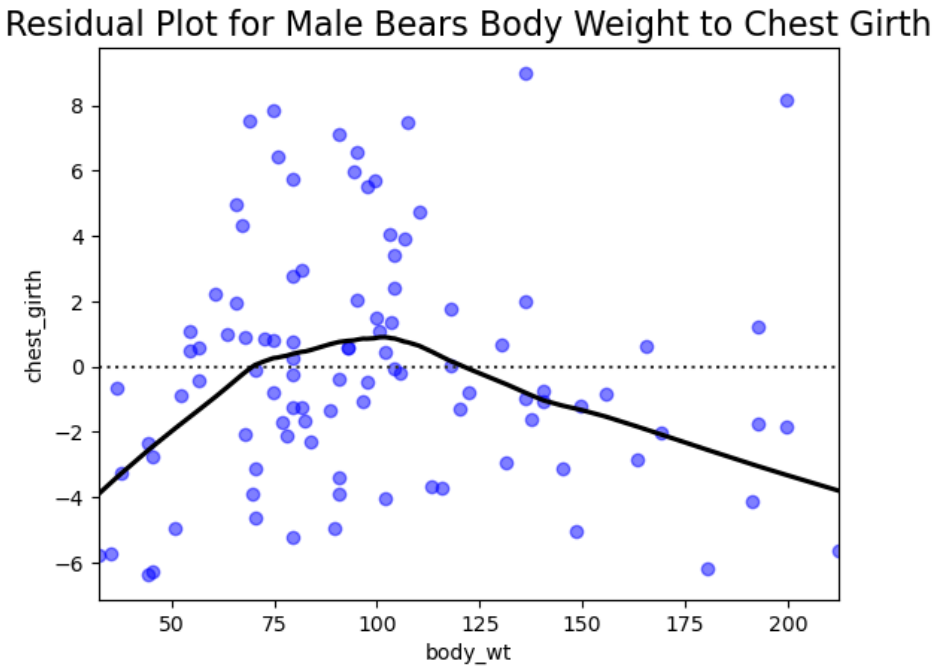
In the linear regression equation for male bears, "y" corresponds to the chest girth value we seek to predict for a male bear. The "b0" value, which stands at 53.17, signifies the chest girth when "x" is equal to 0, essentially representing the baseline chest girth unaffected by the independent variable. Here, "x" denotes the body weight, utilized to forecast the chest girth. "b1" represents the slope, indicating the change in chest girth for each one-unit alteration in body weight.

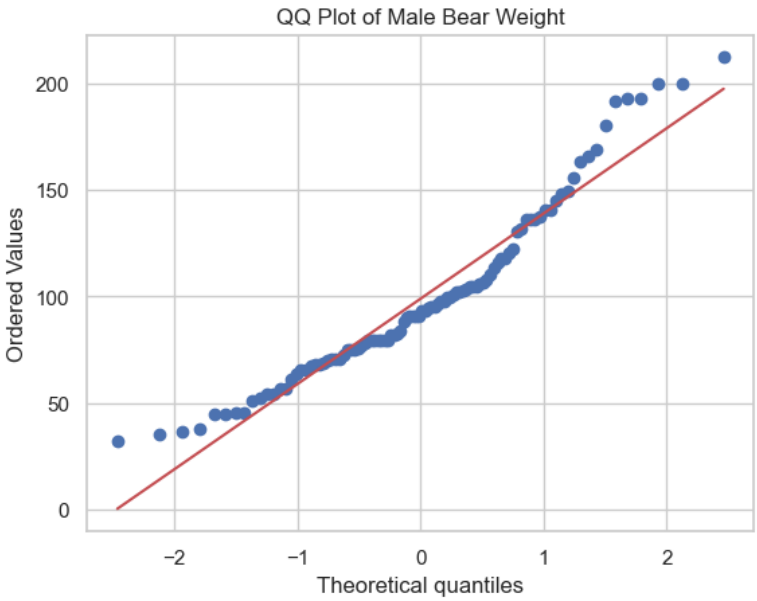
For instance, if we intend to predict the chest girth of a male bear weighing 195 kilograms, we simply substitute 195 for "x" in the equation:

