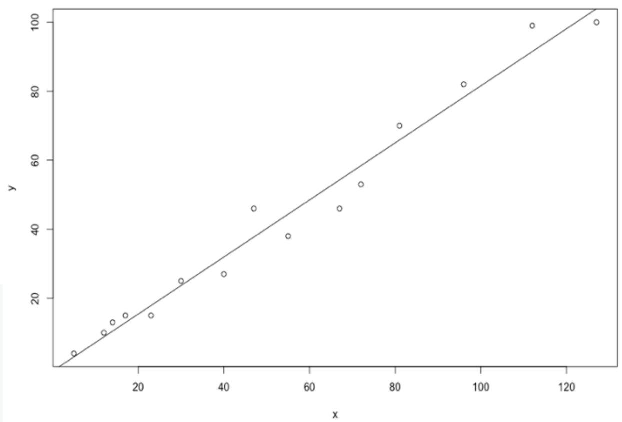
12.2.16 The article **"Characterization of Highway Runoff in Austin, Texas, Area" (*J. of Envir. Engr.*, 1998: 131–137)** gave a scatterplot, along with the least squares line, of x = rainfall volume (m^3) and y = runoff volume (m^3) for a particular location. The accompanying values were read from the plot.

x	5	12	14	17	23	30	40	47	55	67	72	81	96	112	127
y	4	10	13	15	15	25	27	46	38	46	53	70	82	99	100

(a) Does a scatterplot of the data support the use of the simple linear regression model?

```
Console  Terminal
~/
> x<-c(5,12,14,17,23,30,40,47,55,67,72,81,96,112,127)
> y<-c(4,10,13,15,15,25,27,46,38,46,53,70,82,99,100)
> plot(y~x)
> fit<-lm(y~x)
> abline(fit$coefficients)
```



The scatterplot supports the use of the simple linear regression model because the scatterplot does not contain strong curvature and the pattern appears to be roughly linear.

Ans: Yes

(b) Calculate point estimates of the slope and intercept of the population regression line.

```
> fit

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
     -1.128        0.827
```

Ans: $\hat{\boldsymbol{\beta}}_0 = -1.128, \hat{\boldsymbol{\beta}}_1 = 0.827$

(c) Calculate a point estimate of the true average runoff volume when rainfall volume is 50.

```
Console  Terminal
~/
> predict(fit, newdata=data.frame(x=50))
       1
40.22035
```

Ans: $\hat{y} = 40.22035$

(d) Calculate a point estimate of the standard deviation σ.

```
> summary(fit)

Call:
lm(formula = y ~ x)

Residuals:
   Min     1Q Median     3Q    Max 
-8.279 -4.424  1.205  3.145  8.261 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.12830    2.36778  -0.477    0.642    
x            0.82697    0.03652  22.642  7.9e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.24 on 13 degrees of freedom
Multiple R-squared:  0.9753,	Adjusted R-squared:  0.9734 
F-statistic: 512.7 on 1 and 13 DF,  p-value: 7.896e-12
```

Ans: $\boldsymbol{\sigma = 5.24}$

(e) What proportion of the observed variation in runoff volume can be attributed to the simple linear regression relationship between runoff and rainfall?

Ans: $\boldsymbol{r^2 = 0.9753}$

12.3.31 During oil drilling operations, components of the drilling assembly may suffer from sulfide stress cracking. The article **“Composition Optimization of High-Strength Steels for Sulfide Cracking Resistance Improvement” (*Corrosion Science*, 2009: 2878–2884)** reported on a study in which the composition of a standard grade of steel was analyzed. The following data on y = threshold stress (% SMYS) and x = yield strength (MPa) was read from a graph in the article (which also included the equation of the least squares line).

x	635	644	711	708	836	820	810	870	856	923	878	937	948
y	100	93	88	84	77	75	74	63	57	55	47	43	38

$\sum x_i = 10{,}576, \sum y_i = 894, \sum x_i^2 = 8{,}741{,}264, \sum y_i^2 = 66{,}224, \sum x_i y_i = 703{,}192.$

