Quantile regression (Let's have a look?)

Domenico Vistocco
vistocco@unicas.it

Dipartimento di Economia e Giurisprudenza
Università degli Studi di Cassino e del Lazio Meridionale

14 July, 2016

all computation and graphicas were done using the R language
for the slide I used LaTeX (*beamer* library)
and the R Sweave command

Outline I

The Cars93 dataset

The simplest model

A slightly more complex model

The case of a quantitative regressor
- A real data application

Outline I

The Cars93 dataset

The simplest model

A slightly more complex model

The case of a quantitative regressor

A real data application

The Cars93 dataset

```
> library(MASS)
> help(Cars93)
Details

Cars were selected at random from among 1993 passenger car
models that were listed in both the Consumer Reports issue
and the PACE Buying Guide.
Pickup trucks and Sport/Utility vehicles were eliminated
due to incomplete information in the Consumer Reports source.
Duplicate models (e.g., Dodge Shadow and Plymouth Sundance)
were listed at most once.

Further description can be found in Lock (1993).

Source
Lock, R. H. (1993) 1993 New Car Data.
Journal of Statistics Education 1(1).
http://www.amstat.org/publications/jse/v1n1/datasets.lock.html
```

The Cars93 dataset

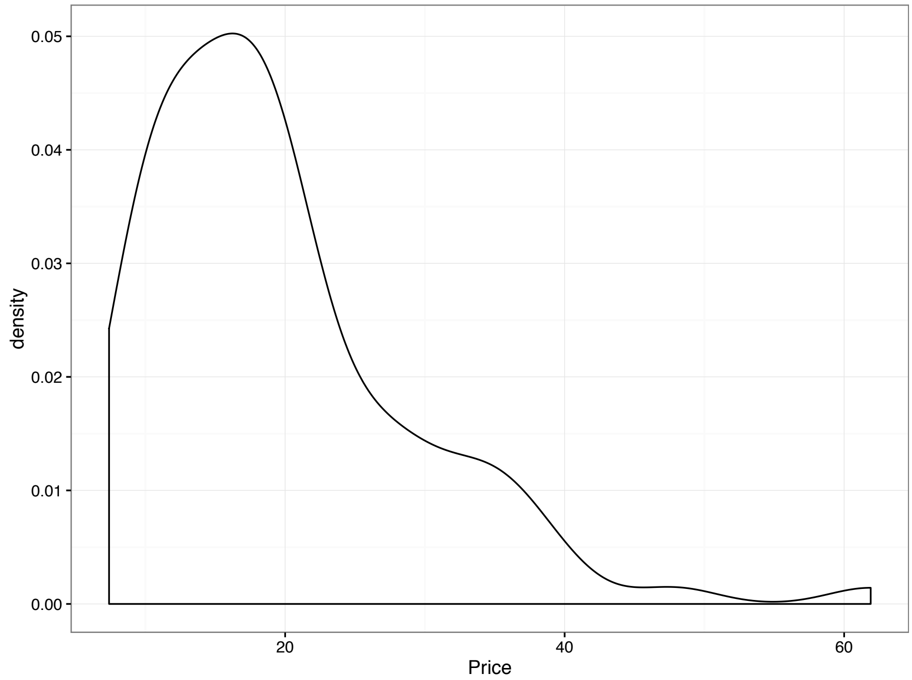
```
>  ls.str(Cars93)

AirBags :  Factor w/ 3 levels "Driver & Passenger",..: 3 1 2 1 2 2 2 2 2 2 ...
Cylinders :  Factor w/ 6 levels "3","4","5","6",..: 2 4 4 4 2 2 4 4 4 5 ...
DriveTrain :  Factor w/ 3 levels "4WD","Front",..: 2 2 2 2 3 2 2 3 2 2 ...
EngineSize :  num [1:93] 1.8 3.2 2.8 2.8 3.5 2.2 3.8 5.7 3.8 4.9 ...
Fuel.tank.capacity :  num [1:93] 13.2 18 16.9 21.1 21.1 16.4 18 23 18.8 18 ...
Horsepower :  int [1:93] 140 200 172 172 208 110 170 180 170 200 ...
Length :  int [1:93] 177 195 180 193 186 189 200 216 198 206 ...
Luggage.room :  int [1:93] 11 15 14 17 13 16 17 21 14 18 ...
Make :  Factor w/ 93 levels "Acura Integra",..: 1 2 4 3 5 6 7 9 8 10 ...
Man.trans.avail :  Factor w/ 2 levels "No","Yes": 2 2 2 2 2 1 1 1 1 1 ...
Manufacturer :  Factor w/ 32 levels "Acura","Audi",..: 1 1 2 2 3 4 4 4 4 5 ...
Max.Price :  num [1:93] 18.8 38.7 32.3 44.6 36.2 17.3 21.7 24.9 26.3 36.3 ...
Min.Price :  num [1:93] 12.9 29.2 25.9 30.8 23.7 14.2 19.9 22.6 26.3 33 ...
Model :  Factor w/ 93 levels "100","190E","240",..: 49 56 9 1 6 24 54 74 73 35 ...
MPG.city :  int [1:93] 25 18 20 19 22 22 19 16 19 16 ...
MPG.highway :  int [1:93] 31 25 26 26 30 31 28 25 27 25 ...
Origin :  Factor w/ 2 levels "USA","non-USA": 2 2 2 2 2 1 1 1 1 1 ...
Passengers :  int [1:93] 5 5 5 6 4 6 6 6 5 6 ...
Price :  num [1:93] 15.9 33.9 29.1 37.7 30 15.7 20.8 23.7 26.3 34.7 ...
Rear.seat.room :  num [1:93] 26.5 30 28 31 27 28 30.5 30.5 26.5 35 ...
Rev.per.mile :  int [1:93] 2890 2335 2280 2535 2545 2565 1570 1320 1690 1510 ...
RPM :  int [1:93] 6300 5500 5500 5500 5700 5200 4800 4000 4800 4100 ...
Turn.circle :  int [1:93] 37 38 37 37 39 41 42 45 41 43 ...
Type :  Factor w/ 6 levels "Compact","Large",..: 4 3 1 3 3 3 2 2 3 2 ...
Weight :  int [1:93] 2705 3560 3375 3405 3640 2880 3470 4105 3495 3620 ...
Wheelbase :  int [1:93] 102 115 102 106 109 105 111 116 108 114 ...
Width :  int [1:93] 68 71 67 70 69 69 74 78 73 73 ...
```


The (small) Cars93 dataset

```
>  miniCars <- Cars93[c("Price", "Origin", "AirBags", "Horsepower")]
>  head(miniCars)

  Price  Origin            AirBags Horsepower
1  15.9 non-USA               None        140
2  33.9 non-USA Driver & Passenger        200
3  29.1 non-USA        Driver only        172
4  37.7 non-USA Driver & Passenger        172
5  30.0 non-USA        Driver only        208
6  15.7     USA        Driver only        110

>  table(miniCars$Origin)

    USA non-USA
     48      45
```

The response variable

```
> library(ggplot2)
> theme_set(theme_bw())
> ggplot(aes(x=Price), data=miniCars) + geom_density()
```



