Rugby Pressure!

By Dr. Wayne Stewart and

Master Reuben-Wayne Stewart (soon to be Professor)

The Data

- This data set was collected as a part of Reuben’s science project. It consists of measurements in inches of the horizontal distance covered when a Rugby ball is hit by a brick. The only variable that was changed was the internal pressure of the Rugby ball. The change in pressure was measured in complete pumps. The pressure of the ball is the independent variable and the length moved by the ball the dependent variable.

How the data was collected

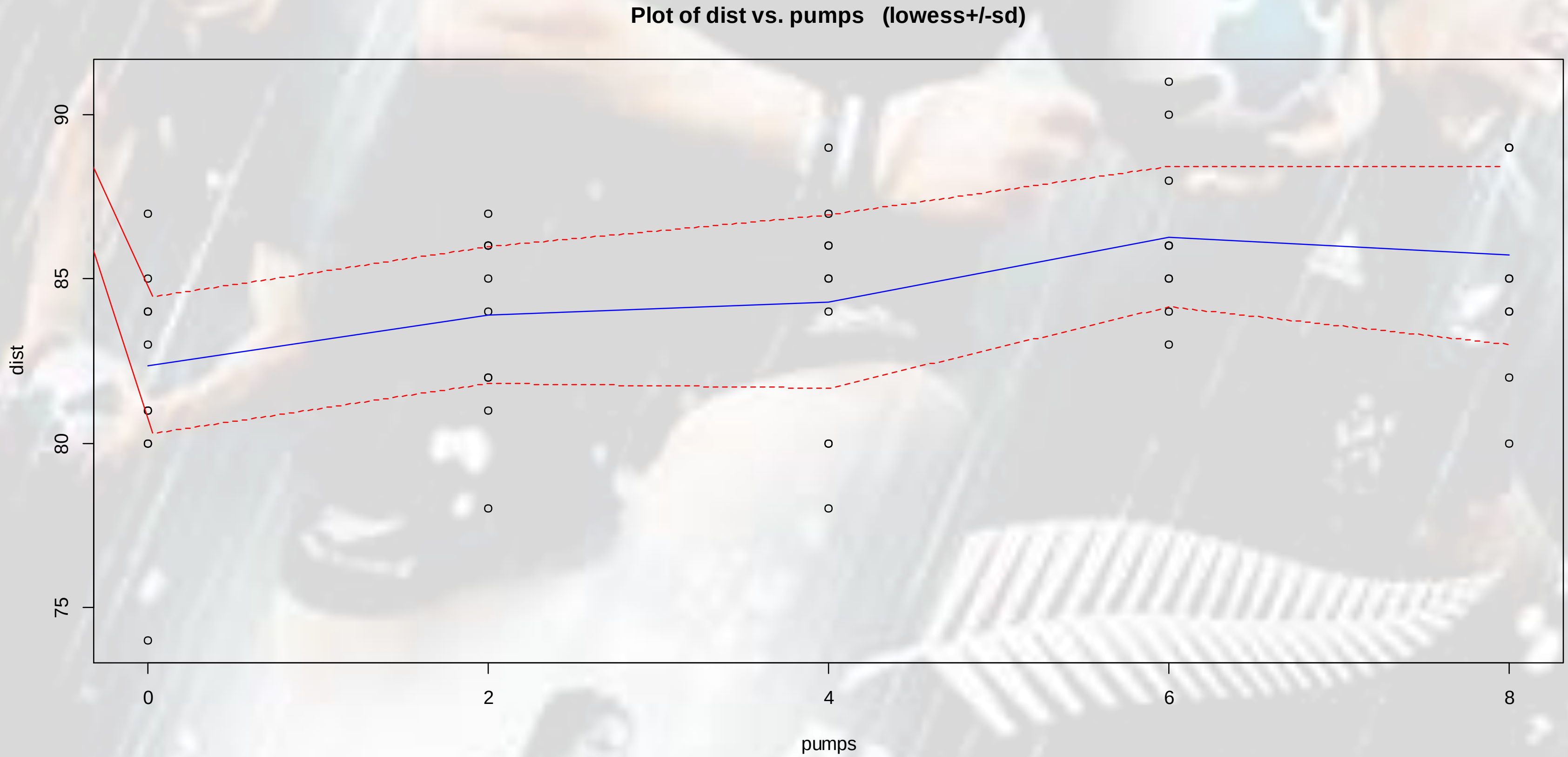
The research question

We were interested in quantifying the change in range for a change in ball pressure given initial conditions.

Assumptions

- The trend is linear.
- The additional assumptions we need to check can be summarized in the following formula
- $\epsilon_i \sim N(0, \sigma^2)$
- This assumes also that the errors are independent.

