

# Applied Statistical Regression

## AS 2013 – Week 06

*Marcel Dettling*

Institute for Data Analysis and Process Design

Zurich University of Applied Sciences

[marcel.dettling@zhaw.ch](mailto:marcel.dettling@zhaw.ch)

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

ETH Zürich, October 21, 2013

# Applied Statistical Regression

## AS 2013 – Week 06

### *What is Regression?*

**The answer to an everyday question:**

How does a target variable of special interest depend on several other (explanatory) factors or causes.

**Examples:**

- growth of plants, depends on fertilizer, soil quality, ...
- apartment rents, depends on size, location, furnishment, ...
- car insurance premium, depends on age, sex, nationality, ...

**Regression:**

- quantitatively describes relation between predictors and target
- high importance, most widely used statistical methodology

# Applied Statistical Regression

## AS 2013 – Week 06

### *Example: Mortality Due to Air Pollution*

Researchers at General Motors collected data on 60 US Standard Metropolitan Statistical Areas (SMSAs) in a study of whether air pollution contributes to mortality.

City	Mortality	JanTemp	JulyTemp	RelHum	Rain	Educ	Dens	NonWhite	WhiteClr	Pop	House	Income	HC	NOx	SO2
Akron, OH	921.87	27	71	59	36	11.4	3243	8.8	42.6	660328	3.34	29560	21	15	59
Albany, NY	997.87	23	72	57	35	11.0	4281	3.5	50.7	835880	3.14	31458	8	10	39
Allentown, PA	962.35	29	74	54	44	9.8	4260	0.8	39.4	635481	3.21	31856	6	6	33
Atlanta, GA	982.29	45	79	56	47	11.1	3125	27.1	50.2	2138231	3.41	32452	18	8	24
Baltimore, MD	1071.29	35	77	55	43	9.6	6441	24.4	43.7	2199531	3.44	32368	43	38	206
Birmingham, AL	1030.38	45	80	54	53	10.2	3325	38.5	43.1	883946	3.45	27835	30	32	72
Boston, MA	934.70	30	74	56	43	12.1	4679	3.5	49.2	2805911	3.23	36644	21	32	62
Bridgeport, CT	899.53	30	73	56	45	10.6	2140	5.3	40.4	438557	3.29	47258	6	4	4
Buffalo, NY	1001.90	24	70	61	36	10.5	6582	8.1	42.5	1015472	3.31	31248	18	12	37
Canton, OH	912.35	27	72	59	36	10.7	4213	6.7	41.0	404421	3.36	29089	12	7	20
Chattanooga, TN	1017.61	42	79	56	52	9.6	2302	22.2	41.3	426540	3.39	25782	18	8	27
Chicago, IL	1024.89	26	76	58	33	10.9	6122	16.3	44.9	606387	3.20	36593	88	63	278
Cincinnati, OH	970.47	34	77	57	40	10.2	4101	13.0	45.7	1401491	3.21	31427	26	26	146
Cleveland, OH	985.95	28	71	60	35	11.1	3042	14.7	44.6	1898825	3.29	35720	31	21	64
Columbus, OH	958.84	31	75	58	37	11.9	4259	13.1	49.6	124833	3.26	29761	23	9	15
Dallas, TX	860.10	46	85	54	35	11.8	1441	14.8	51.2	1957378	3.22	38769	1	1	1

→ see <http://lib.stat.cmu.edu/DASL/Stories/AirPollutionandMortality.html>

# Applied Statistical Regression

## AS 2013 – Week 06

### *Data Preparation: Visualization*

Because we cannot inspect the data in a xy-scatterplot, data visualization and data preparation becomes an important task. We need to identify the necessary variable transformations, mitigate the effect of outliers, ...

**Step 1:** Plotting the marginal distribution (i.e. histograms)

```
> par(mfrow=c(4,4))  
> for (i in 1:15) hist(apm[,i], main= "...")
```

**Step 2:** Identify erroneous and missing values

```
> any(is.na(apm))  
[1] FALSE
```

# Applied Statistical Regression

## AS 2013 – Week 06

### ***Data Preparation: Transformations***

Linear regression and its output are easier to comprehend if one is using an intuitive scale for the variables. Please note that linear transformations do not change the results. However, any non-linear transformation will do so.

**Step 3:** linear transformations  $x' = ax + b$

```
> apm$JanTemp <- (5/9) * (apm$JanTemp - 32)
> apm$JulyTemp <- (5/9) * (apm$JulyTemp - 32)
> apm$Rain <- (2.54) * apm$Rain
```

