Verzani Problem Set

Next are considered the problems from Verzani's book on page 83.

Problem 13.1

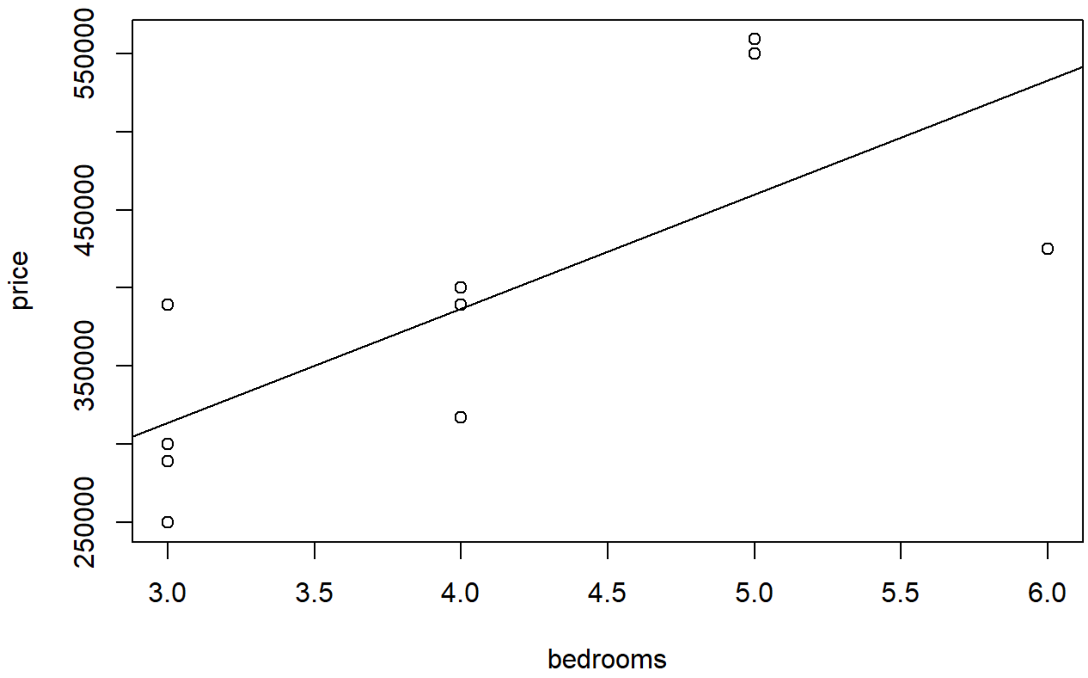
The cost of a home depends on the number of bedrooms in the house. Suppose the following data is recorded for homes in a given town

price (in thousands)	300	250	400	550	317	389	425	289	389	559
Number of bedrooms	3	3	4	5	4	3	6	3	4	5

Make a scatterplot, and fit the data with a regression line. On the same graph, test the hypothesis that an extra bedroom costs $60,000 against the alternative that it costs more.

```
> price <- c(300000, 250000, 400000, 550000, 317000, 389000, 425000, 289000, 389000, 559000)
> bedrooms <- c(3, 3, 4, 5, 4, 3, 6, 3, 4, 5)
> lmResult <- lm(price ~ bedrooms)
> summary(lmResult)

Call:
lm(formula = price ~ bedrooms)

Residuals:
    Min      1Q  Median      3Q     Max 
-108000  -53950   -5750   59775   99100 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)    94400      97983   0.963   0.3635  
bedrooms       73100      23764   3.076   0.0152 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 75150 on 8 degrees of freedom
Multiple R-squared:  0.5419,	Adjusted R-squared:  0.4846 
F-statistic: 9.462 on 1 and 8 DF,  p-value: 0.01521
```

```
> plot(bedrooms, price)
> abline(lmResult)
```



