11.6 We need to minimize $SSE = \sum_{i=1}^{n}(y_i - \hat{\beta}_1 x_i)^2$

Consider $\frac{\partial SSE}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_1 x_i)x_i = 0$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i = \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 \quad \Rightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

OR: $SSE = \sum_{i=1}^{n}(y_i - \hat{\beta}_1 x_i)^2 = \left(\sum_{i=1}^{n} x_i^2\right)\hat{\beta}_1^2 - 2\left(\sum_{i=1}^{n} x_i y_i\right)\hat{\beta}_1 + \left(\sum_{i=1}^{n} y_i^2\right)$

is minimized at $\hat{\beta}_1 = -\frac{-2\sum_{i=1}^{n} x_i y_i}{2\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$

11.8 a) We calculate

$\sum_{i=1}^{n} x_i = 102$, $\sum_{i=1}^{n} y_i = 64.7$, $\sum_{i=1}^{n} x_i^2 = 3940$,

$\sum_{i=1}^{n} x_i y_i = 894.4$, $n = 5$, $\sum_{i=1}^{n} y_i^2 = 949.99$

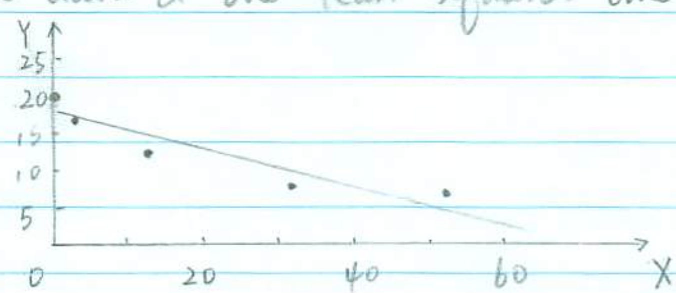
So, $S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) = 894.4 - \frac{1}{5}\cdot 102 \cdot 64.7 = -425.48$

$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 = 3940 - \frac{1}{5}(102)^2 = 1859.2$

$\Rightarrow \hat{\beta}_1 = S_{xy}/S_{xx} = -0.229$

$$\& \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{64.7}{5} - (-.229)\left(\frac{102}{5}\right) = 17.611$$

b) The data & the least squares line are plotted as:



c) The estimate of $E(Y)$ at $x = 20$ is $\hat{y} = 17.611 - .229(20) = 13.031$

In addition, $SSE = S_{yy} - \hat{\beta}_1 S_{xy}$ (refer to the discussion before Example 11.3 on P549)

Here, $S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = 949.99 - \frac{1}{5}(64.7)^2 = 112.772$

$\Rightarrow SEE = 112.772 - (-.229)(-425.48) = 15.34$

Therefore, $S^2 = \frac{SSE}{n-2} = \frac{15.34}{3} = 5.11$

11.22 a) Calculate $\sum x_i = 323.4$, $\sum y_i = 42.6$, $\sum x_i y_i = 2495.08$,

$\sum x_i^2 = 19111.95$, $\sum y_i^2 = 326.06$, $n = 6$

$\therefore S_{xx} = \sum x_i^2 - \frac{1}{n}(\sum x_i)^2 = 19111.95 - \frac{1}{6}(323.4)^2 = 1680.69$

$S_{xy} = \sum x_i y_i - \frac{1}{n}(\sum x_i)(\sum y_i) = 2495.08 - \frac{1}{6}(323.4)(42.6)$