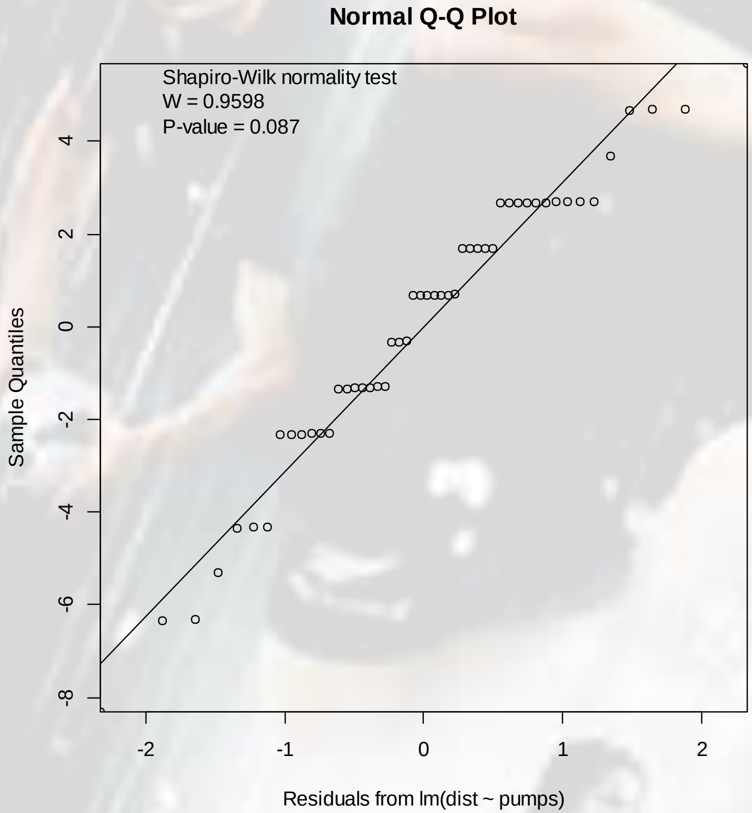
Linear trend



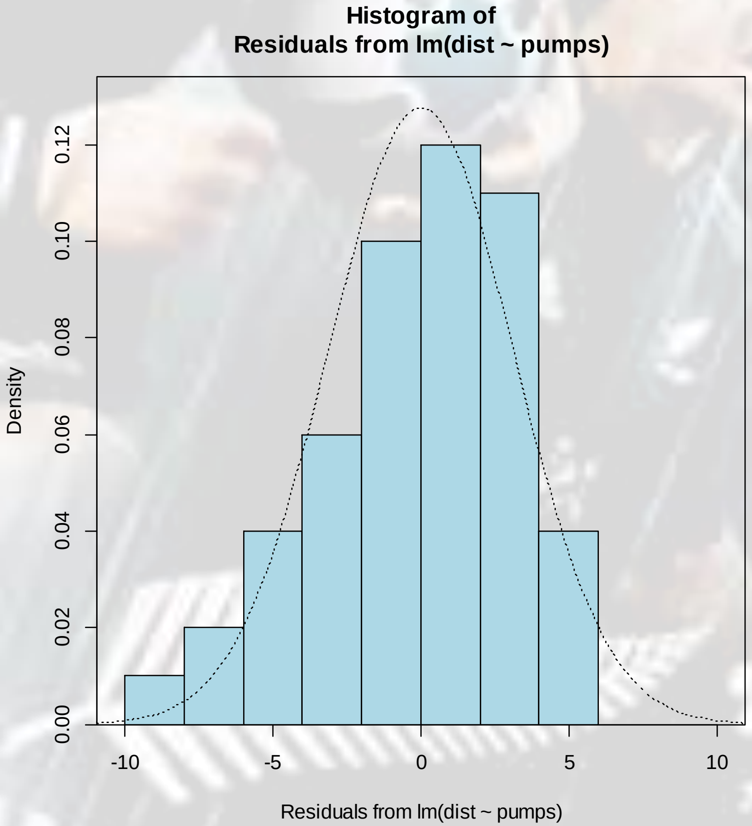
Independence

- The data were collected independently. The outcome of one measurement did not effect the outcome of any after that.

$$N(0, \sigma^2)$$



Normal Q-Q Plot
Shapiro-Wilk normality test
W = 0.9598
P-value = 0.087
Sample Quantiles
-8
-6
-4
-2
0
2
4
-2
-1
0
1
2
Residuals from lm(dist ~ pumps)