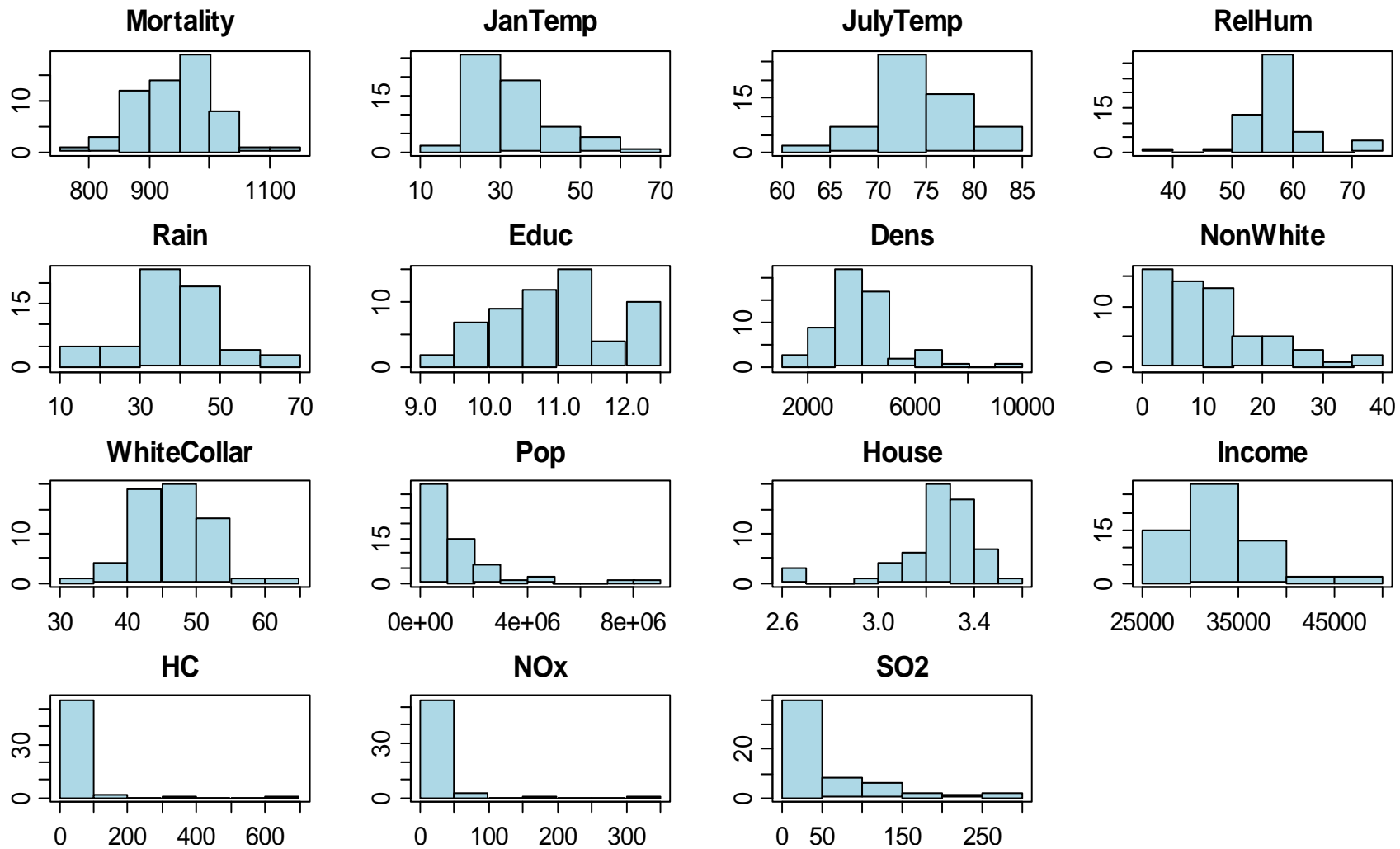
**Step 4:** log-transformation  $x' = \log(x)$

For all variables where it is necessary/beneficial...

# Applied Statistical Regression

## AS 2013 – Week 06

### *Data Preparation: Transformations*



# Applied Statistical Regression

## AS 2013 – Week 06

### *Multiple Linear Regression*

We use linear modeling for a multiple predictor regression:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + E$$

- there are now  $p$  predictors
- the problem cannot be visualized in a scatterplot
- there will be  $n$  observations of response and predictors
- goal: estimating the coefficients  $\beta_0, \beta_1, \dots, \beta_p$  from the data

**IMPORTANT:** simple linear regression of the response on each of the predictors does not equal multiple regression, where *all predictors are used simultaneously*.

# Applied Statistical Regression

## AS 2013 – Week 06

### *Assumptions on the Error Term*

The assumptions are identical to simple linear regression.

- $E[E_i] = 0$ , i.e. the hyper plane is the correct fit
- $Var(E_i) = \sigma_E^2$ , constant scatter for the error term
- $Cov(E_i, E_j) = 0$ , uncorrelated errors
- $E_i \sim N(0, \sigma_E^2)$ , the errors are normally distributed

**Note:** As in simple linear regression, we do not require Gaussian distribution for OLS estimation and certain optimality results, i.e. the Gauss-Markov theorem.

**But:** All tests and confidence intervals rely on the Gaussian, and there are better estimates for non-normal data



# Applied Statistical Regression

## AS 2013 – Week 06

### ***Don't Do Many Simple Regressions***

Doing many simple linear regressions is not equivalent to multiple linear regression. Check the example

x1	0	1	2	3	0	1	2	3
x2	-1	0	1	2	1	2	3	4
yy	1	2	3	4	-1	0	1	2

We have  $Y_i = \hat{y}_i = 2x_{i1} - x_{i2}$ , a perfect fit.

Thus, all residuals are 0 and  $\hat{\sigma}_E^2 = 0$ .

