M7777 Applied Functional Data Analysis
9. Functional Response with Scalar Covariate

Jan Koláček (kolacek@math.muni.cz)

Dept. of Mathematics and Statistics, Faculty of Science, Masaryk University, Brno

Functional Response Models

Functional ANOVA

Just as in the standard ANOVA, let be K ($K \geq 3$) groups and

$$x_{ij}(t) \ldots i\text{-th curve in } j\text{-th group}, i = 1, \ldots, n_j, \; n = \sum_{j=1}^{K} n_j$$

- An over-all mean

$$\hat{\mu}(t) = \bar{x}(t) = \frac{1}{n} \sum_{j=1}^{K} \sum_{i=1}^{n_j} x_{ij}(t)$$

- Effects for each group

$$\hat{\alpha}_j(t) = \frac{1}{n_j} \sum_{i=1}^{n_j} (x_{ij}(t) - \bar{x}(t))$$

- An error process

$$\hat{\varepsilon}_{ij}(t) = x_{ij}(t) - \hat{\alpha}_j(t) - \bar{x}(t)$$

Functional Response Models

Functional ANOVA model

$$x_{ij}(t) = \mu(t) + \alpha_j(t) + \varepsilon_{ij}(t) \qquad (1)$$

Suppose we observe scalar covariates $z_{i1}, \ldots, z_{iK}$

$$z_{ij} = \begin{cases} 1 & \text{if } x_{ij}(t) \text{ belongs to } j\text{-th group} \\ 0 & \text{otherwise} \end{cases}$$

Model (1) can be rewritten as a functional linear model

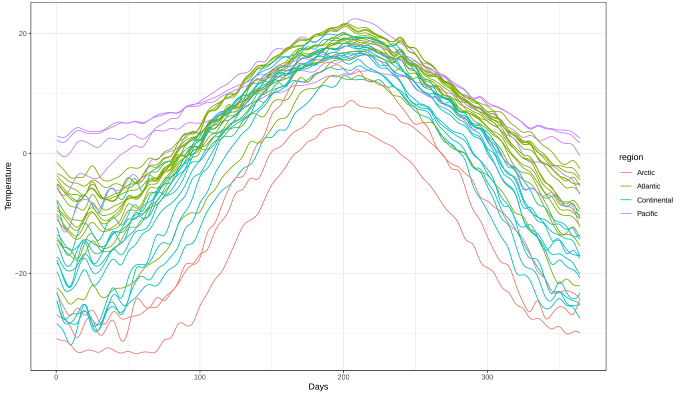
$$y_i(t) = \beta_0(t) + \sum_{j=1}^{K} \beta_j(t) z_{ij} + \varepsilon_i(t)$$

with conditions $\sum_{j=1}^{K} \beta_j(t) = 0$, $E\varepsilon_i(t) = 0$.

Functional Response Models

Canadian Weather

Divide all 35 locations to 4 regions: Atlantic, Continental, Pacific, Arctic



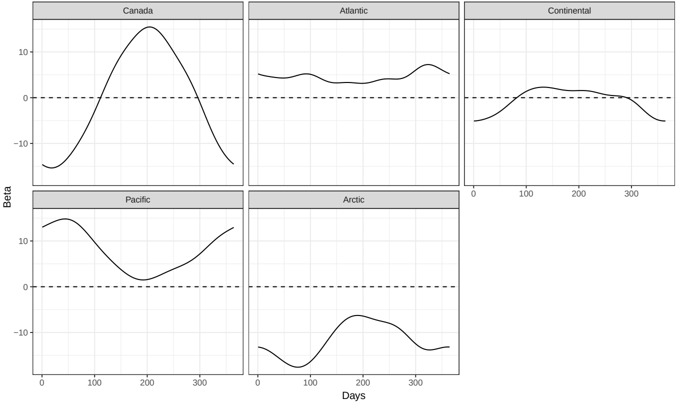
We will study the effect of geographic region on the shape of the temperature curves.

