Example 21
Linear Regression
Setup

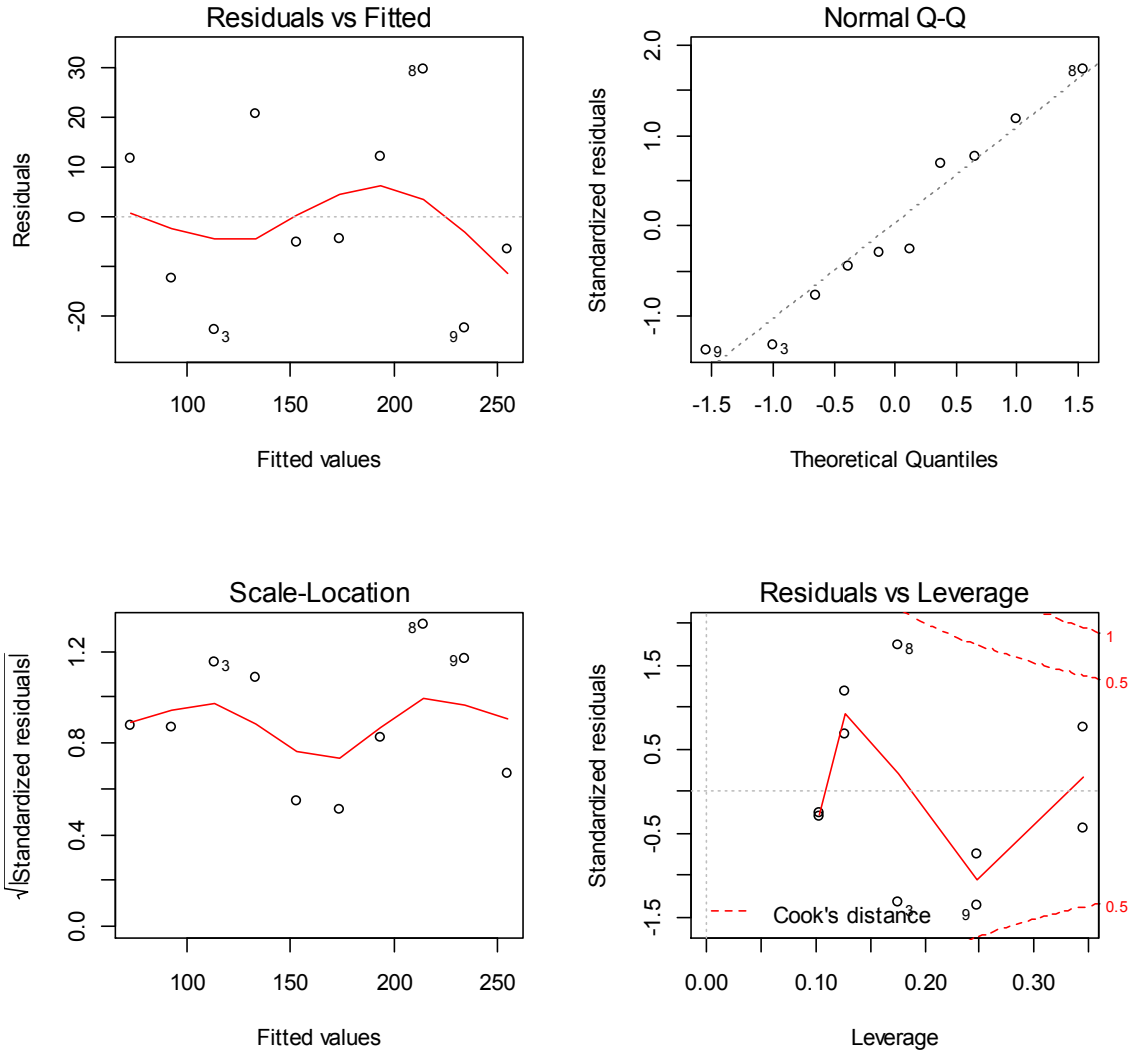
A restoration ecologist is investigating the effect of applying different amounts of fertilizer on reclaimed land. Grass seed is sown uniformly over the area. Ten 1 m^2 plots are located randomly and a different mass of commercial fertilizer (g) is applied evenly to each. Two months later the grass is harvested from each plot, dried, and mass measured (g). Do different levels of fertilizer influence grass growth?

Mass of Fertilizer	Yield of Grass
25	84
50	80
75	90
100	154
125	148
150	169
175	206
200	244
225	212
250	248

Example 21
Linear Regression
Solution

1. State your question: How do different levels of fertilizer influence grass growth.
    a. Is it a good scientific question? Definable, measurable, and controllable.
    b. Identify your population: Growth (mass) of grass
    c. Identify your dependent variable: Growth (mass)
    d. Identify your independent variable: Amount (mass) of fertilizer
2. State your hypothesis set
    a. Verbal hypothesis: The amount of fertilizer applied to the plots will influence the growth of grass.
    b. Statistical hypothesis
    H_0: The amount of fertilizer applied to the plots does not influence the growth of grass.
    H_A: The amount of fertilizer applied to the plots does influence the growth of grass.
3. State your significance level: $\alpha=0.05$
4. Select the appropriate test.
    a. Variable scales
        i. Dependent variable: Ratio
            - Converted or transformed? No
        ii. Independent variable: Ratio
            - Converted or transformed? No
    b. What information is given or available?
        i. Sample data
    c. Number of samples: Many individual measurements
    d. Are the data paired or unpaired? Unpaired
    e. What aspect of the variable do you want to compare?
        i. Causation
    f. State the test to be used: Linear regression
        i. Are the assumptions of the test met? Yes
            - Random samples. – Assumed
            - Independent samples. – Assumed
            - The relationship between the two variables is causal. – Assumed
            - The relationship between the two variables is linear. – graphically assessed (see step 6)
            - The x-variable is under the control of the observer and the X-values are assumed to be exact. – Assumed

- For any value of X there is a theoretical population of y-values that are normally distributed. – visually assessed (see below)
- The variances of Y are equal. – (see below)



5. Conduct your sampling

   10 plots were fertilized with different masses of fertilizer. Equal amounts of grass seed were applied to each plot. The grass was harvested and the biomass of the grass was measured. Data attached.