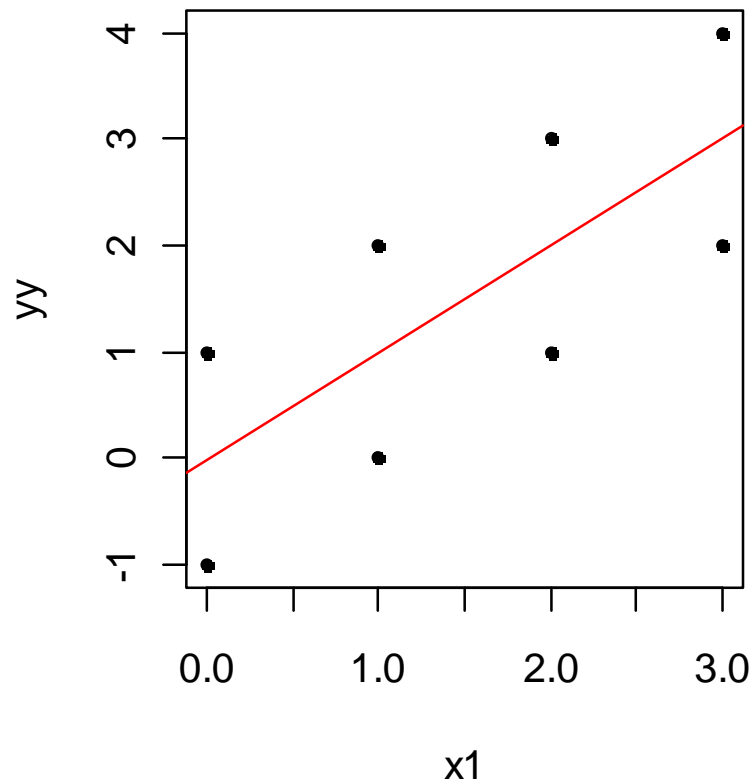
→ *But what is the result from simple linear regressions?*

# Applied Statistical Regression

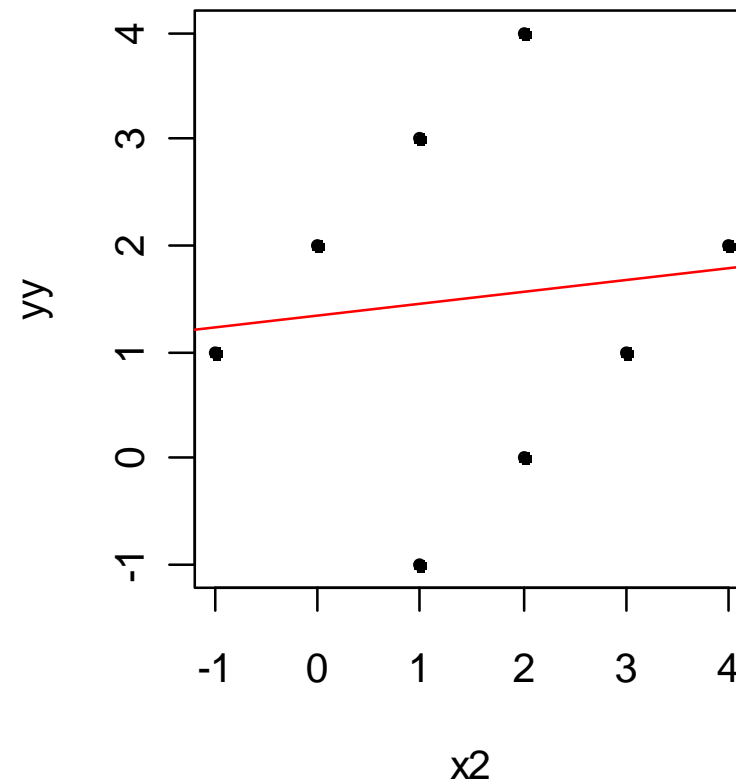
## AS 2013 – Week 06

### *Don't Do Many Simple Regressions*

$yy \sim x1$



$yy \sim x2$



# Applied Statistical Regression

## AS 2013 – Week 06

### *Example: Mortality Due to Air Pollution*

Researchers at General Motors collected data on 60 US Standard Metropolitan Statistical Areas (SMSAs) in a study of whether air pollution contributes to mortality.