Outline I

The Cars93 dataset

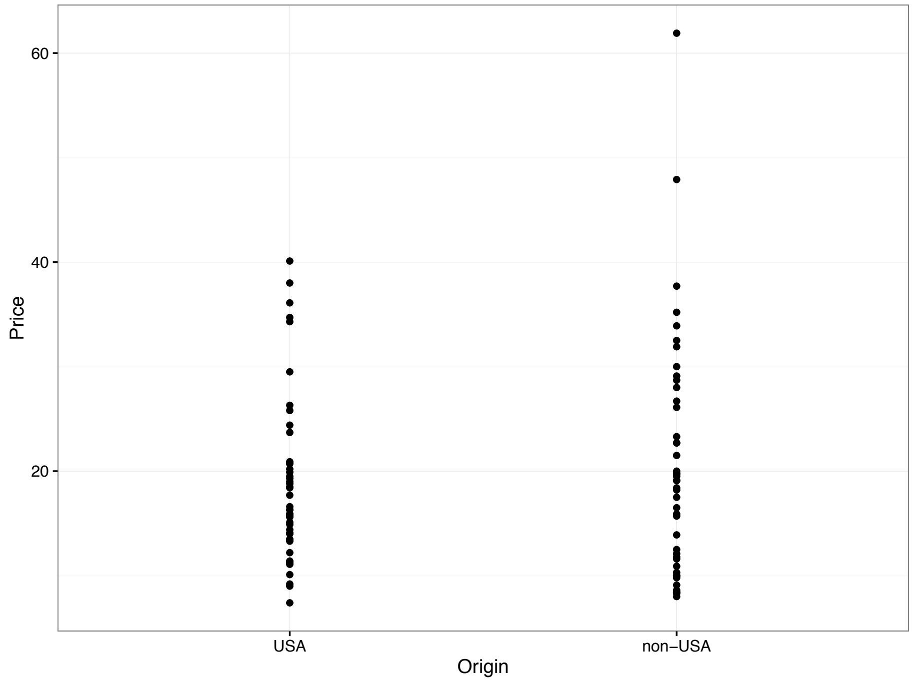
The simplest model

A slightly more complex model

The case of a quantitative regressor
    A real data application

The dummy explicative variable

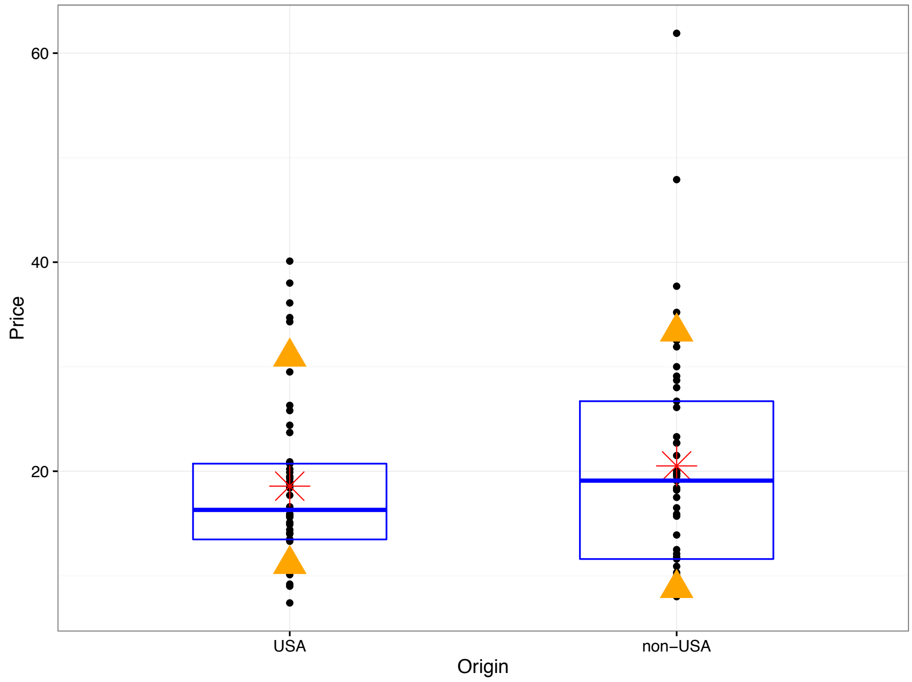
```
>   ggplot(aes(x=Origin, y=Price), data=miniCars) +
+       geom_point()
```



The dummy explicative variable

```
> stat_sum_single <- function(fun, geom="point", colour="red", size=3, ...) {
+    stat_summary(fun.y=fun, colour=colour, geom=geom, size = size, ...)
+ }
> stat_sum_crossbar <- function(fun1, fun2, fun3, geom="crossbar", colour="blue", ...) {
+    stat_summary(fun.y=fun1, fun.ymin=fun2, fun.ymax=fun3, colour=colour, geom=geom, ...)
+ }
> ggplot(aes(x=Origin, y=Price), data=miniCars) +
+     geom_point() +
+     stat_sum_single(mean, shape=8, size=8) +
+     stat_sum_single(function(x) quantile(x, 0.1), colour="orange", shape=17, size=7) +
+     stat_sum_single(function(x) quantile(x, 0.9), colour="orange", shape=17, size=7) +
+     stat_sum_crossbar(function(x) quantile(x, 0.5),
+                        function(x) quantile(x, 0.25),
+                              function(x) quantile(x, 0.75),
+                              geom="crossbar", colour="blue", width=0.5)
```

The dummy explicative variable



Comparing the means: the t–test

```
>   t.test(Price ~ Origin, data=miniCars)

        Welch Two Sample t-test

data:  Price by Origin
t = -0.95449, df = 77.667, p-value = 0.3428
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.974255  2.102311
sample estimates:
    mean in group USA mean in group non-USA
             18.57292              20.50889
```

Comparing the means: the regression model

$$\hat{Price} = \hat{\beta}_0 + \hat{\beta}_1 Origin$$

Using the coding USA = 0 and non–USA = 1 for the Origin indicator variable, the estimated price for the USA cars is then:

$$\hat{Price} = \hat{\beta}_0 + \hat{\beta}_1 \times \mathbf{0} = \hat{\beta}_0 = E[Price|Origin = 0]$$

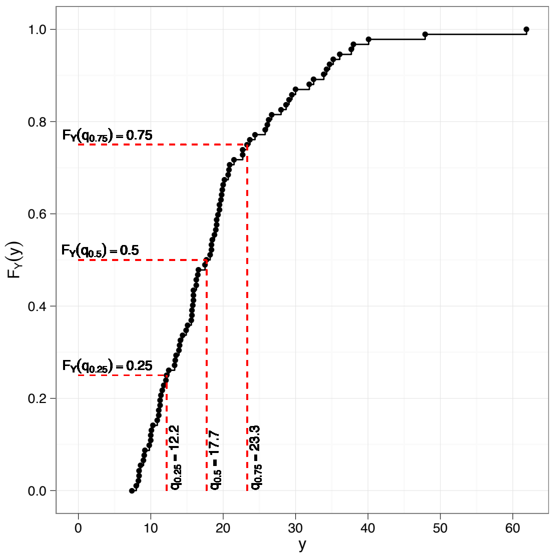
while for the non–USA subset the model becomes:

$$\hat{Price} = \hat{\beta}_0 + \hat{\beta}_1 \times \mathbf{1} = \hat{\beta}_0 + \hat{\beta}_1 = E[Price|Origin = 1]$$

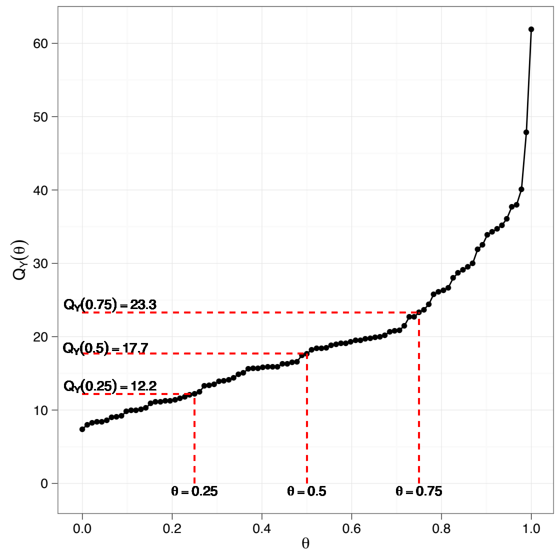
Comparing the means: the regression model

```
>   summary(lm(Price ~ Origin, data=miniCars))

Call:
lm(formula = Price ~ Origin, data = miniCars)

Residuals:
    Min      1Q  Median      3Q     Max
-12.509  -7.173  -2.109   2.791  41.391

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     18.573      1.395  13.316   <2e-16 ***
Originnon-USA    1.936      2.005   0.966    0.337
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 9.663 on 91 degrees of freedom
Multiple R-squared:  0.01014,       Adjusted R-squared:  -0.0007366
F-statistic: 0.9323 on 1 and 91 DF,  p-value: 0.3368
```

