Practice Problems 1a

1. Suppose that the amount of beer a person has consumed (X) and the person's blood alcohol level (Y) in a certain bar follow a bivariate normal distribution with

$$\mu_X = 60 \text{ oz}, \quad \sigma_X = 15 \text{ oz}, \quad \mu_Y = 0.072, \quad \sigma_Y = 0.02, \quad \rho = 0.60.$$

a) What is the probability that the person's blood alcohol level is above 0.08? That is, find $P(Y > 0.08)$.

b) Given that someone has consumed 45 oz of beer, what is the probability that the person's blood alcohol level is above 0.08? That is, find $P(Y > 0.08 \mid X = 45)$.

c) Given that someone has consumed 80 oz of beer, what is the probability that the person's blood alcohol level is above 0.08? That is, find $P(Y > 0.08 \mid X = 80)$.

d) Given that someone has blood alcohol level 0.10, what is the probability that this person has consumed over 90 oz of beer? That is, find $P(X > 90 \mid Y = 0.10)$.

2. **<u>Do NOT</u>** use a computer for this problem.

A student is studying for Exam P (the first Actuarial Science exam). He claims that drinking beer has no effect on the amount of time it takes for him to solve a practice problem. The following data show how many seconds he took to solve a problem after consuming various quantities of beer, measured in ounces:

Beer Consumption, x	Time, y
0	141
12	127
24	141
36	163
48	145
60	179
72	161

$$\sum x = 252, \quad \sum y = 1{,}057, \quad \sum x^2 = 13{,}104, \quad \sum y^2 = 161{,}447, \quad \sum xy = 40{,}068,$$

$$\sum (x - \bar{x})^2 = 4032, \quad \sum (y - \bar{y})^2 = 1{,}840, \quad \sum (x - \bar{x})(y - \bar{y}) = \sum (x - \bar{x}) y = 2{,}016.$$

Consider the model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where ε_i's are i.i.d. $N(0, \sigma^2)$.

a) Find the equation of the least-squares regression line.

b) Calculate the residuals e_i. Does the sum of the residuals equal zero?

c) Give an estimate for σ, the standard deviation of the observations about the true regression line?

d) What proportion of observed variation in time needed to solve a practice problem is explained by a straight-line relationship with the amount of beer consumed?

e) How much time would you expect the student to need to solve a practice problem after consuming 156 ounces of beer.

f) Explain why it may be dangerous to predict the time needed to solve a practice problem for the amount of beer consumed in part (e).

g) Use the F-test to test $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$ (the significance of the regression test) at a 5% significance level. Report the value of the test statistic, critical value(s), and decision.

h) Construct a 95% confidence interval for β_1. Does your answer for part (h) agree with your answer for part (g)?

i) Test the student's claim at a 5% level of significance. That is, test $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 > 0$ at a 5% level of significance. Report the value of the test statistic, critical value(s), and decision. Does the square of your test statistic in part (i) equal your test statistic from part (g)?

j) The student believes that when he is not drinking beer, it takes him on average two minutes to solve a practice problem. What is the p-value (approximately) of the test $H_0: \beta_0 = 120$ vs. $H_1: \beta_0 \neq 120$?

k) Construct a 95% confidence interval for β_0.

l) Construct a 95% confidence interval for σ^2.

m) Construct 95% limits of prediction for the time the student needs to solve a practice problem after consuming 156 ounces of beer.

n) Construct 90% confidence interval for the average time the student needs to solve a practice problem after consuming 156 ounces of beer.

3. **Do** use a computer to double check **2** (g), (h), (k), (m), (n). Please include a printout and mark (circle or highlight) the test statistic and the p-value for part (g), and the four intervals for parts (h), (k), (m), (n).

4. The more beer you drink, the more your blood alcohol level (BAL) rises. The data below shows the number of 12 oz beers consumed (x) and the blood alcohol level (y) for 20 individuals at a local bar.

