What is Quantile Regression and How to use the Rpackage quantreg

Quantile Regression (Part 2)

Lecture Video Slides

Engel Dataset from Rpackage quantreg

- Dataset that records Family Expenditure on Food and Family Income in Belgium 1857.
- Used to show a limitation of Linear Regression and usefulness of Quantile Regression.
- Dataset is in quantreg Rpackage.

	income	foodexp
1	420.1577	255.8394
2	541.4117	310.9587
3	901.1575	485.6800
4	639.0802	402.9974
5	750.8756	495.5608
6	945.7989	633.7978
7	829.3979	630.7566
8	979.1648	700.4409
9	1309.8789	830.9586
10	1492.3987	815.3602

First 10 of 235 records in Engel Dataset.

R Package quantreg

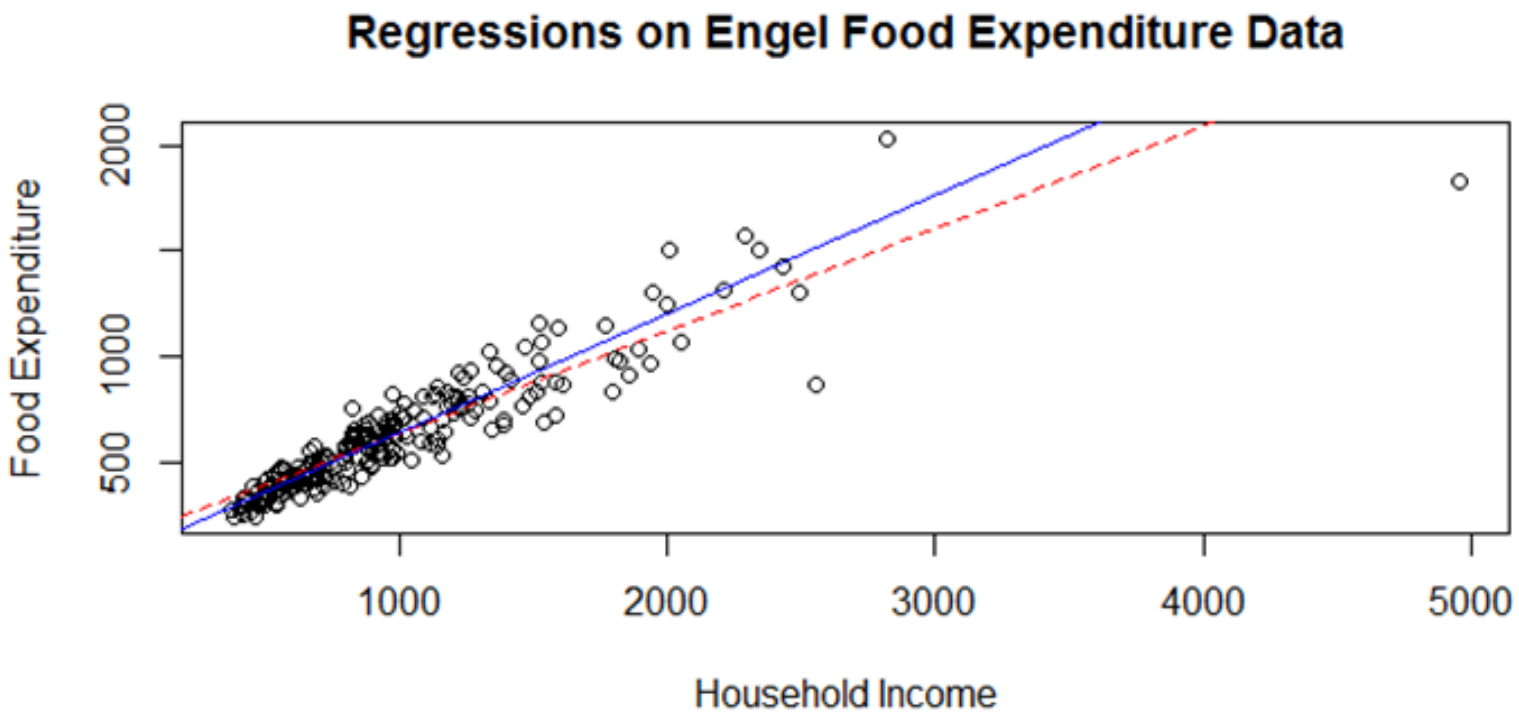
- Install R Package quantreg
- library(quantreg)
- The main function for quantile regression is rq()
- ?rq at Rconsole to view the documentation.