Histogram of
Residuals from lm(dist ~ pumps)
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
-10
-5
0
5
10
Residuals from lm(dist ~ pumps)

Normality

The Qqplot supports the contention that the sample comes from a Normal distribution. Though the data set collected is a little left skew as the histogram shows.

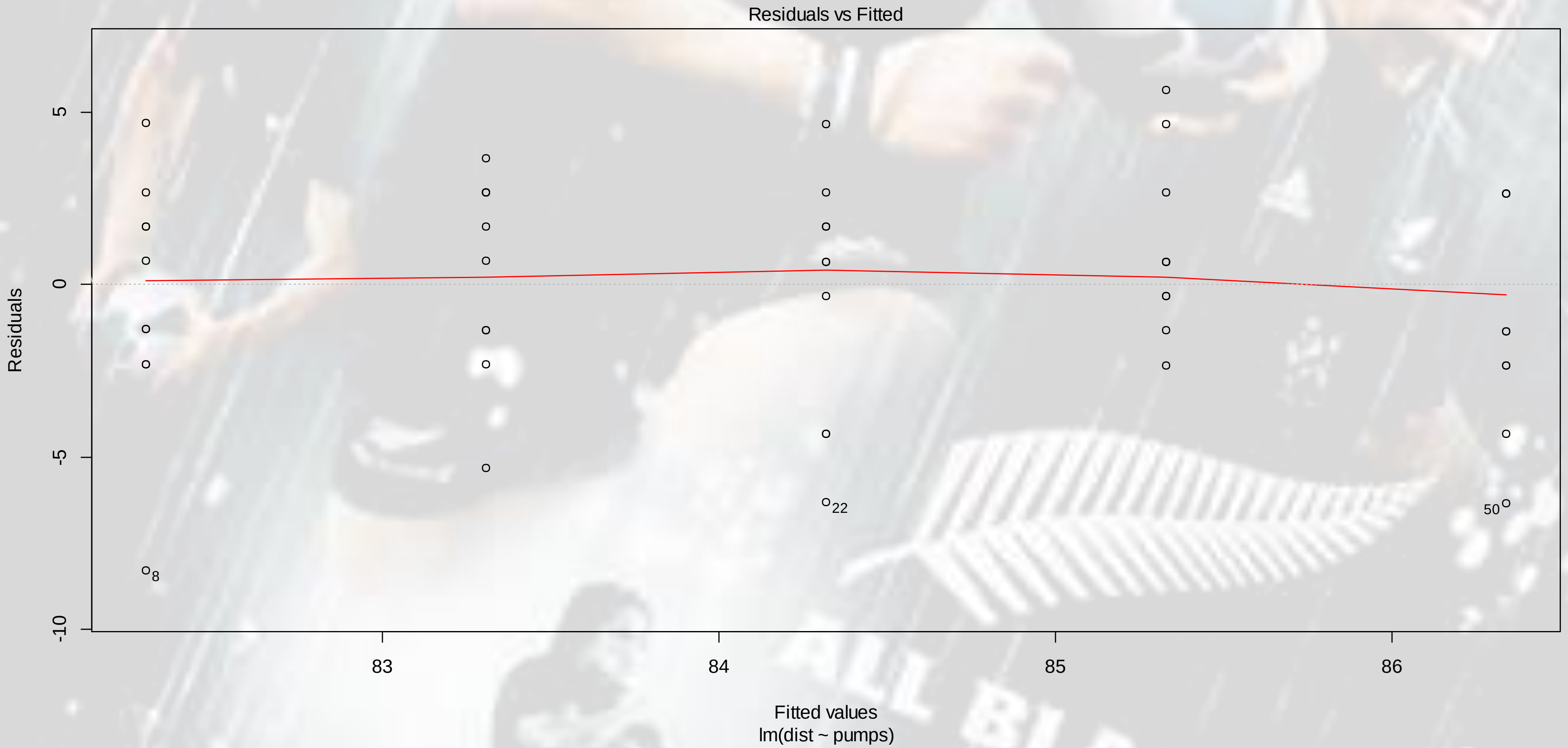
The Shapiro-Wilk test has the NULL hypothesis:

H_0: $\epsilon_i \sim N$ with a P-Value of 0.087 which is not small enough to reject the NULL at the 0.05 level of significance.

Also, since n>30 we could appeal to the CLT to validate the assumption that $\bar{Y} \sim N$

We conclude that the Normality assumption is satisfied.

Constant variance about the line



Residuals

- The residuals are almost constantly distributed about the origin except that last group of the data.
- Our overall impression is that the residuals are consistent with $\epsilon_i \sim N(0, \sigma^2)$

The linear model

We used a SLR model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Which is estimated using

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + r_i$$

Where $\hat{\beta}_i$ is estimated using least squares.

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

How good is the model?

We would not expect the model to predict new dependent values of distance because of the large variability about the regression line.

This is confirmed when we look at the multiple R^2 which has a value of 18%. This means that only 18% of the variation in distance is explained through the pressure variable. The model is a poor predictor.

However, our research question relates only to the slope. And this is answered with our model.

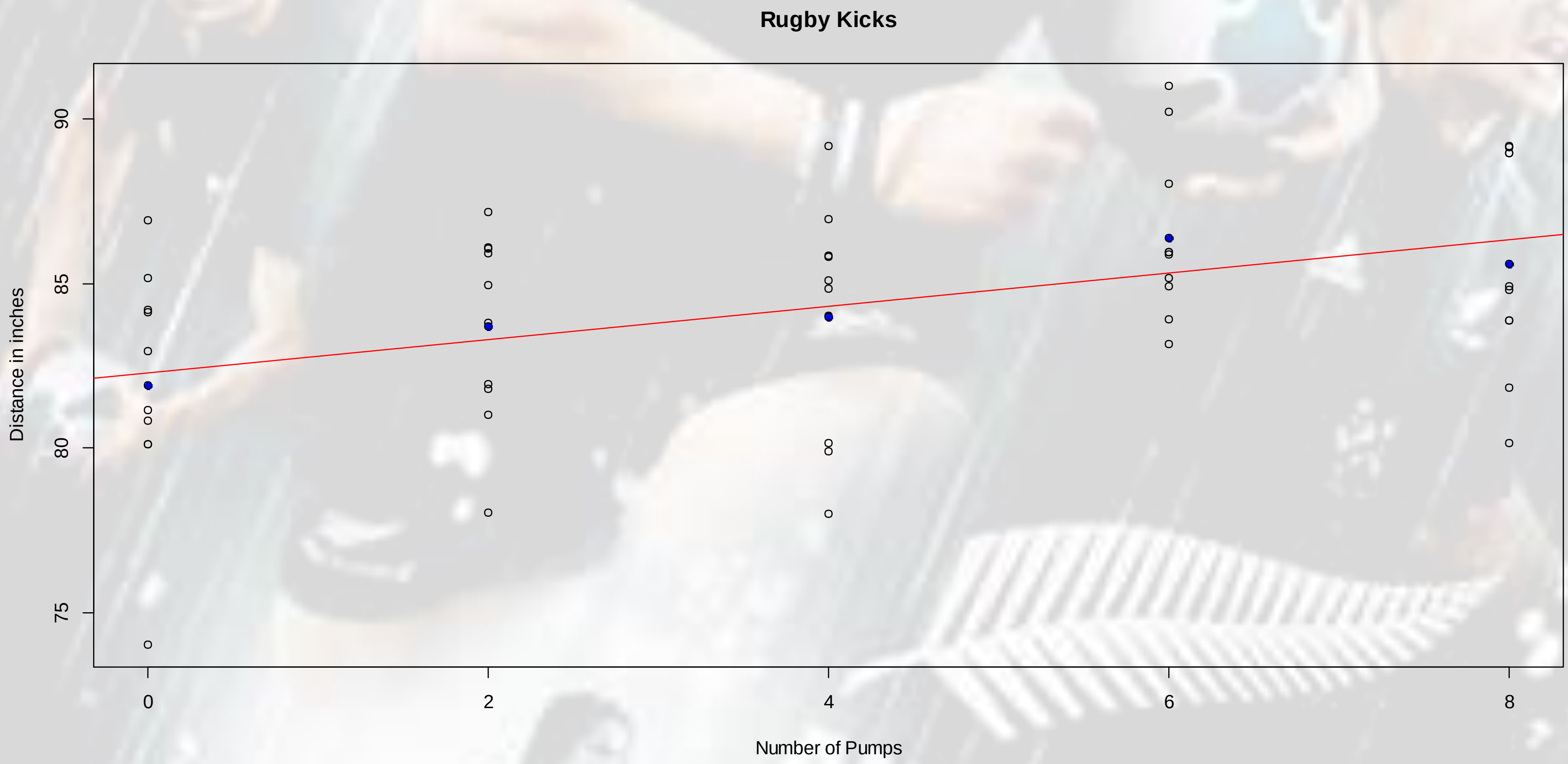
Coefficients and their meaning

```
              Estimate Std.Error t value Pr(>|t|)
(Intercept) 82.3000  0.7737    106.374 < 2e-16 ***
pumps         0.5050  0.1579      3.198 0.00245 **
```

- Both estimates are significantly different from zero and should be included in the model.
- The value we need to answer the research question is $\hat{\beta}_1 = 0.5050 \, \frac{\text{in}}{\text{pump}}$
- The 95% confidence interval estimate is

$$(0.18747, 0.82253)\frac{\text{in}}{\text{pump}}.$$

Least squares regression line



Conclusions

- The model is valid and satisfies the assumptions of the linear model.
- While the data is hugely variable and results in a model that is a poor predictor we can at least ascertain a slope for the mean line.
- We can say with 95% confidence that a change of 1 pump will on average result in an increase in distance of the Rugby ball of between 0.2 and 0.8 inches over the range of the pumps used given the ball pressure at the beginning.
- For each pump the ball would go an extra 0.5 in on average.

Conclusion (cont.)

- The conclusion is only valid for a ball similar to ours with initial conditions the same as our start up values. This is especially true since we did not measure the initial pressure.
- The experiment could be improved by obtaining more accurate distance measurements and using exact pressure measurements over a larger range of pressures, this would distribute the variation across more distances and result in a smaller R^2 assuming the trend could be modelled.