(a) What proportion of observed variation in stress can be attributed to the approximate linear relationship between the two variables?

```
> x<-c(635,644,711,708,836,820,810,870,856,923,878,937,948)
> y<-c(100,93,88,84,77,75,74,63,57,55,47,43,38)
> fit<-lm(y~x)
> summary(fit)

Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max 
-10.4475  -4.3115  -0.1268   4.6093  12.1758 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 211.65503   15.06218  14.052 2.26e-08 ***
x            -0.17563    0.01837  -9.562 1.15e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.806 on 11 degrees of freedom
Multiple R-squared:  0.8926,	Adjusted R-squared:  0.8828 
F-statistic: 91.43 on 1 and 11 DF,  p-value: 1.155e-06
```

Ans: $\boldsymbol{r^2 = 0.8926}$

(b) Compute the estimated standard deviation $s_{\hat{\beta}_1}$.

(see Part (a))

Ans: 0.01837

(c) Calculate a confidence interval using confidence level 95% for the expected change in stress associated with a 1 MPa increase in strength. Does it appear that this true average change has been precisely estimated?

```
> confint(fit, "x")
        2.5 %     97.5 %
x -0.2160637 -0.1352063
```

The true average change appears to have been precisely estimated because the CI is not very wide.

Ans: (–0.2160637, –0.1352063)

12.5.64 The accompanying data on x = UV transparency index and y = maximum prevalence of infection was read from a graph in the article **"Solar Radiation Decreases Parasitism in Daphnia" (*Ecology Letters*, 2012: 47–54):**

x	1.3	1.4	1.5	2.0	2.2	2.7	2.7	2.7	2.8	2.9	3.0	3.6	3.8	3.8	4.6	5.1	5.7
y	16	3	32	1	13	0	8	16	2	1	7	36	25	10	35	58	56

Summary quantities include $S_{xx} = 25.5224, S_{yy} = 5593.0588,$ and $S_{xy} = 264.4882.$

(a) Calculate and interpret the value of the sample correlation coefficient.

```
> x<-c(1.3, 1.4, 1.5, 2.0, 2.2, 2.7, 2.7, 2.7, 2.8, 2.9, 3.0, 3.6, 3.8,
3.8, 4.6, 5.1, 5.7)
> y<-c(16,3,32,1,13,0,8,16,2,1,7,36,25,10,35,58,56)
> cor(x,y)
[1] 0.7000375
```

Ans: r = 0.7000375

(b) If you decided to fit the simple linear regression model to this data, what proportion of observed variation in maximum prevalence could be explained by the model relationship?

```
> fit<-lm(y~x)
> summary(fit)

Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max 
-16.5674 -11.2770   0.1422   9.7429  29.2675 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)  -12.812      8.964  -1.429  0.17343   
x             10.363      2.729   3.797  0.00176 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.79 on 15 degrees of freedom
Multiple R-squared:  0.4901,	Adjusted R-squared:  0.4561 
F-statistic: 14.41 on 1 and 15 DF,  p-value: 0.001755
```

Ans: $r^2 = 0.4901$

(c) If you decided to regress UV transparency index on maximum prevalence (i.e., interchange the roles of x and y), what proportion of observed variation could be attributed to the model relationship?

```
> cor(y,x)
[1] 0.7000375
> fit2<-lm(x~y)
> summary(fit2)

Call:
lm(formula = x ~ y)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.1729 -0.2621  0.1976  0.5457  1.1674 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.15970    0.32506   6.644 7.82e-06 ***
y            0.04729    0.01246   3.797  0.00176 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9315 on 15 degrees of freedom
Multiple R-squared:  0.4901,	Adjusted R-squared:  0.4561 
F-statistic: 14.41 on 1 and 15 DF,  p-value: 0.001755
```

Conclusion: It does not matter how you label variables. The sample correlation coefficient remains unchanged.

Ans: $r^2 = 0.4901$

(d) Carry out a test of $H_0: \rho = .5$ versus $H_a: \rho > .5$ using a significance level of .05. [*Note*: The cited article reported the P-value for testing $H_0: \rho = 0$ versus $H_a: \rho \neq 0$.

```
> (0.5*(log((1+0.7000375)/(1-0.700375)))-0.5*log((1+0.5)/(1-0.5)))/(1/sq
rt(17-3))
[1] 1.192207
> qnorm(.95)
[1] 1.644854
```

Do not reject H_0 because 1.192207 < 1.644854

Ans: Do not reject null hypothesis