Applied Statistical Regression
AS 2012 – Week 03

Marcel Dettling

Institute for Data Analysis and Process Design

Zurich University of Applied Sciences

marcel.dettling@zhaw.ch

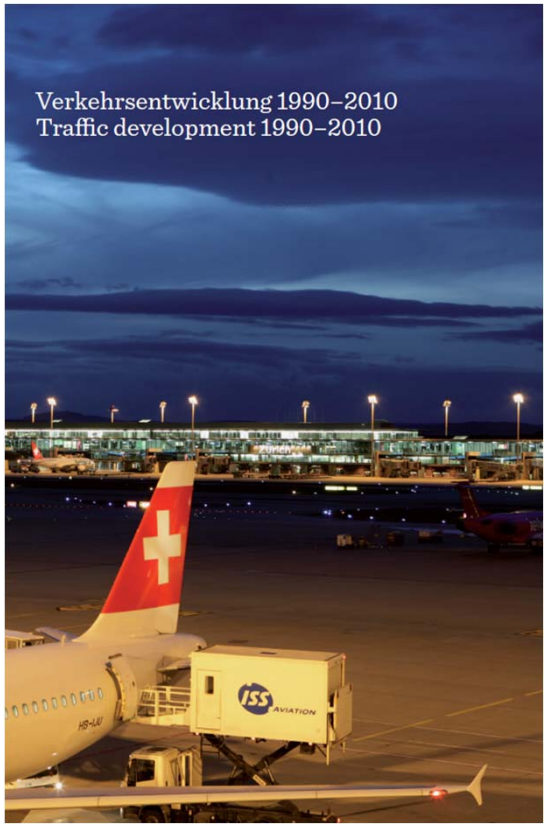
http://stat.ethz.ch/~dettling

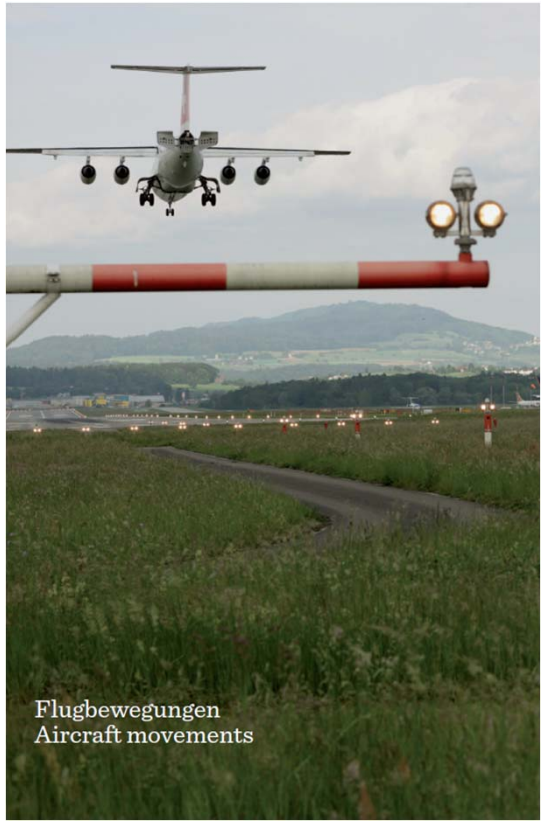
ETH Zürich, October 08, 2012

Applied Statistical Regression
AS 2012 – Week 03

Example: Airline Passengers

Each month, Zurich Airport publishes the number of air traffic movements and airline passengers. We study their relation.

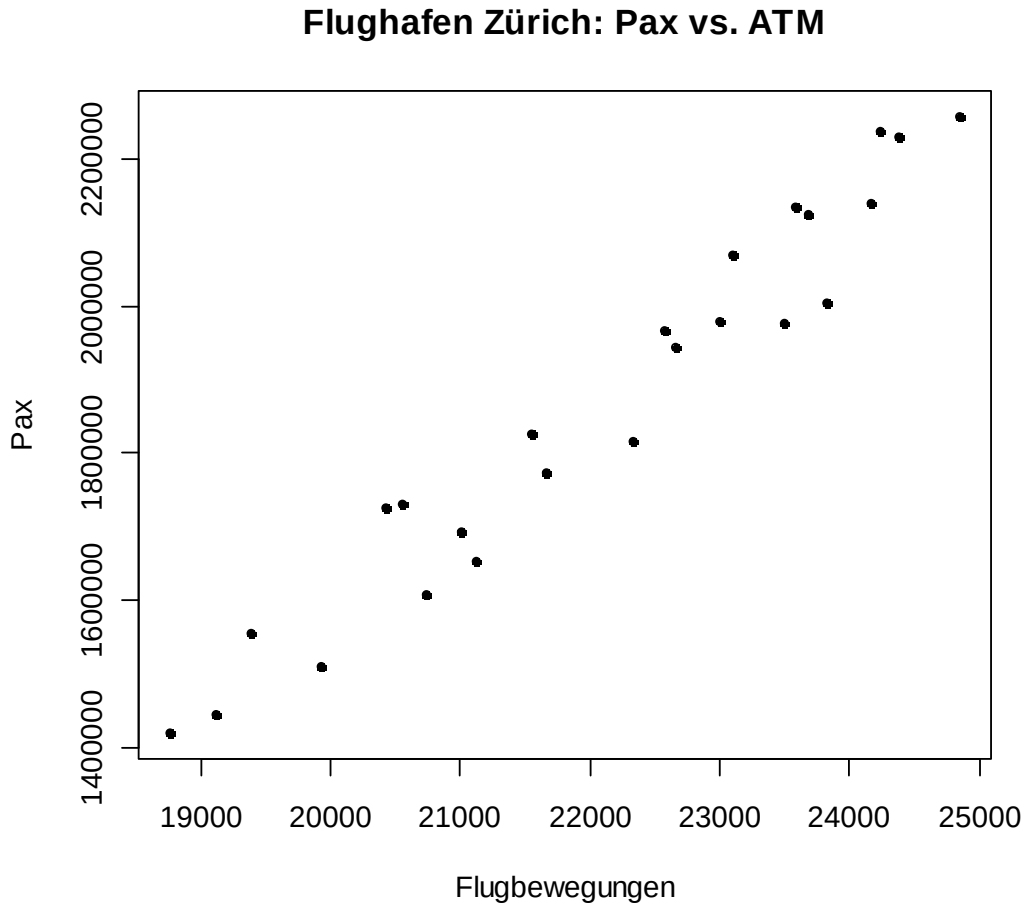




Applied Statistical Regression
AS 2012 – Week 03

Example: Airline Passengers

Month	Pax	ATM
2010-12	1‘730‘629	22‘666
2010-11	1‘772‘821	22‘579
2010-10	2‘238‘314	24‘234
2010-09	2‘139‘404	24‘172
2010-08	2‘230‘150	24‘377
…	…	…

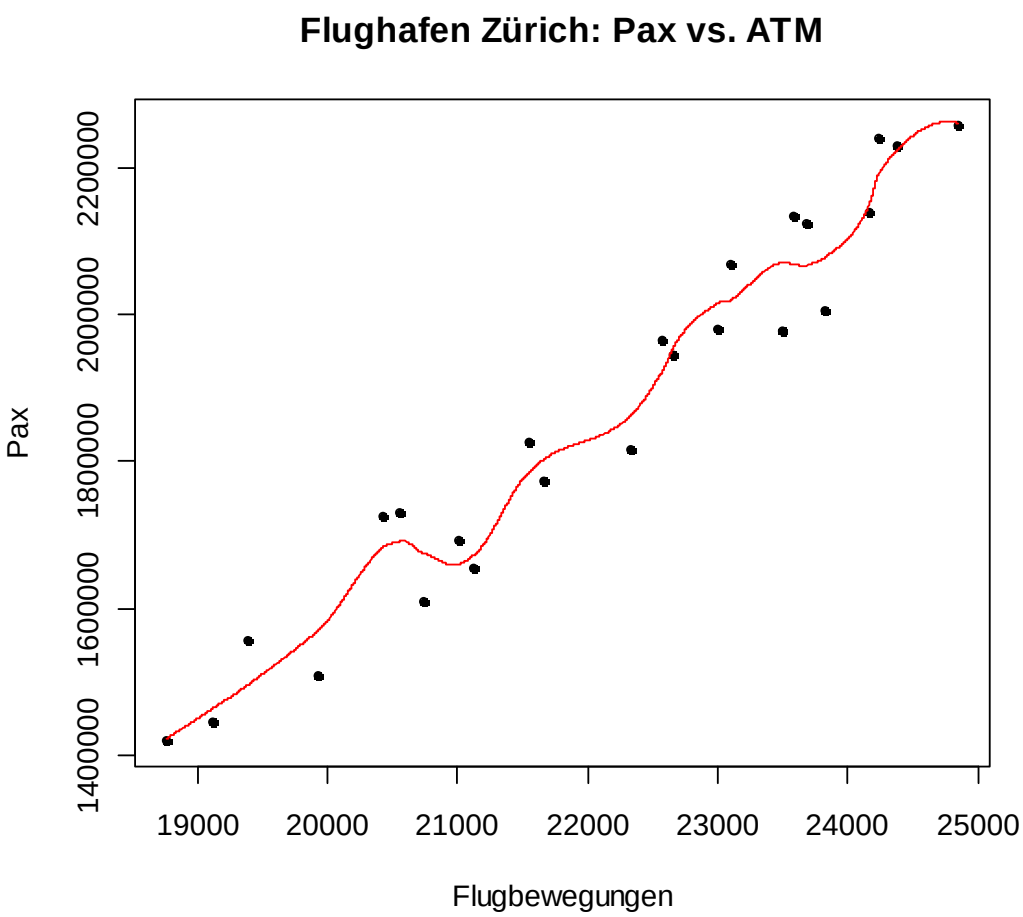


Applied Statistical Regression
AS 2012 – Week 03

Smoothing

We may use an arbitrary smooth function $f(\cdot)$ for capturing the relation between Pax and ATM.

- It should fit well, but not follow the data too closely.
- The question is how the line/function are obtained.

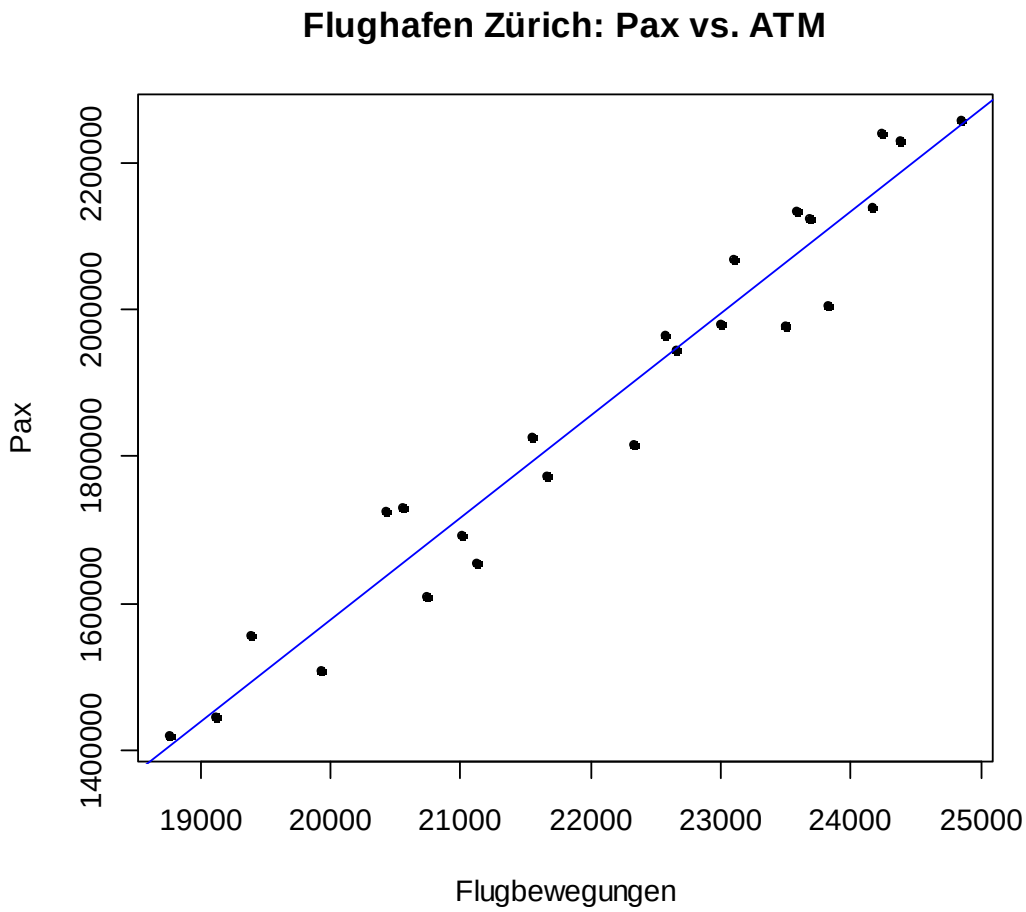


Applied Statistical Regression
AS 2012 – Week 03

Linear Modeling

A straight line represents the systematic relation between Pax and ATM.

- Only appropriate if the true relation is indeed a straight line
- The question is how the line/function are obtained.



Applied Statistical Regression
AS 2012 – Week 03

Simple Linear Regression

The more air traffic movements, the more passengers there are. The relation seems to be linear, which is of course also the mathematically most simple way of describing the relation.

$$f(x) = \beta_o + \beta_1 x, \text{ resp. } Pax = \beta_0 + \beta_1 \cdot ATM$$

Name/meaning of the two parameters in the equation:

β_0 = "Intercept"

β_1 = "Slope"

Fitting a straight line into a 2-dimensional scatter plot is known as **simple linear regression**. This is because:

- there is just one single predictor variable ("*simple*").
- the relation is linear in the parameters ("*linear*").

Model, Data & Random Errors

No we are bringing the data into play. The regression line will not run through all the data points. Thus, there are random errors:

$$y_i = \beta_0 + \beta_1 x_i + E_i, \text{ for all } i = 1, \ldots, n$$

Meaning of variables/parameters:

y_i is the response variable (Pax) of observation i.

x_i is the predictor variable (ATM) of observation i.

