STAC67H3:Regression Analysis
Assign 2
Due: Fri Mar 6, 2020 in class
All relevant work must be shown for credit.

Note 1: The completed assignments must be HANDED IN during the class. You should have your student card at the time you submit the assignment. Please DO NOT EMAIL your assignments. I DO NOT accept assignments through email. Any assignments sent through email will be deleted.

Note 2: In any question, if you are using R, all your R codes and outputs must be included in your answers. You should assume that the reader is not familiar with R outputs and so explain all your findings, quoting necessary values from your outputs.

Note 3: Please note that academic integrity is fundamental to learning and scholarship. You may discuss questions with other students. However, the work you submit should be your own. If I feel suspicious of any assignment (e.g. if your work doesn't appear to be consistent with what we have discussed in class), I will not mark the assignment. Instead, I will ask you to present your work in my office and your grade will be assigned based on your presentation.

Total points for this assignment: 100

1. Suppose X and Y are random variables, $E(Y^2) < \infty$ and $\varepsilon = Y - E(Y|X)$ so that $Y = E(Y|X) + \varepsilon$.

   (a) (5 points) Prove that $E(\varepsilon|X) = 0$ and $E(\varepsilon) = 0$.
   Hint: Theorem of total expectation (Theorem 3.5.2 p177, Evans and Rosenthal(STAB52 textbook)) is helpful.

   Solution: $\varepsilon = Y - E(Y|X) \implies E(\varepsilon|X) = E(Y|X) - E(E(Y|X)|X) = E(Y|X) - E(Y|X) = 0$ (3 points)

   $(E(E(Y|X)|X) = E(Y|X)$ (since $E(Y|X)$ is a function of X.)
   $E(\varepsilon|X) = 0 \implies E(E(\varepsilon|X)) = E(0) = 0$, i.e. $E(\varepsilon) = 0$ (by TTE) (2 points) ■

   Note: $E(Y|X)$ is a random variable and $E(Y) = 0$ does not necessarily imply $E(Y|X) = 0$
   Example: Let $X \sim Uniform\{-1, 1\}$ and $Y|X = x \sim N(x, 1)$. Then $E(Y|X = x) = x \neq 0$ for any value of x but since $E(Y|X) = X$, $E(Y) = E(E(Y|X)) = E(X) = 0$.

   (b) (8 points) Prove that $Cov(\varepsilon, E(Y|X)) = 0$ and deduce that
   $Var(Y) = Var(E(Y|X)) + Var(\varepsilon)$ and $Var(\varepsilon) = E(Var(Y|X))$.
   Hint: Theorem of total expectation (Theorem 3.5.2 p177 and theorem 3.5.6 p180,

Evans and Rosenthal(STAB52/57 textbook)) is helpful.

Note: STAB52/57 textbook (Evans and Rosenthal) can be downloaded from http://www.utstat.toronto.edu/mikevans/jeffrosenthal/

Solution: Using part (a) (i.e. $E(\varepsilon) = 0$), we have
$Cov(\varepsilon, E(Y|X)) = E(\varepsilon E(Y|X))$
and now using TTE, $E(\varepsilon E(Y|X)) = E(E(\varepsilon E(Y|X)|X)) = E(E(Y|X)E(\varepsilon|X))$
$= E(E(Y|X) \times 0) = E(0) = 0$ (4 points)

$Y = E(Y|X)+\varepsilon \implies Var(Y) = Var(E(Y|X))+Var(\varepsilon)+2Cov(\varepsilon, E(Y|X)) = Var(E(Y|X)) + Var(\varepsilon)$ since $Cov(\varepsilon, E(Y|X)) = 0$. (2 points)

By Thm 3.5.6 (Evans and Rosenthal) $Var(Y) = Var(E(Y|X))+E(Var(Y|X))$ and so Var$(\varepsilon) = E(Var(Y|X))$ (2 points) ■

2. In this question we will investigate more properties of residuals (i.e. $e_i = Y_i - \hat{Y}_i, i = 1, \ldots, n$) from the Normal error simple linear regression model that we discussed in class (i.e the model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$ with the assumptions we discussed in class).
   (a) (4 points) Show that $Var(Y_i - \bar{Y}) = \left(1 - \frac{1}{n}\right)\sigma^2$

Solution: $Y_i - \bar{Y} = Y_i - \frac{1}{n}\sum_{j=1}^n Y_j = \left(1 - \frac{1}{n}\right)Y_i - \frac{1}{n}\sum_{\substack{j=1 \\ j\neq i}}^n Y_j$ and so

$$\begin{aligned}
Var(Y_i - \bar{Y}) &= \left(1 - \frac{1}{n}\right)^2 Var(Y_i) + \frac{1}{n^2}\sum_{\substack{j=1 \\ j\neq i}}^n Var(Y_j) \qquad \because Y_i\text{s are independent} \\
&= \left(1 - \frac{1}{n}\right)^2 \sigma^2 + \frac{1}{n^2}(n-1)\sigma^2 \\
&= \left(1 - \frac{1}{n}\right)\sigma^2 \quad \blacksquare
\end{aligned}$$