Functional Response Models

Let us denote model parameters

$\beta_0(t)$ ... Canada $\beta_1(t)$ ... Atlantic $\beta_2(t)$ ... Continental
$\beta_3(t)$ ... Pacific $\beta_4(t)$ ... Arctic

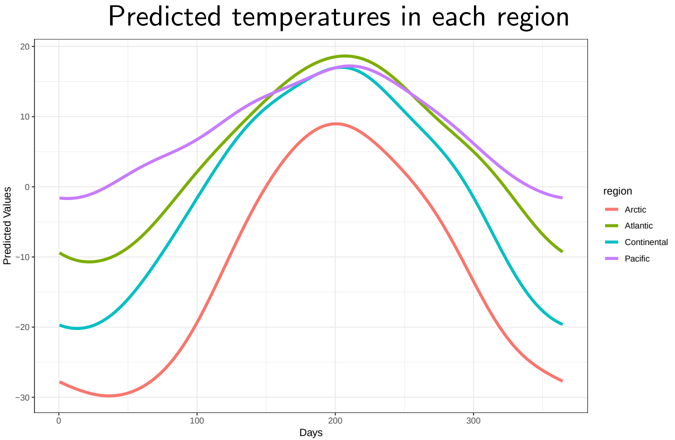
Estimates of $\beta_j(t)$



Functional Response Models

Prediction of the temperature curve in j-th group

$$\hat{y}_j(t) = \hat{\beta}_0(t) + \hat{\beta}_j(t), \qquad j = 1, \ldots, 4.$$



Functional Response Models

Confidence Intervals

Consider a basis representation

$$\beta_j(t) = \mathbf{\Phi}_j(t)\mathbf{c}_j$$

set $\mathbf{b} = (\mathbf{c}_0', \ldots, \mathbf{c}_K')' \Rightarrow \hat{\beta}_j(t)$ depends on $\hat{\mathbf{b}}$.
Let $\mathbf{t} = (t_1, \ldots, t_N)$, $\mathbf{y}_i = (y_i(t_1), \ldots, y_i(t_N))'$, generally, we minimize penalized least squares and get the estimate

$$\hat{\mathbf{b}} = \mathrm{y2cMap}\,\mathbf{y}.$$

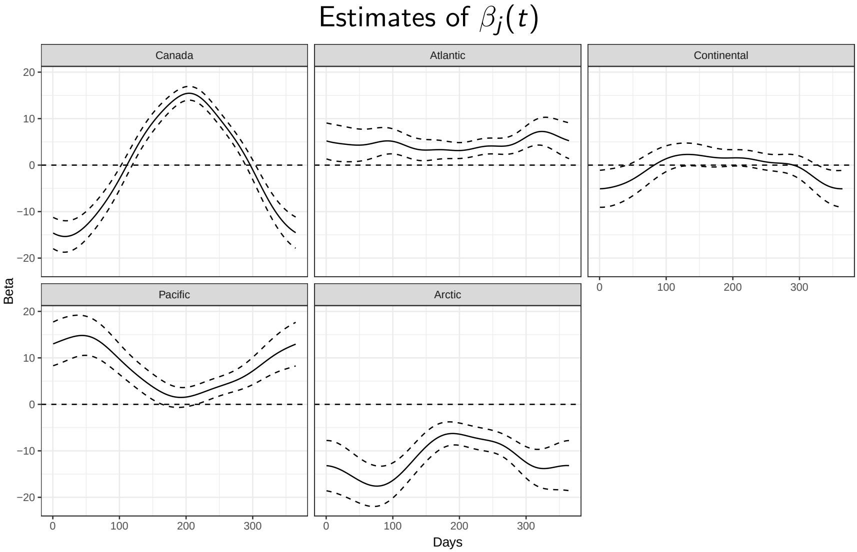
Estimate the covariance matrix

$$\hat{\mathbf{\Sigma}} = \frac{1}{n-K}\sum_{i=1}^{n} \hat{\varepsilon}_i\hat{\varepsilon}_i', \quad \text{where } \hat{\varepsilon}_i = \mathbf{y}_i - \mathbf{z}_i\hat{\beta}(\mathbf{t}).$$