Unconditional and conditional quantiles



$$F_Y(y) = F(y) = P(Y \leq y)$$



$$Q_Y(\theta) = Q(\theta) = F_Y^{-1}(\theta) = \inf\{y : F(y) > \theta\}$$

$$\theta \in [0, 1]$$

NOTE: If $F(.)$ is strictly increasing and continuous, then $F^{-1}(\theta)$ is the unique real nuumber y such that $F(y) = \theta$

Unconditional and conditional quantiles

$$\hat{Price}_\theta = \hat{\beta}_0(\theta) + \hat{\beta}_1 Origin(\theta)$$

Using the coding USA = 0 and non–USA = 1 for the Origin indicator variable, for the USA cars we have:

$$\hat{Price}_\theta = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta) \times \mathbf{0} = \hat{\beta}_0(\theta) = Q_{[Price|Origin=0]}(\theta)$$

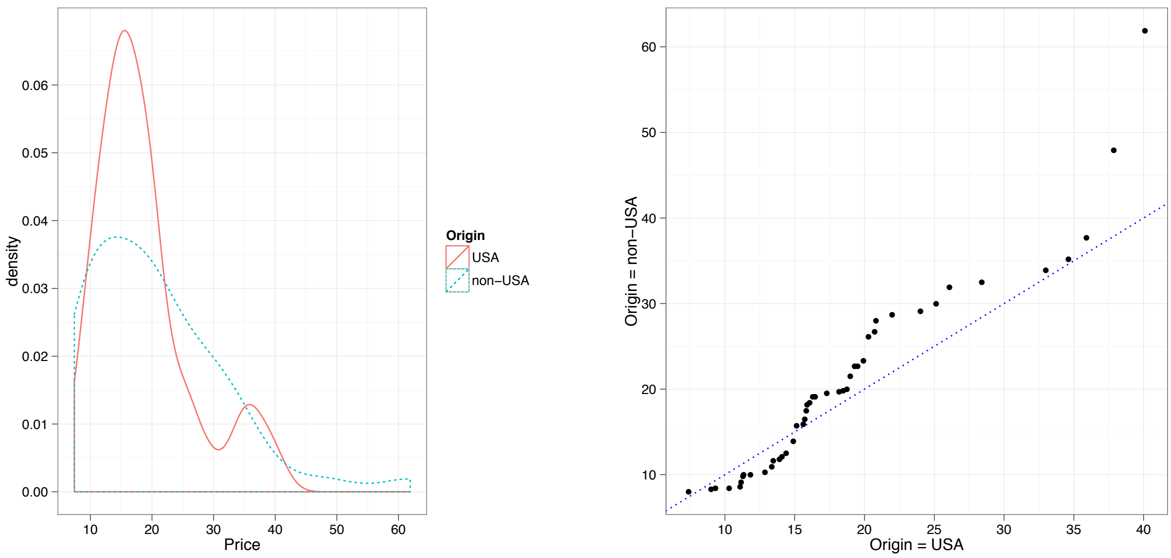
while for the non–USA subset the model becomes:

$$\hat{Price}_\theta = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta) \times \mathbf{1} = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta) = Q_{[Price|Origin=1]}(\theta)$$

Unconditional quantiles

```
> with(miniCars,
+        tapply(Price, Origin, quantile,
+               probs=c(0.1, 0.25, 0.5, 0.75, 0.9)))

$USA
   10%    25%    50%    75%    90%
11.100 13.475 16.300 20.725 30.940

$`non-USA`
  10%   25%   50%   75%   90%
 8.80 11.60 19.10 26.70 33.34
```


The "unconditional" two groups



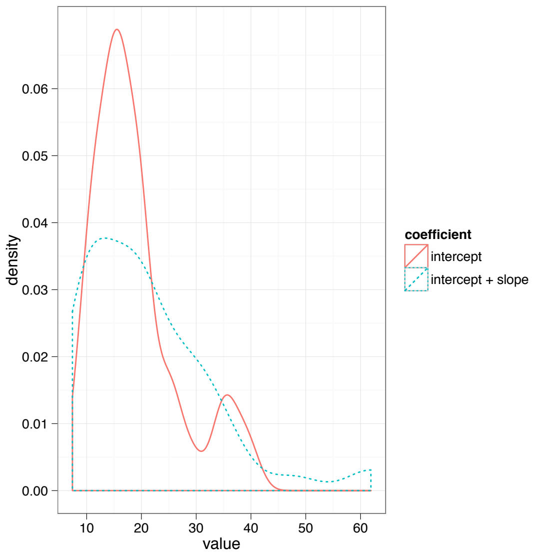
Conditional quantiles

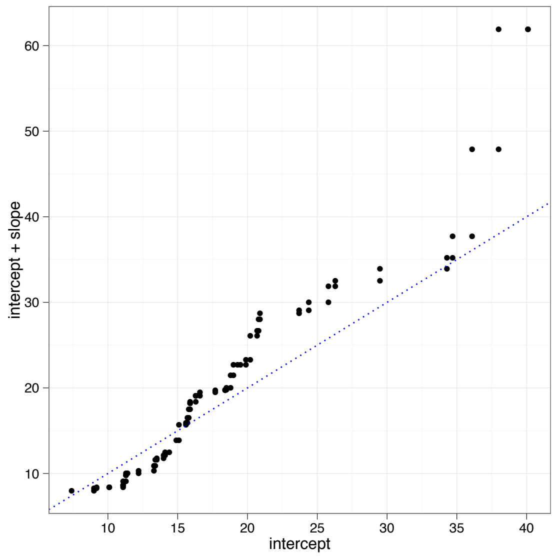
```
> library(quantreg)
> summary(rq(Price ~ Origin, data=miniCars,
+              tau=c(0.1, 0.25, 0.5, 0.75, 0.9)))
```


Conditional quantiles

	$\theta = 0.1$	$\theta = 0.25$	$\theta = 0.5$	$\theta = 0.75$	$\theta = 0.9$
(Intercept)	11.1	13.4	16.3	20.8	34.3
Origin (non–USA)	-2.5	-1.8	2.8	5.9	-0.4
(Intercept)	11.1	13.4	16.3	20.8	34.3
Intercept + Origin (non–USA)	8.6	11.6	19.1	26.7	33.9
Unconditional Price quantiles					
Origin (USA)	11.1	13.5	16.3	20.7	30.9
Origin (non–USA)	8.8	11.6	19.1	26.7	33.3