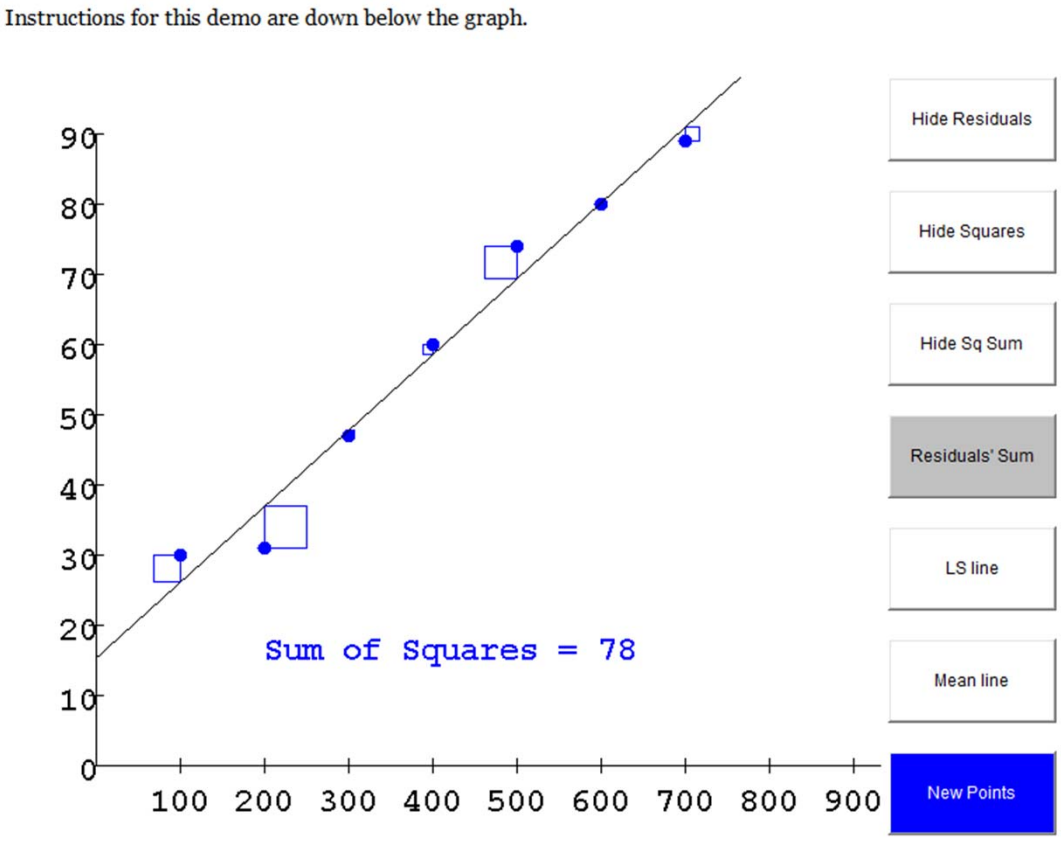
β_0, β_1 are the regression coefficients. They are unknown previously, and need to be estimated from the data.

E_i is the residual or error, i.e. the random difference between observation and regression line.

Applied Statistical Regression
AS 2012 – Week 03

Least Squares Fitting

→ http://hspm.sph.sc.edu/courses/J716/demos/LeastSquares/LeastSquaresDemo.html



We need to fit a straight line that fits the data well.

Many possible solutions exist, some are good, some are worse.

Our paradigm is to fit the line such that the squared errors are minimal.

Least Squares: Mathematics

The paradigm in verbatim...

Given a set of data points $(x_i, y_i)_{i=1,...,n}$, the goal is to fit the regression line such that the sum of squared differences between observed value y_i and regression line is minimal. The function

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta x_i))^2 = \min!$$

measures, how well the regression line, defined by β_0, β_1, fits the data. The goal is to minimize this "quality function".

Solution: → **see next slide...**

Solution Idea: Partial Derivatives

- We are taking partial derivatives on the function $Q(\beta_0, \beta_1)$ with respect to both arguments β_0 and β_1. As we are after the minimum of the function, we set them to zero:

$$\frac{\partial Q}{\partial \beta_0} = 0 \quad \text{and} \quad \frac{\partial Q}{\partial \beta_1} = 0$$

- This results in a linear equation system, which (here) has two unknowns β_0, β_1, but also two equations. These are also known under the name *normal equations*.

- The solution for β_0, β_1 can be written explicitly as a function of the data pairs $(x_i, y_i)_{i=1,\ldots,n}$, **see next slide...**

Least Squares: Solution

According to the least squares paradigm, the best fitting regression line is, i.e. the optimal coefficients are:

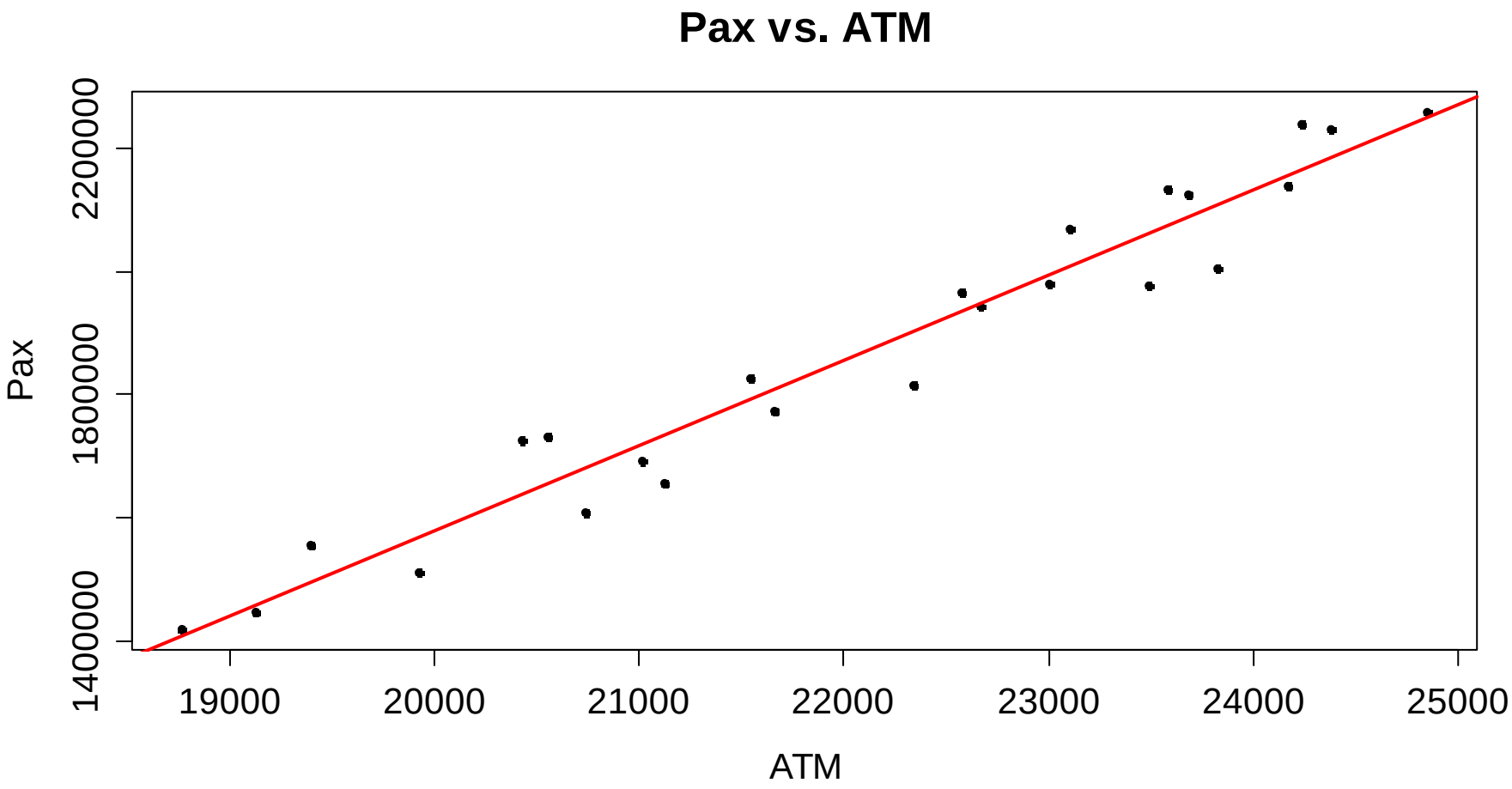
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{und} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

- For a given set of data points $(x_i, y_i)_{i=1,\ldots,n}$, we can determine the solution with a pocket calculator (...or better, with R).
- **The solution for our example Pax vs. ATM:**
  $\hat{\beta}_1 = 138.8, \ \hat{\beta}_0 = -1'197'682$

→ `lm(Pax ~ ATM, data=airpax)`

Applied Statistical Regression
AS 2012 – Week 03

Least Squares Regression Line



Applied Statistical Regression
AS 2012 – Week 03

Is This a Good Model for Predicting the Pax Number from the ATM?

a) Beyond the range of observed data

Unknown, but most likely not...

b) Within the range of observed data

Yes, under the following conditions:

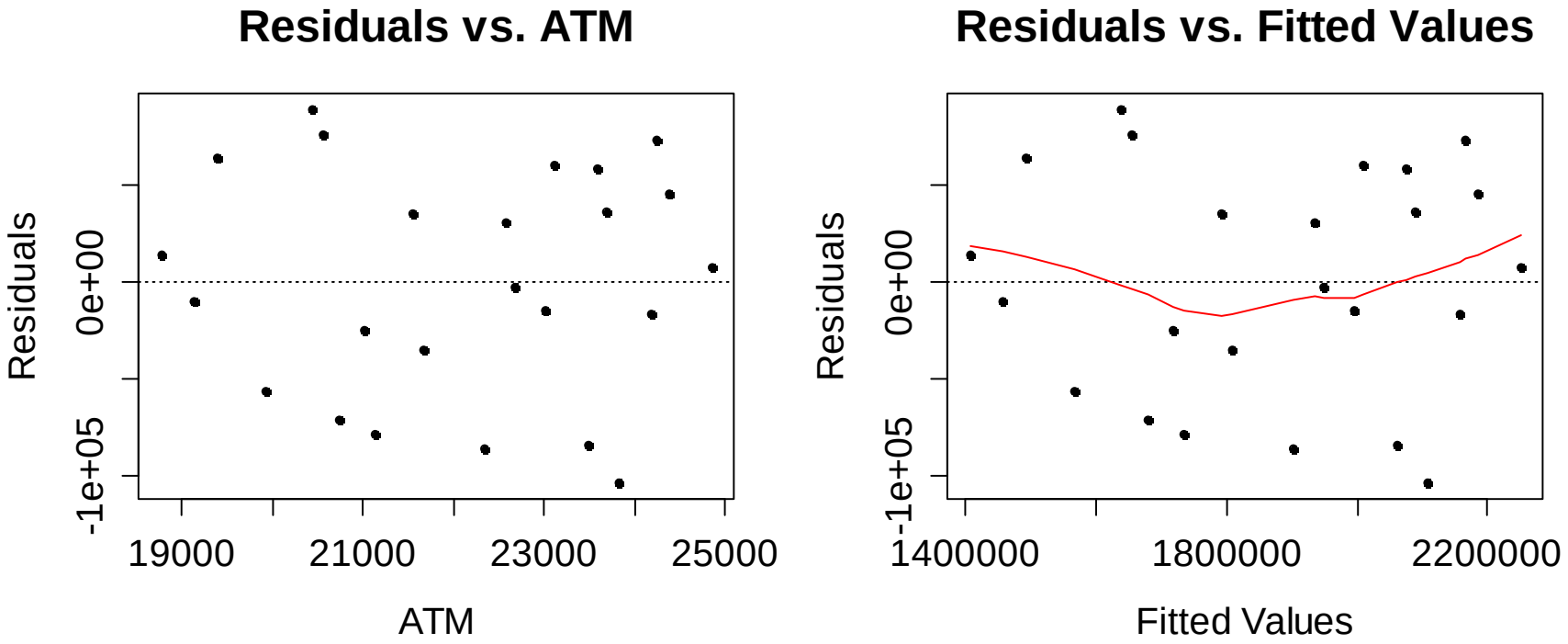
- the relation is in truth a straight line, i.e. $E[E_i] = 0$
- the scatter of the errors is constant, i.e. $Var(E_i) = \sigma^2$
- the data are uncorrelated (from a representative sample)
- the errors are approximately normally distributed

→ **Fodder for thougt: 9/11, SARS, Eyjafjallajökull...?**

Applied Statistical Regression
AS 2012 – Week 03

Model Diagnostics

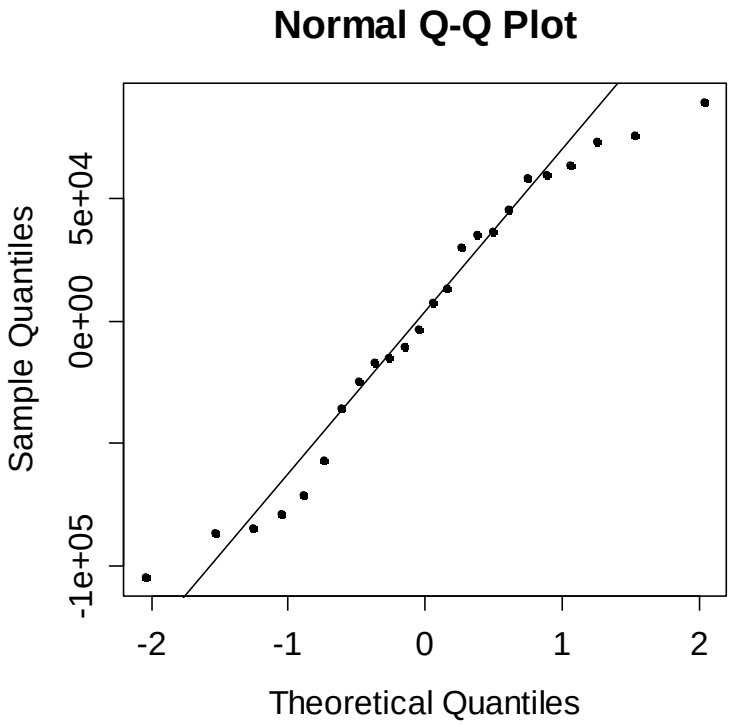
For assessing the quality of the regression line, we need to (at least roughly) check whether the assumptions are met: $E[E_i] = 0$ and $Var(E_i) = \sigma^2$ can be reviewed by:



Applied Statistical Regression
AS 2012 – Week 03

Model Diagnostics

For assessing the quality of the regression line, we need to (at least roughly) check whether the assumptions are met:
Gaussian distribution can be reviewed by:



We will revisit model diagnostics again later in this course, where it will be discussed more deeply.

"Residuals vs. Fitted" and the "Normal Plot" will always stay at the heart of model diagnostics.

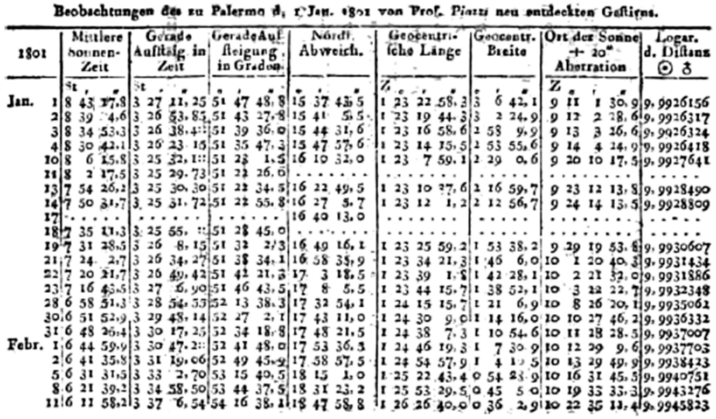
Applied Statistical Regression
AS 2012 – Week 03

Why Least Squares?

