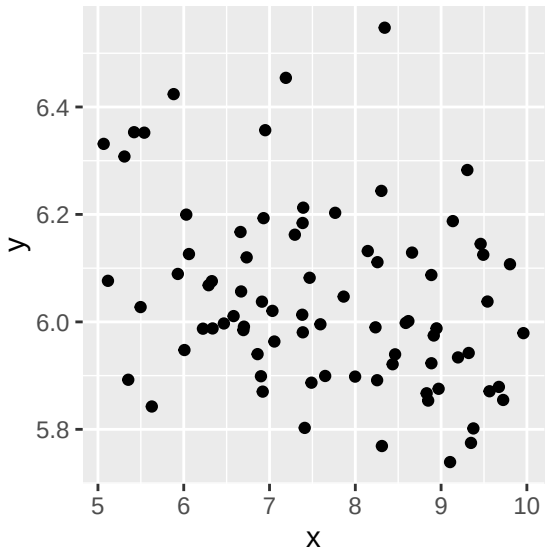
# MIE237 Tutorial

*Neil Montgomery*

*Week of February 22, 2016*

In this tutorial we'll practice fitting components of R regression output together, and also practice interpreting residual plots.

First I'll simulate a dataset with an input variable named $x$ and an output variable named $y$. Here is a summary (sample means and sample variances) and a plot of the data, followed by some R regression output with some entries obscured.




| $\overline{y}$ | $\overline{x}$ | $s_y^2$ | $s_x^2$ |
|---|---|---|---|
| 6.042789 | 7.628518 | 0.0280896 | 1.814987 |

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.34672    0.10335  61.409  < 2e-16 ***
x           -0.03984    *******  ******  ******* **
---