City	Mortality	JanTemp	JulyTemp	RelHum	Rain	Educ	Dens	NonWhite	WhiteCllr	Pop	House	Income	HC	NOx	SO2
Akron, OH	921.87	27	71	59	36	11.4	3243	8.8	42.6	660328	3.34	29560	21	15	59
Albany, NY	997.87	23	72	57	35	11.0	4281	3.5	50.7	835880	3.14	31458	8	10	39
Allentown, PA	962.35	29	74	54	44	9.8	4260	0.8	39.4	635481	3.21	31856	6	6	33
Atlanta, GA	982.29	45	79	56	47	11.1	3125	27.1	50.2	2138231	3.41	32452	18	8	24
Baltimore, MD	1071.29	35	77	55	43	9.6	6441	24.4	43.7	2199531	3.44	32368	43	38	206
Birmingham, AL	1030.38	45	80	54	53	10.2	3325	38.5	43.1	883946	3.45	27835	30	32	72
Boston, MA	934.70	30	74	56	43	12.1	4679	3.5	49.2	2805911	3.23	36644	21	32	62
Bridgeport, CT	899.53	30	73	56	45	10.6	2140	5.3	40.4	438557	3.29	47258	6	4	4
Buffalo, NY	1001.90	24	70	61	36	10.5	6582	8.1	42.5	1015472	3.31	31248	18	12	37
Canton, OH	912.35	27	72	59	36	10.7	4213	6.7	41.0	404421	3.36	29089	12	7	20
Chattanooga, TN	1017.61	42	79	56	52	9.6	2302	22.2	41.3	426540	3.39	25782	18	8	27
Chicago, IL	1024.89	26	76	58	33	10.9	6122	16.3	44.9	606387	3.20	36593	88	63	278
Cincinnati, OH	970.47	34	77	57	40	10.2	4101	13.0	45.7	1401491	3.21	31427	26	26	146
Cleveland, OH	985.95	28	71	60	35	11.1	3042	14.7	44.6	1898825	3.29	35720	31	21	64
Columbus, OH	958.84	31	75	58	37	11.9	4259	13.1	49.6	124833	3.26	29761	23	9	15
Dallas, TX	860.10	46	85	54	35	11.8	1441	14.8	51.2	1957378	3.22	38769	1	1	1

→ see <http://lib.stat.cmu.edu/DASL/Stories/AirPollutionandMortality.html>

# Applied Statistical Regression

## AS 2013 – Week 06

### ***Simple Regression vs. Multiple Regression***

#### **Simple Regressions:**

$$\log(SO_2) : \hat{y} = 886.34 + 16.86 \cdot \log(SO_2)$$

$$\text{NonWhite} : \hat{y} = 887.90 + 4.49 \cdot \text{NonWhite}$$

$$\text{Rain} : \hat{y} = 851.22 + 2.34 \cdot \text{Rain}$$

#### **Multiple Regression:**

```
> lm(Mortality ~ log(SO2) + NonWhite + Rain, data=mortality)
```

```
> Coefficients:
```

> (Intercept)	log(SO2)	NonWhite	Rain
773.020	17.502	3.649	1.763

*The regression coefficient  $\beta_j$  is the increase in the response, if the predictor  $x_j$  increases by 1 unit, but all other predictors remain unchanged.*

# Applied Statistical Regression

## AS 2013 – Week 06

### ***Least Squares Algorithm***

The *paradigm* is to determine the regression coefficients such that the *sum of squared residuals is minimal*. This amounts to minimizing the quality function:

$$Q(\beta_0, \beta_1, \dots, \beta_p) = \sum_{i=1}^n r_i^2 = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}))^2$$

We can take partial derivatives with respect to  $\beta_0, \beta_1, \dots, \beta_p$  and so obtain a linear equation system with  $(p + 1)$  unknowns and the same number of equations.

→ **Mostly (but not always...), there is a unique solution.**

# Applied Statistical Regression

## AS 2013 – Week 06

### ***Matrix Notation***

In matrix notation, the multiple linear regression model can be written as:

$$Y = X\beta + E$$

The elements in this equation are as follows:

→ **see blackboard...**

# Applied Statistical Regression

## AS 2013 – Week 06

### ***Normal Equations and Their Solutions***

The least squares approach leads to the normal equations, which are of the following form:

$$(X^T X)\beta = X^T y$$

- Unique solution if and only if  $X$  has full rank
- Predictor variables need to be linearly independent
- If  $X$  has not full rank, the model is “badly formulated”
- Design improvement mandatory!!!
- Necessary (not sufficient) condition:  $p < n$
- Do not over-parametrize your regression!

# Applied Statistical Regression

## AS 2013 – Week 06

### *Properties of the Estimates*

#### **Gauss-Markov-Theorem:**

The regression coefficients are unbiased estimates, and they fulfill the optimality condition of minimal variance among all linear, unbiased estimators (*BLUE*).

- $E[\hat{\beta}] = \beta$
- $Cov(\hat{\beta}) = \sigma_E^2 \cdot (X^T X)^{-1}$
- $\hat{\sigma}_E^2 = \frac{1}{n - (p + 1)} \sum_{i=1}^n r_i^2$  (note: degrees of freedom!)



# Applied Statistical Regression

## AS 2013 – Week 06

### ***If the Errors are Gaussian...***

While all of the above statements hold for arbitrary error distribution, we obtain some more, very useful properties by assuming i.i.d. Gaussian errors:

$$- \hat{\beta} \sim N\left(\beta, \sigma_E^2 (X^T X)^{-1}\right)$$

$$- \hat{y} \sim N(X\beta, \sigma_E^2 H)$$

$$- \hat{\sigma}_E^2 \sim \frac{\sigma_E^2}{n-p} \chi_{n-p}$$

*What to do if the errors are non-Gaussian?*

# Applied Statistical Regression

## AS 2013 – Week 06

### ***Benefits of Linear Regression***

- **Inference on the relation between  $y$  and  $x_1, \dots, x_p$**

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

- **Prediction of (future) observations**

The regression equation can be employed to predict the response value for any given predictor configuration.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_p x_p$$

However, this mostly will not work well for extrapolation!