The “conditional” two groups





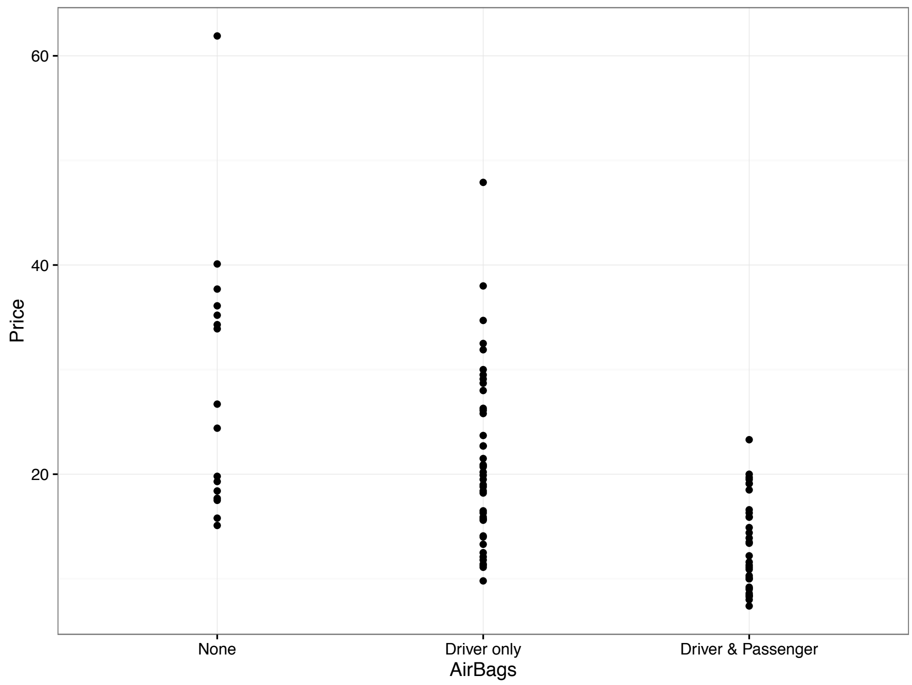
Outline I

A nominal explicative variable

```
>   table(miniCars$AirBags)

Driver & Passenger        Driver only                None
                16                 43                  34

>   levels(miniCars$AirBags) <- c("None","Driver only", "Driver & Passenger")
>   table(miniCars$AirBags)

              None        Driver only Driver & Passenger
                16                 43                  34
```


A nominal explicative variable

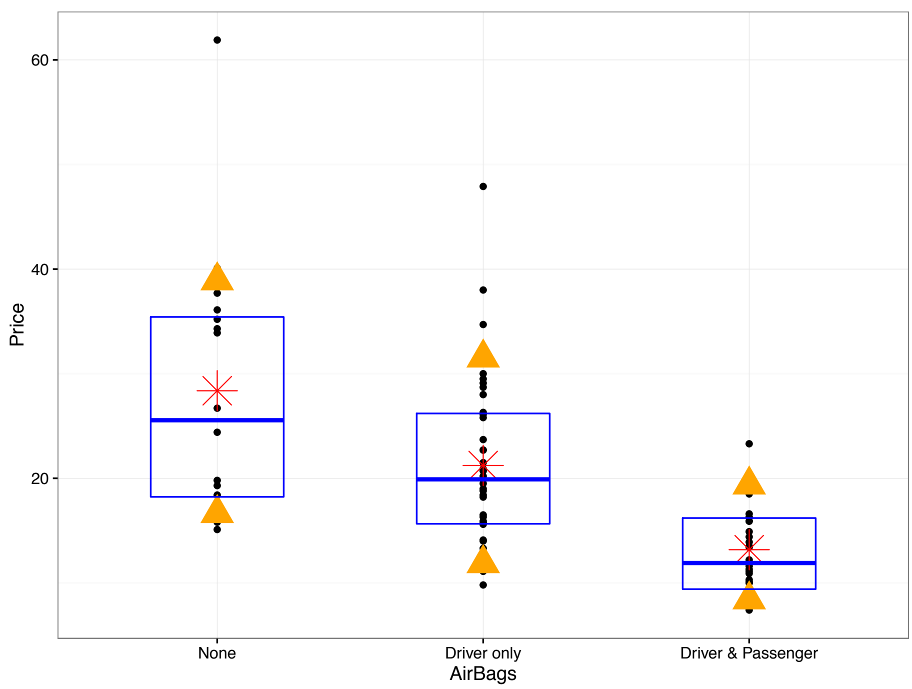
```
> ggplot(aes(x=AirBags, y=Price), data=miniCars) +
+     geom_point()
```



A nominal explicative variable

```
> ggplot(aes(x=AirBags, y=Price), data=miniCars) +
+     geom_point() +
+     stat_sum_single(mean, shape=8, size=8) +
+     stat_sum_single(function(x) quantile(x, 0.1),
+                     colour="orange", shape=17, size=7) +
+     stat_sum_single(function(x) quantile(x, 0.9),
+                     colour="orange", shape=17, size=7) +
+     stat_sum_crossbar(function(x) quantile(x, 0.5),
+                       function(x) quantile(x, 0.25),
+                       function(x) quantile(x, 0.75),
+                       geom="crossbar", colour="blue", width=0.5)
```


A nominal explicative variable



Comparing the means: the ANOVA model

```
>   summary(aov(Price ~ AirBags, data=miniCars))

            Df Sum Sq Mean Sq F value  Pr(>F)
AirBags      2   2747  1373.5   21.18 2.9e-08 ***
Residuals   90   5837    64.9
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```


Comparing the means: the regression model

$$\hat{Price} = \hat{\beta}_0 + \hat{\beta}_1 I(\textit{Driver only}) + \hat{\beta}_2 I(\textit{Driver \& Passenger})$$

where I(*level-name*) is the indicator function returning 1 if the particular unit assumes the value in parenthesis and 0 otherwise

The first level, the so–called reference level, is associated with the intercept. For the *None* level, associated with the intercept, the model reduces to:

$$\hat{Price} = \hat{\beta}_0 + \hat{\beta}_1 \times \mathbf{0} + \hat{\beta}_2 \times \mathbf{0} = \hat{\beta}_0 = E[Price|Origin = None]$$

For the *Drivers only* levels, we have:

$$\hat{Price} = \hat{\beta}_0 + \hat{\beta}_1 \times \mathbf{1} + \hat{\beta}_2 \times \mathbf{0} = \hat{\beta}_0 + \hat{\beta}_1 = E[Price|Origin = Driversonly]$$

and for the *Drivers & Passenger* group, we have:

$$\hat{Price} = \hat{\beta}_0 + \hat{\beta}_1 \times \mathbf{0} + \hat{\beta}_2 \times \mathbf{1} = \hat{\beta}_0 + \hat{\beta}_2 = E[Price|Origin = Drivers\&Passenger]$$

Comparing the means: the regression model

```
>   summary(lm(Price ~ AirBags, data=miniCars))

Call:
lm(formula = Price ~ AirBags, data = miniCars)

Residuals:
    Min      1Q  Median      3Q     Max
-13.269  -4.923  -1.574   5.326  33.531

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                 28.369      2.013  14.090  < 2e-16 ***
AirBagsDriver only          -7.145      2.358  -3.030  0.00319 **
AirBagsDriver & Passenger  -15.195      2.442  -6.224 1.51e-08 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.053 on 90 degrees of freedom
Multiple R-squared:   0.32,        Adjusted R-squared:  0.3049
F-statistic: 21.18 on 2 and 90 DF,  p-value: 2.901e-08
```


Comparing the quantiles

```
> with(miniCars,
+      tapply(Price, AirBags, quantile,
+             probs=c(0.1, 0.25, 0.5, 0.75, 0.9)))