History...

Within a few years (1801, 1805), the method was developed independently by Gauss and Legendre. Both were after solving applied problems in astronomy...

Source: → http://de.wikipedia.org/wiki/Methode_der_kleinsten_Quadrate



Beobachtungen des zu Palermo d. 1. Jan. 1801 von Prof. Piazzi neu entdeckten Gestirns.

1801	Mittlere Sonnen-Zeit	Gerade Aufsteig. in Zeit	Gerade Aufsteigung in Graden	Nördl. Abweich.	Geocentrische Länge	Geocentr. Breite	Ort der Sonne + 20" Abberation	Logar. d. Distanz ☉ ♃
	h ′ ″	h ′ ″	° ′ ″	° ′ ″	♈ ° ′ ″	° ′ ″	♈ ° ′ ″	
Jan. 1	8 43 37,8	3 27 21, 25	51 47 48, 8	15 37 42, 5	1 23 22 58, 2	3 6 42, 1	9 11 1 30, 9	9, 9926156
2	8 39 4,6	3 26 53, 85	51 43 27,8	15 41 5, 5	1 23 19 44, 3	3 2 24, 9	9 12 2 28, 6	9, 9926317
3	8 34 32,3	3 26 38,4	51 39 36, 0	15 44 31, 6	1 23 16 58, 6	2 58 9, 9	9 13 3 26, 6	9, 9926324
4	8 30 42,1	3 26 23, 15	51 35 47, 3	15 47 57, 6	1 23 14 15, 5	2 53 55, 6	9 14 4 24, 9	9, 9926418
10	8 6 15,8	3 25 32, 1	51 23 1, 5	16 10 32, 0	1 23 7 59, 1	2 29 0, 6	9 20 10 17, 5	9, 9927641
11	8 2 17,5	3 25 29, 73	51 22 26, 0	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
13	7 54 26,2	3 25 30, 30	51 22 34, 5	16 22 49, 5	1 23 10 27, 6	2 16 59, 7	9 23 12 13, 8	9, 9928490
14	7 50 31,7	3 25 31, 72	51 22 55, 8	16 27 5, 7	1 23 12 1, 2	2 12 56, 7	9 24 14 13, 5	9, 9928809
17	. . . . .	. . . . .	. . . . .	16 40 13, 0	. . . . .	. . . . .	. . . . .	. . . . .
18	7 35 13,3	3 25 55, 1	51 23 45, 0	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
19	7 31 28,5	3 26 8, 15	51 32 2, 3	16 49 16, 1	1 23 25 59, 2	1 53 38, 2	9 29 19 53, 8	9, 9930607
21	7 24 2,7	3 26 34, 27	51 38 34, 1	16 58 35, 9	1 23 34 21, 3	1 46 6, 0	10 1 20 40, 3	9, 9931434
22	7 20 21,7	3 26 49, 42	51 42 21, 3	17 3 18, 5	1 23 39 1, 8	1 42 28, 1	10 2 21 32, 0	9, 9931886
23	7 16 43,6	3 27 6, 90	51 46 43, 5	17 8 5, 5	1 23 44 15, 7	1 38 52, 1	10 3 22 22, 7	9, 9932348
28	6 58 51,3	3 28 54, 53	52 13 38, 3	17 32 54, 1	1 24 15 15, 7	1 21 6, 9	10 8 26 20, 1	9, 9935061
30	6 51 52,9	3 29 48, 14	52 27 2, 1	17 43 11, 0	1 24 30 9, 0	1 14 16, 0	10 10 27 46, 2	9, 9936332
31	6 48 26,4	3 30 17, 25	52 34 18, 8	17 48 21, 5	1 24 38 7, 3	1 10 54, 6	10 11 28 28, 5	9, 9937007
Febr. 1	6 44 59,9	3 30 47, 21	52 41 48, 0	17 53 36, 2	1 24 46 19, 3	1 7 30, 9	10 12 29 9, 6	9, 9937703
2	6 41 35,8	3 31 19, 06	52 49 45, 9	17 58 57, 5	1 24 54 57, 9	1 4 10, 5	10 13 29 49, 9	9, 9938423
5	6 31 31,5	3 33 2, 70	53 15 40, 5	18 15 1, 0	1 25 22 43, 4	0 54 28, 9	10 16 31 45, 5	9, 9940751
8	6 21 39,2	3 34 58, 50	53 44 37, 5	18 31 23, 2	1 25 53 29, 5	0 45 5, 0	10 19 33 33, 3	9, 9943276
11	6 11 58,2	3 37 6, 54	54 16 38, 1	18 47 58, 8	1 26 26 40, 0	0 36 2, 9	10 22 35 12, 4	9, 9945823

Carl Friedrich Gauss

Adrien-Marie Legendre

Why Least Squares?

Mathematics...

- Least Squares is simple in the sense that the solution is known in closed form as a function of $(x_i, y_i)_{i=1,\ldots,n}$.
- The line runs through the center of gravity $(\bar{x}, \bar{y})$
- The sum of residuals adds up to zero: $\sum_{i=1}^{n} r_i = 0$
- Some deeper mathematical optimality can be shown when analyzing the large sample properties of the estimates $\hat{\beta}_0, \hat{\beta}_1$ This is especially true under the assumption of normally distributed errors E_i.

Gauss-Markov-Theorem

A mathematical optimality result for the Least Squares line

It only holds if the following conditions are met:

- the relation is in truth a straight line, i.e. $E[E_i] = 0$
- the scatter of the errors is constant, i.e. $Var(E_i) = \sigma^2$
- the errors are uncorrelated, i.e. $Cov(E_i, E_j) = 0, \ if \ i \neq j$

Not yet required:

- ~~the errors are normally distributed: $E_i \sim N(0, \sigma_E^2)$~~

Gauss-Markov-Theorem:

- Least Squares yields the *best linear unbiased estimates*

Properties of the Least Square Estimates

Under the conditions above, the estimates are unbiased:

$$E[\hat{\beta}_0] = \beta_0 \quad \text{and} \quad E[\hat{\beta}_1] = \beta_1$$

The variances of the estimates are as follows:

$$Var(\hat{\beta}_0) = \sigma_E^2 \cdot \left(\frac{1}{n} + \frac{\bar{x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right) \quad \text{and} \quad Var(\hat{\beta}_1) = \frac{\sigma_E^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Precise estimates are obtained with:

- a large number of observations n
- a good scatter in the predictor x_i
- an informative/useful predictor, making σ_E^2 small
- (an error distribution which is approximately Gaussian)

Benefits of Linear Regression

- **Inference on the relation between y and x**

The goal is to understand if and how strongly the response variable depends on the predictor. There are performance indicators as well as statistical test adressing the issue.

- **Prediction of (future) observations**

The regression line/equation can be employed to predict the PAX number for any given ATM value.

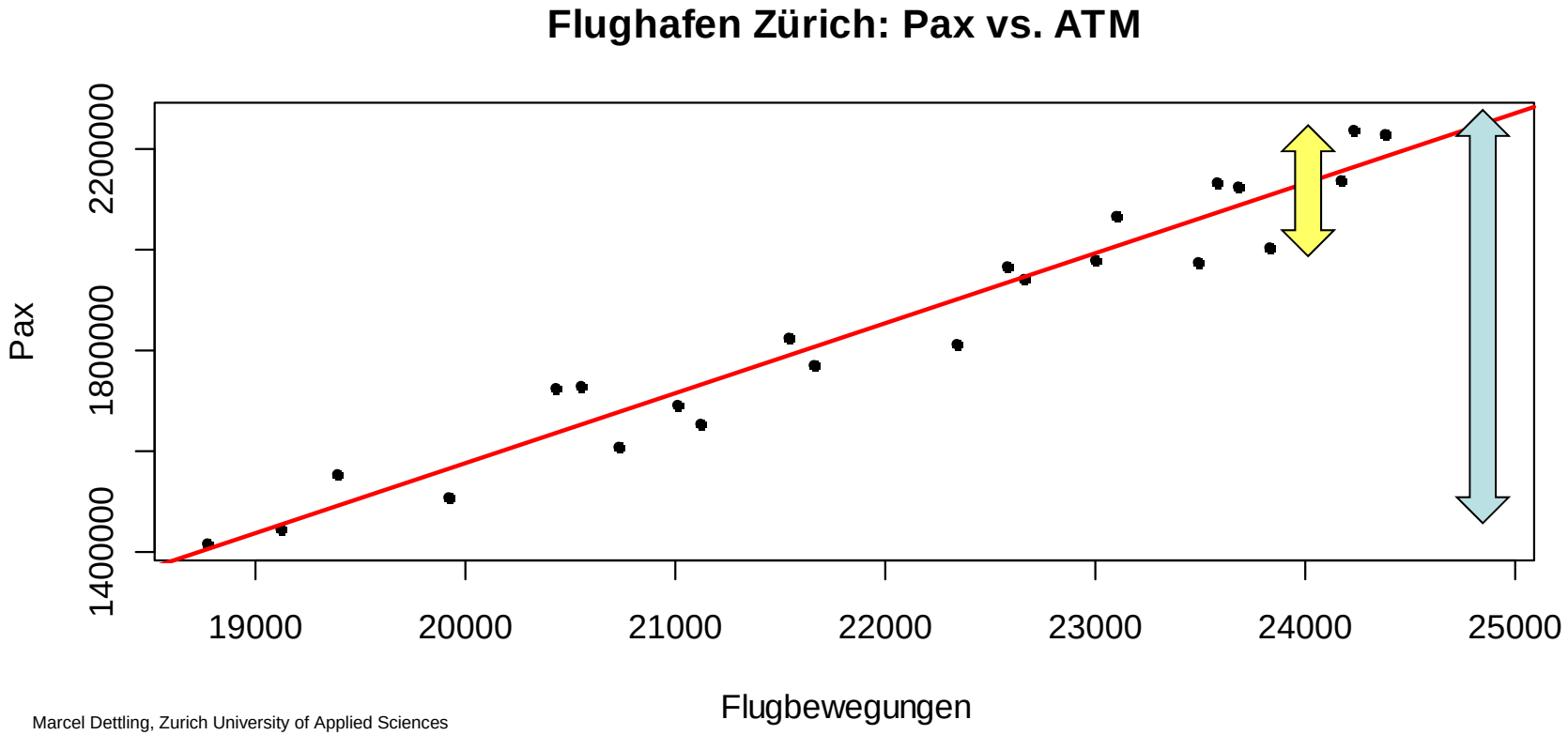
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

However, this mostly will not work well for extrapolation!

Applied Statistical Regression
AS 2012 – Week 03

R^2: *The Coefficient of Determination*

The coefficient of determination R^2 is also known as *multiple R-squared*. It tells which portion of the total variation is accounted for by the regression line.



Applied Statistical Regression
AS 2012 – Week 03

Computation of R^2

R^2 is the portion of the total variation that is explained through regression. It is determined as one minus the quotient of the yellow arrow divided by the blue arrow.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \in [0,1]$$

The closer to 1 the value is, the tighter the datapoints are packed around the regression line. However, there are no formal criteria which R^2 value needs to be met such that the regression can be said to be useful/valid.

Confidence Interval for the Slope β_1

A 95%-CI for the slope β_1 tells which values (besides the point estimate $\hat{\beta}_1$) are plausible, too. The uncertainty is due to estimation/sampling effects.

95%-CI for β_1 : $\hat{\beta}_1 \pm qt_{0.975;n-2} \cdot \hat{\sigma}_{\hat{\beta}_1}$, resp.

$$\hat{\beta}_1 \pm qt_{0.975;n-2} \cdot \sqrt{\hat{\sigma}_E^2 \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

In R:

```
> fit <- lm(Pax ~ ATM, data=unique2010)
> confint(fit, "ATM")
        2.5 %  97.5 %
ATM 124.4983 153.025
```

Testing the Slope β_1

There is a statistical hypothesis test which can be used to check whether the slope is significantly different from zero, or any other arbitrary value b . The null hypothesis is:

$$H_0 : \beta_1 = 0, \text{ bzw. } H_0 : \beta_1 = b$$

One usually tests two-sided on the 95%-level. The alternative is:

$$H_A : \beta_1 \neq 0, \text{ bzw. } H_A : \beta_1 \neq b$$

As a test statistic, we use:

$$T = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}}, \text{ resp. } T = \frac{\hat{\beta}_1 - b}{\hat{\sigma}_{\hat{\beta}_1}}, \text{ both follow a } t_{n-2} \text{ distribution.}$$

Applied Statistical Regression
AS 2012 – Week 03

Reading R-Output

```
> summary(lm(Pax ~ ATM, data=dat))

Call: lm(formula = Pax ~ ATM, data = dat)

Residuals:    Min      1Q  Median      3Q     Max
          -104188  -40885    2099   48588   89154

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.198e+06  1.524e+05  -7.858 7.94e-08 ***
ATM          1.388e+02  6.878e+00  20.176 1.11e-15 ***

---

Residual standard error: 59700 on 22 degrees of freedom
Multiple R-squared: 0.9487, Adjusted R-squared: 0.9464
F-statistic: 407.1 on 1 and 22 DF,  p-value: 1.110e-15
```

→ **Will be explained in detail on the blackboard!**

Testing the Slope β_1

Practical Example:

Use the Pax vs. ATM data and perform a statistical test for the null hypothesis $H_0 : \beta_1 = 150$. The information from the summary on slide 25 can be used as a basis. Then, also answer:

a) *Explain in colloquial language what was just tested. What is the benefit of this test? What claims could motivate the test?*

b) *How does the testing result relate with the 95%-CI that we computed on slide 23? Would we be able to tell the test results from the CI alone?*

→ **See blackboard for the answers**

Testing the Intercept β_0

An analogous test can be done for the intercept.

- No matter what the test result will be, the intercept should generally not be omitted from the regression model.
- The presence of the intercept protects against possible non-linearities and calibration errors of measurement devices. If it is kicked out of the model, the results are generally worse.
- If theory dictates that there should not be an intercept but it is still significant, take this as evidence that the linear relation does not hold when extrapolating to $x = 0$.

Applied Statistical Regression
AS 2012 – Week 03

Prediction

Using the regression line, we can predict the y-value for any desired x-value. The result is the expectation for y given x.

$$E[y \mid x] = \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad \text{a.k.a. } \textbf{\textit{"fitted value"}}$$

Example: With 24'000 air traffic movements, we expect

$$-1'197'682 + 24'000 \cdot 138.8 = 2'133'518 \text{ Passengers}$$

Be careful:

At best, interpolation within the range of observed x-values is trustworthy. Extrapolation with ATM values such as 50'000, 5'000 or even 0 usually produces completely useless results.

Applied Statistical Regression
AS 2012 – Week 03

Prediction with R

We can use the regression fit object for prediction. The syntax for obtaining the fitted value(s) is as follows:

```
> fit <- lm(Pax ~ ATM, data=unique2010)
> dat <- data.frame(ATM=c(24000))
> predict(fit, newdata=dat)
1 2132598
```

The x-values need to be provided in a data frame, where the variable/column name is identical to the predictor name.