   (b) (5 points) Show that $Var(e_i) = \left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{XX}}\right)\sigma^2$

Solution:

$$\begin{aligned}
Var(e_i) &= Var(Y_i - \hat{Y}_i) \\
&= Var(Y_i) + Var(\hat{Y}_i) - 2Cov(Y_i, \hat{Y}_i) \\
&= \sigma^2 + \frac{\sigma^2}{n} + \frac{(x_i - \bar{x})^2}{S_{XX}}\sigma^2 - 2Cov(Y_i, \hat{Y}_i)
\end{aligned}$$

Question 2 continues on the next page...

$$
\begin{aligned}
Cov(Y_i, \hat{Y}_i) &= Cov(Y_i, b_0 + b_1 x_i) \\
&= Cov(Y_i, \bar{Y} - b_1\bar{x} + b_1 x_i) \\
&= Cov(Y_i, \bar{Y} + b_1(x_i - \bar{x})) \\
&= Cov(Y_i, \bar{Y}) + (x_i - \bar{x})Cov(Y_i, b_1) \\
&= Cov(Y_i, \frac{1}{n}\sum_{j=1}^{n} Y_j) + (x_i - \bar{x})Cov(Y_i, \sum_{j=1}^{n} k_j Y_j) \\
&= \frac{\sigma^2}{n} + (x_i - \bar{x})k_i\sigma^2 \quad (\because Y_1, \ldots, Y_n \text{ are independent}) \\
&= \frac{\sigma^2}{n} + (x_i - \bar{x})\frac{x_i - \bar{x}}{S_{XX}}\sigma^2 \\
&= \frac{\sigma^2}{n} + \frac{(x_i - \bar{x})^2}{S_{XX}}\sigma^2
\end{aligned}
$$

Now substitute.

$$
\begin{aligned}
Var(e_i) &= \sigma^2 + \frac{\sigma^2}{n} + \frac{(x_i - \bar{x})^2}{S_{XX}}\sigma^2 - 2Cov(Y_i, \hat{Y}_i) \\
&= \sigma^2 + \frac{\sigma^2}{n} + \frac{(x_i - \bar{x})^2}{S_{XX}}\sigma^2 - 2\left(\frac{\sigma^2}{n} + \frac{(x_i - \bar{x})^2}{S_{XX}}\sigma^2\right) \\
&= \left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{XX}}\right)\sigma^2 \quad \blacksquare
\end{aligned}
$$

(c) (5 points) In class, I mentioned that $E(MSE) = \sigma^2$. The reason for this that I provided is the result $\frac{SSE}{\sigma^2} \sim \chi^2_{n-2}$, which we did not prove. In this part we will show that $E(MSE) = \sigma^2$ without using this result (i.e. without using the result $\frac{SSE}{\sigma^2} \sim \chi^2_{n-2}$.) Use the result in part (b) above to show that $E\left(\sum_{i=1}^{n} e_i^2\right) = (n-2)\sigma^2$.

Solution:

$$
\begin{aligned}
E\left(\sum_{i=1}^{n} e_i^2\right) &= \sum_{i=1}^{n} E(e_i^2) \\
&= \sum_{i=1}^{n}(V(e_i) + (E(e_i))^2) \\
&= \sum_{i=1}^{n} V(e_i) \quad \because E(e_i) = 0 \\
&= \sum_{i=1}^{n}\left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{XX}}\right)\sigma^2 \\
&= (n-1)\sigma^2 - \sigma^2 \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{S_{XX}} \\
&= (n-1)\sigma^2 - \sigma^2 \frac{1}{S_{XX}} \sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= (n-1)\sigma^2 - \sigma^2 \frac{1}{S_{XX}} S_{XX} \\
&= (n-2)\sigma^2 \quad \blacksquare
\end{aligned}
$$

3. (6 points) Consider the Normal error simple linear regression model we discussed in class (i.e. the model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$ satisfyying the usual assumptions). In class we showed that the null hypothesis $H_0 : \beta_1 = 0$ can be tested using the test statistic $T = \frac{b_1 - 0}{\frac{s}{\sqrt{S_{XX}}}}$, which has a t-distribution with $n-2$ degrees of freedom.
   Show that this equation for T can also be written as:

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

Solution: In class, we learned $b_1 = r\frac{s_y}{s_x} = r\frac{\sqrt{S_{YY}}}{\sqrt{S_{XX}}}$, $r = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}}$ and $s = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{S_{YY} - b_1^2 S_{XX}}{n-2}}$
Substituting in the equation for T

$$\begin{aligned}
T = \frac{b_1 - 0}{\frac{s}{\sqrt{S_{XX}}}} &= \frac{b_1\sqrt{S_{XX}}}{s} \\
&= \frac{r\sqrt{S_{YY}}}{\sqrt{\frac{S_{YY} - b_1^2 S_{XX}}{n-2}}} \\
&= \frac{r\sqrt{S_{YY}}\sqrt{n-2}}{\sqrt{S_{YY}}\sqrt{1 - b_1^2 \frac{S_{XX}}{S_{YY}}}} \\
&= \frac{r\sqrt{n-2}}{\sqrt{1 - (r\frac{\sqrt{S_{YY}}}{\sqrt{S_{XX}}})^2 \frac{S_{XX}}{S_{YY}}}} \\
&= \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad \blacksquare
\end{aligned}$$

4. In class, we discussed inverse prediction (or calibration) using simple linear regression (section 4.6) with Normal errors. In particular if a **new** observation Y_h becomes available, and it is desired to estimate level X_h that gave rise to this observation, then we can estimate it by

$$\hat{X}_h = \frac{Y_h - b_0}{b_1}, b_1 \neq 0$$

and an approximate $100(1-\alpha)$ confidence interval for X_h is given by

$$\hat{X}_h \pm t_{1-\alpha/2,n-2}\frac{s}{|b_1|}\sqrt{1 + \frac{1}{n} + \frac{(\hat{X}_h - \bar{x})^2}{s_{XX}}}.$$

I stated this result in class, but didn't tell you why the standard error of $\hat{X}_h$ is $\frac{s}{|b_1|}\sqrt{1 + \frac{1}{n} + \frac{(\hat{X}_h - \bar{x})^2}{s_{XX}}}$. Since $\hat{X}_h$ is a ratio of two Normal random variables, finding its variance is somewhat complicated. For example, it X_1 and X_2 are independent Normal random variables, then X_1/X_2 has a Cauchy distribution for which the varinace is undefined. However we can use asymptotic methods to find an approximate formula for the variance of $\hat{X}_h$. We will investigate that in this question.

(a) (5 points) Let T_1 and T_2 be random variables with means μ_1 and μ_2 respectively and define $T = (T_1, T_2)$ and $\mu = (\mu_1, \mu_2)$. Suppose that $g(t)$ is a differentialbe function (eg an estimator of some parameter) for which we want an approximate estimate of the variance. Let $g'_i(\mu) = \left.\frac{\partial g(t)}{\partial t_i}\right|_{t_1=\mu_1, t_2=\mu_2}, i = 1, 2$. Then the first order Taylor series expansion of g about μ is

$$g(t) \approx g(\mu) + (t_1 - \mu_1)g'_1(\mu) + (t_2 - \mu_2)g'_2(\mu).$$

Question 4 continues on the next page…

Use the above result to show that,

$$Var(g(T)) \approx (g_1'(\mu))^2 Var(T_1) + (g_2'(\mu))^2 Var(T_2) + 2g_1'(\mu)g_2'(\mu)Cov(T_1, T_2).$$

Solution: Using the above result $E(g(T)) \approx g(\mu) + g_1'(\mu)E(T_1 - \mu_1) + g_2'(\mu)E(T_2 - \mu_2) = g(\mu) \quad \because E(T_1 - \mu_1) = E(T_2 - \mu_2) = 0$ and so,

$$\begin{aligned}
Var(g(T)) &\approx E(g(T) - g(\mu))^2 \\
&= E[(g_1'(\mu)E(T_1 - \mu_1) + g_2'(\mu)E(T_2 - \mu_2))^2] \\
&= ((g_1'(\mu))^2(T_1 - \mu_1)^2 + ((g_2'(\mu))^2(T_2 - \mu_2)^2 \\
&\quad +2g_1'(\mu)g_2'(\mu)E[(T_1 - \mu_1)(T_2 - \mu_2)] \\
&= ((g_1'(\mu))^2 Var(T_1) + ((g_2'(\mu))^2 Var(T_2) + 2g_1'(\mu)g_2'(\mu)Cov(T_1, T_2) \quad \blacksquare
\end{aligned}$$

(b) (4 points) If $g(T) = \frac{T_1}{T_2}$, use the result in part (a) to show that,

$$Var(g(T)) \approx \frac{Var(T_1)}{(E(T_2))^2} + \frac{(E(T_1))^2}{(E(T_2))^4} Var(T_2) - 2\frac{E(T_1)}{(E(T_2))^3} Cov(T_1, T_2).$$