12 oz beers, x	4	6	3	4	7	5	4	2	4	2
BAL (%), y	0.07	0.12	0.04	0.08	0.14	0.11	0.09	0.03	0.09	0.03

12 oz beers, x	3	4	6	4	5	3	3	6	2	3
BAL (%), y	0.03	0.05	0.11	0.08	0.09	0.03	0.06	0.09	0.02	0.04

Consider the model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where ε_i's are i.i.d. $N(0, \sigma^2)$.

a) Make a scatterplot with [0, 8] range on the x-axis and [0, 0.15] range on the y-axis. Add the least-squares regression line and 95% confidence intervals around the regression line to the scatterplot.

b) Test the hypothesis that one 12 oz beer raises your BAL by 0.02% (on average) against the alternative that it raises it more. Use a 5% level of significance. Find the p-value for this test. That is, test $H_0 : \beta_1 = 0.02$ vs. $H_1 : \beta_1 > 0.02$.

c) Test the hypothesis that the y-intercept is 0 against the two-sided alternative. Use a 5% level of significance. Find the p-value for this test.

d) Construct a 90% prediction interval for the blood alcohol level for an individual who consumed 8 (eight) 12 oz beers.

Answers:

1. Suppose that the amount of beer a person has consumed (X) and the person's blood alcohol level (Y) in a certain bar follow a bivariate normal distribution with

$$\mu_X = 60 \text{ oz}, \qquad \sigma_X = 15 \text{ oz}, \qquad \mu_Y = 0.072, \qquad \sigma_Y = 0.02, \qquad \rho = 0.60.$$

a) What is the probability that the person's blood alcohol level is above 0.08? That is, find $P(Y > 0.08)$.

Y has Normal distribution with mean $\mu_Y = 0.072$ and standard deviation $\sigma_Y = 0.02$.

$$P(Y > 0.08) = P\left(Z > \frac{0.08 - 0.072}{0.02}\right) = P(Z > 0.40) = \mathbf{0.3446}.$$

b) Given that someone has consumed 45 oz of beer, what is the probability that the person's blood alcohol level is above 0.08? That is, find $P(Y > 0.08 \mid X = 45)$.

Given X = 45, Y has Normal distribution

with mean $0.072 + 0.6 \cdot \frac{0.02}{15} \cdot (45 - 60) = 0.06$

and variance $\left(1 - 0.6^2\right) \cdot 0.02^2 = 0.000256$ (standard deviation 0.016).

$$P(Y > 0.08 \mid X = 45) = P\left(Z > \frac{0.08 - 0.06}{0.016}\right) = P(Z > 1.25) = \mathbf{0.1056}.$$

c) Given that someone has consumed 80 oz of beer, what is the probability that the person's blood alcohol level is above 0.08? That is, find $P(Y > 0.08 \mid X = 80)$.

Given X = 80, Y has Normal distribution

with mean $0.072 + 0.6 \cdot \frac{0.02}{15} \cdot (80 - 60) = 0.088$

and variance $\left(1 - 0.6^2\right) \cdot 0.02^2 = 0.000256$ (standard deviation 0.016).

$$P(Y > 0.08 \mid X = 80) = P\left(Z > \frac{0.08 - 0.088}{0.016}\right) = P(Z > -0.50) = \mathbf{0.6915}.$$

d) Given that someone has blood alcohol level 0.10, what is the probability that this person has consumed over 90 oz of beer? That is, find $P(X > 90 \mid Y = 0.10)$.

Given Y = 0.10, X has Normal distribution

with mean $60 + 0.6 \cdot \dfrac{15}{0.02} \cdot (0.10 - 0.072) = 72.6$

and variance $\left(1 - 0.6^2\right) \cdot 15^2 = 144$ (standard deviation 12).

$$P(X > 90 \mid Y = 0.10) = P\left(Z > \frac{90 - 72.6}{12}\right) = P(Z > 1.45) = \mathbf{0.0735}.$$

2. **<u>Do NOT</u>** use a computer for this problem.