$None
   10%    25%    50%    75%    90%
16.650 18.225 25.550 35.425 38.900

$`Driver only`
  10%   25%   50%   75%   90%
11.86 15.65 19.90 26.20 31.52

$`Driver & Passenger`
  10%   25%   50%   75%   90%
 8.40  9.40 11.90 16.20 19.38
```

Comparing the quantiles: the regression model

$$\hat{Price}_\theta = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)I(\textit{Driver only}) + \hat{\beta}_2(\theta)I(\textit{Driver \& Passenger})$$

where $I(\textit{level-name})$ is the indicator function returning 1 if the particular unit assumes the value in parenthesis and 0 otherwise

For the *None* level, associated with the intercept, the model reduces to:

$$\hat{Price}_\theta = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta) \times \mathbf{0} + \hat{\beta}_2(\theta) \times \mathbf{0} = \hat{\beta}_0(\theta) = Q_{[Price|Origin=None]}(\theta)$$

For the *Drivers only* group, we have:

$$\hat{Price}_\theta = \hat{\beta}_0 + \hat{\beta}_1(\theta) \times \mathbf{1} + \hat{\beta}_2(\theta) \times \mathbf{0} = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta) = Q_{[Price|Origin=Driversonly]}(\theta)$$

and for the *Drivers & Passenger* level, we have:

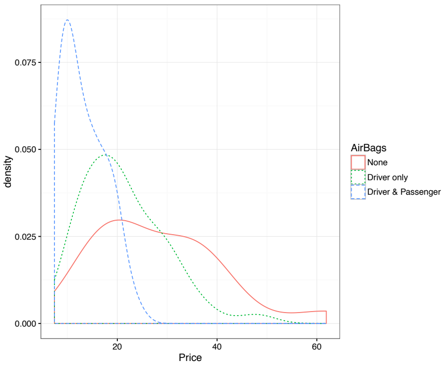
$$\hat{Price}_\theta = \hat{\beta}_0 + \hat{\beta}_1 \times \mathbf{0} + \hat{\beta}_2 \times \mathbf{1} = \hat{\beta}_0 + \hat{\beta}_2 = Q_{[Price|Origin=Drivers\&Passenger]}(\theta)$$

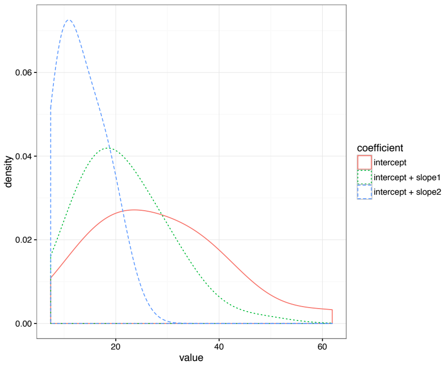
Comparing the quantiles: the regression model

```
> library(quantreg)
> summary(rq(Price ~ AirBags, data=miniCars,
+             tau=c(0.1, 0.25, 0.5, 0.75, 0.9)))
```

Comparing the quantiles: the regression model

	$\theta = 0.1$	$\theta = 0.25$	$\theta = 0.5$	$\theta = 0.75$	$\theta = 0.9$
(Intercept)	8.4	9.2	12.2	16.3	19.5
AirBags (*Driver only*)	3.4	6.4	7.7	10.0	12.4
AirBags (*Driver & Passenger*)	7.4	8.5	12.2	18.9	20.6
(Intercept)	8.4	9.2	12.2	16.3	19.5
Intercept + AirBags (*Driver only*)	11.8	15.6	19.9	26.3	31.9
Intercept + AirBags (*Driver & Passenger*)	15.8	17.7	24.4	35.2	40.1
Unconditional Price quantiles					
AirBags (*None*)	8.4	9.4	11.9	16.2	19.4
AirBags (*Driver only*)	11.9	15.6	19.9	26.2	31.5
AirBags (*Driver & Passenger*)	16.6	18.2	25.5	35.4	38.9





Outline I

The Cars93 dataset

The simplest model

A slightly more complex model

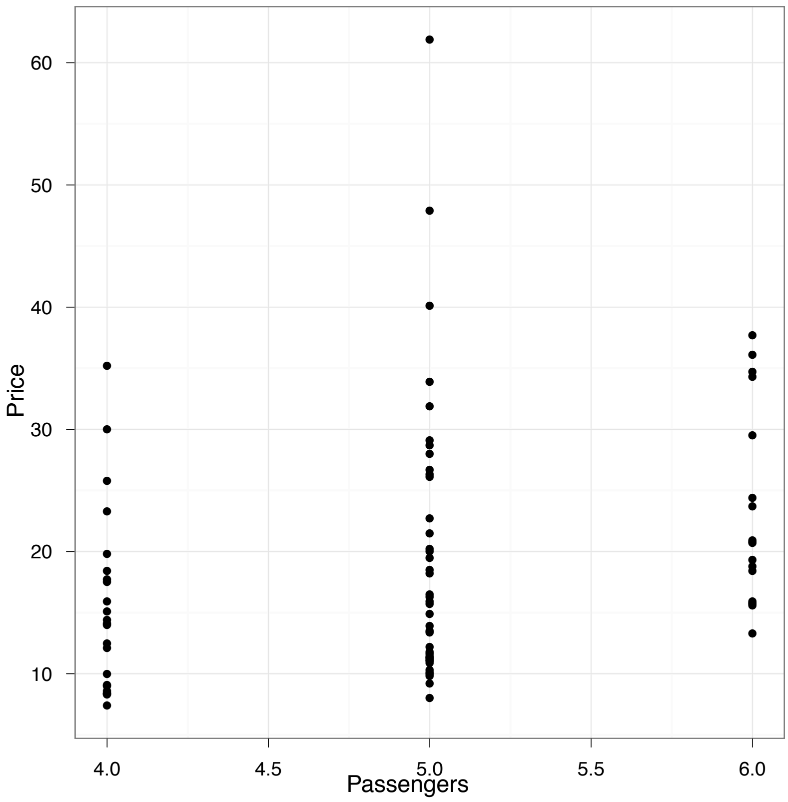
The case of a quantitative regressor

A real data application

A quantitative explicative variable

```
> miniCars <- subset(Cars93, Passengers %in% 4:6)
> ggp <- qplot(data= miniCars, Passengers, Price,
+              geom="point", group=Passengers)
> ggp
```

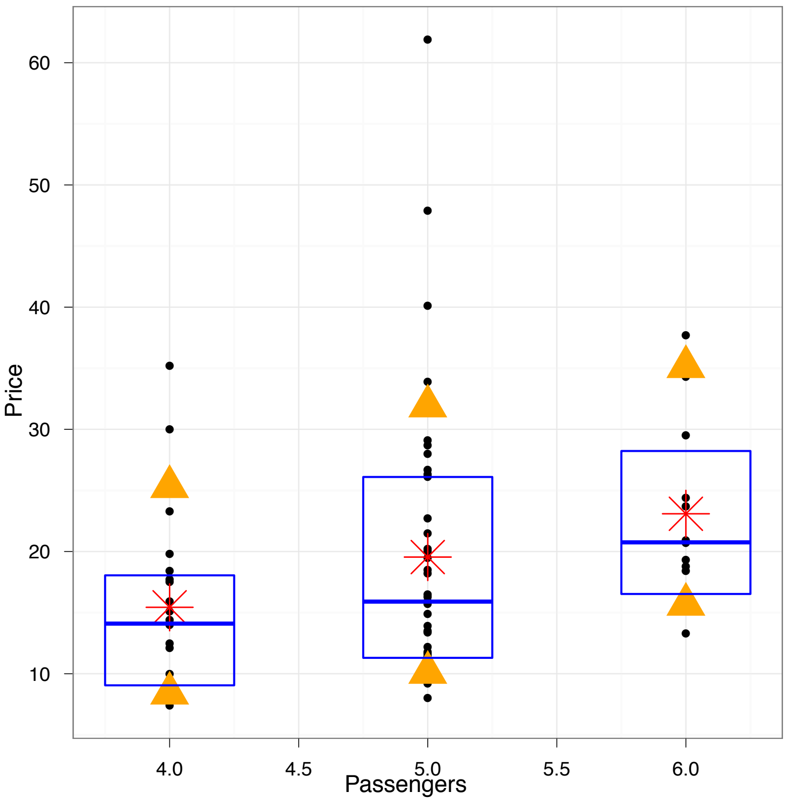

A quantitative explicative variable



A quantitative explicative variable