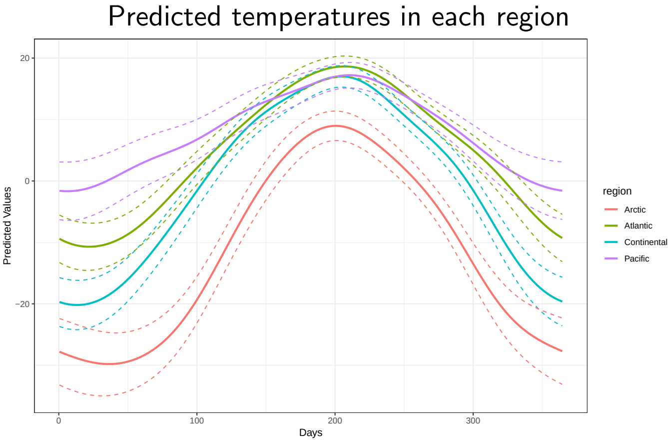
Thus (formally)

$$\mathrm{Var}\hat{\beta}_j(t) = \mathbf{\Phi}_j(t)\mathrm{y2cMap}\hat{\mathbf{\Sigma}}\mathrm{y2cMap}'\mathbf{\Phi}_j(t)'.$$

Functional Response Models

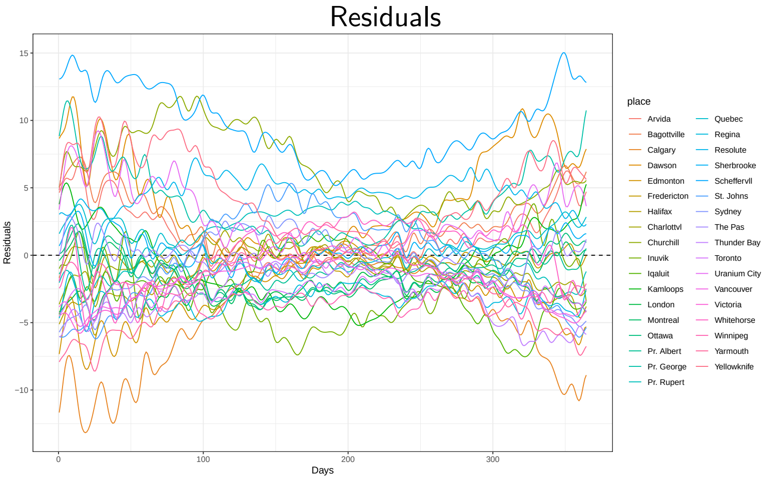


Functional Response Models



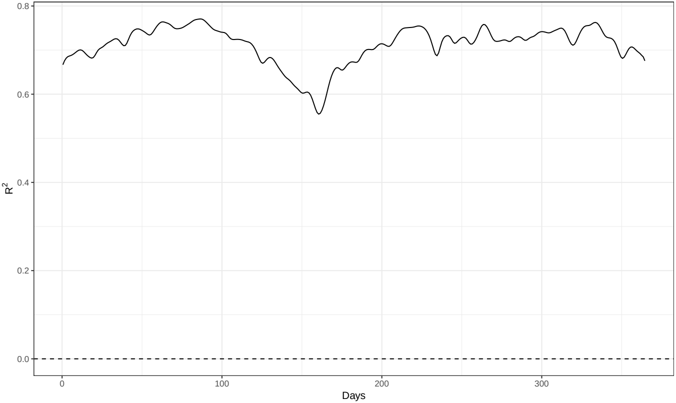
Functional Response Models

Assessing the fit of the fANOVA



Functional R^2

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i(t) - \hat{y}_i(t))^2}{\sum_{i=1}^{n} (y_i(t) - \bar{y}(t))^2}$$



F-statistic

To test significance, we can define a pointwise F-statistic

$$F(t) = \frac{\mathrm{Var}(\hat{\mathbf{y}}(t))}{\sum_{i=1}^{n} (y_i(t) - \hat{y}_i(t))^2 / n}$$

indicates where there is a large amount of signal relative to variance.