Then, the `predict()` procedure is invoked with the regression fit and the new x-values as arguments.

Applied Statistical Regression
AS 2012 – Week 03

Confidence Interval for $E[y \mid x]$

We just computed the fitted value $\hat{\beta}_0 + \hat{\beta}_1 x$, i.e. the expected number of passengers for 24'000 ATMs. This is not a deterministic value, but an estimate that is subject to variability.

A 95%-CI for the fitted value at position x is given by:

$$\hat{\beta}_0 + \hat{\beta}_1 x \pm qt_{0.975;n-2} \cdot \hat{\sigma}_E \cdot \sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}}$$

In R:
```
> predict(fit, newdata=dat, interval="confidence")
        fit     lwr     upr
1 2132598 2095450 2169746
```

Applied Statistical Regression
AS 2012 – Week 03

Prediction Interval for $\hat{y}$

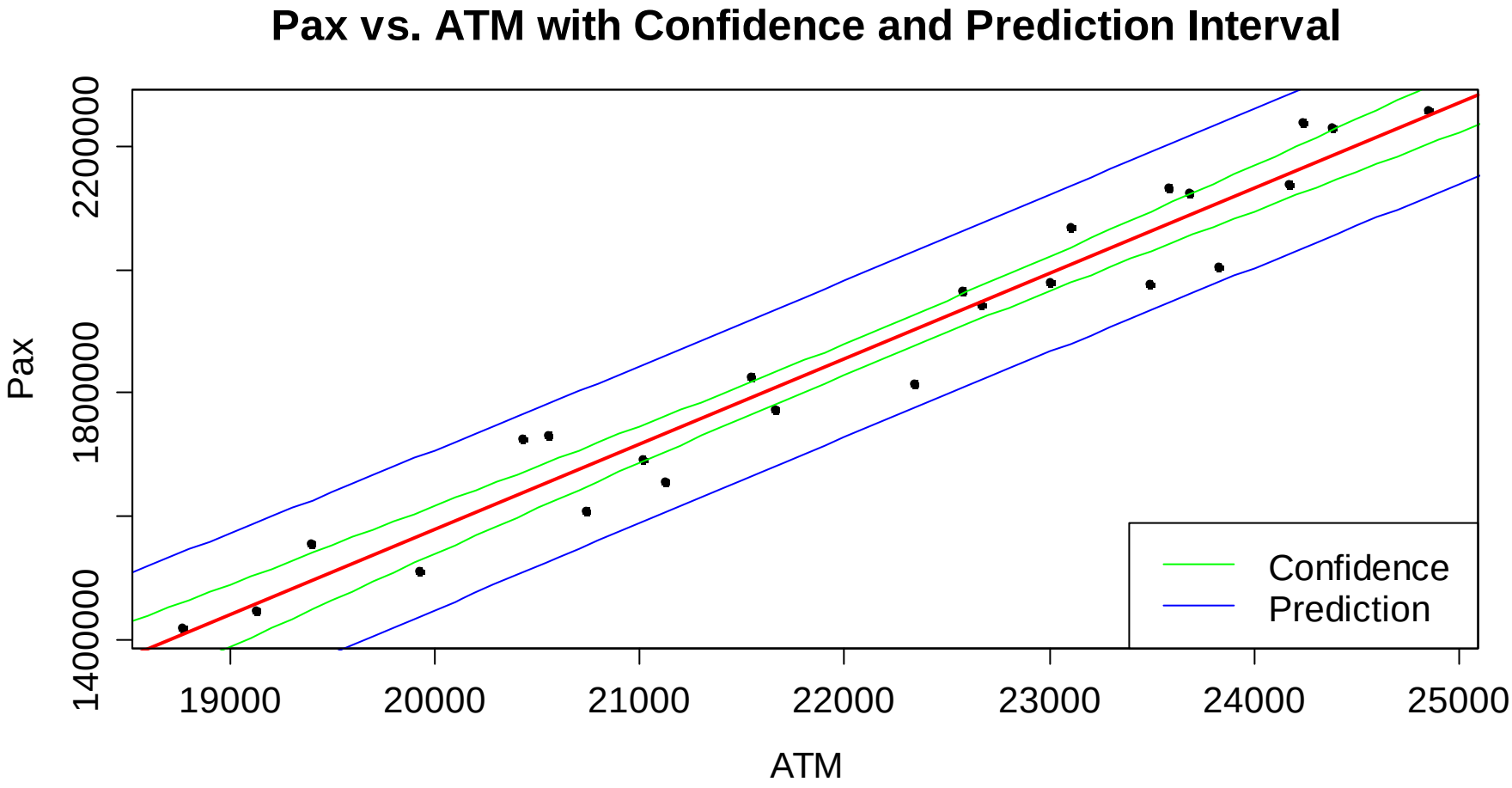
The confidence interval for $E[y \mid x]$ tells about the variabilty of the fitted value. It does not account for the scatter of the data points around the regression line and thus does not define a region where we have to expect the observed value. A 95% prediction interval at position x is given by:

$$\hat{\beta}_0 + \hat{\beta}_1 x \pm qt_{0.975;n-2} \cdot \hat{\sigma}_E \cdot \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

In R:
```
> predict(fit,newdata=dat,interval="prediction")
        fit     lwr     upr
1 2132598 2003343 2261853
```

Applied Statistical Regression
AS 2012 – Week 03

Confidence and Prediction Interval



Applied Statistical Regression
AS 2012 – Week 03

Confidence and Prediction Interval

Note:

Visualizing the confidence and prediction intervals in R is not straightforward, but requires some tedious handwork.

R-Hints:

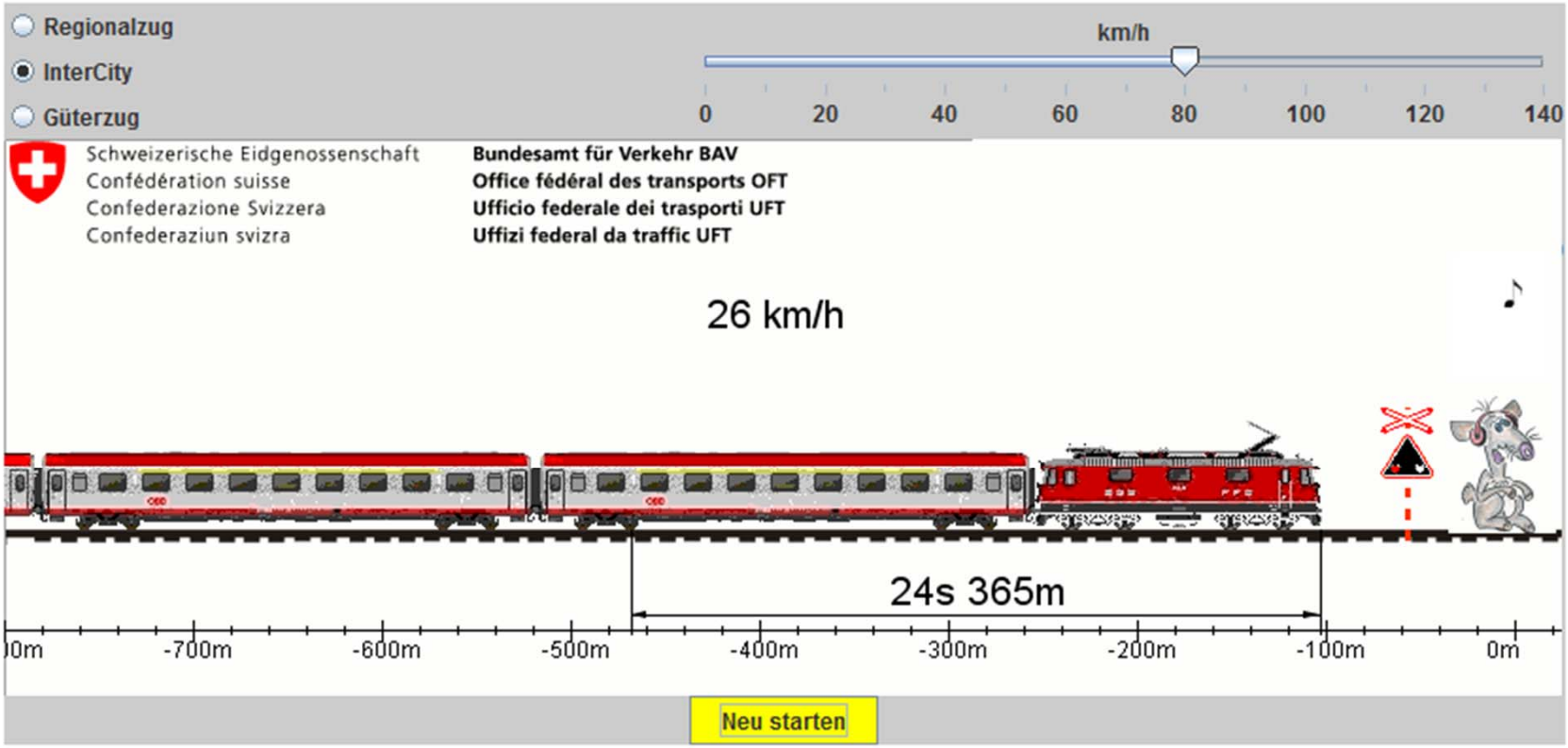
```
dat  <- data.frame(xx=seq(..., ..., length=200))
pred <- predict(fit, newdata=dat, interval=...)
plot(..., ..., main="...")
lines(dat$xx, pred[,2], col=...)
lines(dat$xx, pred[,3], col=...)
```