$= 198.94$

$S_{yy} = 326.06 - \frac{1}{6}(42.6)^2 = 23.6$

$\Rightarrow \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{198.94}{1680.69} = 0.1184$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{1}{6}(42.6) - 0.1184 \cdot \frac{323.4}{6} = 0.7182$

b) $SSE = S_{yy} - \hat{\beta}_1 S_{xy} = 23.6 - (.1184)(198.94) = .052$

you may get seemingly different results because of round error

$\Rightarrow S^2 = SSE/n-2 = 0.052/4 = 0.013$

95% CI for β_1 is $\hat{\beta}_1 \pm t_{.025,4} \cdot s\sqrt{C_{11}}$

$= .1184 \pm 2.776\sqrt{0.013}\sqrt{\frac{1}{1680.69}}$

$= .1184 \pm .0077$

c) At $x=0$, $E(Y) = \beta_0 + \beta_1 \cdot 0 = \beta_0$

Thus, the hypothesis to be tested is

$H_0: \beta_0 = 0$ VS. $H_a: \beta_0 \neq 0$

Test statistic is $t = \dfrac{\hat{\beta_0}}{s\sqrt{C_{00}}} = \dfrac{.7182}{\sqrt{.013}\sqrt{\dfrac{1911.95}{6 \times 1680.69}}} = 4.58$

Compare to $t_{.025,4} = 2.776 < 4.58$, we should reject H_0 at significant level $\alpha = .05$.

11.27 Using the coding $x = \dfrac{year - 1971.5}{.5}$, we get the following:

you may get different $\hat{\beta_1}$ because of different choice of x, but SSE should be the same.

$\Sigma x_i = 0$, $\Sigma y_i = 215.9$, $\Sigma x_i y_i = -174.9$

$\Sigma x_i^2 = 330$, $\Sigma y_i^2 = 4760.43$, $n = 10$,

$S_{xy} = -174.9$, $S_{xx} = 330$, $S_{yy} = 99.149$

$\Rightarrow \hat{\beta_1} = S_{xy} / S_{xx} = -174.9 / 330 = -.53$

$SSE = 99.149 - (-0.53)(-174.9) = 6.452$

& $s^2 = SSE / n-2 = 6.452 / 8 = 0.8065$

To test the hypothesis $H_0: \beta_1 = 0$ VS. $H_a: \beta_1 < 0$, we use the test statistic $t = \dfrac{\hat{\beta_1}}{\sqrt{\dfrac{s^2}{S_{xx}}}} = -10.72$

The RR with $\alpha = .05$ is $t < -1.86$ and H_0 is rejected. We conclude that the rate of tuberculosis is decreasing with time

5.

(a)

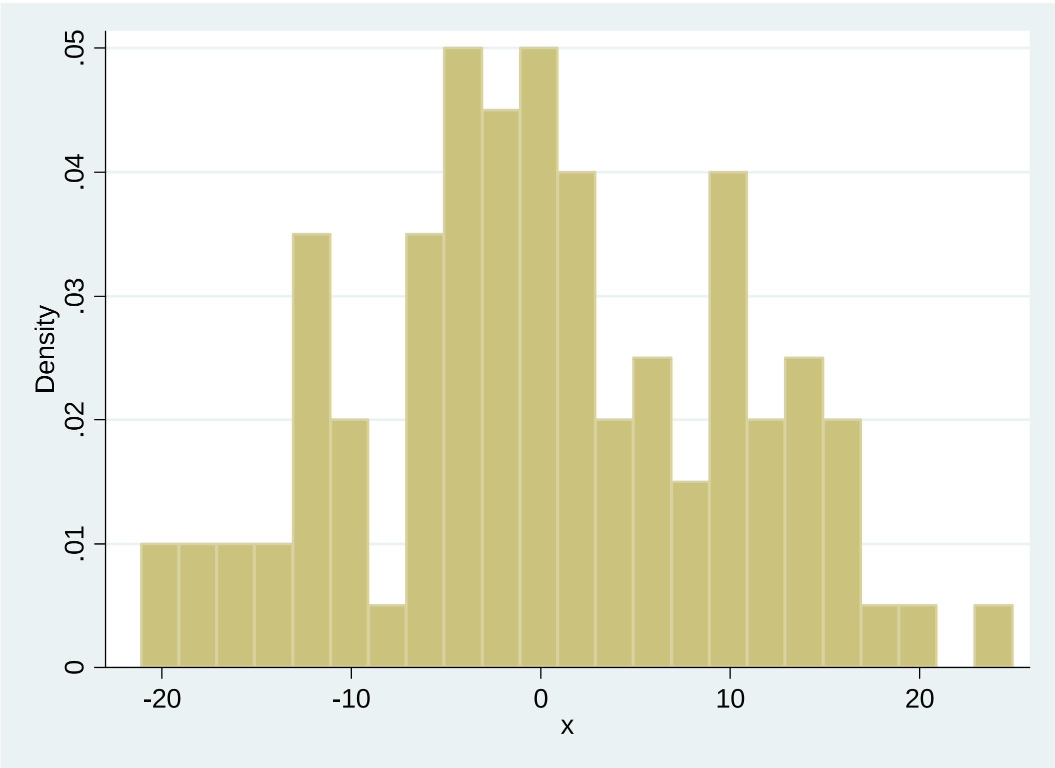
```
. set obs 100
obs was 0, now 100

. gen x0=invnormal(uniform())

. gen x=x0*10
```

