Narges Goudarzi

250993028

Narges Goudarzi 250993028

1) $\bar{X} = \sum_{i=1}^{9} x_i / 9 = 14.50111$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \Rightarrow S = \sqrt{\frac{\sum (x_i - 14.50111)}{8}} = 2.07370$$

$$\text{T-value} = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = -2.16842$$

using t-table, p-value is obtained as 0.03099 which is less than 0.05. we have evidence to reject null hypothesis $\mu_0 = 16$ so we can reject it. $H_A: \mu < 16$ can be accepted.

2) a) $$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{(-2022)}{102} = -19.82$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 x = -90 + 19.82(5) = 9.1$$

b) $y = 9.1 - 19.82(x)$ at $x = 3$ $y = -50.36$

c) $$\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 47.13$$

$$\sum (y_i - \hat{y}_i)^2 = 47.13$$

$$\sigma^2 = \sum \frac{e_i^2}{n-2} = \frac{47.13}{8} = 5.89$$

d) 95% confidence interval for $E[Y/X=3]$

$\hat{y} = -50.36 \quad \bar{Y} = -90$

$$\hat{Y} \sim N\left(\beta_0 + \beta_1 x,\ \sigma^2\left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right)\right)$$

variation for $\hat{Y}$ is calcutated here:

$$\sigma^2 = 5.89\left[\frac{1}{10} + \frac{(3-5)^2}{102}\right] = 0.82 \quad \Rightarrow \sigma = 0.9$$

Using Z-table 95% CI $Z = 0.975$ is 1.96

Y is in the interval $-50.36 \pm 1.96\,(0.9)$

$$-52.124 < Y < 48.596$$

3. a) we want to show that $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ passes through $(\bar{X}, \bar{Y})$

$x_i = \bar{x}$ at the model $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ will give us:

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$, knowing that $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \Rightarrow$

$\hat{y}_i = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x}$, so $\hat{y}_i = \bar{y}$ Finally we found that putting $X_i = \bar{X}$ will give us $\hat{y}_i = \bar{y}$

b) show that $SST = SSE + SSR$

$SST = \sum (y_i - \bar{y})^2 \quad SSE = \sum (y_i - \hat{y}_i)^2 \quad SSR = \sum (\hat{y}_i - \bar{y})^2$

Let's begin by SST

$$SST = \sum (y_i - \bar{y})^2 = \sum \Big(\underbrace{y_i - \hat{y}}_{A} + \underbrace{\hat{y} - \bar{y}}_{B} \Big)^2$$

we know that $\sum (A+B)^2 = \sum A^2 + \sum B^2 + 2\sum AB$

$$SST = \sum (y_i - \hat{y})^2 + \sum (\hat{y}_i - \bar{y})^2 + 2\cancelto{0}{\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})}$$

why this part is zero? $\left[\sum \hat{e}_i \hat{y}_i = 0 \; , \; \sum y_i = \sum \hat{y}_i \right]$

$$\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum (y_i - \hat{y}_i)\hat{y}_i - \bar{y} \sum (y_i - \hat{y}_i)$$

$$= \sum \hat{e}_i \hat{y}_i - \bar{y} \sum y_i - \bar{y} \sum \hat{y}_i = 0$$

So $SST = SSR + SSE$

4-

a) > length(which(x1 > 6))
[1] 26

b) > sum(with(hw1_data, x1>6 & x2=="H"))
[1] 23

c) > newdata <- subset(hw1_data,x1&x2=="H")
> mean(newdata$x1)
[1] 5.832919

> sd(newdata$x1)
[1] 1.790704

d) Conducting hypothesis test $\begin{cases} H_0 : \mu = 4 \\ H_1 : \mu \neq 4 \end{cases}$. $\overline{X}$ = 4.435 , s = 2.52925 .

t value = 1.7191, p-value = 0.0887. p_value > 0.05, we can't reject H0.

```
> t_stat = (mean(x1) - 4)/(sd(x1)/sqrt(100))
> t_stat # test statistic
[1] 1.719151
```

```
>      A = 1 - pt(abs(t_stat),df=99) # p_value
>      p_val = 2*A
>      p_val
[1] 0.08871225
```

e) Conducting hypothesis test $\begin{cases} H_0 : \mu = 4 \\ H_1 : \mu > 4 \end{cases}$. $\overline{X}$ = 5.832919 , s = 1.790704.

t value= 7.7278, p-value = 0.000 . p_value < 0.05, we can't reject H0.

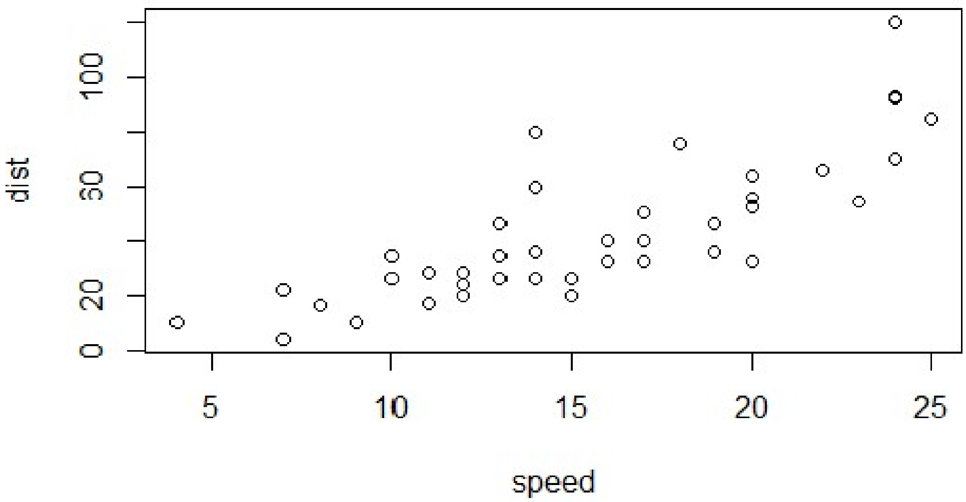
```
> t.test((newdata$x1),mu=4,alternative="greater")

        One Sample t-test
```

```
data:  (newdata$x1)
t = 7.7278, df = 56, p-value = 1.086e-10
alternative hypothesis: true mean is greater than 4
95 percent confidence interval:
 5.436223      Inf
sample estimates:
mean of x
 5.832919
```