Applied Statistical Regression
AS 2012 – Week 03

Example: Braking Distance for Railway

Animation „Bremsweg Eisenbahn“

Zurück zur Übersicht Bremsweg Eisenbahn

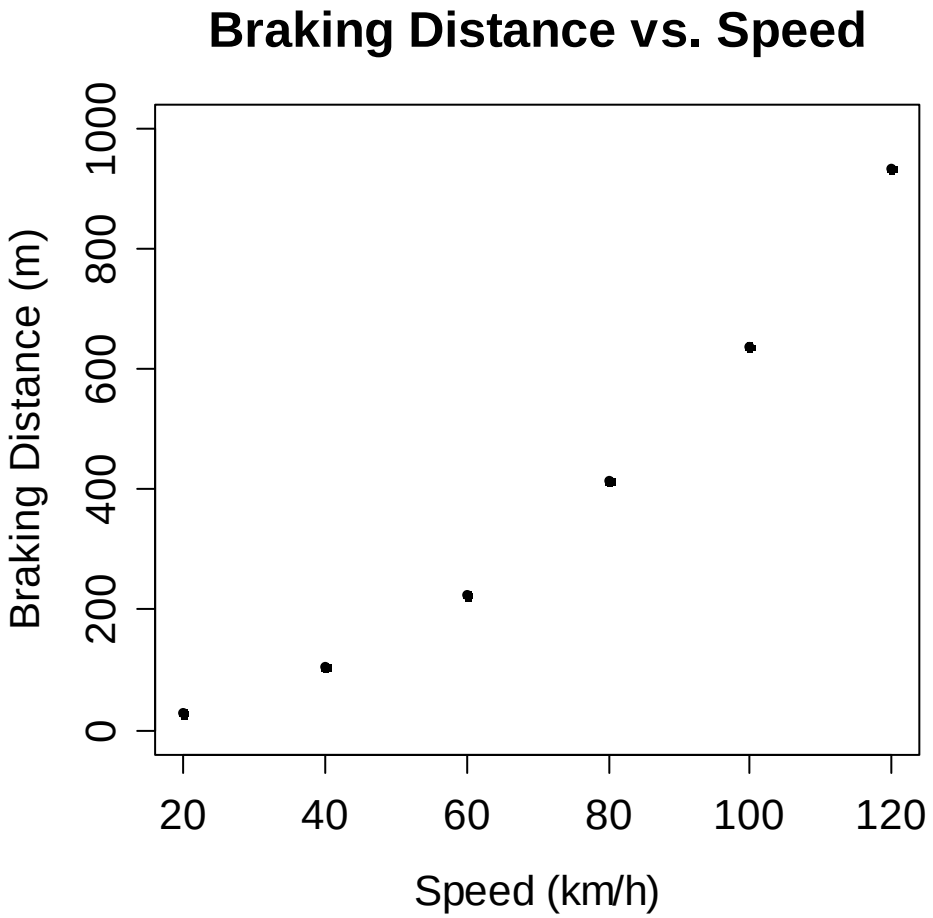


Source: http://www.bav.admin.ch/aktuell/bremswege/index.html?lang=de

Applied Statistical Regression
AS 2012 – Week 03

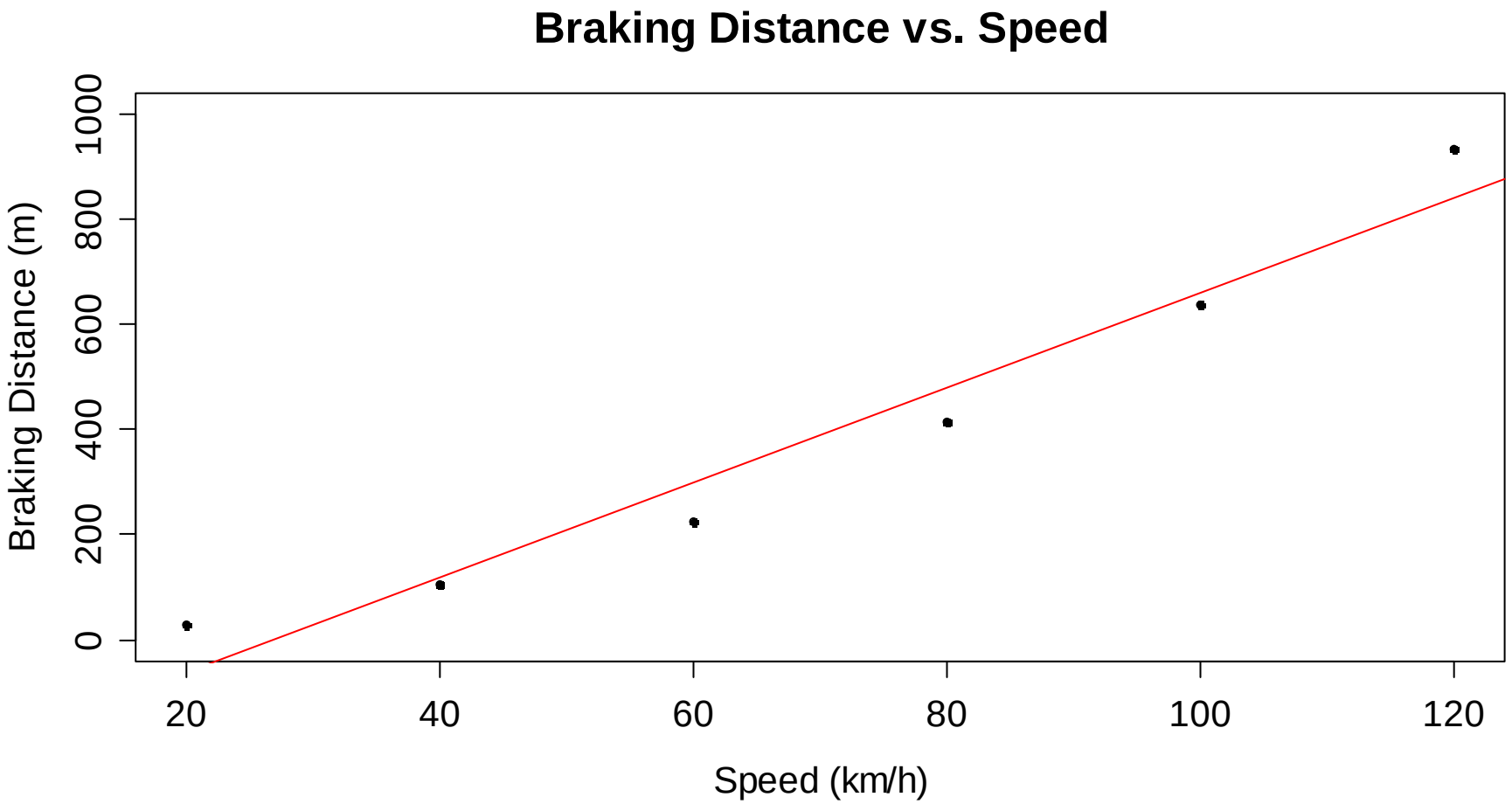
Braking Distance: Data

Beob.	Speed	Weg
1	20	27
2	40	104
3	60	226
4	80	414
5	100	638
6	120	934



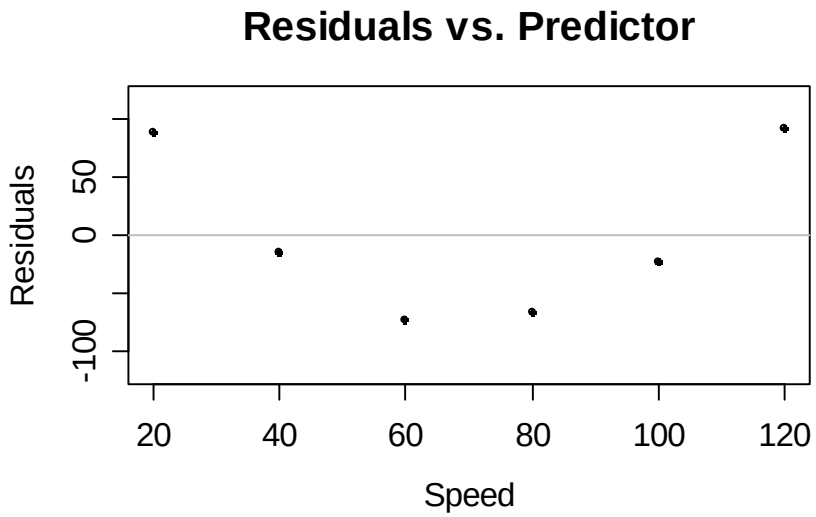
Applied Statistical Regression
AS 2012 – Week 03

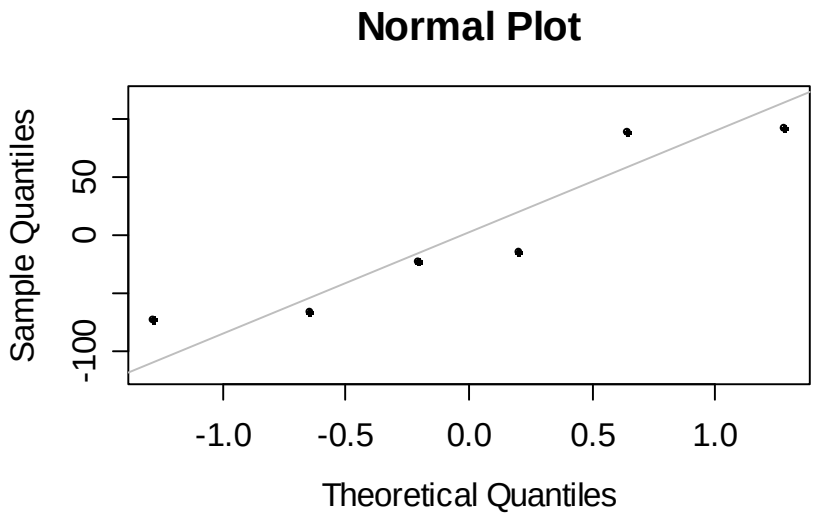
Braking Distance: Fitting a Straight Line

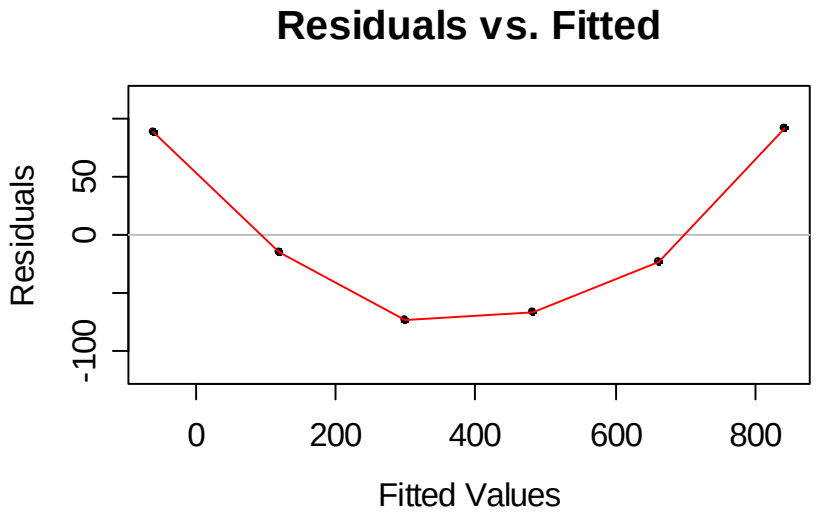


Applied Statistical Regression
AS 2012 – Week 03

Braking Distance: Residual Diagnostics







Braking Distance: Facts

Conclusion from the residual plots:

- A straight line does not correctly describe the relation braking distance vs. speed relation. Energy is proportional to the square of speed, thus we need a quadratic function.

$$BrDist_i = \beta_0 + \beta_1 \cdot Speed_i^2 + E_i$$

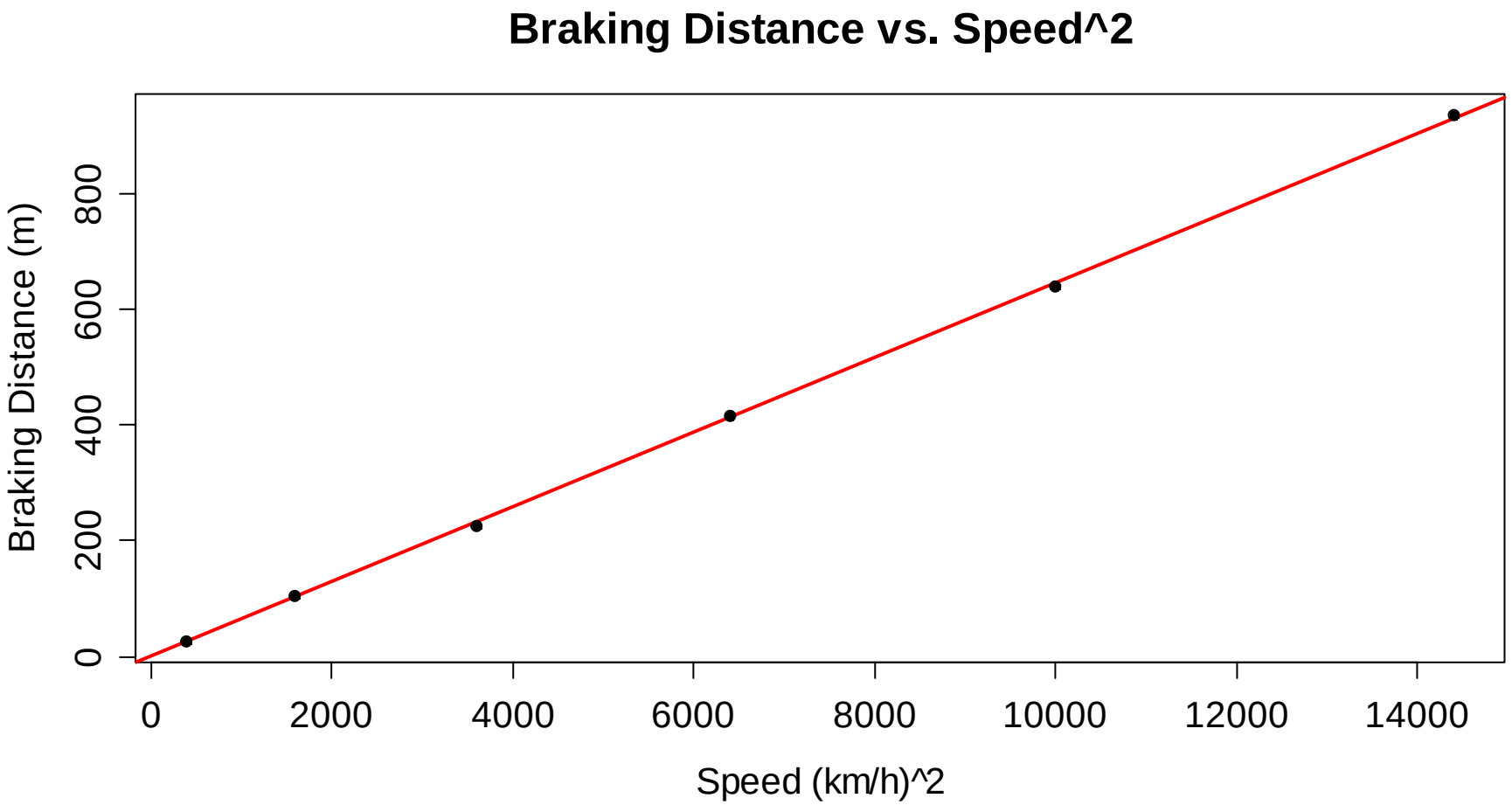
bzw. $Y_i = \beta_0 + \beta_1 \cdot x_i' + E_i$, wobei $x_i' = x_i^2 = Speed_i^2$

- The above model still is a simple linear regression problem. There is only one single predictor, the optimal coefficients $\hat{\beta}_0, \hat{\beta}_1$ can be explicitly determined by using the ordinary LS-algorithm.

Applied Statistical Regression
AS 2012 – Week 03

Braking Distance vs. Speed^2

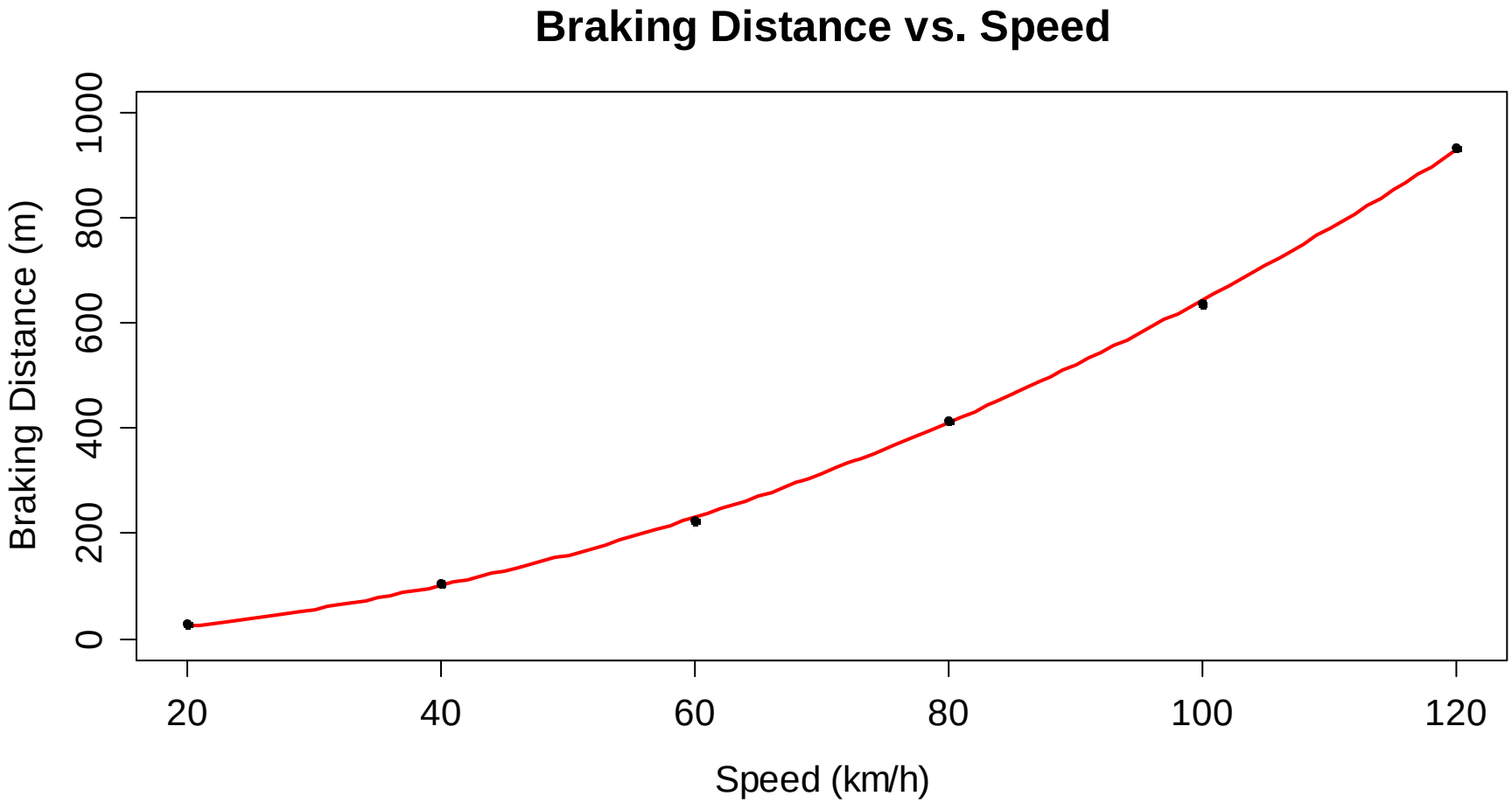
```
> fit02 <- lm(weg ~ I(speed^2))
```



Applied Statistical Regression
AS 2012 – Week 03

Braking Distance: Fitting a Parabola

```
> fit02 <- lm(weg ~ I(speed^2))
```



Curvilinear Regression

We have seen that simple linear regression offers more than just fitting straight lines. Using the ordinary LS-algorithm, we can fit any curvilinear relation, e.g.:

- $Y_i = \beta_0 + \beta_1 \cdot \ln(x_i) + E_i$
- $Y_i = \beta_0 + \beta_1 \cdot \sqrt{x} + E_i$
- $Y_i = \beta_0 + \beta_1 \cdot x^{-1} + E_i$

Note that the predictor gets transformed, i.e. $x_i' = \ln(x_i)$, $x_i' = \sqrt{x_i}$ or $x_i' = (x_i)^{-1}$. Using this notation, it is obvious that these are simple linear regressions.