(b)

```
. histogram x, width(2)
(bin=23, start=-21.093739, width=2)
```



(c)

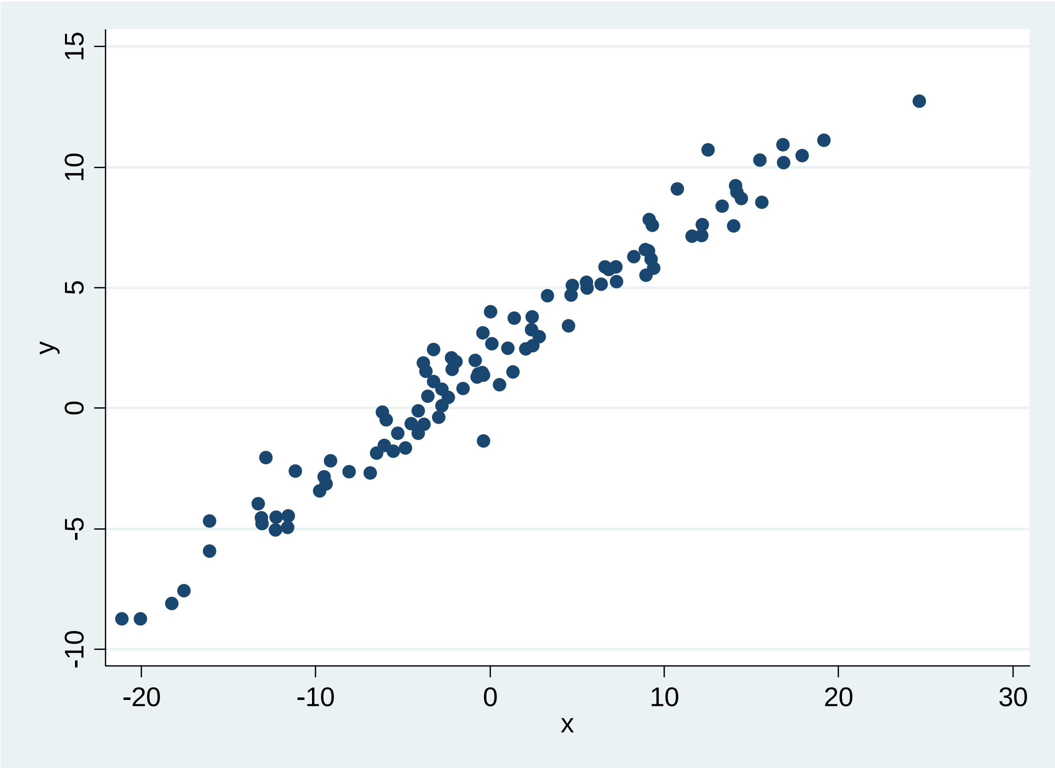
```
. gen z=invnormal(uniform())
```

(f)

```
. gen y=2+0.5*x+z
```

(g)