Solution: If $g(T) = \frac{T_1}{T_2}$, then $g_1'(\mu) = \left.\frac{\partial g(t)}{\partial t_1}\right|_{t_1=\mu_1, t_2=\mu_2} = \frac{1}{\mu_2} = \frac{1}{E(T_2)}$ and $g_2'(\mu) = \left.\frac{\partial g(t)}{\partial t_2}\right|_{t_1=\mu_1, t_2=\mu_2} = -\frac{\mu_1}{\mu_2^2} = -\frac{E(T_1)}{(E(T_2))^2}$.
Substituting these in the result in part (a), we have

$$\begin{aligned}
Var(g(T)) &\approx (g_1'(\mu))^2 Var(T_1) + (g_2'(\mu))^2 Var(T_2) + 2g_1'(\mu)g_2'(\mu)Cov(T_1, T_2) \\
&= \left(\frac{1}{E(T_2)}\right)^2 Var(T_1) + \left(-\frac{E(T_1)}{(E(T_2))^2}\right)^2 Var(T_2) \\
&\quad +2\left(\frac{1}{E(T_2)}\right)\left(-\frac{E(T_1)}{(E(T_2))^2}\right) Cov(T_1, T_2) \\
&= \frac{Var(T_1)}{(E(T_2))^2} + \frac{(E(T_1))^2}{(E(T_2))^4} Var(T_2) - 2\frac{E(T_1)}{(E(T_2))^3} Cov(T_1, T_2) \quad \blacksquare
\end{aligned}$$

(c) (3 points) Now let's consider $\hat{X}_h = \frac{Y_h - b_0}{b_1}$.
Show that $Var(Y_h - b_0) = \sigma^2 \left(1 + \frac{1}{n} + \frac{\bar{x}^2}{s_{XX}}\right)$.

Solution:

$$\begin{aligned}
Var(Y_h - b_0) &= Var(Y_h) + Var(b_0) \quad \because Y_h \text{ is a new obn } y_h \perp b_0 \\
&= \sigma^2 + \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{s_{XX}}\right) \\
&= \sigma^2\left(1 + \frac{1}{n} + \frac{\bar{x}^2}{s_{XX}}\right) \quad \blacksquare
\end{aligned}$$

Question 4 continues on the next page...

(d) (3 points) Show that $Cov(Y_h - b_0, b_1) = \frac{\bar{x}\sigma^2}{s_{XX}}$.

Solution:

$$\begin{aligned} Cov(Y_h - b_0, b_1) &= -Cov(b_0, b_1) \quad \because \ Y_h \text{ is a new obn } y_h \perp b_0 \\ &= \frac{\bar{x}\sigma^2}{s_{XX}} \quad \because Cov(b_0, b_1) = -\frac{\bar{x}\sigma^2}{s_{XX}} \text{ (showed in class)} \quad \blacksquare \end{aligned}$$

(e) (5 points) Use the results in parts (b), (c) and (d) to show that,
$Var(\hat{X}_h) \approx \frac{s^2}{b_1^2}\left(1 + \frac{1}{n} + \frac{(\hat{X}_h - \bar{x})^2}{s_{XX}}\right)$.

Solution: Substituting $E(T_1) = E(Y_h - b_0) = \beta_0 + \beta_1 x_h - \beta_0 = \beta_1 X_h$, $E(T_2) = E(b_1) = \beta_1$, $Var(T_1) = Var(Y_h - b_0) = \sigma^2\left(1 + \frac{1}{n} + \frac{\bar{x}^2}{s_{XX}}\right)$, $Var(T_2) = Var(Yb_1) = \frac{\sigma^2}{s_{XX}}$ and $Cov(T_1, T_2 = Cov(Y_h - b_0, b_1) = \frac{\bar{x}\sigma^2}{s_{XX}}$ in the result in part (b), we have

$$\begin{aligned} Var(\hat{X}_h) &\approx \frac{Var(T_1)}{(E(T_2))^2} + \frac{(E(T_1))^2}{(E(T_2))^4}Var(T_2) - 2\frac{E(T_1)}{(E(T_2))^3}Cov(T_1, T_2) \\ &= \frac{\sigma^2\left(1 + \frac{1}{n} + \frac{\bar{x}^2}{s_{XX}}\right)}{\beta_1^2} + \frac{(\beta_1 X_h)^2}{\beta_1^4}\frac{\sigma^2}{s_{XX}} - 2\frac{\beta_1 X_h}{\beta_1^3}\frac{\bar{x}\sigma^2}{s_{XX}} \\ &= \frac{\sigma^2}{\beta_1^2}\left(1 + \frac{1}{n} + \frac{\bar{x}^2 + X_h^2 - 2X_h\bar{x}}{s_{XX}}\right) \\ &= \frac{\sigma^2}{\beta_1^2}\left(1 + \frac{1}{n} + \frac{(X_h - \bar{x})^2}{s_{XX}}\right) \end{aligned}$$

and now replace σ^2 and β_1 by their estimates $s^2 = MSE$ and b_1 $\blacksquare$