6. Graph the data

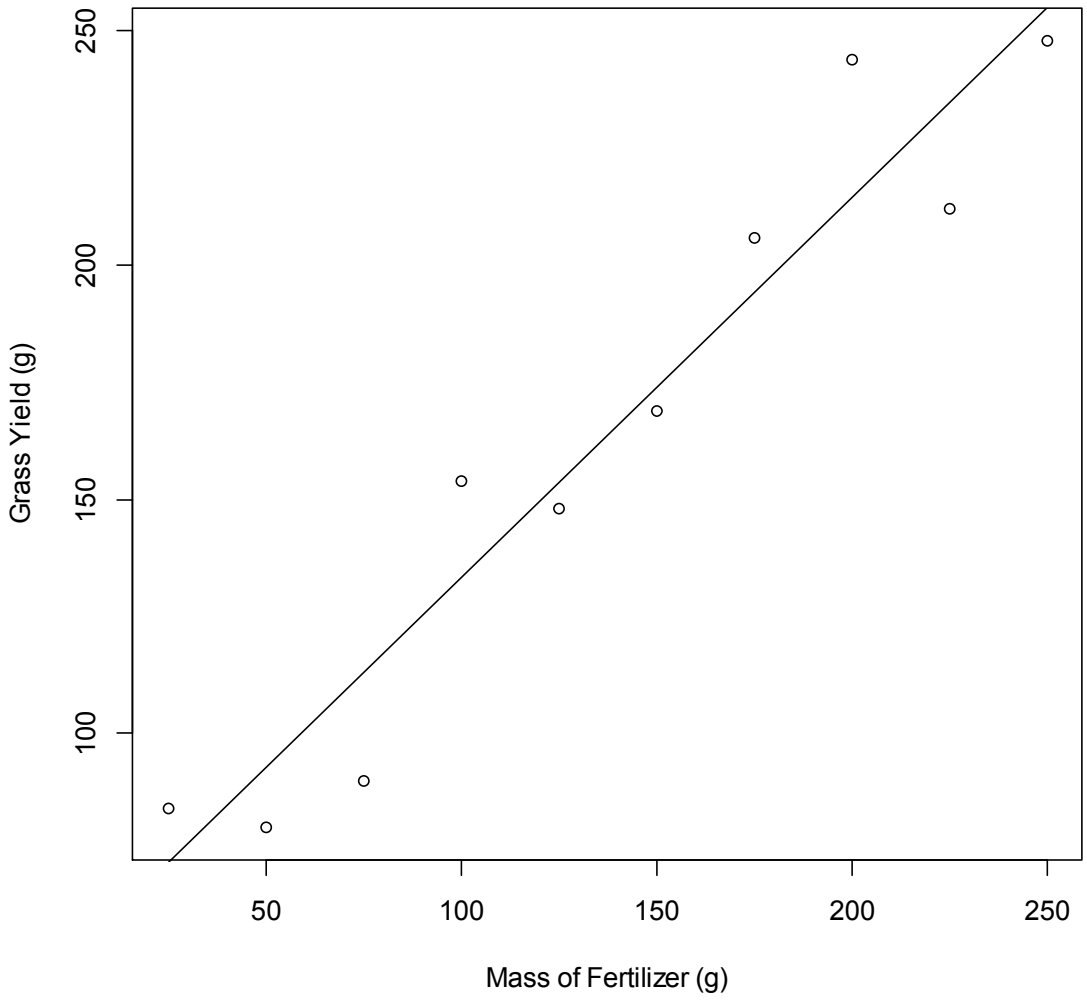


Figure 1. The mass of grass grown (g) on plots treated with different masses of fertilizer (g).

7. Summarize the data

```
Call:
lm(formula = Grass$Yield ~ Grass$Fertilizer)

Coefficients:
    (Intercept)  Grass$Fertilizer
        51.9333            0.8114
```

Linear equation:

$Y = 51.9 + 0.8114 * X$

Mass of Grass = 51.9 + 0.8114 * Mass of Fertilizer

8. Calculate the test statistic

Call:
lm(formula = Grass$Yield ~ Grass$Fertilizer)

Residuals:

Min	1Q	Median	3Q	Max
-22.79	-11.07	-5.00	12.00	29.79

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	51.93333	12.97904	4.001	0.00394 **
Grass$Fertilizer	0.81139	0.08367	9.697	1.07e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19 on 8 degrees of freedom
Multiple R-squared: 0.9216, Adjusted R-squared: 0.9118
F-statistic: 94.04 on 1 and 8 DF, p-value: 1.067e-05

9. Retain or reject your null hypothesis based on your test statistic.
   Since the calculated p-value (1.067e-5) is less than the significance value (0.05), we reject the null hypothesis and retain the alternate hypothesis.
10. Interpret the results in biological terms.
    The mass of fertilizer significantly affects the mass of grass grown (R^2= 0.91, F=94.04, df =1, 8, $p<0.001$).