Example 1: (CANCELLED)

Are the strengths of the Minnesota tornadoes dependent on the locations of occurrence? In this problem, we examine the following measurements for each of the 1,363 tornadoes that had made touchdowns in Minnesota between 1950 and 2006:

- `Strength`: The Fujita scale of the tornado (F0, F1, F2, and F3+.)
- `Location`: The region of Minnesota in which the touchdown occurred (NW = Northwest, N&NE = North Central and Northeast, W = West Central, C = Central, E = East Central, SW = Southwest, S = South Central, and SE = Southeast.)

The joint distribution of `Strength` and `Location` is summarized in the following contingency table:

		Strength				Total
		F0	**F1**	**F2**	**F3+**	
Location	**NW**	136	65	24	4	229
	N&NE	52	38	8	8	106
	W	86	64	26	7	183
	C	128	76	30	9	243
	E	54	42	17	14	127
	SW	62	59	20	9	150
	S	90	61	31	9	191
	SE	77	36	18	3	134
Total		685	441	174	63	1,363

(Source: NOAA/NWS Storm Prediction Center)
Use the following code to load the above contingency table into R:

```
tornado_table <- read.table("http://users.stat.umn.edu/~wuxxx725/tornadoes.txt",
                            header = T, check.names = F)
```

a. State the explanatory and response variables.

b. Use the following R codes to conduct the five-step hypothesis test for the association between the `Strength`s and `Location`s of Minnesota tornadoes at the significance level $\alpha = 0.05$.

```
mytest <- chisq.test(tornado_table)
mytest
mytest$expected
```

c. Use the `pchisq()` function to reproduce the p-value from the test statistic you obtained in part b).

d. Find the estimated risks for a tornado to be F2 or higher in East Central Minnesota (E) and in North central & Northeast Minnesota (N&NE), respectively.

e. Calculate and interpret the relative risk for a tornado to be F2 or higher in East Central Minnesota (E) vs in North central & Northeast Minnesota (N&NE).

Example 2:

The R dataset `trees` in the `datasets` package contains the following measurements of 31 black cherry trees:

- `Girth`: Diameter of the tree at the height of 4 ft 6 in, measured in inches
- `Volume`: Lumber volume, measured in ft^3

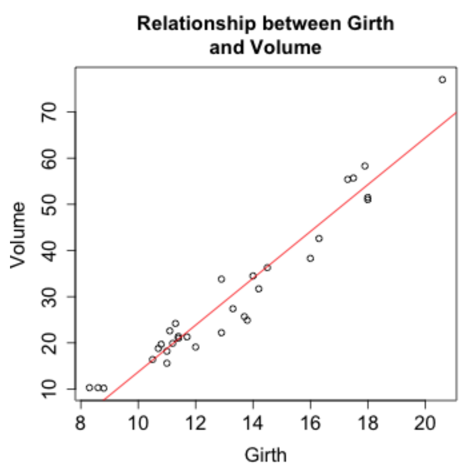
Please use the following command to load the dataset:

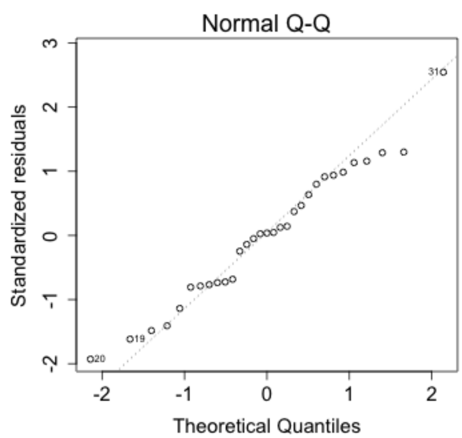
```
attach(trees)
```

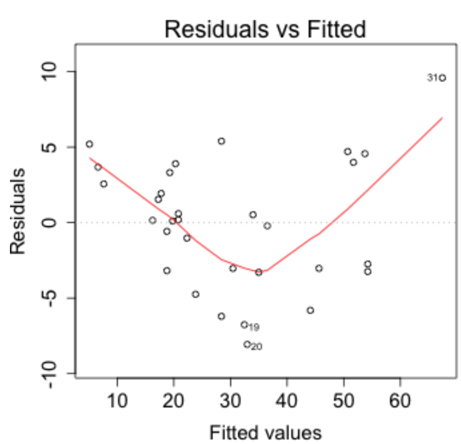
If the dataset fails to load, please copy and paste the following R codes to load the data manually:

```
Girth <- c( 8.3,  8.6,  8.8, 10.5, 10.7, 10.8, 11.0, 11.0, 11.1, 11.2,
           11.3, 11.4, 11.4, 11.7, 12.0, 12.9, 12.9, 13.3, 13.7, 13.8,
           14.0, 14.2, 14.5, 16.0, 16.3, 17.3, 17.5, 17.9, 18.0, 18.0, 20.6)
Volume <- c(10.3, 10.3, 10.2, 16.4, 18.8, 19.7, 15.6, 18.2, 22.6, 19.9,
            24.2, 21.0, 21.4, 21.3, 19.1, 22.2, 33.8, 27.4, 25.7, 24.9,
            34.5, 31.7, 36.3, 38.3, 42.6, 55.4, 55.7, 58.3, 51.5, 51.0, 77.0)
```

In this problem, we consider the linear regression model for `Volume` on `Girth`. The scatterplot for `Volume` vs `Girth`, the normal Q-Q plot for the errors, and the residual plot are given below.







Please answer the following questions:

a. Fit a linear regression model for `Volume` on `Girth` and obtain its summary using the `lm()` and `summary()` functions.

b. State and interpret the value of r^2 from the model summary output in part a).

c. Calculate the correlation r between `Girth` and `Volume`, and state the strength and the direction of the correlation.

d. State the estimated regression equation in the form $V\hat{olume} = a + b(Girth)$.

e. Interpret the slope b.

f. Explain why it does not make sense to interpret the intercept a.

g. Tree #15 has a `Girth` of 12 inches. Predict its lumber `Volume` using the estimated regression equation in part d).

h. The actual lumber volume of tree #15 is 19.1 cubic feet. Find the residual for tree #15.

i. Is it appropriate to use the estimated regression equation in part d) to predict the lumber volume of a black cherry tree with a girth of 25 inches? If so, give the estimated lumber volume. If not, please explain the reason.

j. How would the correlation r change if `Girth` were given in centimeters and `Volume` were given in liters? Please explain. (1 in = 2.54 cm; 1 ft^3 = 28.3 L)

k. Conduct a five-step hypothesis test on whether the true population slope β is different from 0.

l. Check the linearity, normal error, and constant variance assumptions using the diagnostic plots.