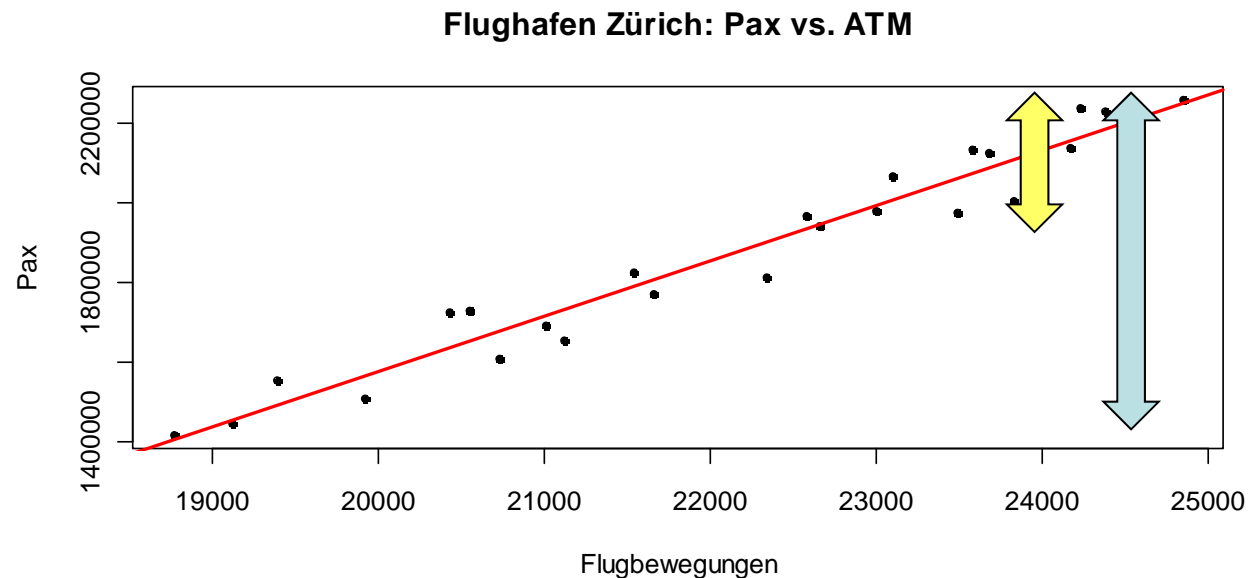
# Applied Statistical Regression

## AS 2013 – Week 06

### $R^2$ : The Coefficient of Determination

The coefficient of determination  $R^2$  tells which portion of the total variation is accounted for by the regression hyperplane.

- For multiple linear regression, visualization is impossible!
- The number of predictor used should be taken into account.



# Applied Statistical Regression

## AS 2013 – Week 06

### ***Coefficient of Determination***

The coefficient of determination, also called *multiple R-squared*, is aimed at describing the goodness-of-fit of the multiple linear regression model:

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

It shows the proportion of the total variance which has been explained by the predictors. The extreme cases 0 and 1 mean:....

# Applied Statistical Regression

## AS 2013 – Week 06

### ***Adjusted Coefficient of Determination***

If we add more and more predictor variables to the model, R-squared will always increase, and never decreases

*Is that a realistic goodness-of-fit measure?*

→ **NO, we better adjust for the number of predictors!**

The adjusted coefficient of determination is defined as:

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

# Applied Statistical Regression

## AS 2013 – Week 06

### ***Confidence Interval for Coefficient $\beta_j$***

We can give a 95%-CI for the regression coefficient  $\beta_j$ . It tells which values, besides the point estimate  $\hat{\beta}_j$ , are plausible too.

**Note:** This uncertainty comes from sampling effects

**95%-VI for  $\beta_j$ :**  $\hat{\beta}_j \pm qt_{0.975;n-(p+1)} \cdot \hat{\sigma}_{\hat{\beta}_j}$

```
In R: > fit <- lm(Mortality ~ ., data=mt)
      > confint(fit, "Educ")
           2.5 %      97.5 %
Educ -31.03177  4.261925
```

# Applied Statistical Regression

## AS 2013 – Week 06

### ***Testing the Coefficient $\beta_j$***

There is a statistical hypothesis test which can be used to check whether  $\hat{\beta}_j$  is significantly different from zero, or different from any other arbitrary value  $b$ . The null hypothesis is:

$$H_0 : \beta_j = 0, \text{ resp. } H_0 : \beta_j = b$$

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

$$H_A : \beta_j \neq 0, \text{ resp. } H_A : \beta_j \neq b$$

As a test statistic, we use:

$$T = \frac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}}, \text{ resp. } T = \frac{\hat{\beta}_j - b}{\hat{\sigma}_{\hat{\beta}_j}}, \text{ both follow a } t_{n-(p+1)} \text{ distribution.}$$

# Applied Statistical Regression

## AS 2013 – Week 06

### *Reading R-Output*

```
> summary(fit.orig)
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	1496.4915	572.7205	2.613	0.01224	*
JanTemp	-2.4479	0.8808	-2.779	0.00798	**
...					
Dens	11.9490	16.1836	0.738	0.46423	
NonWhite	326.6757	62.9092	5.193	5.09e-06	***
WhiteCollar	-146.3477	112.5510	-1.300	0.20028	

