**Statistical Report Writing Sample No.3.**

**Introduction.** This dataset provides information on body fat, thigh circumference, and mid-arm circumference for twenty healthy females aged 20 to 34. In general, it is expensive and cumbersome to determine the body fat in humans as it involves immersion of the person in water. Thus, it is desirable if a model could provide reliable predictions of the amount of body fat, since the measurements needed for circumference are easy to obtain.

|  |  |  |
| --- | --- | --- |
| thigh | midarm | fat |
| 43.1 | 29.1 | 11.9 |
| 49.8 | 28.2 | 22.8 |
| 51.9 | 37 | 18.7 |
| 54.3 | 31.1 | 20.1 |
| 42.2 | 30.9 | 12.9 |
| 53.9 | 23.7 | 21.7 |
| 58.5 | 27.6 | 27.1 |
| 52.1 | 30.6 | 25.4 |
| 49.9 | 23.2 | 21.3 |
| 53.5 | 24.8 | 19.3 |
| 56.6 | 30 | 25.4 |
| 56.7 | 28.3 | 27.2 |
| 46.5 | 23 | 11.7 |
| 44.2 | 28.6 | 17.8 |
| 42.7 | 21.3 | 12.8 |
| 54.4 | 30.1 | 23.9 |
| 55.3 | 25.7 | 22.6 |
| 58.6 | 24.6 | 25.4 |
| 48.2 | 27.1 | 14.8 |
| 51 | 27.5 | 21.1 |

In the table, the first two columns thigh and midarm show the measurement of thigh and mid-arm circumference. The last column fat measures the actual amount of body fat. Here we present the summary of statistics for each variable.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Mean | S.D | Min | L.Quartile | Median | U.Quartile | Max |
| thigh | 51.17 | 5.234612 | 42.2 | 47.775 | 52 | 54.625 | 58.6 |
| midarm | 27.62 | 3.647147 | 21.3 | 24.75 | 27.9 | 30.025 | 37 |
| fat | 20.195 | 5.106186 | 11.7 | 17.05 | 21.2 | 24.275 | 27.2 |

In this study we want to determine whether regression model could provide reliable predictions of the amount of body fat.

**Data analysis.**  Pairwise scatter plots for all possible combinations of variables are used to see any correlation between them visually. The following figures show the plots for thigh versus fat (the first row left), midarm versus fat (the first row right), and thigh versus midarm (the second row).



These figures suggest that the thigh circumference (thigh) and the amount of body fat (fat) are correlated. In particular, it is important to observe that the thigh circumference (thigh) and the mid-arm circumference (midarm) does not seem correlated (the bottom-left plot). This observation helps us lead the following conclusion:

1. Two potential covariates (thigh and midarm) are not dependent, avoiding “collinearity” (see Section 8.1 of Ekstrom and Sorensen).
2. Since the mid-arm circumference is also “uncorrelated” to the amount of body fat, the thigh circumference may be sufficient to predict the body fat.

We now start with a multiple regression with two covariates, thigh and midarm, to model the response of body fat. And we set a hypothesis test for whether the regression parameter to each covariate is significantly different from zero or not.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Predictor | Estimate | Std.error | test.statistic | p.value |
| (Intercept) | -25.99695 | 6.997321 | -3.715272 | 0.001719844 |
| thigh | 0.8508817 | 0.1124482 | 7.566874 | 7.72E-07 |
| midarm | 0.09602947 | 0.1613927 | 0.5950052 | 0.5596775 |

The results summarized in the table above suggests that the parameter to the covariate midarm is not significantly different from zero, and therefore, that the variable midarm be removed from the model. We fit the reduced model with one covariate “thigh,” and obtain the following table for the simple linear regression.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameter | Estimate | Std.error | test.statistic | p.value |
| Intercept | -23.63449 | 5.657414 | -4.177614 | 5.66E-04 |
| Slope | 0.8565466 | 0.1100156 | 7.785681 | 3.60E-07 |

Note that the parameters for intercept (p-value = 0.00057) and slope (p-value = 0.00000036) are significantly different from zero. The following residual plot was produced for the final model. It does not indicate any systematic patterns nor detect outliers on the residuals.



**Conclusion.** Our analysis indicates that the explanatory variable “thigh” is sufficient to predict the amount of body fat. We have obtained the formula

Fat = −23.63 + (0.85)(thigh)

For example, the measurement of thigh circumference is 55 then the amount of body fat can be predicted by calculating Fat = −23.63 + (0.85)(55) ≈ 23.5.