Console
Terminal
Jobs
R 4.1.2 · ~/
> ?rq
>
Files
Plots
Packages
Help
Viewer
Refresh He
R: Quantile Regression
Find in Topic
Quantile Regression
Description
Returns an object of class "rq" "rqs" or "rq.process" that represents a quantile regression fit.
Usage
rq(formula, tau=.5, data, subset, weights, na.action,
method="br", model = TRUE, contrasts, ...)
Arguments
formula
a formula object, with the response on the left of a ~ operator, and the terms, separated by + operators, on the right.
tau
the quantile(s) to be estimated, this is generally a number strictly between 0 and 1, but if

Parameter tau in rq() function controls the Percentile of Y

```
library(quantreg)
# Fit 50th Percentile Line (i.e. Median)
fit.p.5 <- rq(engel$foodexp ~ engel$income, tau=.5)
abline(fit.p.5, col="blue")
```

Parameter tau = 0.5



Median represented as Blue Line.
Mean represented as dotted red line.

Run the name of the model to get model coefficients

```
> fit.p.5
Call:
rq(formula = foodexp ~ income, tau = 0.5)

Coefficients:
(Intercept)      income
 81.4822474   0.5601806

Degrees of freedom: 235 total; 233 residual
```

50th Percentile: Food Expenditure = 81.48 + 0.56*income

Use summary() to get model coefficients and confidence intervals

```
> summary(fit.p.5)

Call: rq(formula = foodexp ~ income, tau = 0.5)

tau: [1] 0.5

Coefficients:
             coefficients lower bd  upper bd
(Intercept)   81.48225     53.25915 114.01156
income         0.56018      0.48702   0.60199
```

Q: What is the critical number that we want to check in the 95% confidence interval?

A: Check that the interval excludes 0.

Use summary(..., se = “nid”) to get p-values

```
> summary(fit.p.5, se = "nid")

Call: rq(formula = foodexp ~ income, tau = 0.5)

tau: [1] 0.5

Coefficients:
             Value    Std. Error t value  Pr(>|t|)
(Intercept) 81.48225 19.25066    4.23270  0.00003
income       0.56018  0.02828   19.81032  0.00000
```