$H_0 : \beta_1 = 60000$
$H_1 : \beta_1 > 60000$

$$S_\varepsilon^2 = \frac{\sum_{i=1}^n \varepsilon_i^2}{n-2}$$

$$SE(\beta_1) = \frac{S_\varepsilon}{\sqrt{\sum_{i=1}^n (X_i - \overline{X}_n)^2}}$$

$$\left(\frac{\hat{\beta}_1 - b_1}{Se(\beta_1)} \,|\, H_0 \right) \in t(n-2)$$

Given α the critical area is

$$W_\alpha = \left\{ \frac{\hat{\beta}_1 - b_1}{SE(\beta_1)} \geq t_{1-\alpha;n-2} \right\}$$

```
> e <- resid(lmResult); e
     1      2      3      4      5      6      7      8      9     10
-13700 -63700  13200  90100 -69800  75300 -108000 -24700   2200  99100
```

```
> n <- length(e)
> Seps <- sqrt(sum(e^2) / (n - 2)); Seps
[1] 75149.43
> beta1hat <- coef(lmResult)[['bedrooms']]; beta1hat
[1] 73100
> SEbeta1 <- Seps / sqrt(sum((bedrooms - mean(bedrooms))^2)); SEbeta1
[1] 23764.34
> b1 <- 60000
> t <- (beta1hat - b1) / SEbeta1; t
[1] 0.5512462
> pvalue <- pt(t, n - 2, lower.tail = FALSE);pvalue
[1] 0.2982602
```

The $p-value = 0.2982602 > 0.05 = \alpha$, so we have no evidence to reject H_0. An extra bedroom costs $\$60{,}000$.

Some of the values, for example S_ε, $\hat{\beta}_1$, $SS(\beta_1)$ could be taken from the output of summary(lmResult).

Problem 13.2

It is well known that the more beer you drink, the more your blood alcohol level rises. Suppose we have the following data on student beer consumption

Student	1	2	3	4	5	6	7	8	9	10
Beers	5	2	9	8	3	7	3	5	3	5
BAL	0.10	0.03	0.19	0.12	0.04	0.095	0.07	0.06	0.02	0.05

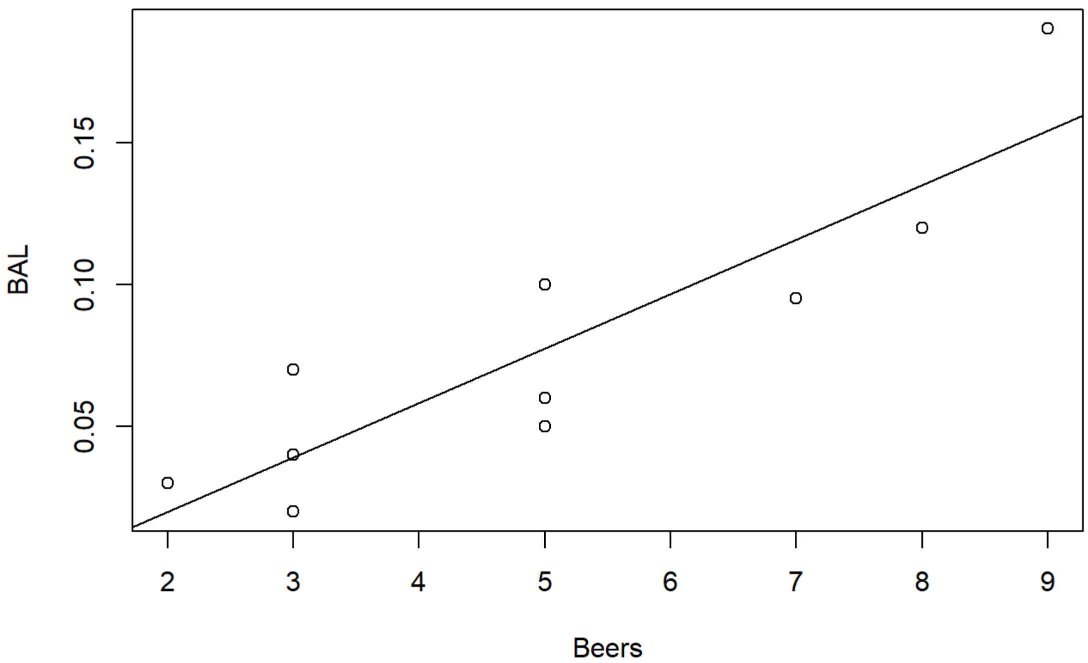
Make a scatterplot and fit the data with a regression line. Test the hypothesis that another beer raises your BAL by 0.02 percent against the alternative that it is less.

```
> Beers <- c(5, 2, 9, 8, 3, 7, 3, 5, 3, 5)
> BAL <- c(0.10, 0.03, 0.19, 0.12, 0.04, 0.095, 0.07, 0.06, 0.02, 0.05)
```