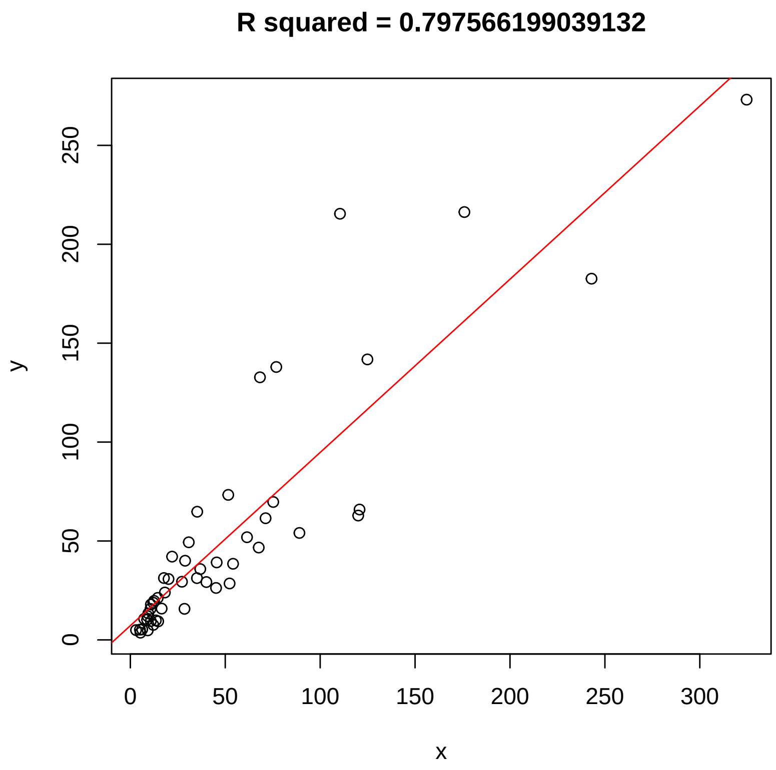
5. The data file dataxy.txt on quercus contains a data set with two variables x and y. We will use this dataset to investigate heteroscedasticity (i.e. non-conatant variance).

   (a) (9 points) Fit a simple linear regression model for y on x and obtain the residuals for the model. Make a plot of residuals versus x and a scatterplot of y versus x. Comment on your plots.

Solution:

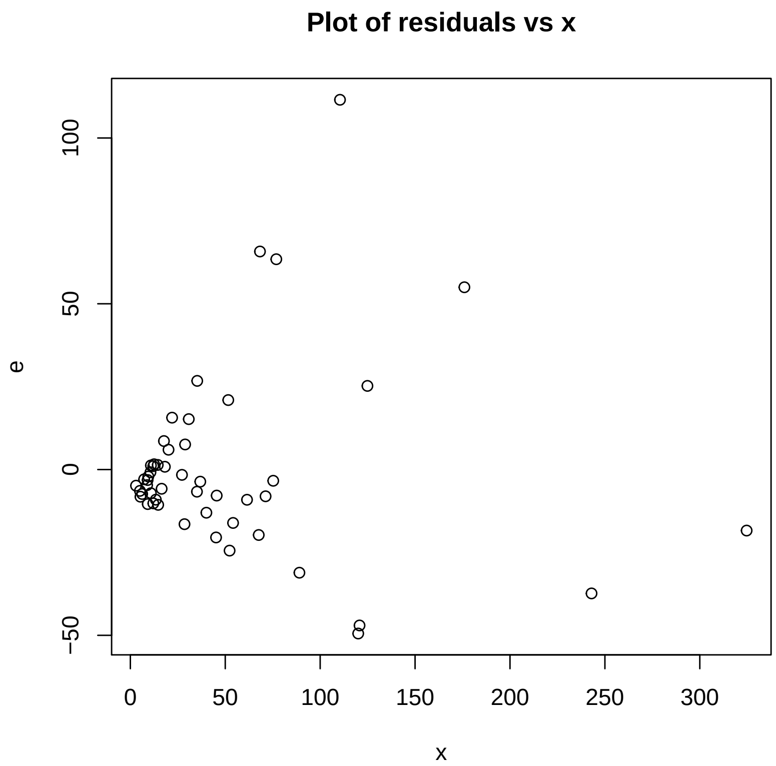
```
> xy <- read.table("dataxy.txt", header = T)
> x <- xy$x
> y <- xy$y
> n <- nrow(xy)
> fit <- lm(y ~ x)
> plot(x,y,main=paste("R squared =",summary.lm(fit)[8]))
> abline(coefficients(fit),col=2)
```

Question 5 continues on the next page. . .




```
> e <- fit$residuals
> plot(x, e, main = "Plot of residuals vs x")
```

Question 5 continues on the next page. . .