Test over-all regression significance based on

$$F^* = \max F(t).$$

Permutation Test

We would like to test the null hypothesis

$$H_0 : Ey(t) = 0 \quad \forall t \in [t_1, t_N]$$

Do B times

1. Permute indexes $1, \dots, n$ to get $i_1, \dots, i_n$, leaving the design unchanged.
2. Define $y_j^b(t) = y_{i_j}(t)$.
3. Estimate the model using $\mathbf{y}^b(t)$ as the response.
4. Measure F_b^* and set $I_b = \begin{cases} 1 & \text{if } F_b^* > F^* \\ 0 & \text{if } F_b^* \leq F^* \end{cases}$

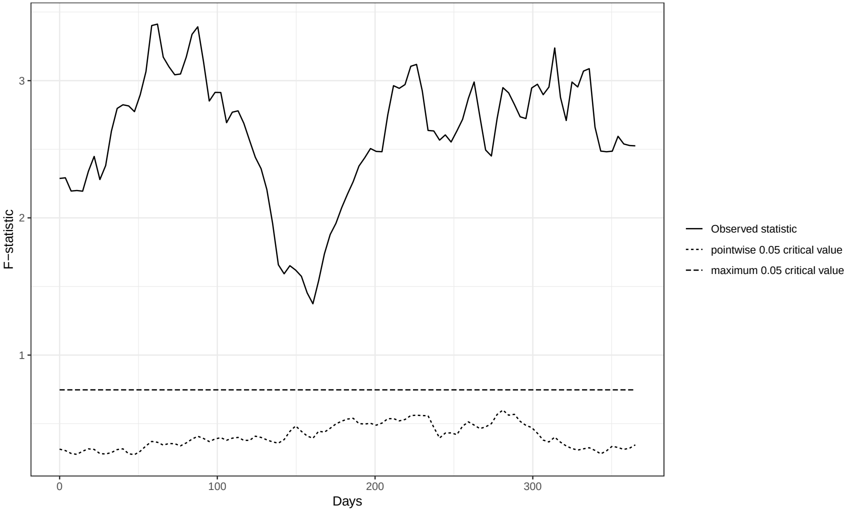
Then p-value for the test

$$p_B = \frac{1}{B} \sum_{b=1}^{B} I_b$$

Functional Response Models

Canadian Weather

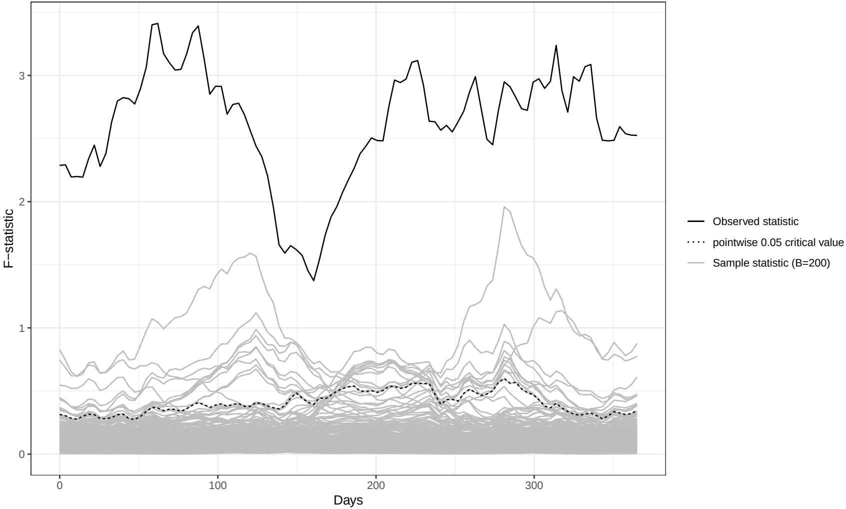
$B = 200, p_B = 0$



Functional Response Models

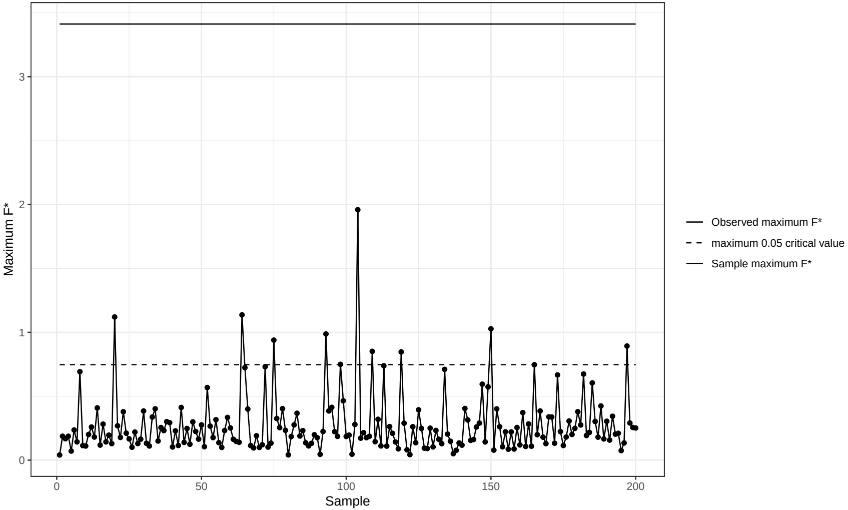
Canadian Weather

detailed test results



Canadian Weather

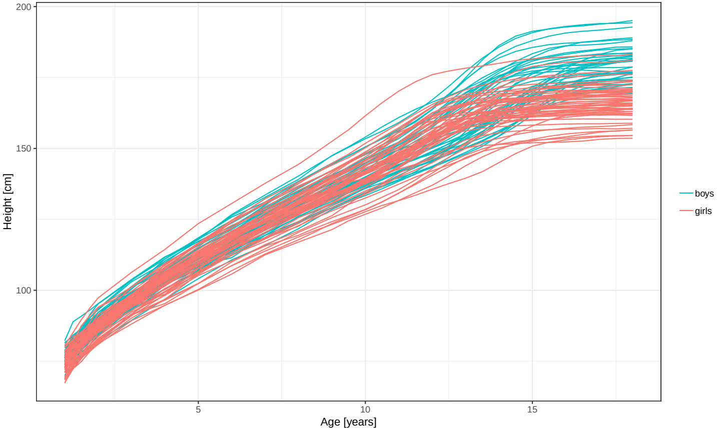
$F^* = 3.41, F^*_{0.95} = 0.747$