$$H_0 : \beta_1 = 0.02$$
$$H_A : \beta_1 < 0.02$$

```
> lmResult <- lm(BAL ~ Beers)
> summary(lmResult)

Call:
lm(formula = BAL ~ Beers)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.0275 -0.0187 -0.0071  0.0194  0.0357 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) -0.018500   0.019230  -0.962 0.364200
Beers        0.019200   0.003511   5.469 0.000595 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02483 on 8 degrees of freedom
Multiple R-squared:  0.789, Adjusted R-squared:  0.7626
F-statistic: 29.91 on 1 and 8 DF,  p-value: 0.0005953
> plot(Beers, BAL)
> abline(lmResult)
```




```
> e <- resid(lmResult); e
      1       2       3       4       5       6       7       8       9      10
 0.0225  0.0101  0.0357 -0.0151  0.0009 -0.0209  0.0309 -0.0175 -0.0191 -0.0275
> n <- length(e)
> Seps <- sqrt(sum(e^2) / (n - 2)); Seps
[1] 0.02482564
> beta1hat <- coef(lmResult)[['Beers']]; beta1hat
[1] 0.0192
> SEbeta1 <- Seps / sqrt(sum((Beers - mean(Beers))^2)); SEbeta1
[1] 0.003510876
> b1 <- 0.02
> t <- (b1 - beta1hat) / SEbeta1; t
[1] 0.2278634
> pvalue <- pt(t, n - 2, lower.tail = TRUE);pvalue
[1] 0.5872658
```

The $p-value = 0.5872658 > 0.05 = \alpha$, so we have no evidence to reject H_0. Another beer raises your BAL by 0.02 percent or more.

Problem 13.3

For the same Blood alcohol data, do a hypothesis test that the intercept is 0 with a two-sided alternative.

```
> Beers <- c(5, 2, 9, 8, 3, 7, 3, 5, 3, 5)
> BAL <- c(0.10, 0.03, 0.19, 0.12, 0.04, 0.095, 0.07, 0.06, 0.02, 0.05)
```

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

```
> lmResult <- lm(BAL ~ Beers)
> summary(lmResult)

Call:
lm(formula = BAL ~ Beers)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.0275 -0.0187 -0.0071  0.0194  0.0357 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.018500   0.019230  -0.962 0.364200    
Beers        0.019200   0.003511   5.469 0.000595 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02483 on 8 degrees of freedom
Multiple R-squared:  0.789,	Adjusted R-squared:  0.7626 
F-statistic: 29.91 on 1 and 8 DF,  p-value: 0.0005953
```

The $p-value = 0.3642 > 0.05 = \alpha$, so we have no evidence to reject H_0.

Problem 13.4

The lapse rate is the rate at which temperature drops as you increase elevation. Some hardy students were interested in checking empirically if the lapse rate of 9.8 degrees C/km was accurate for their hiking. To investigate, they grabbed their thermometers and their Suunto wrist altimeters and found the following data on their hike

elevation (ft)	**600**	**1000**	**1250**	**1600**	**1800**	**2100**	**2500**	**2900**
temperature (F)	56	54	56	50	47	49	47	45