→ **BUT...** see next slide!

Braking Distance: Some Thoughts

Curvilinear models rarely yield a good fit:

- For modelling braking distance, reaction time should also be accounted for. This yields a multiple regression:

$$BrDist_i = \beta_0 + \beta_1 \cdot Speed_i + \beta_2 \cdot Speed_i^2 + E_i$$

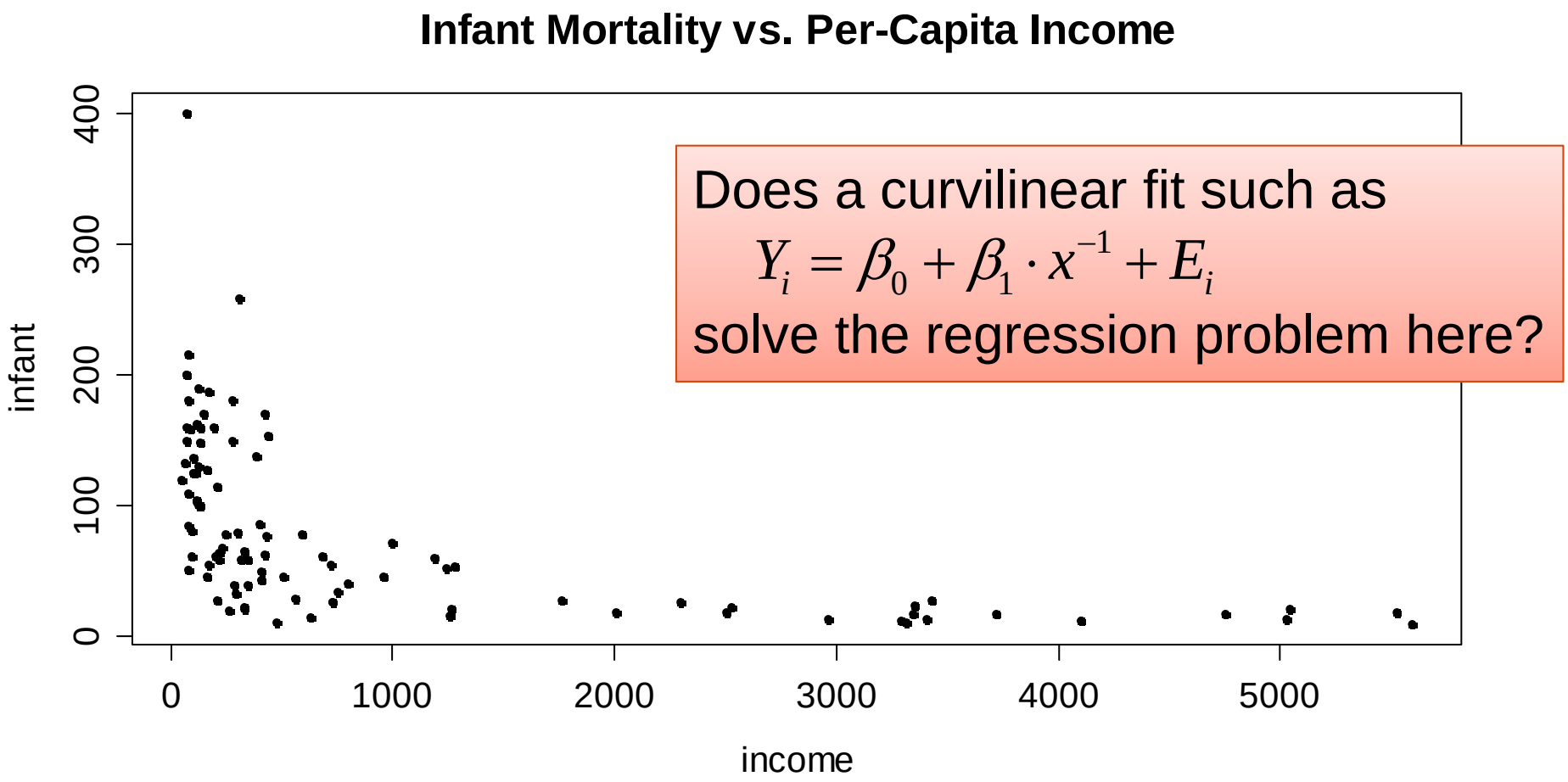
- The scatter of the errors is not constant: the higher the speed, the higher the variation in braking distance.

- There are many applications where the exponent is unknown and needs to be estimated from data:

$$Y_i = \beta_0 + \beta_1 \cdot x^{\beta_2} + E_i$$

Applied Statistical Regression
AS 2012 – Week 03

Infant Mortality vs. Per-Capita Income



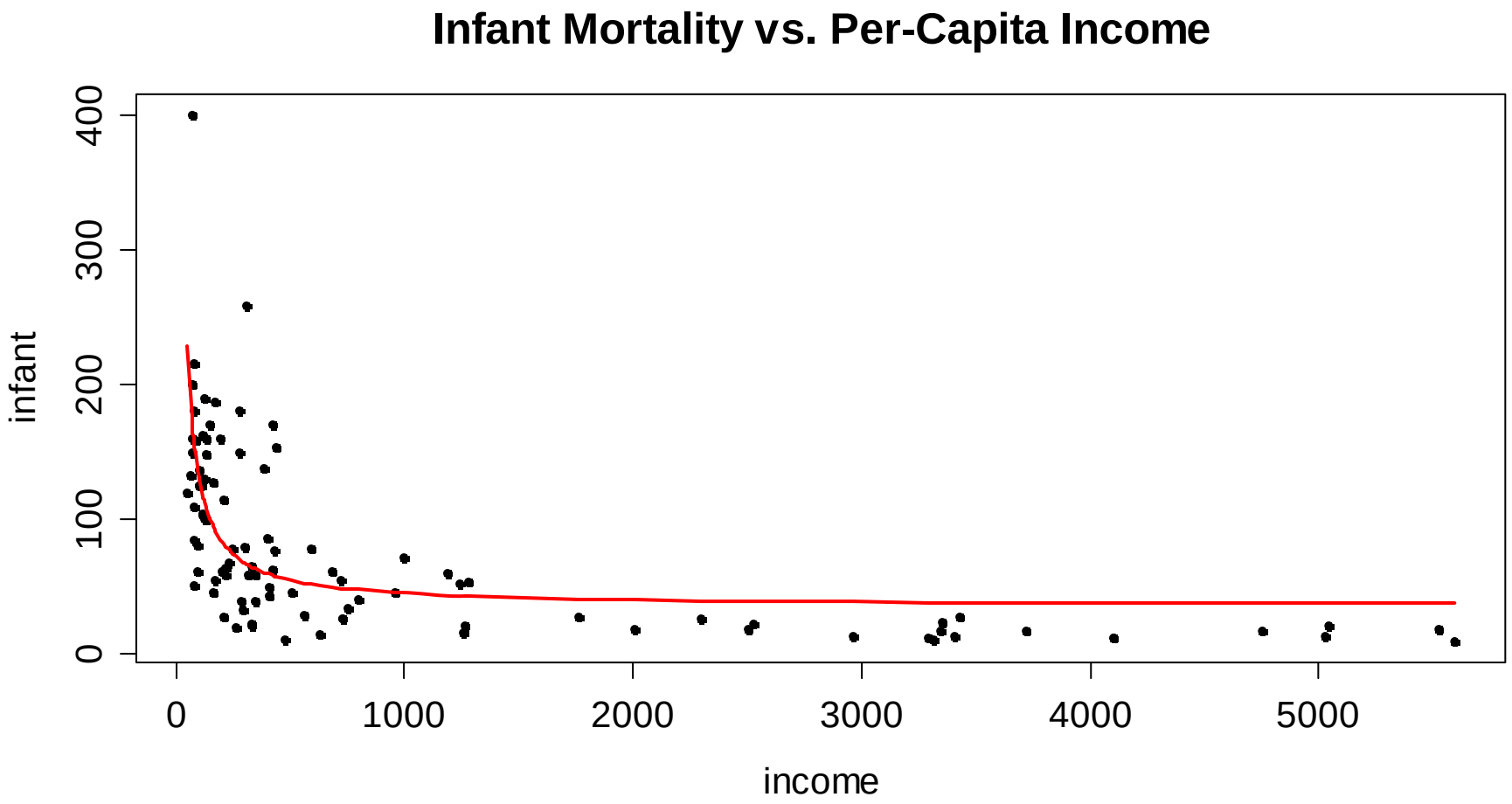
Does a curvilinear fit such as

$$Y_i = \beta_0 + \beta_1 \cdot x^{-1} + E_i$$

solve the regression problem here?

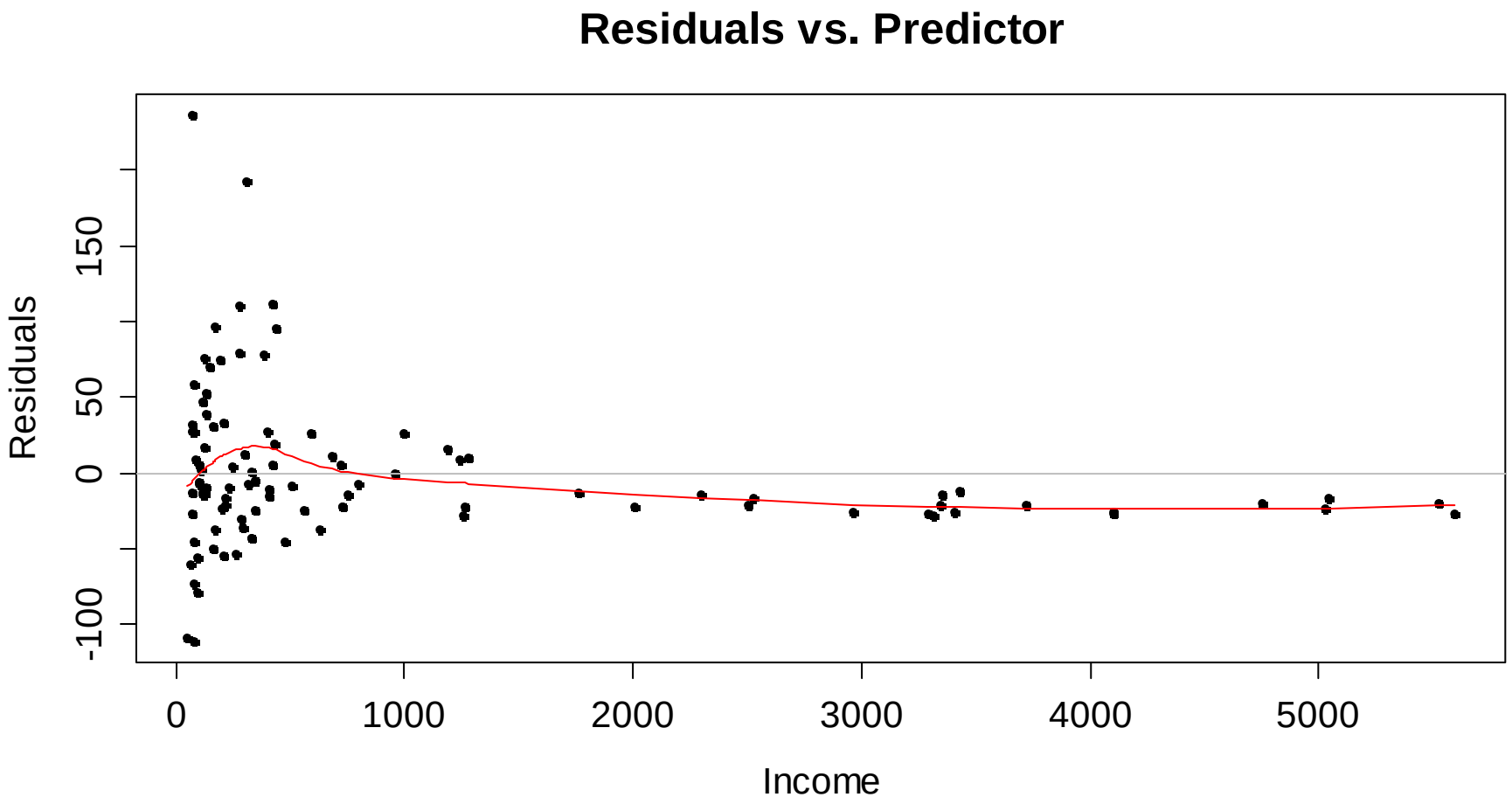
Applied Statistical Regression
AS 2012 – Week 03

Infant Mortality vs. Per-Capita Income



Applied Statistical Regression
AS 2012 – Week 03

Residual Analysis of Hyperbolic Fit



Conclusions from Residual Analysis

The problem with the previous fit is that the the **power** x^{-1} is **not correct**. Adjusting it by hand is very laborious. If we try a model such as

$$Y_i = \beta_0 + \beta_1 \cdot x_i^{\beta_2} + E_i$$

we do **no longer** have a **linear** problem in the parameters. Thus, least squares fitting cannot be applied for such models.