5-

a) It seems there is a linear relationship between speed and dist.

```
> plot(dist~speed, data=cars2)
```



b) $\hat{\beta}_0 = -17.237,\ \hat{\beta}_1 = 3.882,\ \hat{\sigma}_2 = \dfrac{15.85}{38} = 0.4171$

```
> fit<-lm(dist~speed,data=cars2)
  fit

Coefficients:
(Intercept)       speed
   -17.237        3.882
```

```
> summary(fit)

Call:
lm(formula = dist ~ speed, data = cars2)

Residuals:
    Min      1Q  Median      3Q     Max
-28.403  -8.904  -3.285   6.818  44.069
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.2369     7.7336  -2.229   0.0318 *
speed         3.8820     0.4698   8.264 5.15e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.85 on 38 degrees of freedom
Multiple R-squared:  0.6425,	Adjusted R-squared:  0.6331
F-statistic: 68.29 on 1 and 38 DF,  p-value: 5.152e-10
```

c)

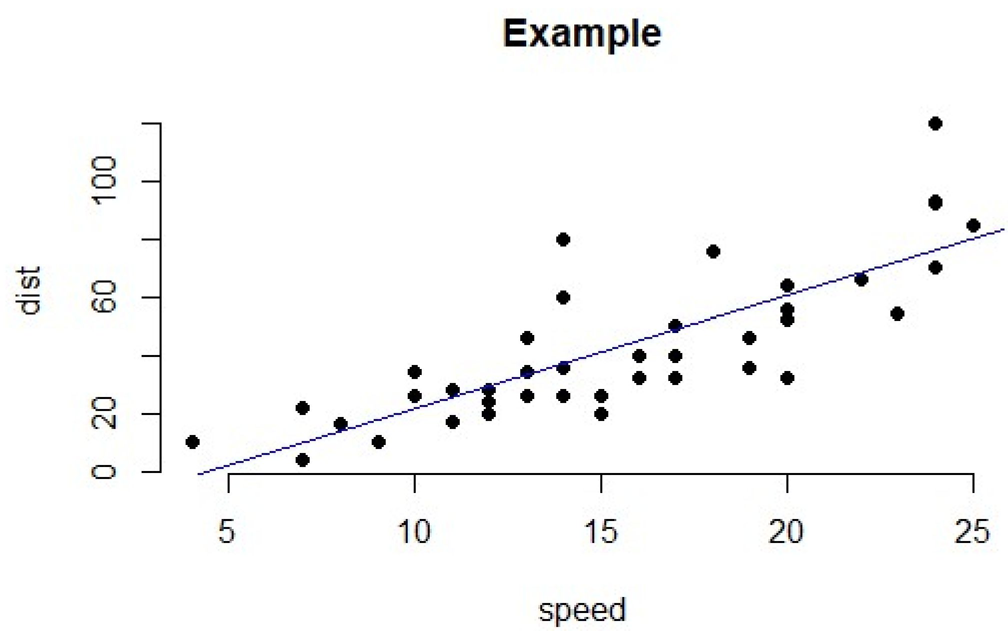
```
> resid(fit)

> cars2$resid=resid(fit)

cars2[4,c(3)]
[1] -10.5208

> cars2[7,c(3)]
[1] 1.243172

> cars2[10,c(3)]
[1] 2.181033
```

e) It is close to zero

```
> sum(resid(fit))
[1] -2.953193e-14
```

f)

```
> plot(speed, dist, main = "Example",
    xlab = "speed", ylab = "dist",
    pch = 19, frame = FALSE)
abline(lm(dist ~ speed, data = cars2), col = "blue")
```




```
> predict(fit,newdata=data.frame(speed=17),interval="prediction",level=0.95)
       fit      lwr      upr
1   48.75683 16.25272 81.26094
```

g) It has calculated in section b.

Multiple R-squared: 0.6425

h) From section (b) we can put $\hat{\beta}_0 = -17.237,\ \hat{\beta}_1 = 3.882, x = 100$ in the regression model. $y = \hat{\beta}_0 + \hat{\beta}_1 x_i = -17.237 - 3.882(100) = 370.963$ this value is distance taken t o stop.

i) > confint(fit,level=0.90)

```
                 5 %      95 %
(Intercept) -30.275356 -4.198463
speed         3.089992  4.673977
```

$$P(3.089992 < \beta_1 < 4.673977) = 0.90$$

j) > predict(fit,newdata=data.frame(speed=15),interval="confidence",level=0.95)

fit lwr upr

40.99286 35.89159 46.09413

$$35.89159 < \beta_1 < 46.09413$$

.