. scatter y x



(h)

. regress y x

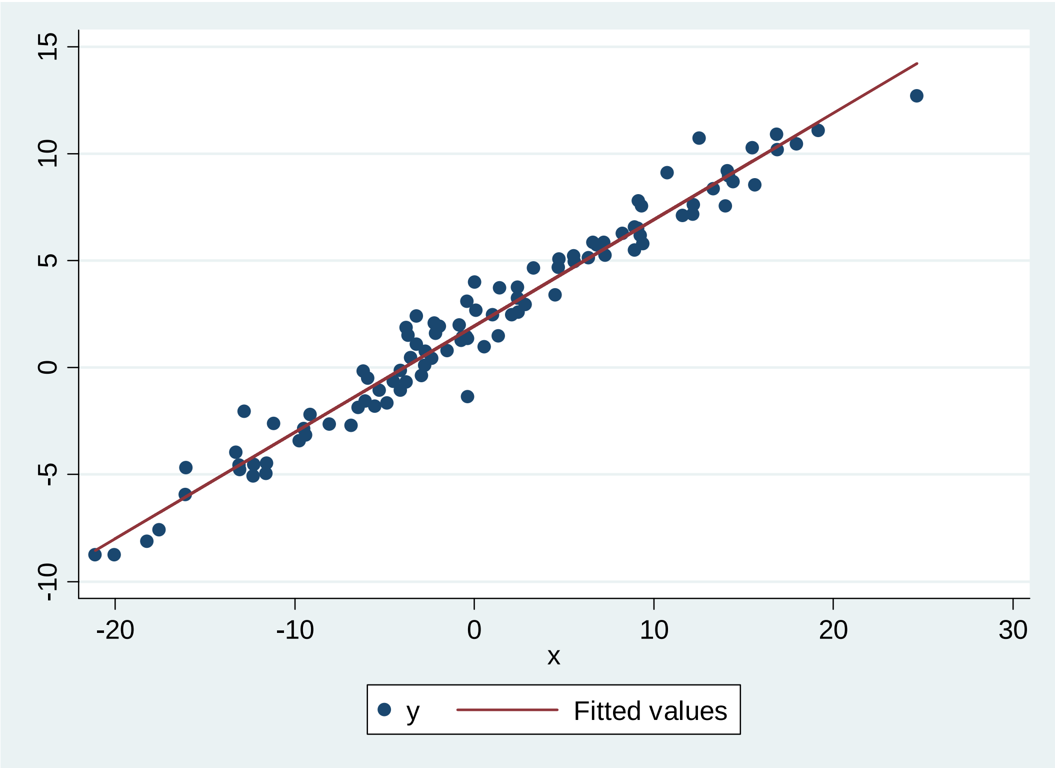
```
      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  1,    98) = 2663.16
       Model |  2315.78895     1  2315.78895           Prob > F      =  0.0000
    Residual |  85.2173022    98  .869564308           R-squared     =  0.9645
-------------+------------------------------           Adj R-squared =  0.9641
       Total |  2401.00626    99  24.2525884           Root MSE      =   .9325

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .4975553   .0096415    51.61   0.000     .4784221    .5166885
       _cons |   1.953626   .0933189    20.93   0.000     1.768437    2.138814
------------------------------------------------------------------------------
```

. predict yhat

(option xb assumed; fitted values)

. scatter y x || line yhat x



The fitted least square regression line is: $\hat{y} = 1.954 + 0.498x$.

The estimates are: $\hat{\beta}_0 = 1.954, \hat{\beta}_1 = 0.498, SSE = 85.217$, and the 95% CI for $\hat{\beta}_0$ is [1.768, 2.139], for $\hat{\beta}_1$ is [0.478, 0.517].