A student is studying for Exam P (the first Actuarial Science exam). He claims that drinking beer has no effect on the amount of time it takes for him to solve a practice problem. The following data show how many seconds he took to solve a problem after consuming various quantities of beer, measured in ounces:

Beer Consumption, x	Time, y
0	141
12	127
24	141
36	163
48	145
60	179
72	161

$\sum x = 252$, $\sum y = 1{,}057$, $\sum x^2 = 13{,}104$, $\sum y^2 = 161{,}447$, $\sum xy = 40{,}068$,

$\sum (x - \bar{x})^2 = 4032$, $\sum (y - \bar{y})^2 = 1{,}840$, $\sum (x - \bar{x})(y - \bar{y}) = \sum (x - \bar{x})\, y = 2{,}016$.

Consider the model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where ε_i's are i.i.d. $N(0, \sigma^2)$.

a) Find the equation of the least-squares regression line.

$$\bar{x} = \frac{\sum x}{n} = \frac{252}{7} = 36. \qquad \bar{y} = \frac{\sum y}{n} = \frac{1057}{7} = 151.$$

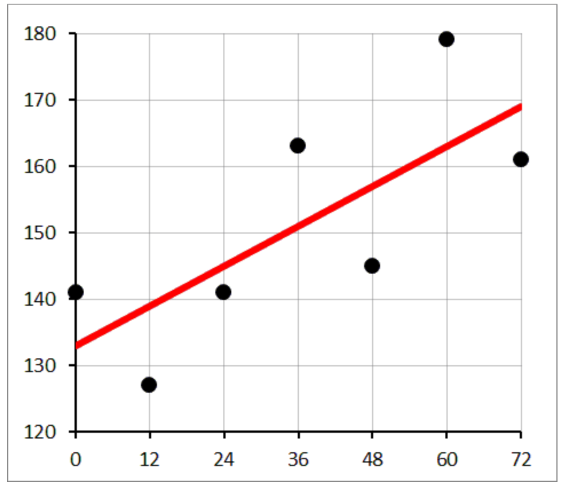
SXX = 4032,   SXY = 2016.

$$\hat{\beta}_1 = \frac{\text{SXY}}{\text{SXX}} = \frac{2016}{4032} = \mathbf{0.5}.$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x} = 151 - 0.5 \cdot 36 = \mathbf{133}.$$

Least-squares regression line:

$$\hat{y} = \mathbf{133 + 0.5}\, \boldsymbol{x}.$$



b) Calculate the residuals e_i. Does the sum of the residuals equal zero?

x	y	$\hat{y}$	$e = y - \hat{y}$	e^2
0	141	133	**8**	64
12	127	139	**– 12**	144
24	141	145	**– 4**	16
36	163	151	**12**	144
48	145	157	**– 12**	144
60	179	163	**16**	256
72	161	169	**– 8**	64
Sum:			0	832

SSResid

The sum of the residuals does equal zero.

c) Give an estimate for σ, the standard deviation of the observations about the true regression line?

$$s_e^2 = \frac{1}{n-2}\sum\left(y_i - \hat{y}_i\right)^2 = \frac{\text{SSResid}}{n-2} = \frac{832}{5} = 166.4. \qquad s_e = \sqrt{166.4} = \mathbf{12.9}.$$

OR

$$\hat{\sigma}^2 = \frac{1}{n}\sum\left(y_i - \hat{y}_i\right)^2 = \frac{\text{SSResid}}{n} = \frac{832}{7} = 118.857. \qquad \hat{\sigma} = \sqrt{118.857} = \mathbf{10.9}.$$

d) What proportion of observed variation in time needed to solve a practice problem is explained by a straight-line relationship with the amount of beer consumed?