Functional t-test

Just 2 groups of curves $(x_{ij}(t), x_{i2}(t))$: Is the difference statistically significant?

Example. Berkeley Growth Study (39 boys, 54 girls)



Functional t-statistic

To test significance, we can define a pointwise t-statistic

$$T(t) = \frac{|\bar{x}_1(t) - \bar{x}_2(t)|}{\sqrt{\frac{1}{n_1}\mathrm{Var}[x_1(t)] + \frac{1}{n_2}\mathrm{Var}[x_2(t)]}}$$

indicates where there is a large mean difference relative to variance.

Test over-all significance based on

$$T^* = \max T(t).$$

Permutation Test

We would like to test the null hypothesis

$$H_0 : Ex_1(t) = Ex_2(t) \quad \forall t \in [t_1, t_N]$$

Do B times

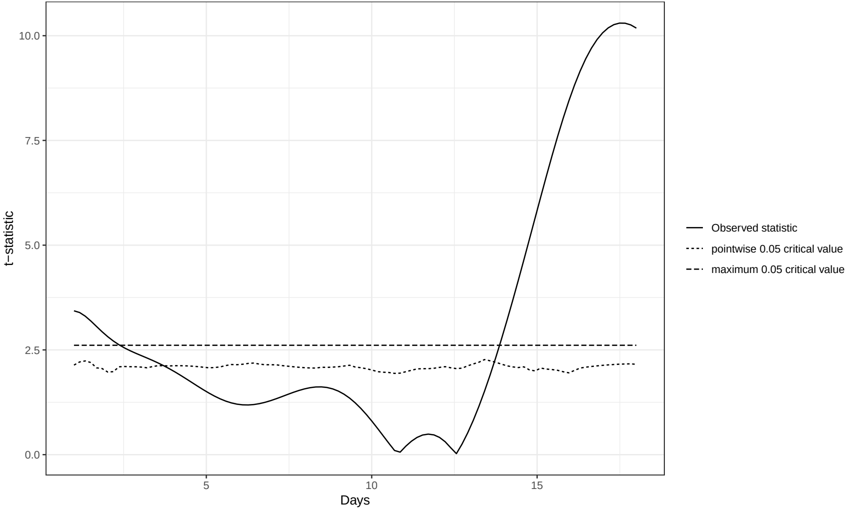
1. Randomly shuffle the labels of the curves.
2. Calculate the t-statistic $T_b(t)$ with the new labels.
3. Measure T_b^* and set $I_b = \begin{cases} 1 & \text{if } T_b^* > T^* \\ 0 & \text{if } T_b^* \leq T^* \end{cases}$