Scatterplot of y on x shows that the variance in y appears to increase with x. (2 points)
Plot of residuals vs x shows that the variance of residuals appears to increase with x. (2 points)
3 points for the plot of residuals vs x (TA: Students don't have to show the regression line on the plot)
2 points for the scaatterplot of y on x.

(b) Use Brown-Forsythe test that we discussed in class (on section 3.6, page 116 in textbook) to test for the constancy of the error variance. Show delails of your work. As we discussed in class, this test procedure requires the dataset to be split into two parts based on the values of x. Make sure that each part has exactly 25 residuals. You should use R for this question. Show the specific parts of your R code that did the following tasks and the values they calculated. The notations are as we discussed in class (and in textbook).

i. (4 points) Calculate $\tilde{e}_1$ and $\tilde{e}_2$.

Solution:

```
> #Brown-Forsythe Test for Constancy of Error Variance sec 3.6, p116
> table <- cbind(x, e)
> orderedbyx <- table[order(x),]
> orderedbyx <- data.frame(orderedbyx)
```

```
> eorderedbyx <- orderedbyx$e
> n <- nrow(table)
> n1 <- ceiling(n/2)
> n1

[1] 25

> n2 <- n - n1
> e1 <- eorderedbyx[1:n1]
> e2 <- eorderedbyx[c((n1+1):n)]
> etilde1 <- median(e1)
> etilde1

[1] -3.158003

> etilde2 <- median(e2)
> etilde2

[1] -7.849612
```

ii. (4 points) Calculate $\bar{d}_1$ and $\bar{d}_2$.

Solution:

```
> #Brown-Forsythe Test for Constancy of Error Variance sec 3.6, p116
> d1 <- abs(e1-etilde1)
> dbar1 <- mean(d1)
> dbar1

[1] 5.206294

> d2 <- abs(e2-etilde2)
> dbar2 <- mean(d2)
> dbar2

[1] 26.88054
```

iii. (3 points) Calculate t_{BF}.

Solution:

```
> #Brown-Forsythe Test for Constancy of Error Variance sec 3.6, p116
> s_sq <- (sum((d1-dbar1)^2)+sum((d2-dbar2)^2))/(n-2)
> s_sq

[1] 426.6666

> s <- sqrt(s_sq)
> tBF <- (dbar1-dbar2)/(s*sqrt(1/n1+1/n2))
> tBF
```

Question 5 continues on the next page. . .

```
[1] -3.709835
```

iv. (3 points) Calculate the p-value. Give your conclusion based your p-value. Use $\alpha = 0.05$.

Solution:

```
> #Brown-Forsythe Test for Constancy of Error Variance sec 3.6, p116
> p_value <- 2*(1-pt(abs(tBF),n-2))
> p_value

[1] 0.0005382865
```

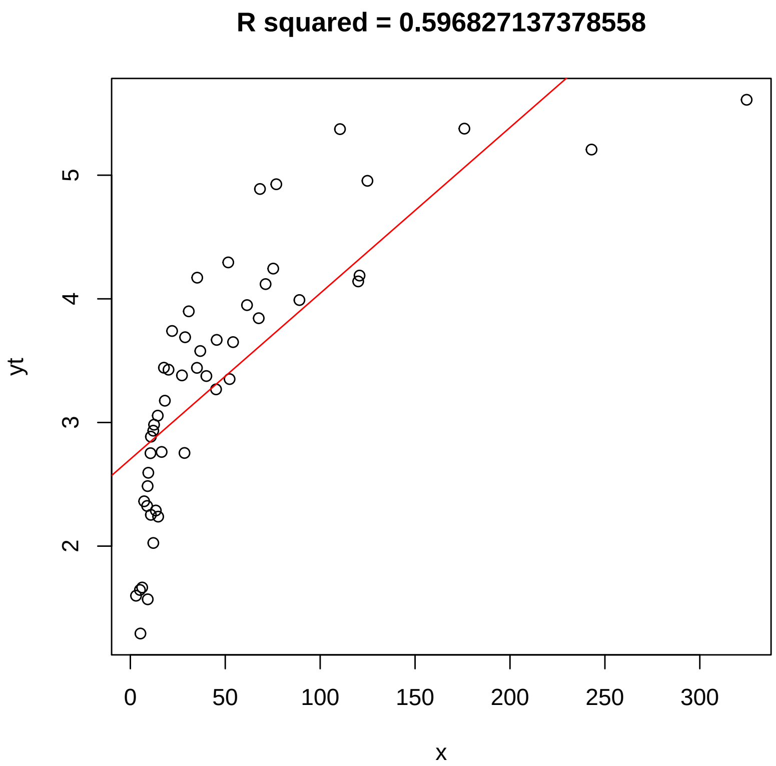
p-value < 0.05 indicating significant heteroscedasticity.