Need $R^2 = ?$ (coefficient of determination)

$$R^2 = 1 - \frac{\text{SSResid}}{\text{SYY}} = 1 - \frac{832}{1840} = \mathbf{0.547826}. \qquad \mathbf{54.7826\%}.$$

e) How much time would you expect the student to need to solve a practice problem after consuming 156 ounces of beer.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 133 + 0.5 \cdot 156 = \mathbf{211} \text{ sec.}$$

f) Explain why it may be dangerous to predict the time needed to solve a practice problem for the amount of beer consumed in part (e).

Extrapolation.

$x = 156$ oz is outside of the observed data range. We do not have any guarantee that the relationship would stay the same for larger values of x.

g) Use the F-test to test $\text{H}_0 : \beta_1 = 0$ vs. $\text{H}_1 : \beta_1 \neq 0$ (the significance of the regression test) at a 5% significance level. Report the value of the test statistic, critical value(s), and decision.

SSRegr = SYY – RSS = 1840 – 832 = 1008.

OR $\qquad$ $\text{SSRegr} = \hat{\beta}_1^2 \sum (x_i - \bar{x})^2 = \hat{\beta}_1^2 \text{SXX} = 0.5^2 \times 4032 = 1008.$

Source	SS	DF	MS	F
Regression	1008	1	1008	**6.0577**
Residuals	832	$n - 2 = 5$	166.4	
Total	1840	$n - 1 = 6$		

Rejection Region:

Reject H_0 if $F > F_{0.05}(1,5)$ $\qquad$ $F_{0.05}(1,5) = \mathbf{6.61}$.

The Test Statistic F is NOT in the Rejection Region.

Do NOT Reject $\mathbf{H_0}$ at $\alpha = 0.05$.

h) Construct a 95% confidence interval for β_1. Does your answer for part (h) agree with your answer for part (g)?

$$\hat{\beta}_1 \pm t_{\alpha/2} \cdot \frac{s_e}{\sqrt{SXX}} \qquad 7 - 2 = 5 \text{ degrees of freedom,} \qquad t_{0.025} = 2.571.$$

$$0.5 \pm 2.571 \cdot \frac{12.9}{\sqrt{4032}} \qquad \mathbf{0.5 \pm 0.5223} \qquad \mathbf{(-0.0223,\ 1.0223)}$$

95% confidence interval for β_1 covers 0 $\Leftrightarrow$ Do NOT Reject $H_0 : \beta_1 = 0$ at $\alpha = 0.05$.

i) Test the student's claim at a 5% level of significance. That is, test $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 > 0$ at a 5% level of significance. Report the value of the test statistic, critical value(s), and decision. Does the square of your test statistic in part (i) equal your test statistic from part (g)?

$$T = \frac{\hat{\beta}_1 - \beta_{10}}{s_e / \sqrt{SXX}} = \frac{0.5 - 0}{12.9 / \sqrt{4032}} = \mathbf{2.46124}. \qquad n - 2 = 5 \text{ degrees of freedom.}$$

Rejection Region: $T > t_{\alpha}(n-2)$. $\qquad$ $\alpha = 0.05$ $\qquad$ $t_{0.05}(5) = \mathbf{2.015}$.

The Test Statistic T IS in the Rejection Region. $\qquad$ **Reject $\mathbf{H_0}$** at $\alpha = 0.05$.

p-value = right tail. $\qquad$ $0.025 <$ p-value < 0.05.

(p-value ≈ 0.02857.)

Indeed, $T^2 = F$.

j) The student believes that when he is not drinking beer, it takes him on average two minutes to solve a practice problem. What is the p-value (approximately) of the test $H_0 : \beta_0 = 120$ vs. $H_1 : \beta_0 \neq 120$?

$$T = \frac{\hat{\beta}_0 - \beta_{00}}{s_e \cdot \sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{\text{SXX}}}} = \frac{133 - 120}{12.9 \cdot \sqrt{\frac{1}{7} + \frac{(36)^2}{4032}}} = 1.479.$$

$n - 2 = 5$ degrees of freedom.

p-value = 2 tails. $0.05 <$ left tail < 0.10. right tail ≈ 0.10.