```
...  
---
```

```
Residual standard error: 34.23 on 44 degrees of freedom  
Multiple R-squared: 0.7719, Adjusted R-squared: 0.6994  
F-statistic: 10.64 on 14 and 44 DF, p-value: 6.508e-10
```

**Note:** due to space constraints, this is only a part of the output!



# Applied Statistical Regression

## AS 2013 – Week 06

### *Individual Parameter Tests*

These tests quantify the effect of the predictor  $x_j$  on the response  $y$  after having subtracted the linear effect of all other predictor variables on  $y$ .

**Be careful, because of:**

- a) The *multiple testing problem*: when doing many tests, the total type I error increases. By how much?  
→ **See blackboard...**
- b) It can happen that all individual tests do not reject the null hypothesis, although some predictors have a significant effect on the response. **Reason: correlated predictors!**

# Applied Statistical Regression

## AS 2013 – Week 06

### *Individual Parameter Tests*

These tests quantify the effect of the predictor  $x_j$  on the response  $y$  after having subtracted the linear effect of all other predictor variables on  $y$ .

**Be careful, because of:**

- c) The p-values of the individual hypothesis tests are based on the assumption that the other predictors remain in the model and do not change. Therefore, you must not drop more than one single non-significant predictor at a time!

**Solution:** *drop one, re-evaluate the model, drop one, ...*

# Applied Statistical Regression

## AS 2013 – Week 06

### *Simple Variable Selection*

**Goal:** Dropping all predictors from the regression model which are not necessary, i.e. do not show a significant impact on the response.

**How:** In a step-by-step manner, the least significant predictor is dropped from the model, as long as its p-value still exceeds the value of 0.05.

**In R:**

```
> fit <- update(fit, . ~ . - RelHum)
> summary(fit)
```

→ **Exercise: try do to this for the Mortality Data**

# Applied Statistical Regression

## AS 2013 – Week 06

### *Comparing Hierarchical Models*

**Idea:** Correctly comparing two multiple linear regression models when the smaller has  $>1$  predictor less than the bigger.

**Where and why do we need this?**

- for the 3 pollution variables in the mortality data.
- soon also for the so-called factor/dummy variables.

**Idea:** We compare the residual sum of squares (RSS):

$$\text{Big model: } y = \beta_0 + \beta_1 x_1 + \dots + \beta_q x_q + \beta_{q+1} x_{q+1} + \dots + \beta_p x_p$$

$$\text{Small model: } y = \beta_0 + \beta_1 x_1 + \dots + \beta_q x_q$$

The big model must contain all the predictors from the small model, else they are not hierarchical and the test does not apply.

### ***Comparing Hierarchical Models***

**Null hypothesis:**

$H_0 : \beta_{q+1} = \beta_{q+2} = \dots = \beta_p = 0$ , versus the alternative hypothesis that at least one  $\beta_j \neq 0$ ,  $j = q + 1, \dots, p$

The test compares the RSS of the big and the small model:

$$F = \frac{n - (p + 1)}{p - q} \cdot \frac{RSS_{Small} - RSS_{Big}}{RSS_{Big}} \sim F_{p-q, n-(p+1)}$$

→ If the  $F$ -value is small ( $p \geq 0.05$ ), there is no evidence against the null, and we can work well with the smaller model.

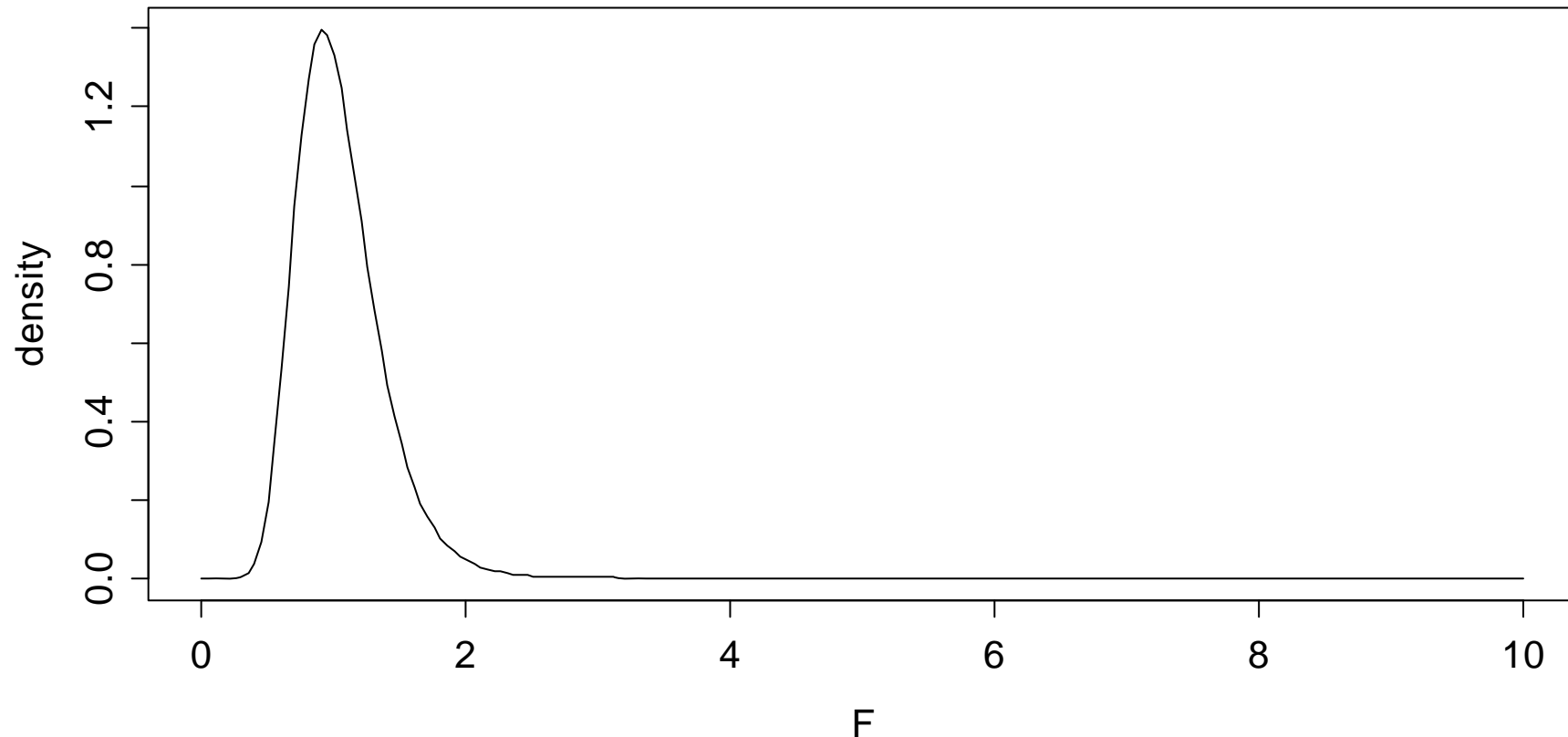
→ ***The p-value answers: is the big model significantly better?***

# Applied Statistical Regression

## AS 2013 – Week 06

### *Density Function of the F-distribution*

The F-distribution with 44 and 47 degrees of freedom



# Applied Statistical Regression

## AS 2013 – Week 06

### *Comparing Hierarchical Models in R*