(c) (7 points) Transform the variable y by the log transformation, i.e. use yt = log(y) and fit the simple linear regression model of yt on x. Make a plot of residuals versus x and a scatterplot of yt versus x. Comment on your plots.

Solution:

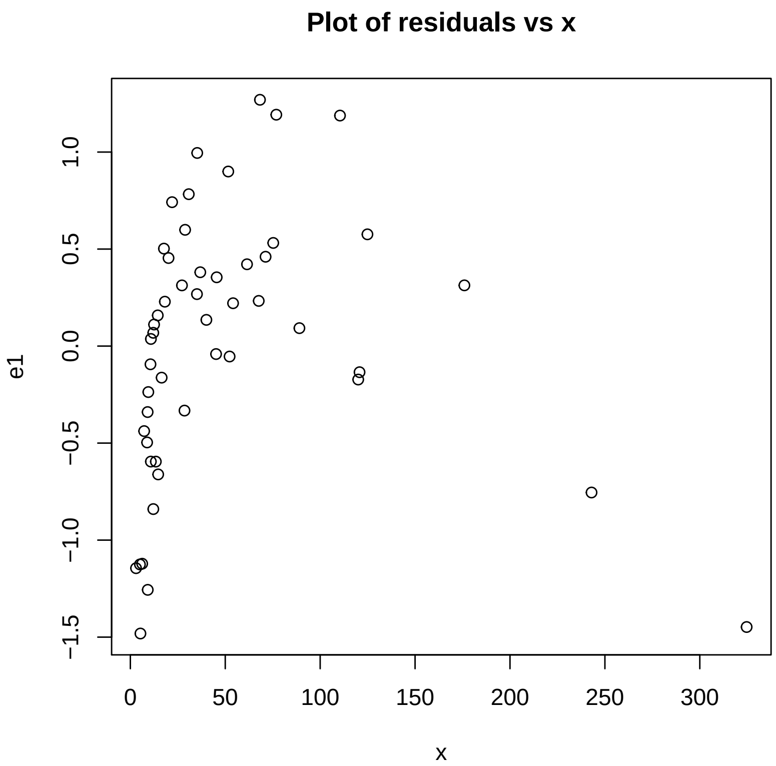
```
> yt <- log(y)
> fit1 <- lm(yt ~ x)
> plot(x,yt,main=paste("R squared =",summary.lm(fit1)[8]))
> abline(coefficients(fit1),col=2)
```

Question 5 continues on the next page. . .




```
> e1 <- fit1$residuals
> plot(x, e1, main = "Plot of residuals vs x")
```

Question 5 continues on the next page. . .

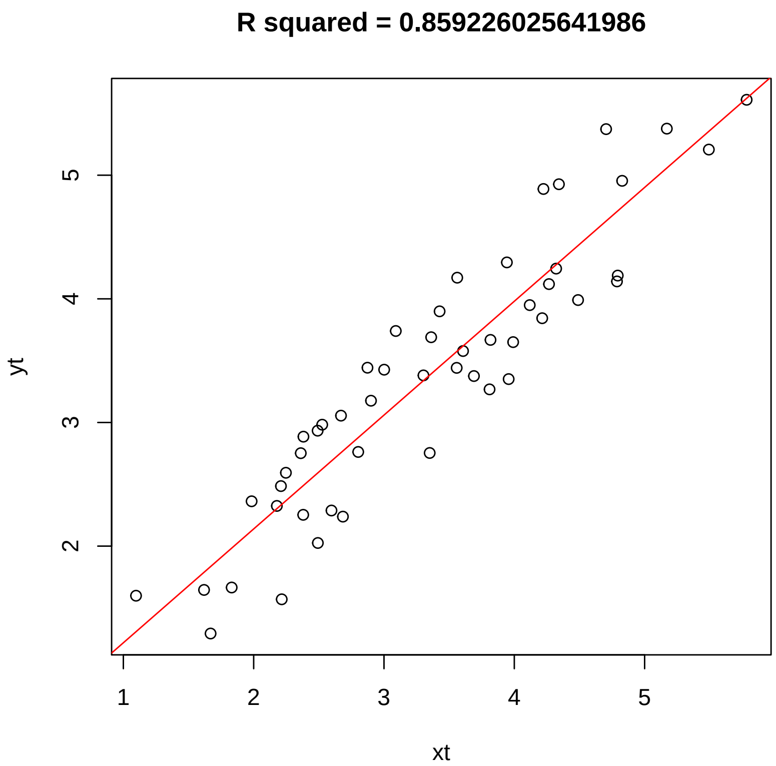


Both these plots show curving patterns, indicating that a linear model is not apprpriate. (2 points)
(2 points for each plot and 1 point for using the log transformation propertly (in R code)).
TA: Students don't have to show the regression line on scatterplot.