0.10 < p-value < **0.20**. p-value ≈ **0.20**.

(p-value ≈ 0.1992.)

k) Construct a 95% confidence interval for β_0.

$$\hat{\beta}_0 \pm t_{\alpha/2} \cdot s_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SXX}}$$

5 degrees of freedom, $t_{0.025} = 2.571$.

$$133 \pm 2.571 \cdot 12.9 \cdot \sqrt{\frac{1}{7} + \frac{36^2}{4032}}$$

133 ± 22.6

(110.4 , 155.6)

l) Construct a 95% confidence interval for σ^2.

$$\left(\frac{(n-2)s_e^2}{\chi^2_{\alpha/2}}, \frac{(n-2)s_e^2}{\chi^2_{1-\alpha/2}} \right)$$

$\chi^2_{0.025}(5 \text{ df}) = 12.83$,

$\chi^2_{0.975}(5 \text{ df}) = 0.831$.

$$\left(\frac{5 \cdot 166.4}{12.83}, \frac{5 \cdot 166.4}{0.831} \right)$$

(64.848 , 1001.203)

m) Construct 95% limits of prediction for the time the student needs to solve a practice problem after consuming 156 ounces of beer.

Limits of prediction for y: $\hat{y} \pm t_{\alpha/2} \cdot s_e \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\text{SXX}}}$

$\hat{y} = \hat{\alpha} + \hat{\beta}\, x = 133 + 0.5 \cdot 156 = 211$ sec.

$\alpha = 0.05$. $\alpha/2 = 0.025$. $n - 2 = 5$ degrees of freedom. $t_{0.025}(5) = 2.571$.

$211 \pm 2.571 \cdot 12.9 \cdot \sqrt{1 + \frac{1}{7} + \frac{(156-36)^2}{4032}}$ **211 ± 72** **$(139, 283)$**

n) Construct 90% confidence interval for the average time the student needs to solve a practice problem after consuming 156 ounces of beer.

Confidence interval for $\mu(x)$: $\hat{y} \pm t_{\alpha/2} \cdot s_e \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\text{SXX}}}$

$\hat{y} = \hat{\alpha} + \hat{\beta}\, x = 133 + 0.5 \cdot 156 = 211$ sec.

$\alpha = 0.10$. $\alpha/2 = 0.05$. $n - 2 = 5$ degrees of freedom. $t_{0.05}(5) = 2.015$.

$211 \pm 2.015 \cdot 12.9 \cdot \sqrt{\frac{1}{7} + \frac{(156-36)^2}{4032}}$ **211 ± 50.1** **$(160.9, 261.1)$**

3. <u>**Do**</u> use a computer to double check **2** (g), (h), (k), (m), (n). Please include a printout and mark (circle or highlight) the test statistic and the p-value for part (g), and the four intervals for parts (h), (k), (m), (n).

```
x <- c(  0, 12, 24, 36, 48, 60, 72)
y <- c(141,127,141,163,145,179,161)

fit <- lm(y ~ x)

summary(fit)

Call:
lm(formula = y ~ x)

Residuals:
  1   2   3   4   5   6   7
  8 -12  -4  12 -12  16  -8

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 133.0000     8.7896  15.132 2.28e-05 ***
x             0.5000     0.2031   2.461   0.0571 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.9 on 5 degrees of freedom
Multiple R-squared: 0.5478,     Adjusted R-squared: 0.4574
F-statistic: 6.058 on 1 and 5 DF,  p-value: 0.05714                  (g)


confint(fit, "x", level=0.95)
          2.5 %     97.5 %
x   -0.02221319   1.022213                                           (h)

confint(fit, "(Intercept)", level=0.95)
                 2.5 %     97.5 %
(Intercept)   110.4056   155.5944                                    (k)

predict.lm(fit,data.frame(x=156),interval=c("prediction"),level=0.95)
    fit        lwr        upr
1   211   139.0027   282.9973                                        (m)

predict.lm(fit,data.frame(x=156),interval=c("confidence"),level=0.90)
    fit        lwr        upr
1   211   160.9044   261.0956                                        (n)
```

4. The more beer you drink, the more your blood alcohol level (BAL) rises. The data below shows the number of 12 oz beers consumed (x) and the blood alcohol level (y) for 20 individuals at a local bar.