```
> ggp2 <- ggp + stat_sum_single(mean, shape=8, size=8)
> ggp2 <- ggp2 +
+   stat_sum_single(function(x) quantile(x, 0.1),
+                   colour="orange", shape=17, size=7) +
+   stat_sum_single(function(x) quantile(x, 0.9),
+                   colour="orange", shape=17, size=7)
> ggp2 <- ggp2 +
+   stat_sum_crossbar(function(x) quantile(x, 0.5),
+                     function(x) quantile(x, 0.25),
+                     function(x) quantile(x, 0.75),
+                     geom="crossbar", colour="blue", width=0.5)
> ggp2
```

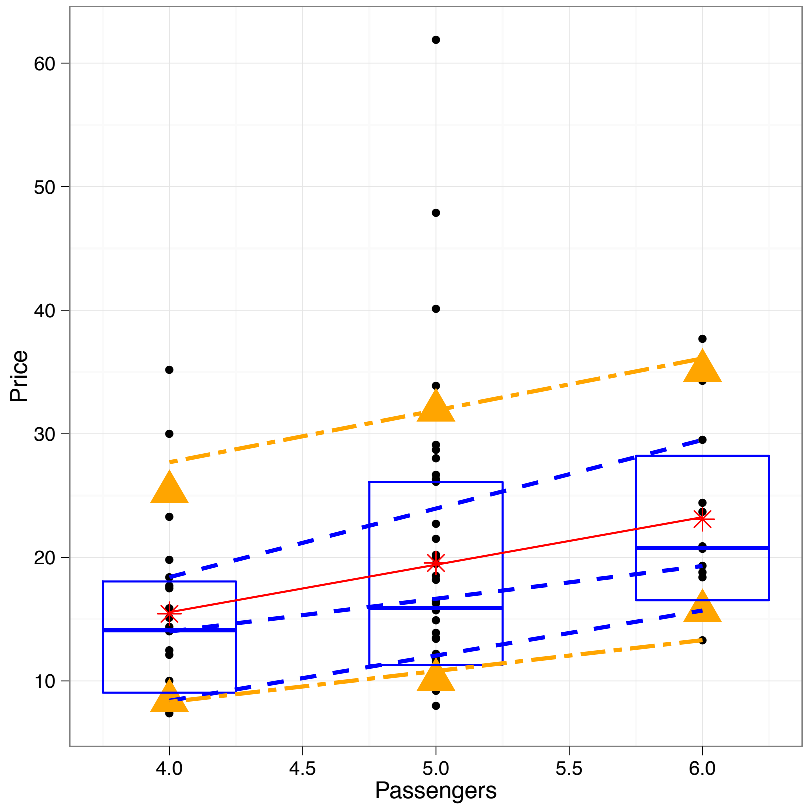

A quantitative explicative variable



A quantitative explicative variable

```
> ggp3 <- ggplot(miniCars,aes(Passengers,Price)) +
+   geom_point() +
+   stat_summary(fun.y ="mean",colour="red",size=4,shape=8,geom="point") +
+   stat_summary(fun.y = function(x) quantile(x,.9),
+                colour="orange", shape=17, size=7, geom="point") +
+   stat_summary(fun.y = function(x) quantile(x,.1),
+                colour="orange", shape=17, size=7, geom="point") +
+   stat_summary(aes(Passengers,Price,group=Passengers),
+                fun.y = function(x) quantile(x,.5),
+                fun.ymax = function(x) quantile(x,.75),
+                fun.ymin = function(x) quantile(x,.25),
+                width=0.5,geom="crossbar",colour="blue") +
+   stat_quantile(colour="blue", size=1, linetype=2) +
+   stat_quantile(colour="orange",
+                 size=1,
+                 linetype="twodash",
+                 quantiles=c(0.1, 0.9)) +
+   stat_smooth(method=lm, se=FALSE, colour="red")
> ggp3
```


A quantitative explicative variable



A quantitative explicative variable

QR model for a given conditional quantile θ (linear regression):

$$Q_\theta(\mathbf{Y}|\mathbf{X}) = \mathbf{X}\beta(\theta)$$

where

- $0 < \theta < 1$
- $Q_\theta(.|.)$ denotes the conditional quantile function for the θ^{th} quantile

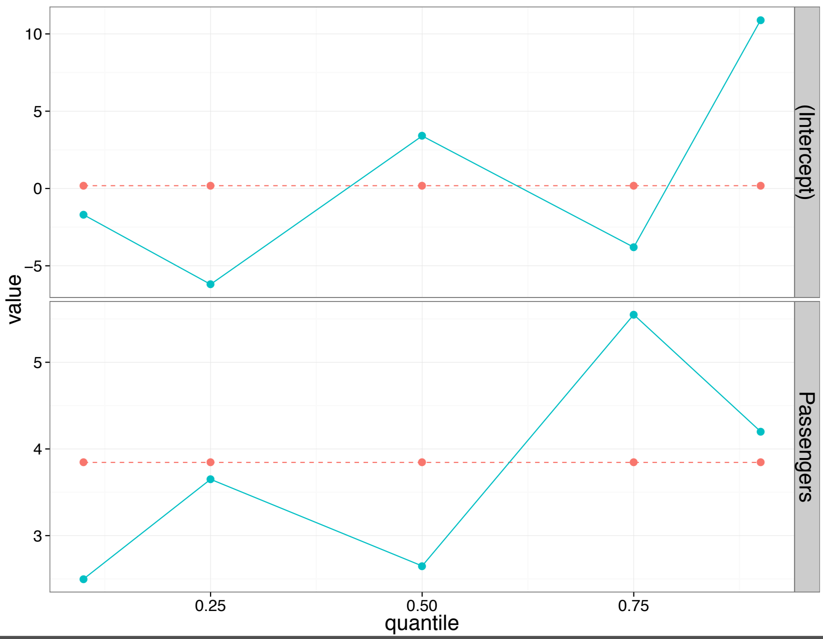
Interpretation

$$\beta_i(\theta) = \frac{\partial Q_\theta(\mathbf{Y}|\mathbf{X})}{\partial \mathbf{x}_i}$$

Rate of change of the θ^{th} quantile of the dependent variable distribution per unit change in the value of the i^{th} regressor

A quantitative explicative variable

	OLS	θ=0.1	θ=0.25	θ=0.5	θ=0.75	θ=0.9
(Intercept)	0.95	-1.7	-6.20	3.40	-3.80	10.9
Passengers	2.01	2.5	3.65	2.65	5.55	4.2



The QR conditional and marginal responses

Using an adeguate grid of quantiles, it is possible to "reconstruct" both the conditional response distributions and the marginal response distribution:

1. estimate the QR model using a grid of quantiles
2. combine the different estimates for the different values of θ
3. select the "best" model for each unit according to the deviation between the observed value and the estimated values
4. predict the response variable using the "best" models determined at the previous step

The QR conditional and marginal responses

```
> tau2use <- seq(0.01,0.99,by=0.05)
> regLS <- lm(Price ~ Passengers, data=miniCars)
> regQR <- rq(Price ~ Passengers, data=miniCars, tau=tau2use)
> yLS <- predict(regLS)
> yQR <- predict(regQR)
> colnames(yQR) <- tau2use
> head(yQR)
```

