Name: Rufus Petrie

Final Problem 8: Body Fat Study

Part (a) - First Predictor

From the program, I found abdomen and bodyfat measurements have the greatest correlation.

```
firstReg = Transpose[{abdomen, bf}];
Clear[x]; Remove[f];
lm = LinearModelFit[firstReg, {1, x}, x];
Sqrt[lm["EstimatedVariance"]]
```

```
4.82324
```

Using abdomen measurements to predict bodyfat percentage results in a common standard deviation of 4.823.

Part (b) - Second Predictor

After using abdomen measurements to predict bodyfat percentage, I found that wrist measurements had the next greatest correlation.

```
firstReg = Transpose[{abdomen, wrist, bf}];
Clear[x1, x2]; Remove[f];
lm = LinearModelFit[firstReg, {1, x1, x2}, {x1, x2}];
Sqrt[lm["EstimatedVariance"]]
```

```
4.54239
```

Using both abdomen and wrist measurements to predict bodyfat percentage results in a common standard deviation of 4.542.

Part (c) - Third Predictor

After using abdomen and wrist measurements to predict bodyfat percentage, I found that age measurements had the next greatest correlation.

```
firstReg = Transpose[{abdomen, wrist, age, bf}];
Clear[x1, x2, x3]; Remove[f];
lm = LinearModelFit[firstReg, {1, x1, x2, x3}, {x1, x2, x3}];
Sqrt[lm["EstimatedVariance"]]
```

```
4.38781
```

Using abdomen, wrist, and age metrics to predict bodyfat percentage results in a common standard deviation of 4.388.

Part (d) - ANOVA Table

```
ANOVATable[lm]
lm["RSquared"]
```

	DF	SS	MS	F-Stat	P-Value
Model	3	10122.1	3374.03	175.248	0.
(x1)			9289.45	482.496	9.50277×10^{-55}
(x2)			541.44	28.1225	3.05737×10^{-7}
(x3)			291.189	15.1244	0.00013777
Error	196	3773.57	19.2529		
Total	199	13895.6			

```
0.728435
```

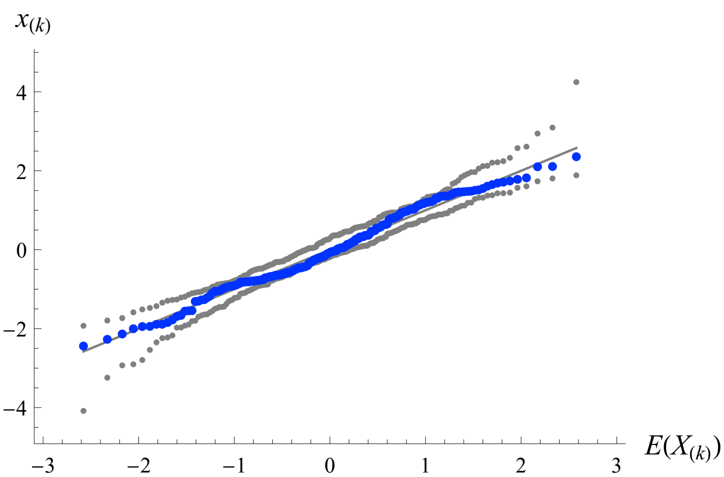
From the results of the ANOVA table, we can see that the measurements for abdomen, wrist, and age all have a statistically significant correlations with bodyfat percentage. Furthermore, because the model has a P-value of zero, we can assume that the entire model is a good predictor of bodyfat percentage for the sample. As we can see from the R-squared value, the model manages to explain about 72.8% of the variation in the data, which indicates that the model performs well as a predictor despite the fact that it only has three explanatory variables.

Part (e) - Predicted Outcome

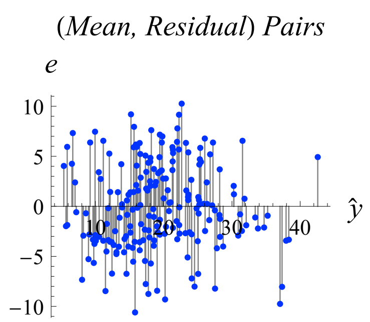
Predictor Values (x1, x2, x3)	*Predicted Body Fat*	*95 % CI for Mean Body Fat*
96.4, 18.2, 35	21.53	[20.64, 22.43]

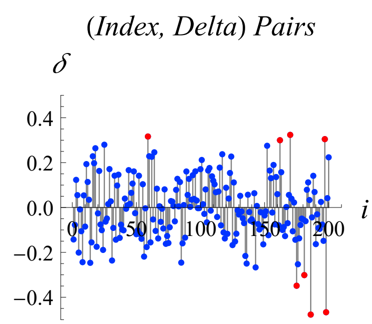
Part (f) - Discussion

```
standardized = lm["StandardizedResiduals"];
QuantileComparisonPlot[NormalDistribution[0, 1], standardized,
 SimulationBands → True]
```




```
DiagnosticSummary[lm]
```






```
Mean-Residual Pairs with Min/Max Residuals:
   Minimum (Case 175): (ŷ,e)=(15.8053,-10.6053)
   Maximum (Case 162): (ŷ,e)=(22.6372,10.2628)

Index-Delta Pairs with |δ|>0.282843:
(59,0.316417)       (162,0.299919)      (170,0.323728)
(175,-0.348784)     (181,-0.30138)      (186,-0.477366)
(197,0.30473)       (198,-0.466885)
```

As we can see from the results of the graphs above, it appears that we can safely assume normally distributed errors for this study. In the first graph, we don't see any significant deviations from the predicted range of residuals. Furthermore, the graph of residuals against predicted y values in the second table appears reveals that the levels of the residuals don't have a clear trend with respect to y. There are some odd looking residuals for high predicted values of y, but this probably stems from the fact that we have relatively few observations where the model would predict very high bodyfat percentages.

By choosing regressors that would reduce the standard errors to the greatest extent, I found that abdomen, wrist, and age measurements had the greatest statistical relationship with bodyfat percentage in that order. By predicting the relationship, I found:

FittedModel[$-12.8385 + 0.777153\,x1 - 2.41496\,x2 + 0.0972699\,x3$]

where x1 corresponds to abdomen measurement, x2 corresponds to wrist measurement, and x3 corresponds to age measurements. Interestingly enough, we observe that abdomen and age measurements have positive correlations with bodyfat percentage while wrist measurements have a negative correlation. This makes sense intuitively because people’s metabolisms naturally slow down as they age in their adult years, which points to a positive correlation between age and bodyfat. Furthermore, much more fat accumulates on the abdomen that on the wrists than joint areas, so the main reason why people may have larger wrists is because they work out. This helps to explain why abdomen measurements have a positive correlation with bodyfat percentage and wrist measurements have a negative correlation.
From the ANOVA table, it appears that we can be fairly confident with out results. From the low model P-value, we see that the entire model does a good job predicting the true values. Furthermore, from the R-squared value of 72.8%, we can see that the model accounts for over two thirds of the entire variation in the sample. If we were to incorporate more of our variables into the model, we could probably increase this number even more.