12 oz beers, x	4	6	3	4	7	5	4	2	4	2
BAL (%), y	0.07	0.12	0.04	0.08	0.14	0.11	0.09	0.03	0.09	0.03

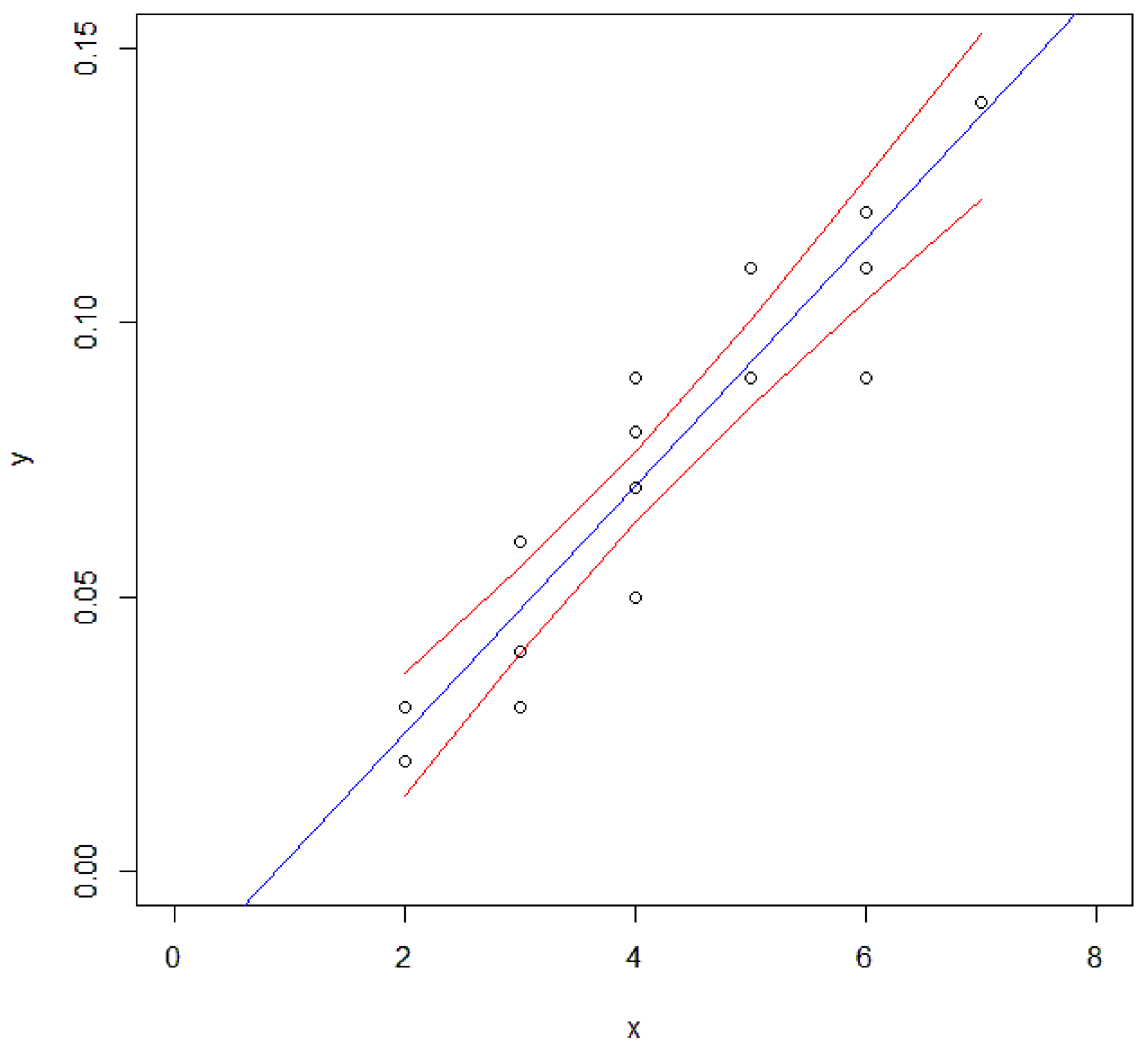
12 oz beers, x	3	4	6	4	5	3	3	6	2	3
BAL (%), y	0.03	0.05	0.11	0.08	0.09	0.03	0.06	0.09	0.02	0.04