```
> fit.big    <- lm(Mortality ~ ., data=mt)
> fit.small <- update(fit.big, .~.-HC-NOx-SO2)

> anova(fit.big, fit.small)
```

Analysis of Variance Table

Model 1: Mortality ~ JanTemp + JulyTemp + RelHum + Rain +  
Educ + Dens + NonWhite + WhiteCollar + Pop +  
House + Income + HC + NOx + SO2

Model 2: Mortality ~ JanTemp + JulyTemp + RelHum + Rain +  
Educ + Dens + NonWhite + WhiteCollar + Pop +  
House + Income

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	44	51543				
2	47	61244	-3	-9700.8	2.7604	0.0533 .

# Applied Statistical Regression

## AS 2013 – Week 06

### ***The Global F-Test***

*Idea: is there any relation between response and predictors?*

This is another hierarchical model comparison. The full model is tested against a small model with only the intercept, but without any predictors.

We are testing the null  $H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$  against the alternative  $H_A : \beta_j \neq 0$  for at least one predictor  $x_j$ . This test is again based on comparing the RSS:

$$F = \frac{n - (p + 1)}{p} \cdot \frac{RSS_{Small} - RSS_{Big}}{RSS_{Big}} \sim F_{p, n - (p + 1)}$$

→ Test statistic and p-value are shown in the R summary!



# Applied Statistical Regression

## AS 2013 – Week 06

### *Reading R-Output*

```
> summary(fit.orig)
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	1496.4915	572.7205	2.613	0.01224	*
JanTemp	-2.4479	0.8808	-2.779	0.00798	**
...					
Dens	11.9490	16.1836	0.738	0.46423	
NonWhite	326.6757	62.9092	5.193	5.09e-06	***
WhiteCollar	-146.3477	112.5510	-1.300	0.20028	

```
...
```

```
---
```

```
Residual standard error: 34.23 on 44 degrees of freedom
```

```
Multiple R-squared: 0.7719, Adjusted R-squared: 0.6994
```

```
F-statistic: 10.64 on 14 and 44 DF, p-value: 6.508e-10
```