Residual standard error: 0.1598 on 78 degrees of freedom

Analysis of Variance Table

Response: y
          Df Sum Sq  Mean Sq F value   Pr(>F)
x         ** ****** ********  (OMIT)   (OMIT)
Residuals ** ****** ********
```

Tutorial tasks: fill in all the blanks `*******` (not the ones that say `(OMIT)` because that hasn't been covered.)

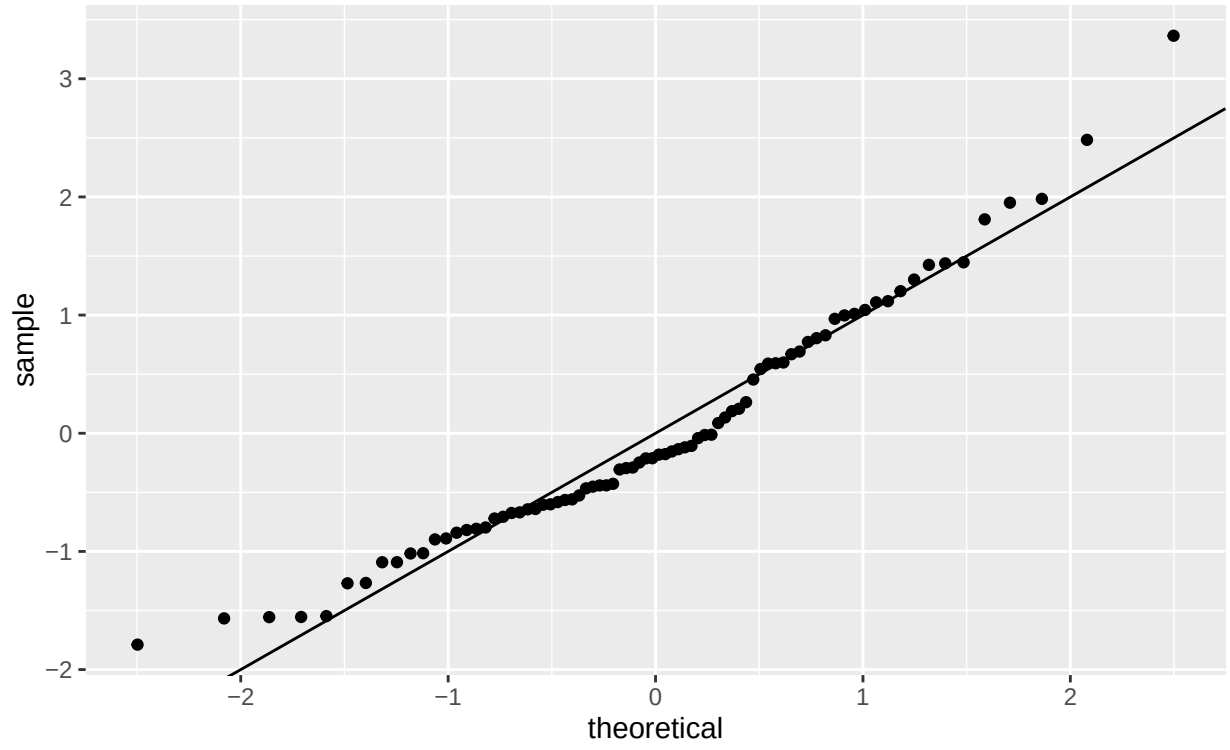
```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.28000 -0.10509 -0.03096  0.10706  0.53320
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.34672    0.10335  61.409  < 2e-16 ***
## x           -0.03984    0.01334  -2.986  0.00378 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1598 on 78 degrees of freedom
## Multiple R-squared:  0.1026, Adjusted R-squared:  0.09106
## F-statistic: 8.914 on 1 and 78 DF,  p-value: 0.003779
```

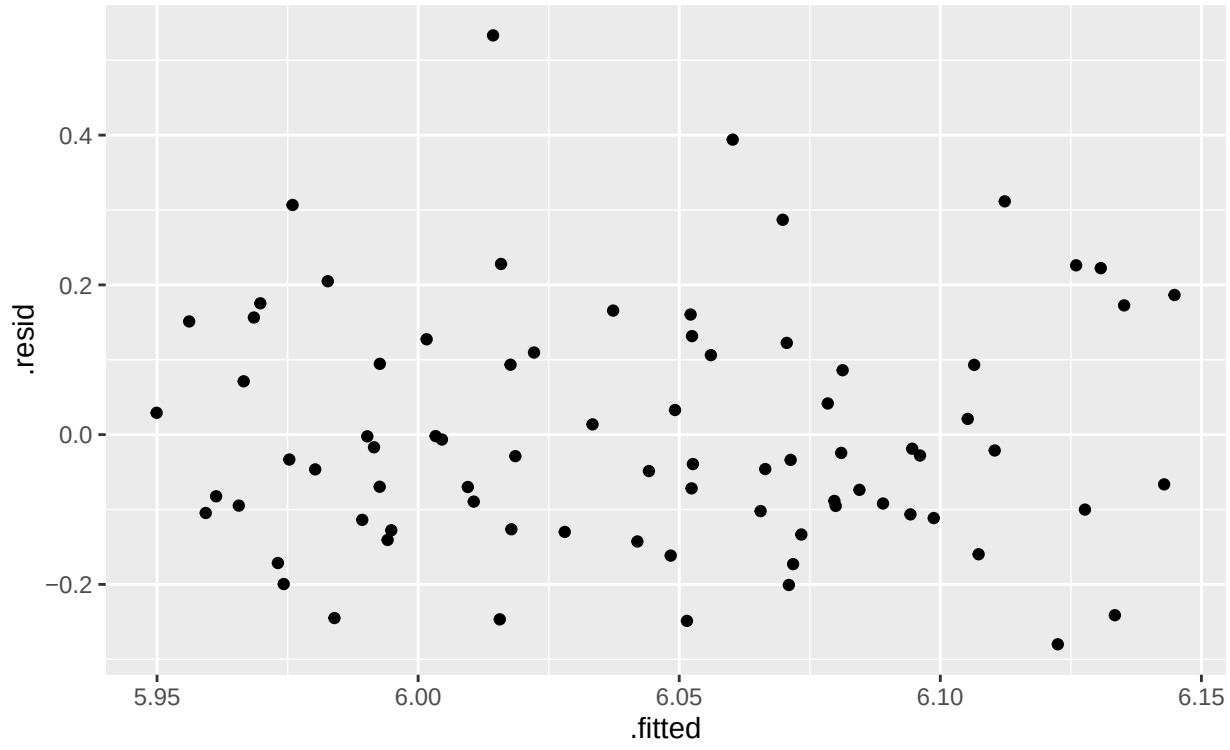
```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq  Mean Sq F value   Pr(>F)
## x          1 0.2276 0.227598  8.9143 0.003779 **
## Residuals 78 1.9915 0.025532
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Procedure:

1. `Residual standard error` is $\sqrt{\text{MSE}}$ with 78 degrees of freedom, so the `Residuals` line in the ANOVA table can be filled in.
2. We are given $s_y^2$ and the error degrees of freedom is 78, so the sample size is $n = 80$. So we can calculate the (not shown) total sum of squares $(n-1)s_y^2$ and from that get the `x` or regression sum of squares. Its degrees of freedom is 1. So we can fill in the first line of the ANOVA table.
3. The standard error $\hat{\beta}_1$ is $\sqrt{\text{MSE}/S_{xx}}$. We have the MSE and $S_{xx} = (n-1)s_x^2$. Then we can get the `t value` and the p-value from a table (or the computer).

Here are the residual plots. Comment on them, and on any possible issues with the model assumptions/calculations.