Consider the model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where ε_i's are i.i.d. $N(0, \sigma^2)$.

a) Make a scatterplot with [0, 8] range on the x-axis and [0, 0.15] range on the y-axis. Add the least-squares regression line and 95% confidence intervals around the regression line to the scatterplot.

```
x <- c(   4,   6,   3,   4,   7,   5,   4,   2,   4,   2,
          3,   4,   6,   4,   5,   3,   3,   6,   2,   3)
y <- c(0.07,0.12,0.04,0.08,0.14,0.11,0.09,0.03,0.09,0.03,
       0.03,0.05,0.11,0.08,0.09,0.03,0.06,0.09,0.02,0.04)

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

Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max
-0.02500 -0.00750  0.00125  0.01000  0.02000

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.020000   0.009152  -2.185   0.0423 *
x            0.022500   0.002157  10.431 4.65e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01364 on 18 degrees of freedom
Multiple R-squared: 0.8581,     Adjusted R-squared: 0.8502
F-statistic: 108.8 on 1 and 18 DF,  p-value: 4.647e-09
```

```
plot(x,y, xlim=c(0,8), ylim=c(0,0.15))
abline(fit$coefficients, col="blue")

xx <- 2:7
yy <- predict.lm(fit, data.frame(x=xx),
                  interval=c("confidence"), level=0.95)
lines(xx,yy[,2], col="red")
lines(xx,yy[,3], col="red")
```



b) Test the hypothesis that one 12 oz beer raises your BAL by 0.02% (on average) against the alternative that it raises it more. Use a 5% level of significance. Find the p-value for this test. That is, test $H_0: \beta_1 = 0.02$ vs. $H_1: \beta_1 > 0.02$.

$H_0: \beta_1 = 0.02$ vs. $H_1: \beta_1 > 0.02$

```
s2 <- sum(fit$residuals^2)/18
s2
[1] 0.0001861111
SXX <- sum((x-mean(x))^2)
SXX
[1] 40
t = (0.0225-0.02) / (sqrt(s2)/sqrt(SXX))
t
[1] 1.159001
1-pt(t,18)
[1] 0.1308005
```

p-value = 0.1308 > 0.05 = α **Do NOT Reject H_0** at $\alpha = 0.05$.

c) Test the hypothesis that the *y*-intercept is 0 against the two-sided alternative. Use a 5% level of significance. Find the p-value for this test.

$H_0: \beta_0 = 0$ vs. $H_1: \beta_0 \neq 0$

```
(Intercept) -0.020000   0.009152  -2.185   0.0423 *
```

p-value = 0.0423 < 0.05 = α **Reject H_0** at $\alpha = 0.05$.

d) Construct a 90% prediction interval for the blood alcohol level for an individual who consumed 8 (eight) 12 oz beers.

```
predict.lm(fit, data.frame(x=8), interval=c("prediction"),
            level=0.90)
     fit        lwr         upr
1   0.16   0.1315138   0.1884862
```