(d) (7 points) Let's now try transforming x as well. i.e. use xt = log(x) and fit the simple linear regression model of yt on xt. Make a plot of residuals versus xt and a scatterplot of yt versus xt. Comment on your plots.

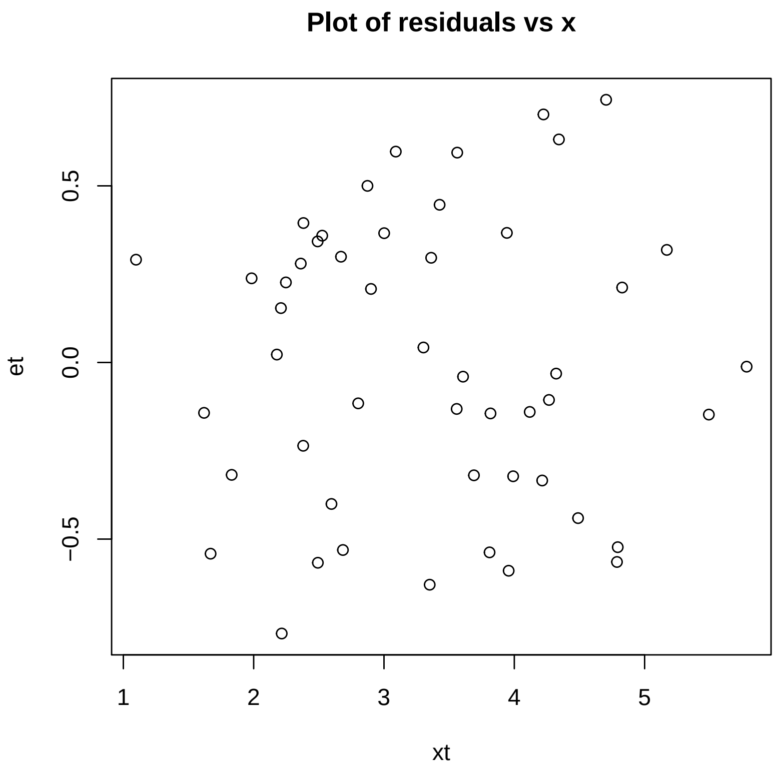
Solution:

```
> xt <- log(x)
> fit2 <- lm(yt ~ xt)
> plot(xt,yt,main=paste("R squared =",summary.lm(fit2)[8]))
> abline(coefficients(fit2),col=2)
```




```
> et <- fit2$residuals
> plot(xt, et, main = "Plot of residuals vs x")
```

Question 5 continues on the next page. . .



Now the scatter of yt on xt looks close to a straight line and the plot looks more random. (2 points)
(2 points for each plot and 1 point for using the log transformation properly (in R code)).
TA: Students don't have to show the regression line on scatterplot.

(e) (10 points) Now repeat part (b) for this transformed model, i.e. regression of yt on xt, testing whether this model shows heteroscedasticity (i.e. non-constant variance).
Note: Part (b) has four subparts and this part requires answers to each of those four parts for the regression of yt on xt.

Solution:

```
> #Brown-Forsythe Test for Constancy of Error Variance sec 3.6, p116
> x <- xt
> e <- e2
> table <- cbind(x, e)
> orderedbyx <- table[order(x),]
> orderedbyx <- data.frame(orderedbyx)
> eorderedbyx <- orderedbyx$e
> n <- nrow(table)
> n1 <- ceiling(n/2)
```

```
> n1
[1] 25
> n2 <- n - n1
> e1 <- eorderedbyx[1:n1]
> e2 <- eorderedbyx[c((n1+1):n)]
> etilde1 <- median(e1)
> etilde1
[1] -13.0189
> etilde2 <- median(e2)
> etilde2
[1] -3.38633
> d1 <- abs(e1-etilde1)
> dbar1 <- mean(d1)
> dbar1
[1] 24.32981
> d2 <- abs(e2-etilde2)
> dbar2 <- mean(d2)
> dbar2
[1] 27.89853
> s_sq <- (sum((d1-dbar1)^2)+sum((d2-dbar2)^2))/(n-2)
> s_sq
[1] 835.4075
> s <- sqrt(s_sq)
> tBF <- (dbar1-dbar2)/(s*sqrt(1/n1+1/n2))
> tBF
[1] -0.4365341
> p_value <- 2*(1-pt(abs(tBF),n-2))
> p_value
[1] 0.6644066
```

3 points for $\tilde{e}_1$ and $\tilde{e}_2$
3 points for $\bar{d}_1$ and $\bar{d}_2$
2 points for t_{BF}
2 points for p-value and conclusion.