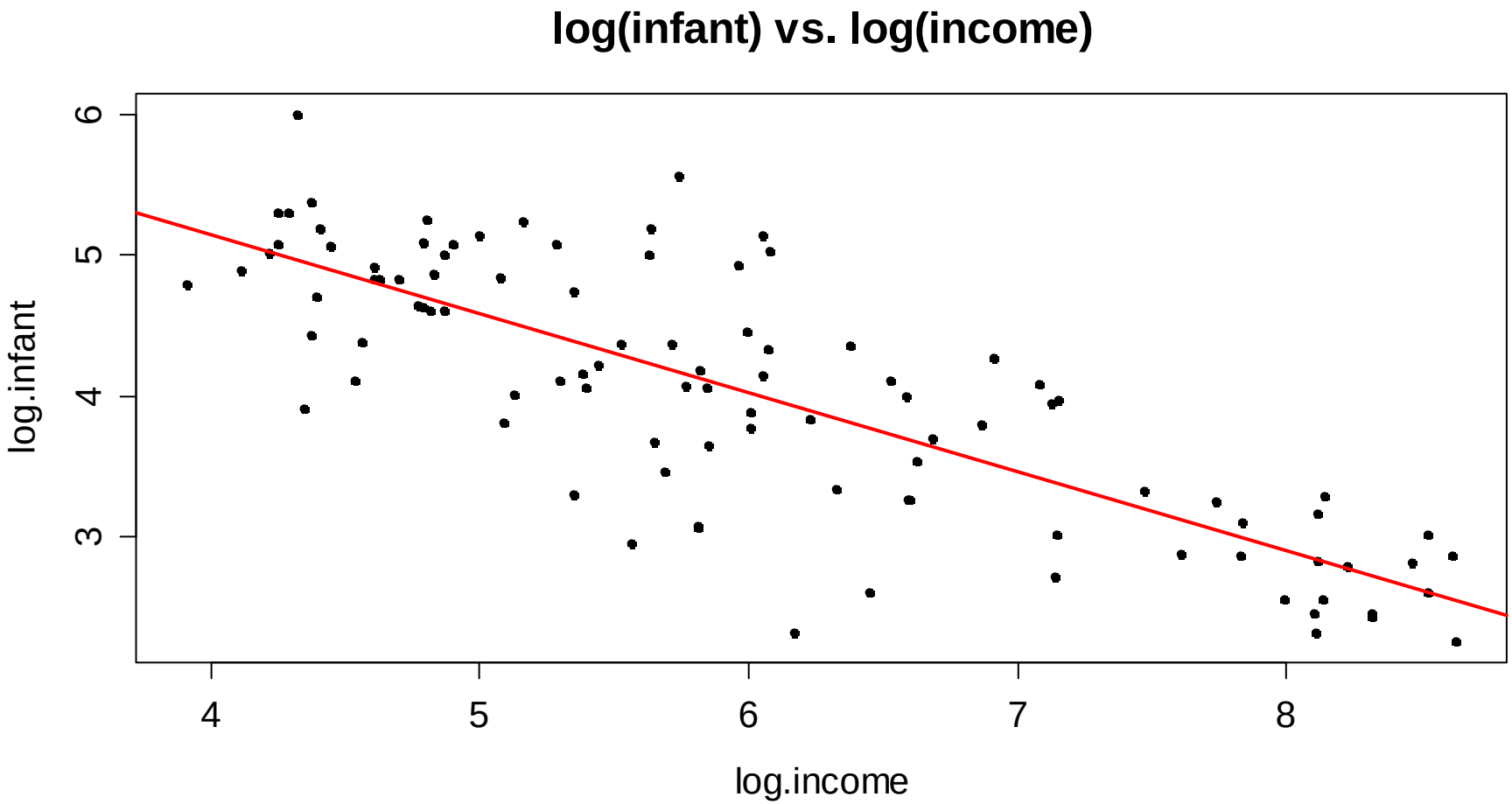
Yet there is a simple, but very powerful trick that often helps:

$$y_i' = \log(y_i), \;\; x_i' = \log(x_i)$$

solves the problem, **see next slide and the blackboard...**

Applied Statistical Regression
AS 2012 – Week 03

Log-Transformation Helps!



Model and Coefficients

If a straight line is fitted on the log-log-scale,

$$y_i' = \beta_0' + \beta_1 \cdot x_i' + E_i', \text{ where } y_i' = \log(y_i),\ x_i' = \log(x_i)\ ,$$

this means fitting the following relation on the original scale:

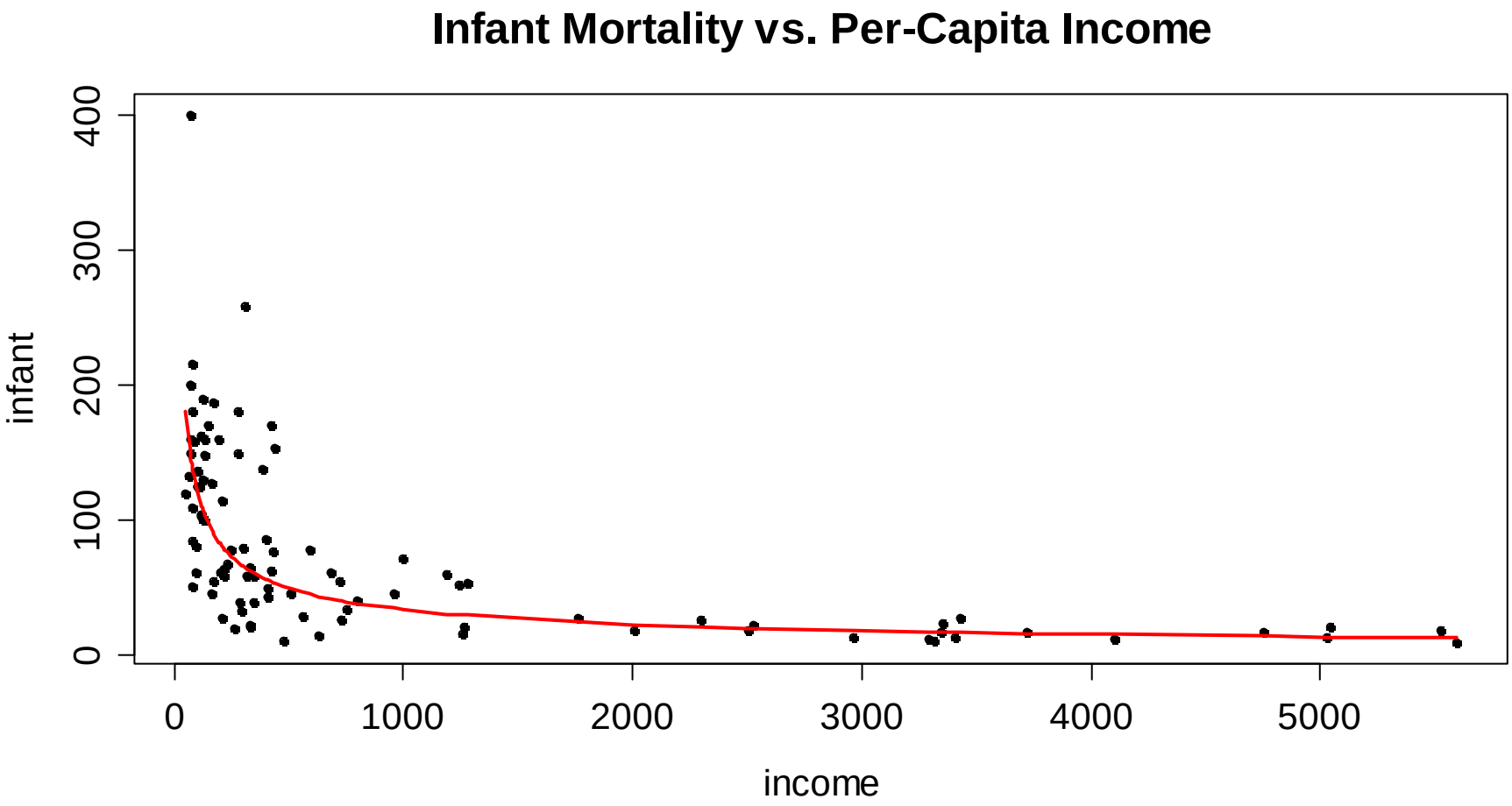
$$y_i = \beta_0 \cdot x_i^{\beta_1} \cdot E_i$$

The meaning of the parameter β_1 is as follows:

If x, i.e. the income increases by 1%, then y, i.e. the mortality decreases by $\hat{\beta}_1 = 0.56\%$. In other words: β_1 characterizes the relative change in y per unit of relative change in x .

Applied Statistical Regression
AS 2012 – Week 03

Fitting on the Original Scale



Fitted Values and Intervals

- To obtain a prediction on the original scale, we can just re-exponentiate to invert the log-transformation.

$$\hat{y} = \exp(\hat{y}')$$

- **Be careful**: that result is an estimate for the median, but not for the mean $E[y \mid x]$. If we require *unbiased estimation*, we need to use a correction factor:

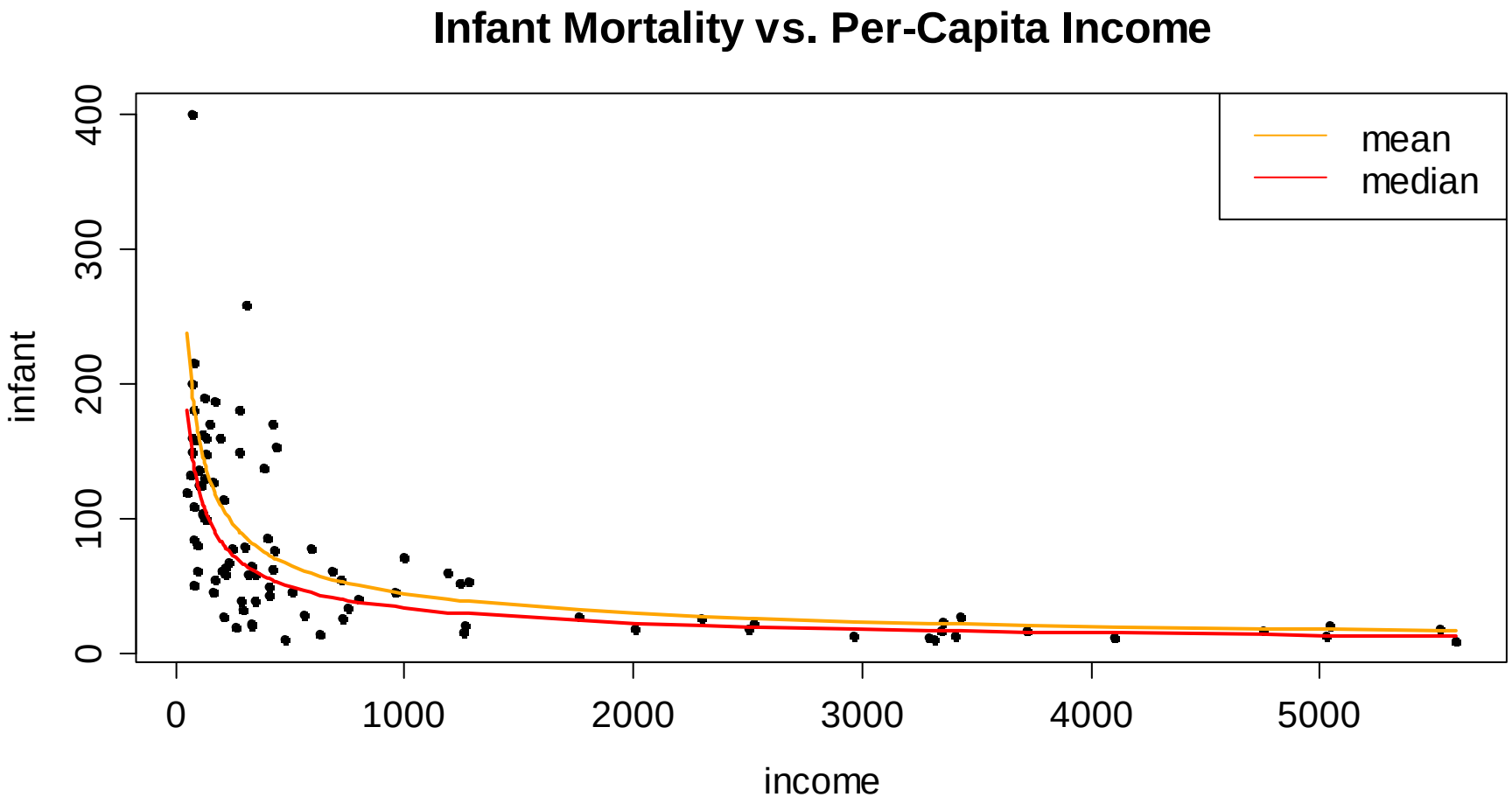
$$\hat{y} = \exp(\hat{y}' + \hat{\sigma}_E^2 / 2)$$

- The confidence and prediction intervals are simply:

$$[l, u] \rightarrow [\exp(l), \exp(u)]$$

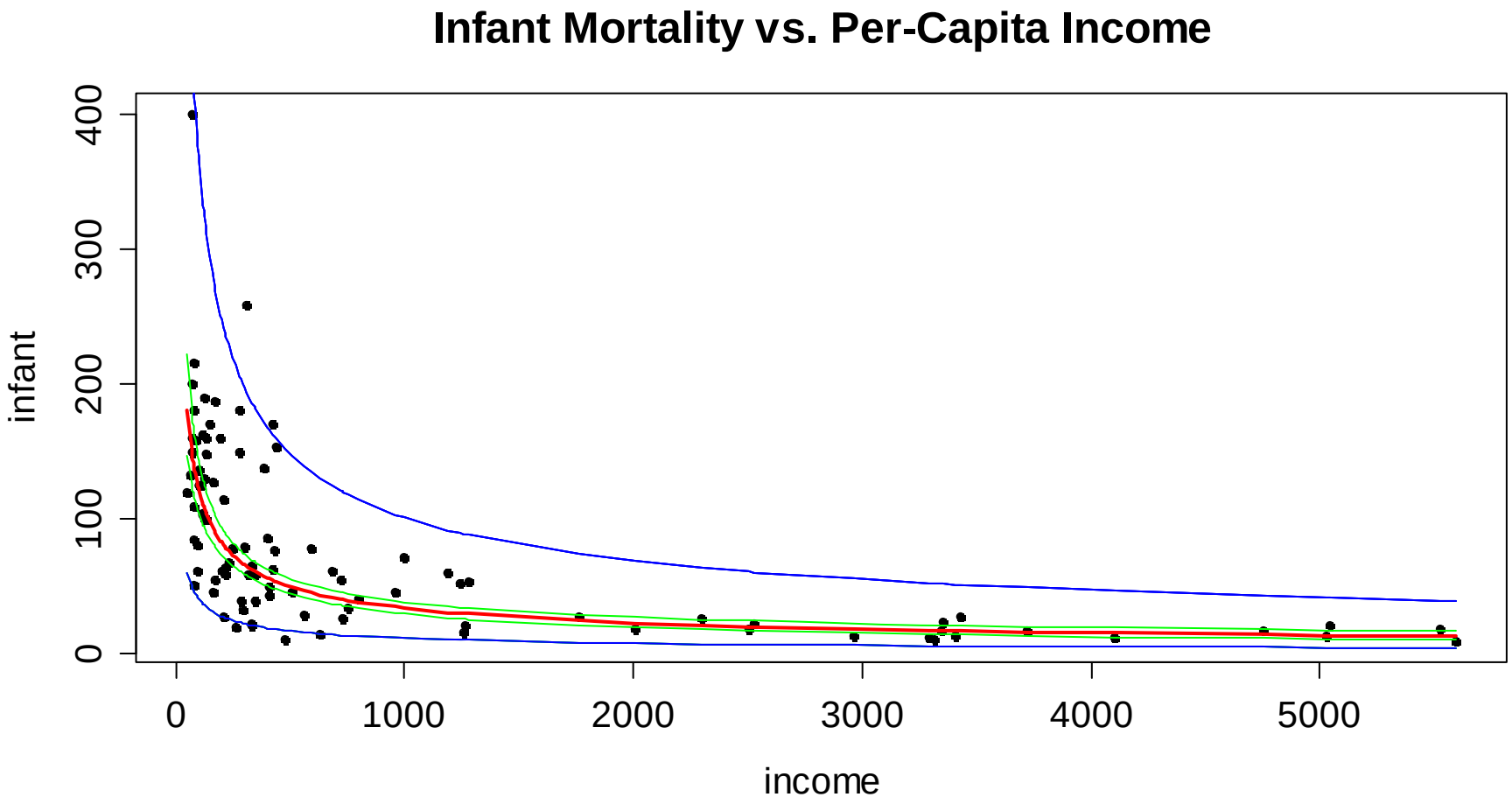
Applied Statistical Regression
AS 2012 – Week 03

Mean and Median



Applied Statistical Regression
AS 2012 – Week 03

Confidence and Prediction Interval



What to Do if y=0 and/or x=0?