```
     0.01 0.06 0.11 0.16 0.21 0.26  0.31 0.36  0.41 0.46  0.51  0.56 0.61 0.66
[1,]  8.0 10.1 10.9 11.3 12.0 12.2 12.45 13.9 15.45 15.9 17.35 17.65 18.5 20.2
[2,]  8.0 10.1 10.9 11.3 12.0 12.2 12.45 13.9 15.45 15.9 17.35 17.65 18.5 20.2
[3,]  8.0 10.1 10.9 11.3 12.0 12.2 12.45 13.9 15.45 15.9 17.35 17.65 18.5 20.2
[4,]  8.6 12.8 13.5 14.3 15.6 15.8 15.90 18.4 18.80 19.3 20.70 20.90 21.1 23.7
[5,]  7.4  7.4  8.3  8.3  8.4  8.6  9.00  9.4 12.10 12.5 14.00 14.40 15.9 16.7
[6,]  8.6 12.8 13.5 14.3 15.6 15.8 15.90 18.4 18.80 19.3 20.70 20.90 21.1 23.7
     0.71 0.76  0.81 0.86  0.91  0.96
[1,] 21.5 26.1 27.05 29.0 33.05 36.45
[2,] 21.5 26.1 27.05 29.0 33.05 36.45
[3,] 21.5 26.1 27.05 29.0 33.05 36.45
[4,] 25.3 33.8 34.30 34.7 36.10 37.70
[5,] 17.7 18.4 19.80 23.3 30.00 35.20
[6,] 25.3 33.8 34.30 34.7 36.10 37.70
```

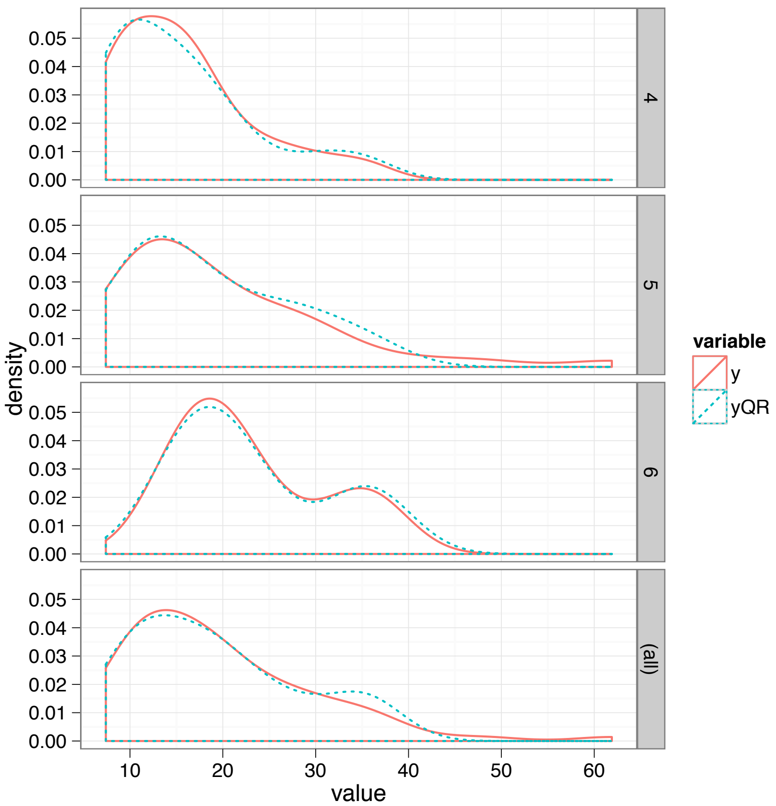
The QR conditional and marginal responses

```
> y <- miniCars$Price
> units2quantili <- as.numeric(colnames(yQR)[apply(abs(y - yQR), 1, which.min)])
> table(units2quantili)

units2quantili
0.01 0.06 0.11 0.16 0.21 0.26 0.31 0.36 0.41 0.46 0.51 0.56 0.61 0.66 0.71 0.76
   2    5    5    4    4    4    3    5    3    6    3    4    3    5    5    3
0.81 0.86 0.91 0.96
   4    5    4    5

> t(table(units2quantili, miniCars$Passengers))

   units2quantili
    0.01 0.06 0.11 0.16 0.21 0.26 0.31 0.36 0.41 0.46 0.51 0.56 0.61 0.66 0.71
  4    1    0    1    0    2    1    2    1    1    1    2    2    1    0    2
  5    1    5    3    4    1    1    0    3    1    4    0    0    2    3    2
  6    0    0    1    0    1    2    1    1    1    1    1    2    0    2    1
   units2quantili
    0.76 0.81 0.86 0.91 0.96
  4    1    1    2    1    1
  5    2    2    2    2    3
  6    0    1    1    1    1
```


The QR conditional and marginal responses

```
> yQR <- yQR[cbind(1:nrow(yQR),apply(abs(y - yQR), 1, which.min))]
> head(data.frame(y, yQR, miniCars$Passengers))

     y   yQR miniCars.Passengers
1 15.9 15.90                   5
2 33.9 33.05                   5
3 29.1 29.00                   5
4 37.7 37.70                   6
5 30.0 30.00                   4
6 15.7 15.80                   6