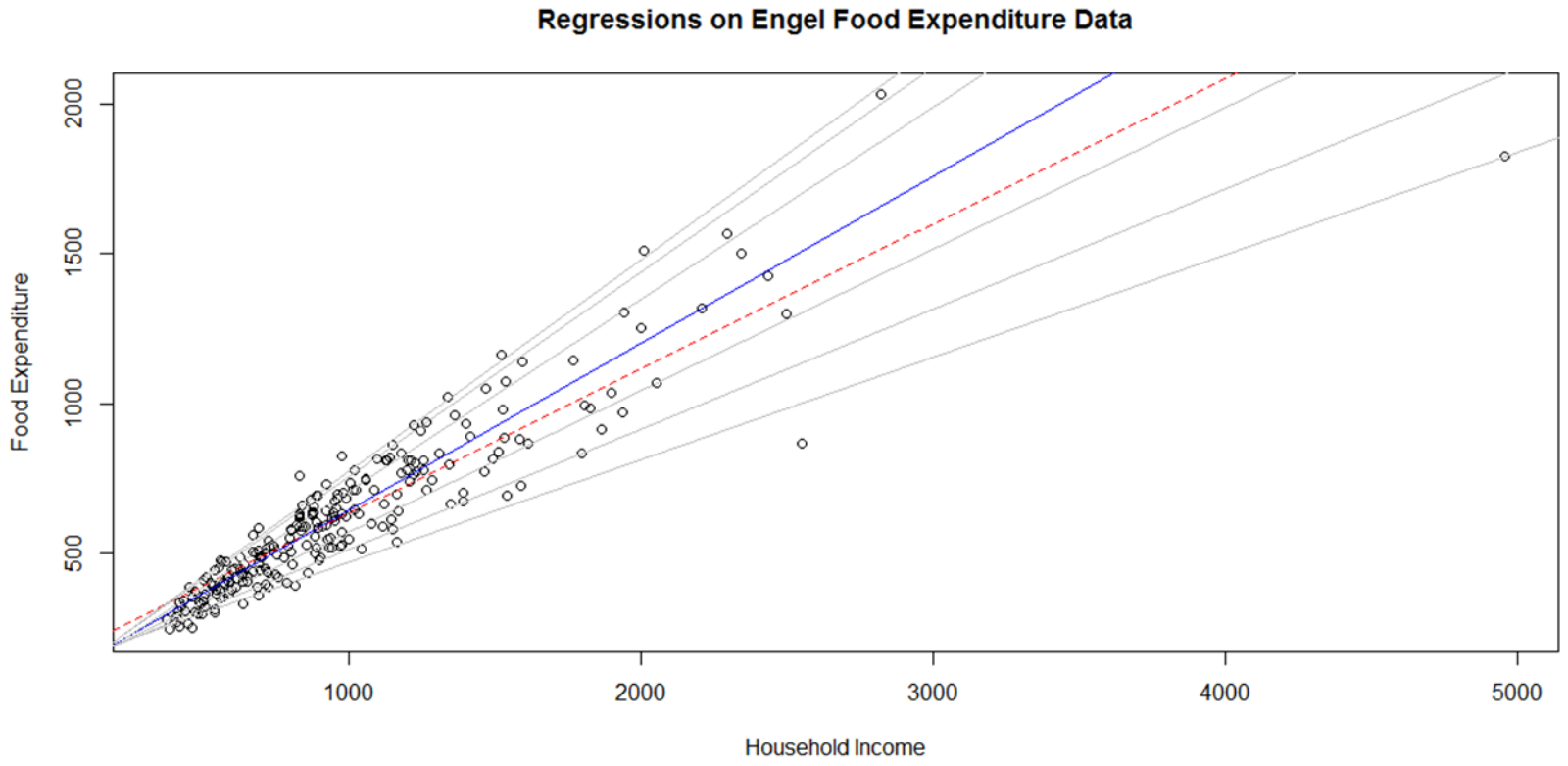
Q: What is the cut-off that we want to check in the p-values?