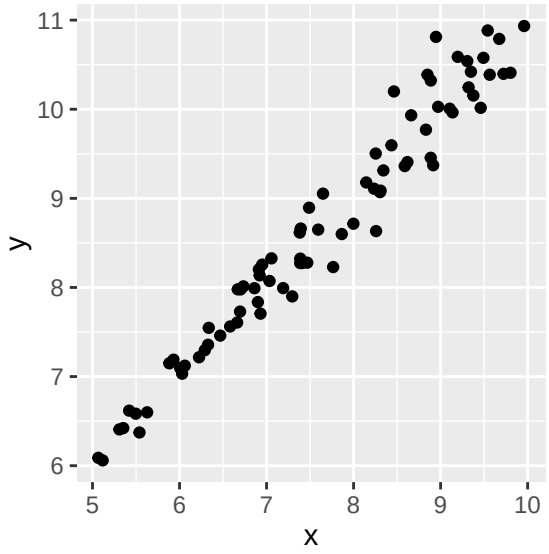


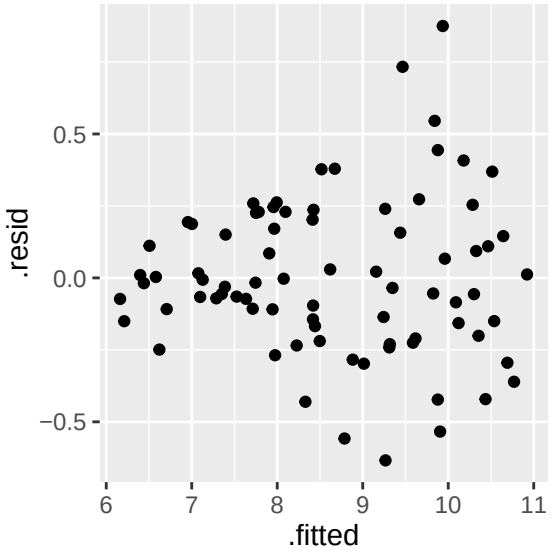



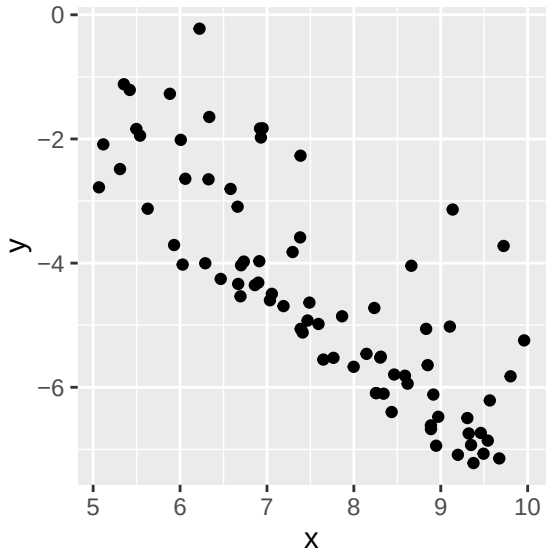
(Normal quantile plot has a bit of a curve to it, which might not have even been so clear had I not put the line on it, but with a sample of size 80 the p-value above is probably still accurate. There is no clear non-linear pattern in the other plot. Also, the amount of variation looks even when going from left to right, so the equal variance assumption is also satisfied.)

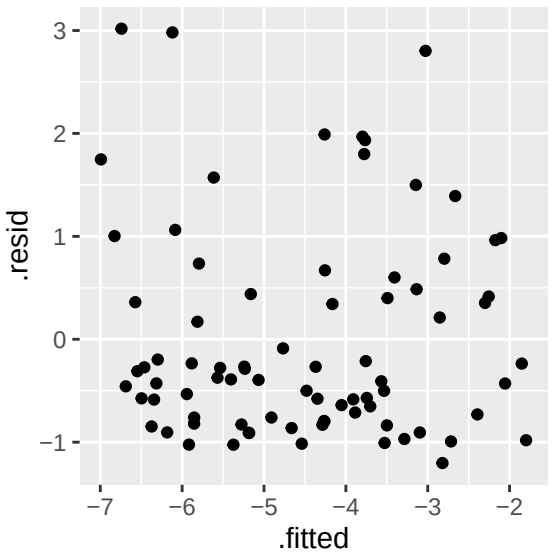
Now I'll simulate some other regression datasets for the purpose of practicing the interpretation of the residual vs. fitted values plot. Each of the following pairs of plots has the original data plotted on the left and the plot of residuals vs. fitted values on the right. Match up the following 4 plots to the these correct interpretations:

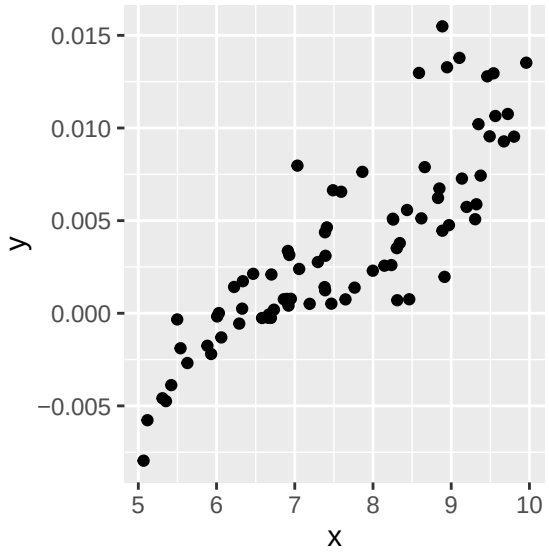
1. No clear non-linear pattern and no violation of the equal variance assumption.
2. A clear non-linear pattern but no violation of the equal variance assumption.
3. No clear non-linear pattern but a violation of the equal variance assumption.
4. A clear non-linear pattern and a violation of the equal variance assumption.

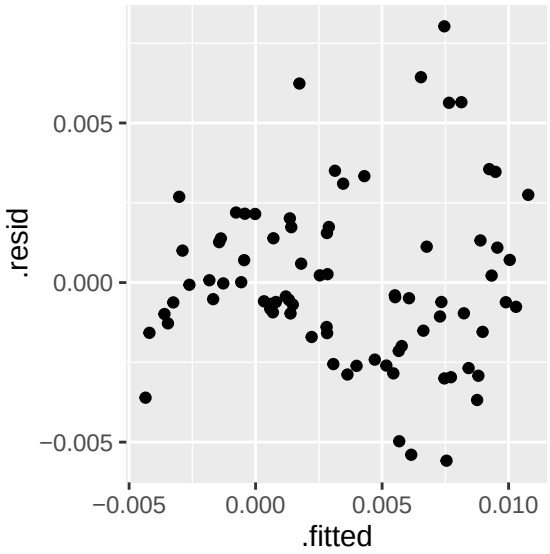


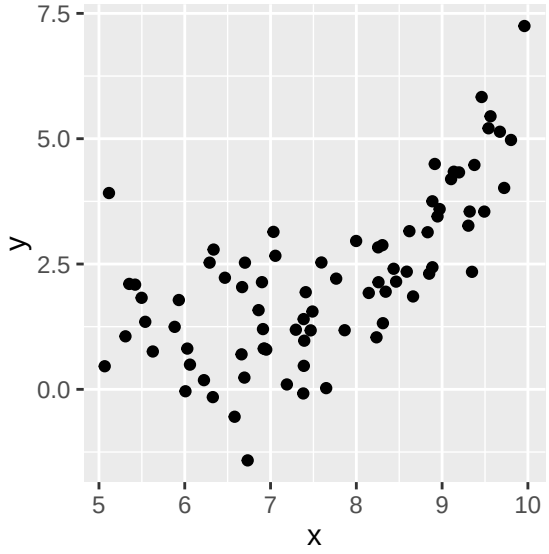



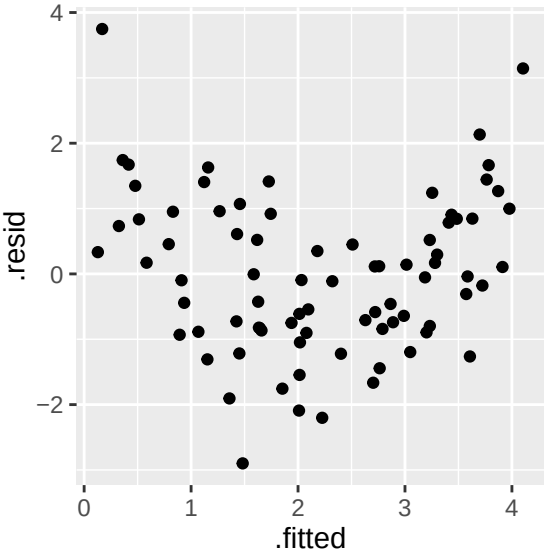




0.015
0.010
0.005
0.000
−0.005
y
5
6
7
8
9
10
x


0.005
0.000
−0.005
.resid
−0.005
0.000
0.005
0.010
.fitted


7.5
5.0
2.5
0.0
y
5
6
7
8
9
10
x


4
2
0
−2
.resid
0
1
2
3
4
.fitted

First pair: 3.

Second pair: 1.

Third pair: 4.

Fourth pair: 2.