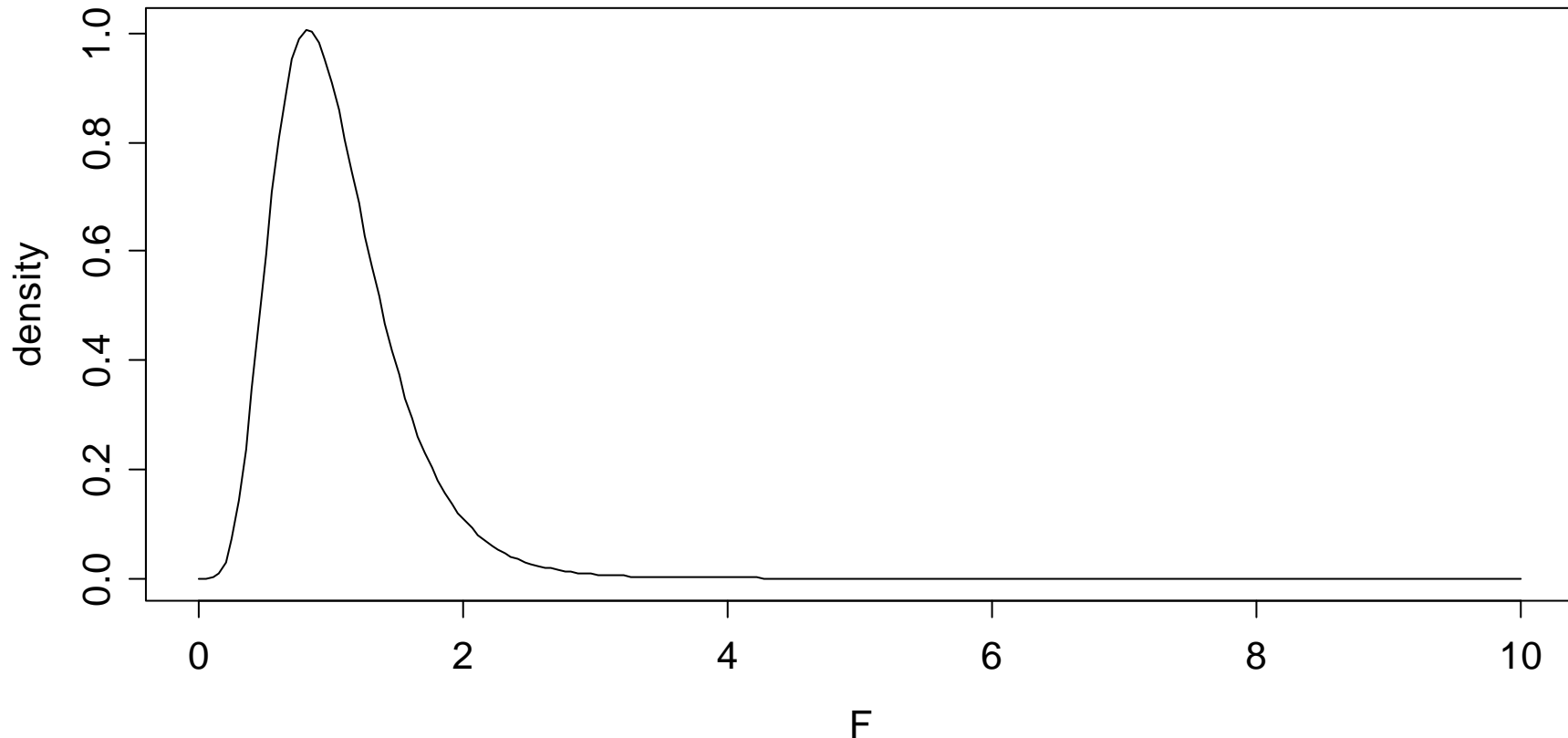
**Note:** due to space constraints, this is only a part of the output!

# Applied Statistical Regression

## AS 2013 – Week 06

### *Density Function of the F-distribution*

The F-distribution with 14 and 47 degrees of freedom



# Applied Statistical Regression

## AS 2013 – Week 06

### ***Prediction***

The regression equation can be employed to predict the response value for any given predictor configuration.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_{.1} + \hat{\beta}_2 x_{.2} + \dots + \hat{\beta}_p x_{.p}$$

#### **Note:**

This can be a predictor configuration that was not part of the original data. For example a (new) city, for which only the predictors are known, but the mortality is not.

#### **Be careful:**

Only interpolation, i.e. prediction within the range of observed y-values works well, extrapolation yields non-reliable results.

# Applied Statistical Regression

## AS 2013 – Week 06

### *Prediction in R*

We can use the regression fit for predicting new observations.

The syntax is as follows

```
> fit.big <- lm(Mortality ~ ., data=mt)
> dat      <- data.frame(JanTemp=..., ...)
> predict(fit.big, newdata=dat)
1 932.488
```

The x-values need to be provided in a data frame. The variable (column) names need to be identical to the predictor names. Of course, all predictors need to be present.

Then, it is simply applying the `predict()`-procedure.

# Applied Statistical Regression

## AS 2013 – Week 06

### ***Confidence- and Prediction Interval***

The confidence interval for the fitted value and the prediction interval for future observation also exist in multiple regression.

a) 95%-CI for the fitted value  $E[y | x]$

```
> predict(fit, newdata=dat, "confidence")
```

b) 95%-PI for a future observation  $\hat{y}$ :

```
> predict(fit, newdata=dat, "prediction")
```

- The visualization of these intervals is no longer possible in the case of multiple regression
- It is possible to write explicit formulae for the intervals using the matrix notation. We omit them here.

# Applied Statistical Regression

## AS 2013 – Week 06

### *Reading R-Output*

```
> summary(fit.orig)
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	1496.4915	572.7205	2.613	0.01224	*
JanTemp	-2.4479	0.8808	-2.779	0.00798	**
...					
Dens	11.9490	16.1836	0.738	0.46423	
NonWhite	326.6757	62.9092	5.193	5.09e-06	***
WhiteCollar	-146.3477	112.5510	-1.300	0.20028	

```
...
```

```
---
```

```
Residual standard error: 34.23 on 44 degrees of freedom
```

```
Multiple R-squared: 0.7719, Adjusted R-squared: 0.6994
```

```
F-statistic: 10.64 on 14 and 44 DF, p-value: 6.508e-10
```

**Note:** due to space constraints, this is only a part of the output!