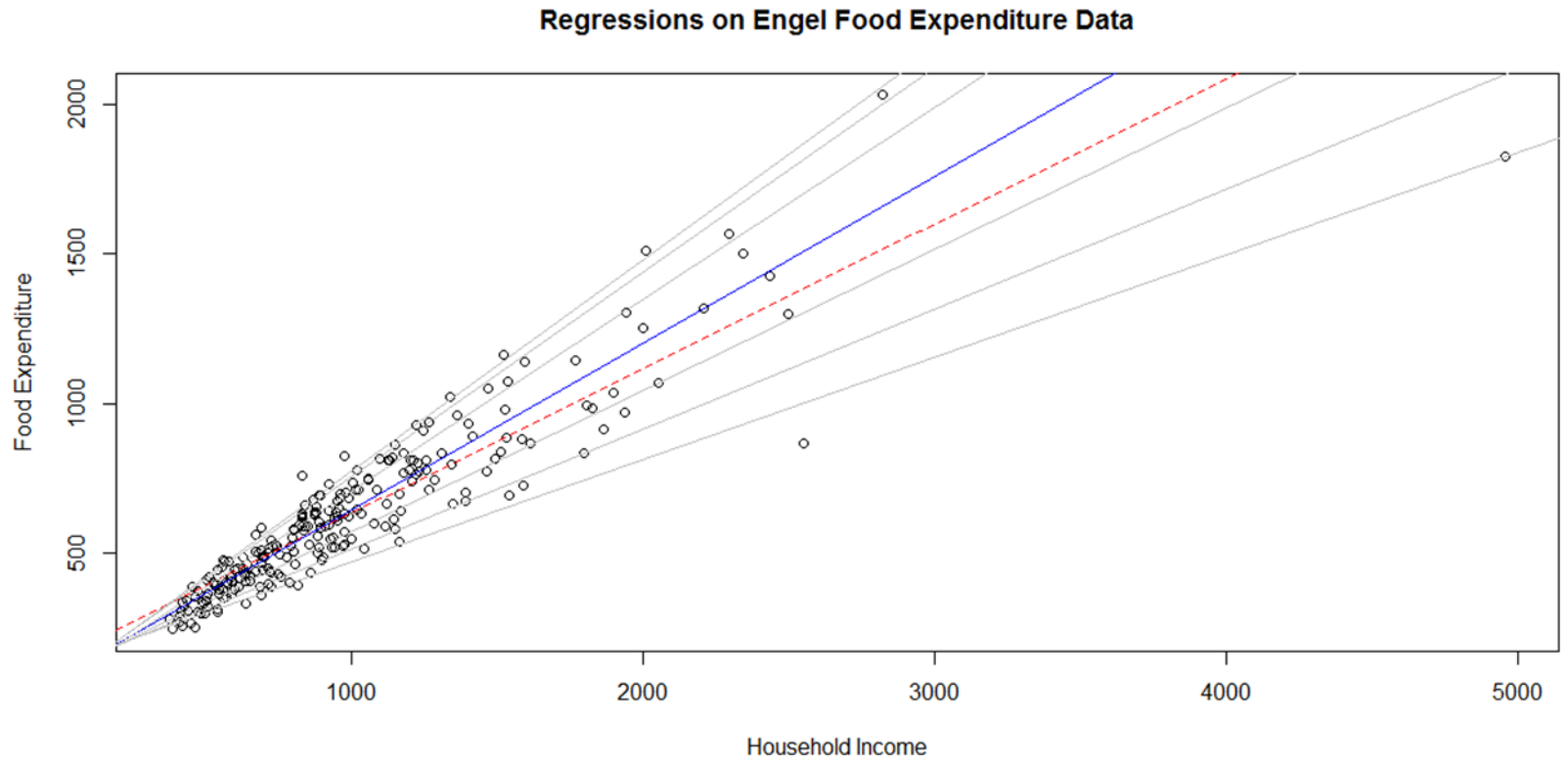
A: Check p-value < 0.05.

Fit & Plot Six Percentiles with a for loop

```
# 5th, 10th, 25th, 75th, 90th, 95th percentiles.
taus <- c(.05, .1, .25, .75, .90, .95)

# Plot the 6 percentile grey lines
for( i in 1:length(taus)){
  abline(rq(engel$foodexp~engel$income,tau=taus[i]), col = "grey")
}
```





Q: How does the quantile regression lines contribute to the data analysis?

A: They quantify the relationship between income and food expenditure at various percentiles. They show precisely how the variance in food expenditure increases as income increases.

Quantile Regression in Python

- statsmodels
  - Includes Engel dataset
  - https://www.statsmodels.org/dev/examples/notebooks/generated/quantile_regression.html

- scikit-learn
  - https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.QuantileRegressor.html

sklearn.linear_model.QuantileRegressor

class sklearn.linear_model.QuantileRegressor(*, *quantile=0.5*, *alpha=1.0*, *fit_intercept=True*, *solver='interior-point'*, *solver_options=None*)

How were Quantreg “lines” determined?

- Check Function.
- Hands-on Exercise.