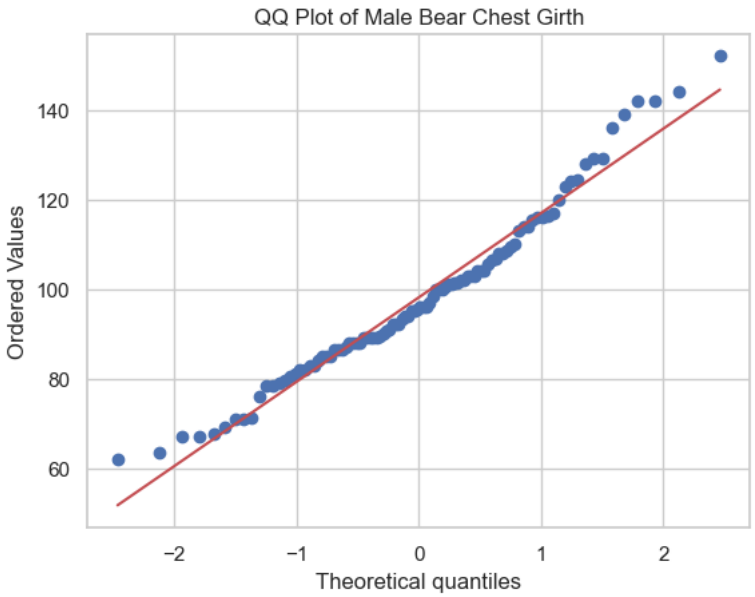
$y = 53.17 + 0.45(195)$
$y = 140.92$

This computation implies that we can anticipate the chest girth of a male bear weighing 195 kilograms to measure approximately 140.92 centimeters. In this manner, the regression equation facilitates the prediction of chest girth values for male bears based on their body weight.

The R^2 value of 0.96 for male bears suggests that the regression model is of exceptional quality and provides an excellent fit to the sample data. This high R^2 value signifies that the model effectively explains a substantial proportion of the variation in the data, reinforcing its reliability in describing the relationship between the variables being studied.







Female Bear Liner Regression Equation: $y = 49.35 + 0.52x$
Female Bear Coefficient of Determination: $R^2 = 0.92$

In the linear regression equation for female bears, "y" corresponds to the chest girth value we seek to predict for a female bear. The "b0" value, which stands at 49.35, signifies the chest girth when "x" is equal to 0, essentially representing the baseline chest girth unaffected by the independent variable. Here, "x" denotes the body weight, utilized to forecast the chest girth. "b1" represents the slope, indicating the change in chest girth for each one-unit alteration in body weight.

For instance, if we intend to predict the chest girth of a female bear weighing 155 kilograms, we simply substitute 155 for "x" in the equation:

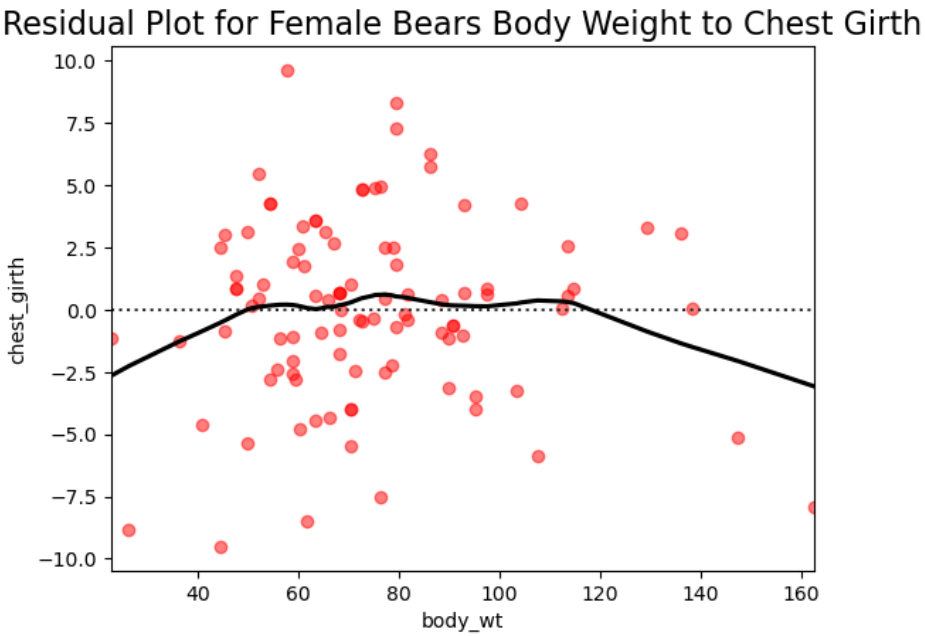
$y = 49.35 + 0.52(155)$
$y = 129.95$

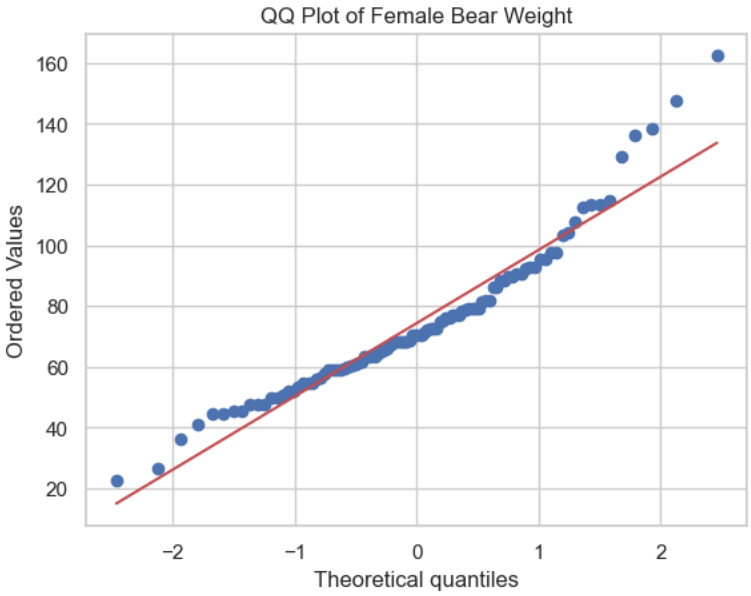
This computation implies that we can anticipate the chest girth of a female bear weighing 155 kilograms to measure approximately 129.95 centimeters. In this manner, the regression equation facilitates the prediction of chest girth values for female bears based on their body weight.

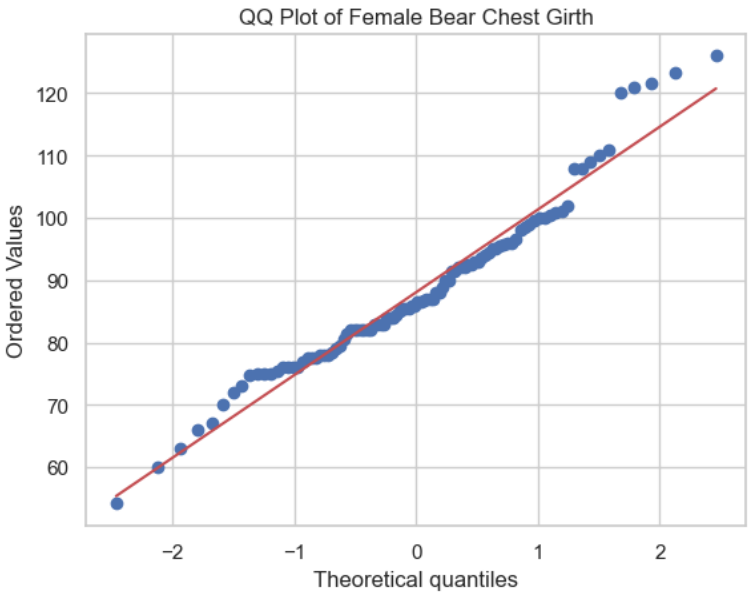
The R^2 value of 0.92 for female bears underscores the quality of the regression model and its strong alignment with the sample data. This high R^2 value signifies that the model adeptly accounts for a significant portion of the variance in the data, affirming its capacity to provide an accurate and reliable representation of the relationship between the variables under examination.

Important Note:

It's crucial to bear in mind that the regression model or equation is specifically applicable within the range of values used to develop it. In practical terms, this means that attempting to predict the chest girth of a male bear weighing less than 32.21 kilograms or exceeding 212.28 kilograms using the provided regression equation is not advisable. Similarly, when dealing with the female bear regression model or equation, it's unwise to predict the chest girth of a female bear weighing less than 22.68 kilograms or more than 162.39 kilograms. The validity and reliability of these predictions are limited to the specified weight ranges.







Summary

Upon examining the residual plots and QQ plots for both male and female bears, it becomes evident that the data does not conform to the key assumptions of linear regression analysis. Specifically, the assumptions of "mean of zero," "constant variance," and "normality" are not met.

Despite the initial impression of a robust linear relationship, as suggested by the scatter plot, the high correlation coefficient, and a substantial coefficient of determination, these observations do not necessarily validate the appropriateness of the regression model in this context.

To address these issues, it is advisable to consider the following steps:

Reassess the Sample: Firstly, it is crucial to thoroughly reevaluate the sample itself. This might involve examining data collection methods, identifying potential outliers, or verifying the representativeness of the sample in relation to the broader population of interest.

Segmentation by Age Group: One potential strategy is to segment the data by age group. Such a stratified analysis can help to reveal if the relationship between the variables under consideration varies across different age groups. This can be particularly informative in cases where the assumption of constant variance is violated.

Explore More Complex Models: If the linear regression model does not fit the data well, it may be worthwhile to explore more complex regression models.

In summary, while the initial analysis indicated a strong linear relationship, it is essential to exercise caution when interpreting regression results, especially when underlying assumptions are not satisfied. By reassessing the data, considering segmentation by age group, or exploring alternative regression models, a more accurate and robust analysis can be conducted to better align with the underlying characteristics of the data.