> library(reshape2)
> dfm <- melt(data.frame(y, yQR, miniCars$Passengers), id=3)
> ggp4 <- ggplot(aes(x=value, group=variable), data=dfm) +
+   geom_density(aes(colour=variable, linetype=variable)) +
+   facet_grid(miniCars.Passengers ~ ., margins=TRUE)
```

The QR conditional and marginal responses

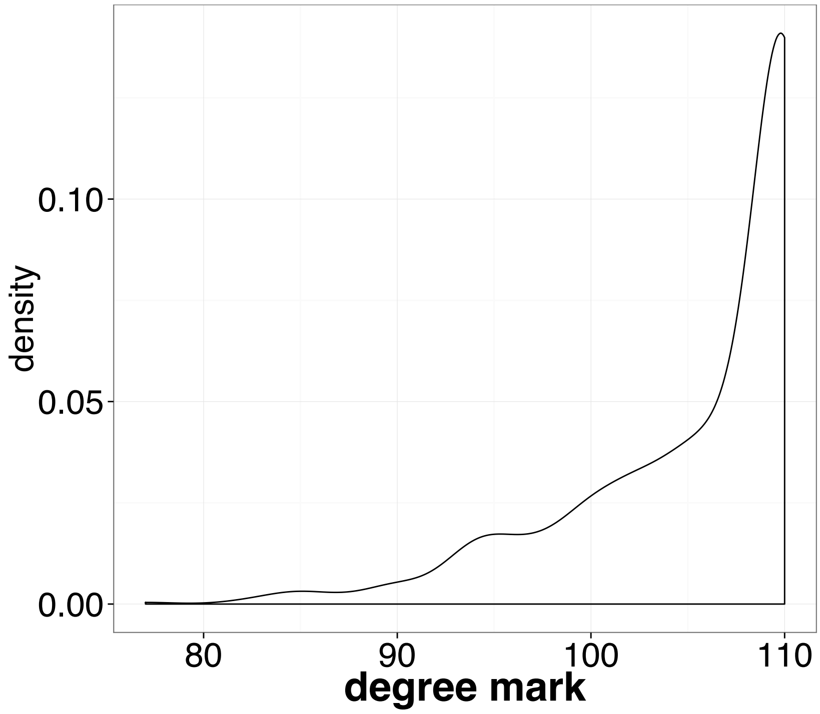


The dataset

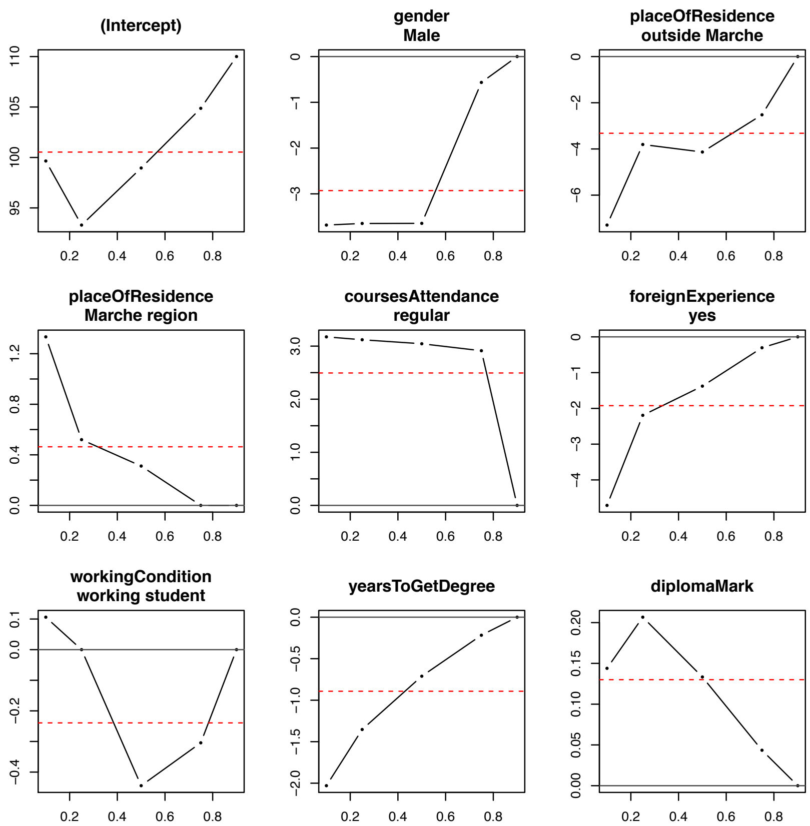
The evaluation of University educational processes

- random sample of 685 students graduated at University of Macerata (Italy)
- dependent variable: degree mark
- 7 regressors related to the student profile:
  - gender
  - place of residence during University (Macerata and its province, Marche region, outside Marche)
  - course attendance (no attendance, regular)
  - foreign experience (yes, no)
  - working condition (full time student, working student)
  - number of years to get a degree
  - diploma mark

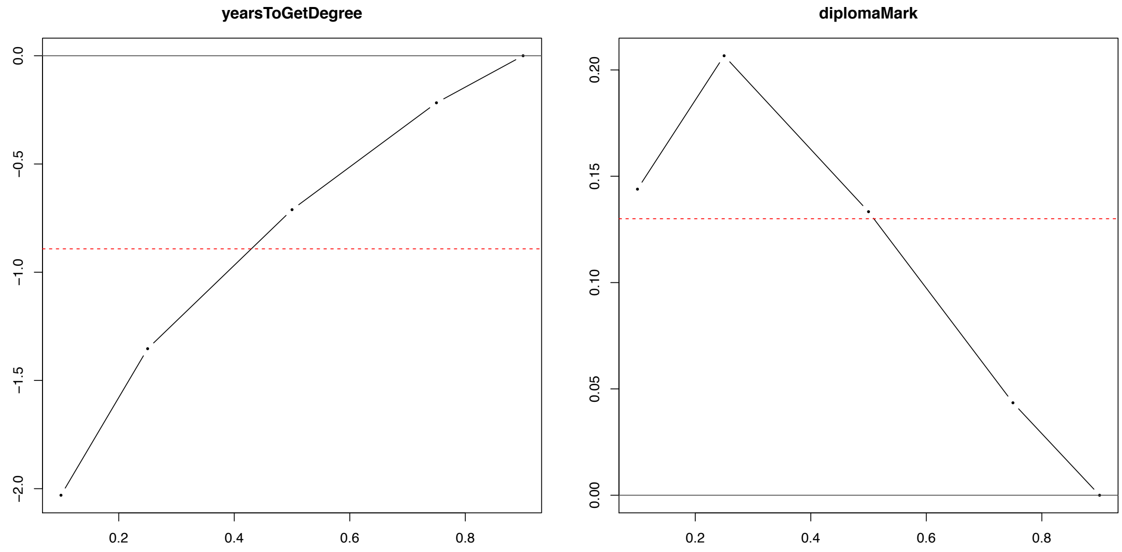
The response variable



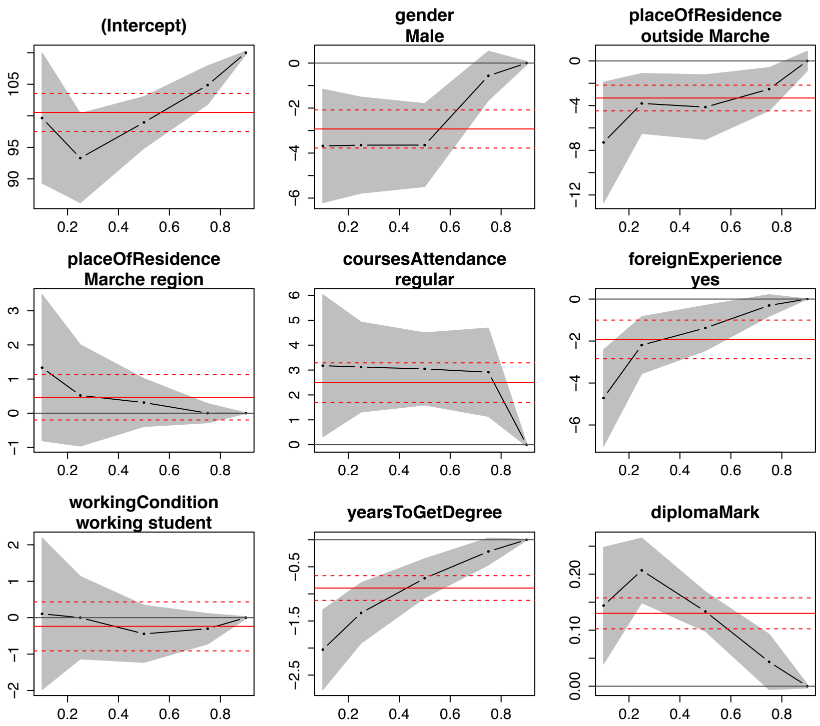
The QR estimates



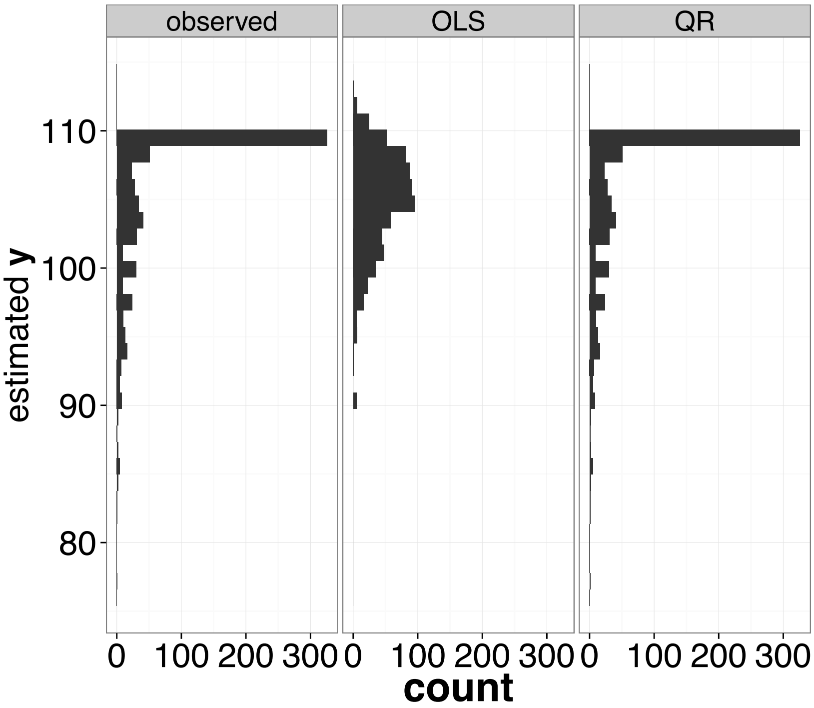
The QR estimates (zoom)



The QR estimates



The estimated response variable



A practical guide: a summary

- While classical regression gives only information on the conditional expectation, QR extends the viewpoint on the whole conditional distribution of the response variable
- QR provides location, scale and shape shift information on the conditional distribution of the response variable
- Using a dense set of quantiles, QR allows to approximate the conditional and marginal distribution of the response

Note: Through QR, it is possible to conduct inference on the comparison among quantiles in two or more groups in a non parametric context