- We can only take logarithms if $x, y > 0$. In cases where the response and/or predictor takes negative values, we should not log-transform. If zero's occur, they need treatment.

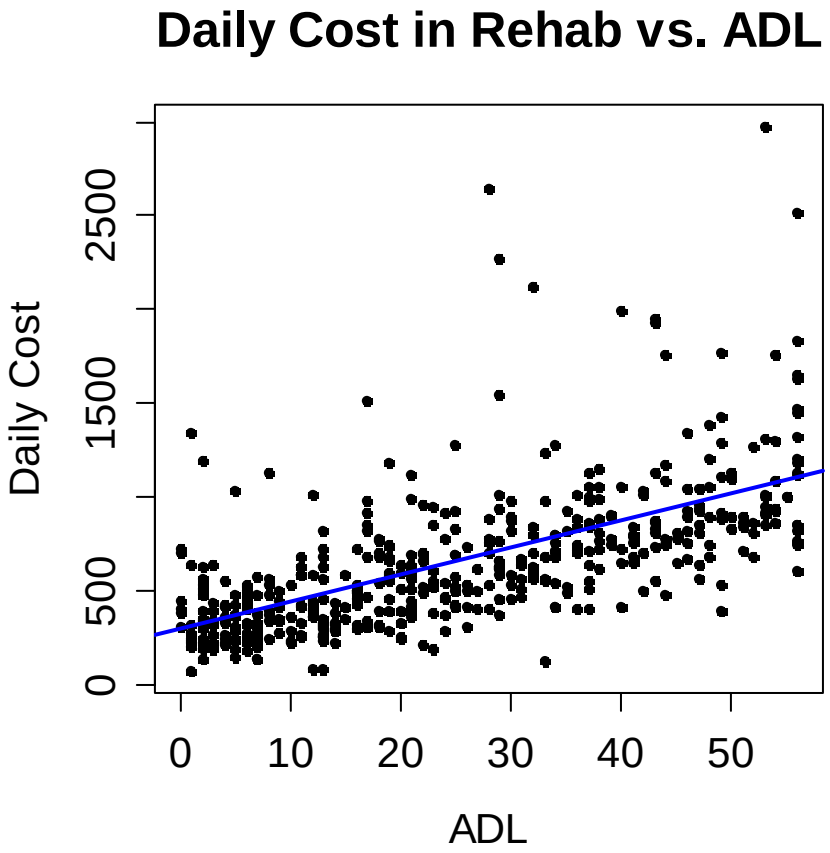
- What do we do with either $x = 0$ or $y = 0$?
  - do never exclude such data points!
  - adding a constant value is allowed!

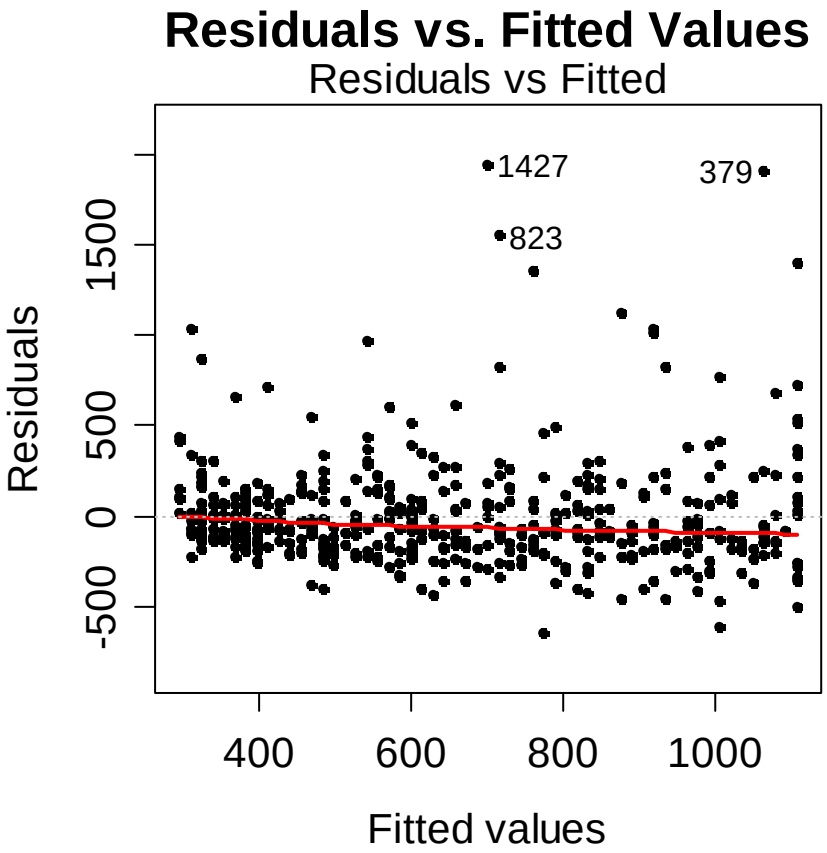
- What about the choice of the constant?
  - standard choice: $c = 1$
  - scale dependent, thus not recommended!

→ **Set $c =$ smallest value > 0!**

Applied Statistical Regression
AS 2012 – Week 03

Another Example: Daily Cost in Rehab





Logged Response Model

We *transform* the *response* variable and try to explain it using a linear model with our previous predictors:

$$Y' = \log(Y) = \beta_0 + \beta_1 x + E$$

In the *original scale*, we can write the logged response model using the same predictors:

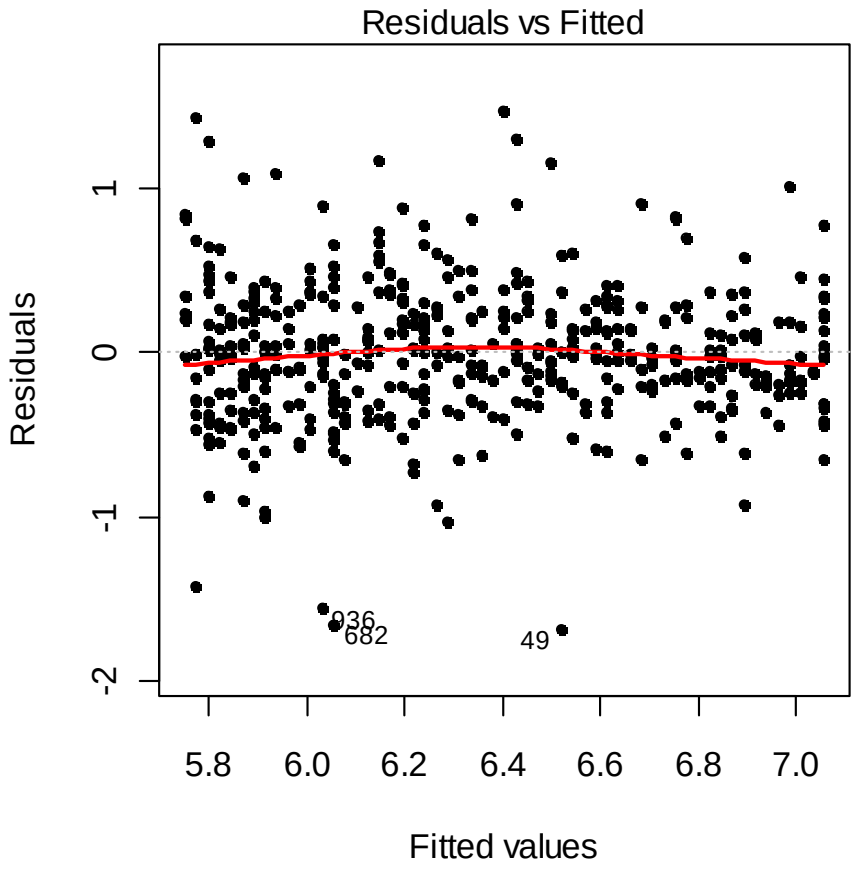
$$Y = \exp(\beta_0 + \beta_1 x) \cdot \exp(E)$$

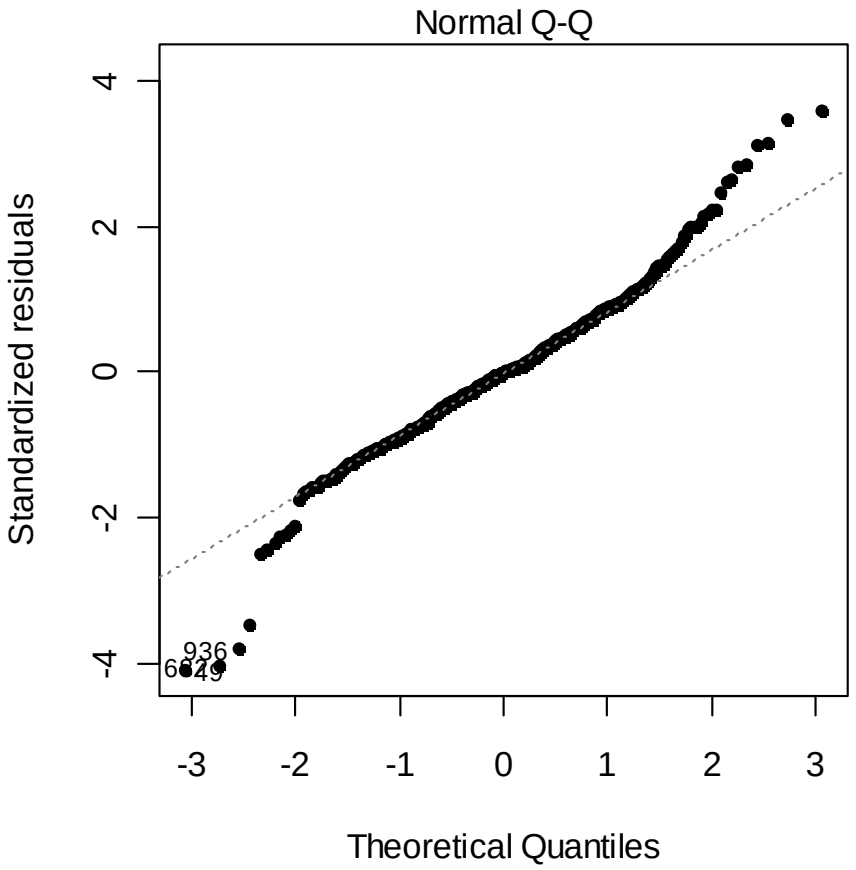
→ **Multiplicative model**

→ $E \sim N(0, \sigma_E^2)$, and thus, $\exp(E)$ has a *lognormal distribution*

Applied Statistical Regression
AS 2012 – Week 03

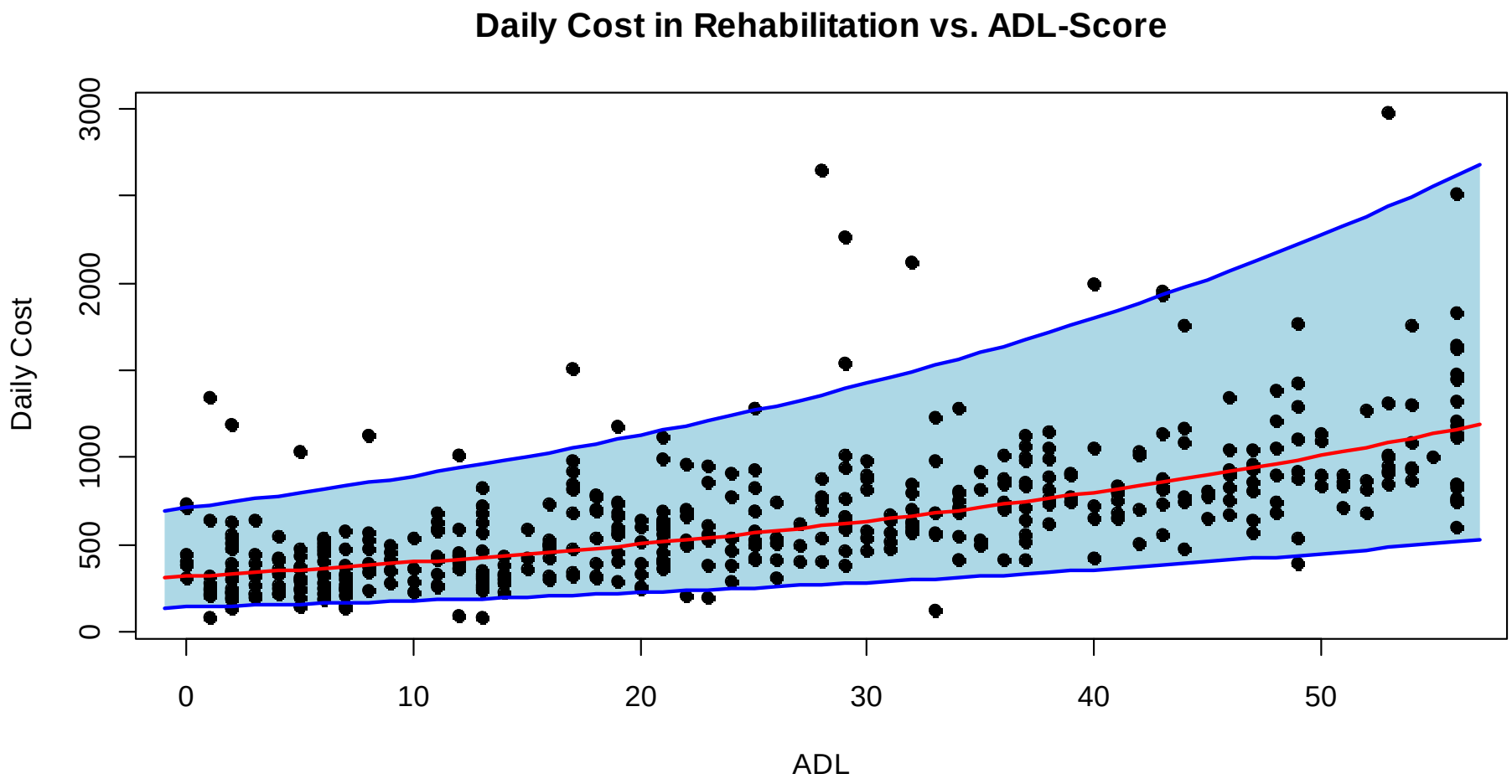
Also This Transformation Works!





Applied Statistical Regression
AS 2012 – Week 03

Fitted Values on the Original Scale



Interpretation of the Coefficients

Important: There is no back transformation for the coefficients to the original scale, but still a good interpretation

$$\begin{aligned} \log(\hat{y}) &= \hat{\beta}_0 + \hat{\beta}_1 x_1 \\ \hat{y} &= \exp(\hat{\beta}_0)\exp(\hat{\beta}_1 x_1) \end{aligned}$$

An increase by one unit in x_1 would multiply the fitted value in the original scale with $\exp(\hat{\beta}_1)$.

→ **Coefficients are interpreted multiplicatively!**