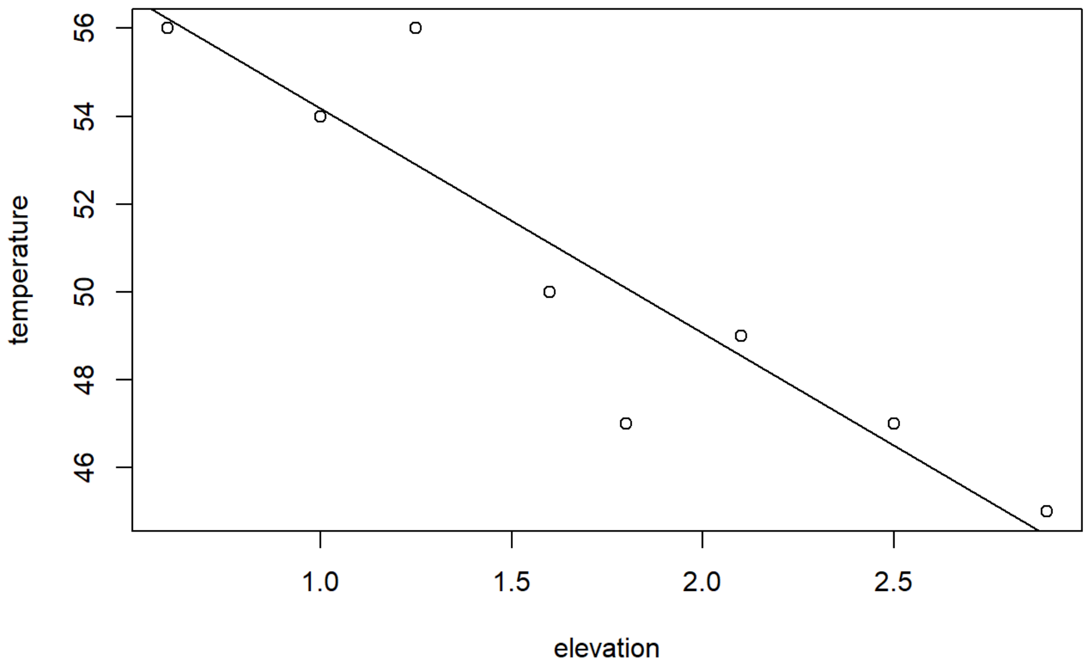
Draw a scatter plot with regression line, and investigate if the lapse rate is 9.8 C/km. (First, it helps to convert to the rate of change in Fahrenheit per feet with is 5.34 degrees per 1000 feet.) Test the hypothesis that the lapse rate is 5.34 degrees per 1000 feet against the alternative that it is less than this.

According to the conditions we have to check if the regression equation is

$$y = \beta_0 - 5.34x$$

Let us first build up our simple linear regression model and then to test the hypothesis if the slope is $\beta_1 = -5.34$.

```
> elevation <- c(600, 1000, 1250, 1600, 1800, 2100, 2500, 2900) / 1000
> temperature <- c(56, 54, 56, 50, 47, 49, 47, 45)
> lmResult <- lm(temperature ~ elevation)
> plot(elevation, temperature)
> abline(lmResult)
```



We test

$$H_0 : \beta_1 = -5.34$$
$$H_A : \beta_1 < -5.34$$

```
> summary(lmResult)

Call:
lm(formula = temperature ~ elevation)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.0844 -0.4433  0.1369  0.5072  3.1025 
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  59.2907     1.7173  34.525 3.93e-08 ***
elevation    -5.1146     0.9214  -5.551  0.00144 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.879 on 6 degrees of freedom
Multiple R-squared:  0.837, Adjusted R-squared:  0.8099
F-statistic: 30.81 on 1 and 6 DF,  p-value: 0.001445
```

We can take β_1 and $SE(\beta_1)$ values from the output of the function summary however the rest of those in

$$W_\alpha = \left\{ \frac{\hat{\beta}_1 - b_1}{SE(\beta_1)} < t_{\alpha;n-2} \right\}, SE(\beta_1) := \frac{S_\varepsilon}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X}_n)^2}}$$

we have to compute them step by step

```
> e <- resid(lmResult)
> n <- length(e)
> beta1hat <- (coef(lmResult))[['elevation']]; beta1hat
[1] -5.114567
> Seps <- sqrt(sum(e^2) / (n-2))
> SEbeta1 <- Seps / sqrt(sum((elevation - mean(elevation))^2)); SEbeta1
[1] 0.9213715
> alpha <- 0.05
> tquantile <- qt(alpha, n - 2, lower.tail = TRUE); tquantile
[1] -1.94318
> const <- -5.34
> temp <- (const - beta1hat) / SEbeta1; temp
[1] -0.2446712
> pvalue <- pt(temp, n - 2, lower.tail = TRUE); pvalue
[1] 0.4074318
```

See also the outputs of `summary(lmResult)`.

The $p-value = 0.4074318 > 0.05 = \alpha$, so we have no evidence to reject H_0 and according to the sample the difference between $\hat{\beta}_1 = -5.114567$ and the tested value -5.34 is not statistically significant.