Then p-value for the test

$$p_B = \frac{1}{B} \sum_{b=1}^{B} I_b$$

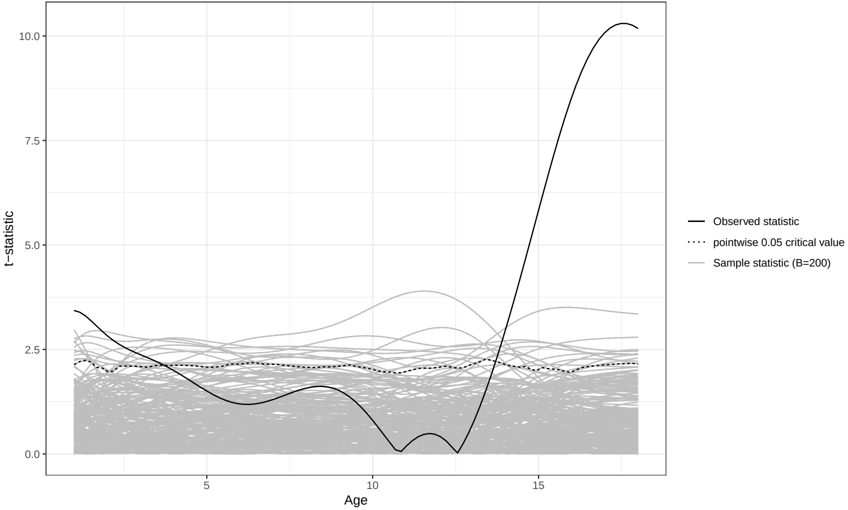
Berkeley Growth Study

$B = 200, p_B = 0$



Berkeley Growth Study

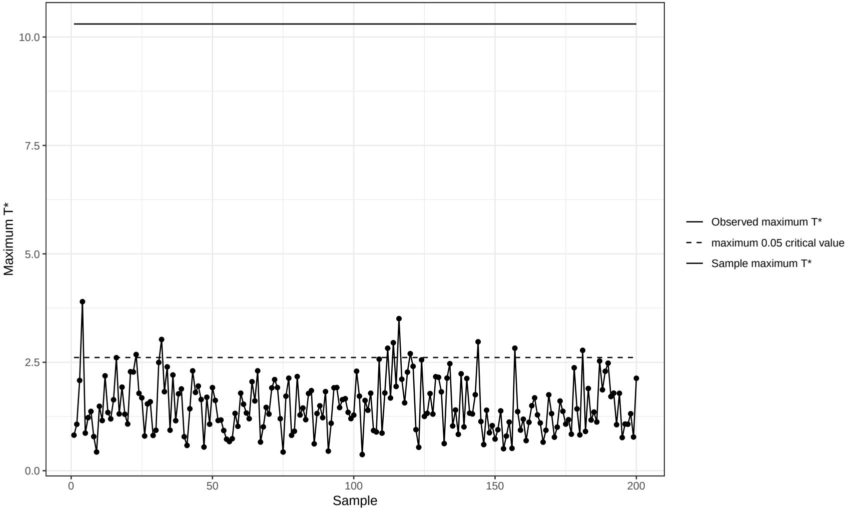
detailed test results



Functional Response Models

Berkeley Growth Study

$T^* = 10.3$, $T^*_{0.95} = 2.61$



Problems to solve

1. Sound Intensity Data
   Load the variable `rat3` from the `rat3.RData` file. The variable `rat3` contains observations of a rat neural activity evoked by sound intensity. The evoked potential (EPI) was measured in dependence on 19 sound intensities for 5 days. The dataset contains 79 repetitions for each day.
   - Smooth the data by B-spline bases with second-derivative penalties and plot the result with color-day specification (see Figure 1).
   - Conduct a study of the effect of the day on the shape of the EPI curves. Consider the fANOVA model with days as covariates. Plot estimated parameters with its pointwise confidence bands (see Figure 2).
   - Plot predictions for each day with its pointwise confidence bands (see Figure 3).
   - Plot functional R^2 of the model (see Figure 4) and interpret it.
   - Asses the model by the permutation test for F-statistic and plot the result (see Figure 5).

Problems to solve

2. Sound Intensity Data
    - Consider just days `SS4` and `SS5` and plot the EPI estimates with color-day specification (see Figure 6).
    - Is the difference between days statistically significant? Conduct the functional t-test.
    - Asses the model by the permutation test for t-statistic, plot the result (see Figure 7) and interpret it.

Problems to solve

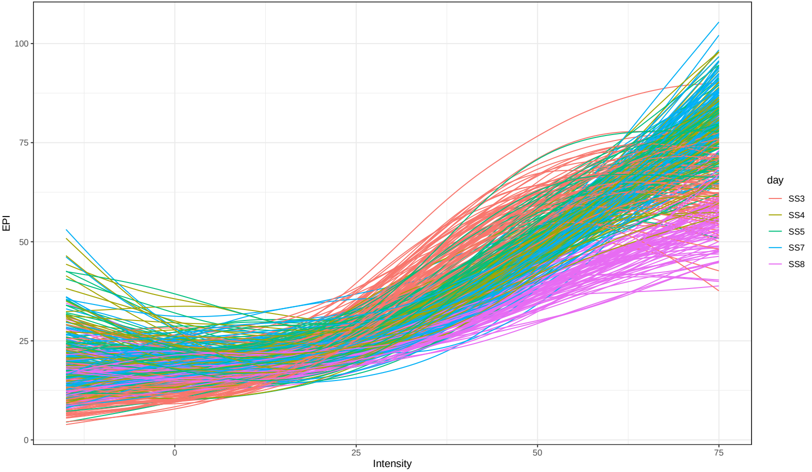


Figure 1.

Problems to solve

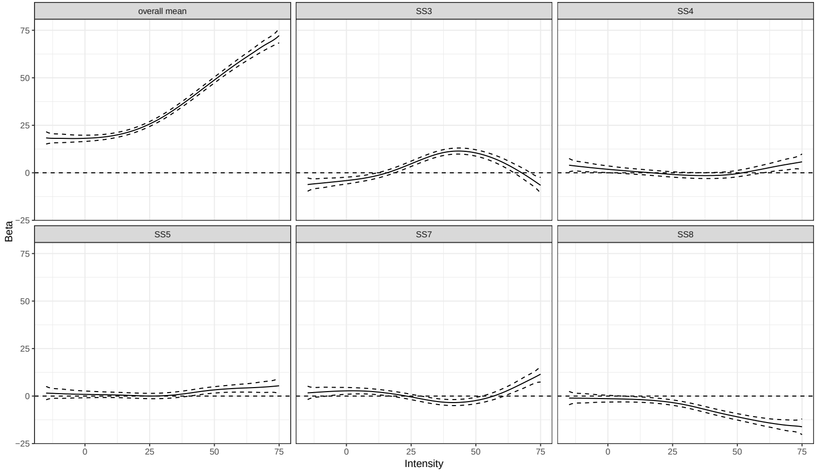


Figure 2.

Problems to solve

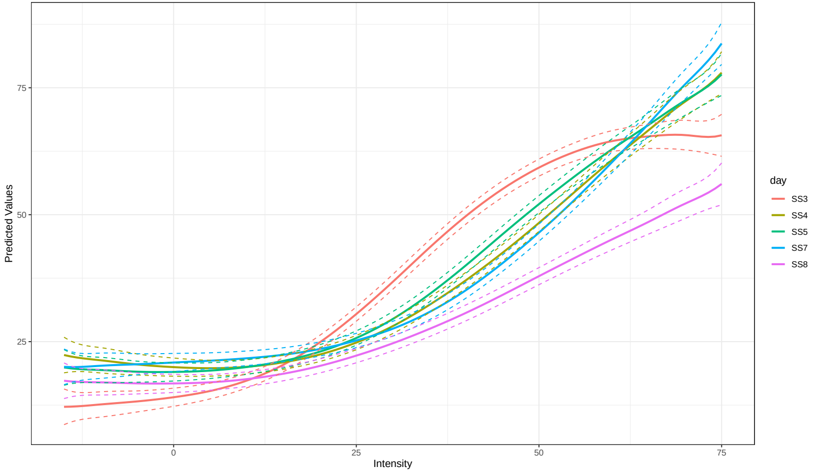


Figure 3.

Problems to solve

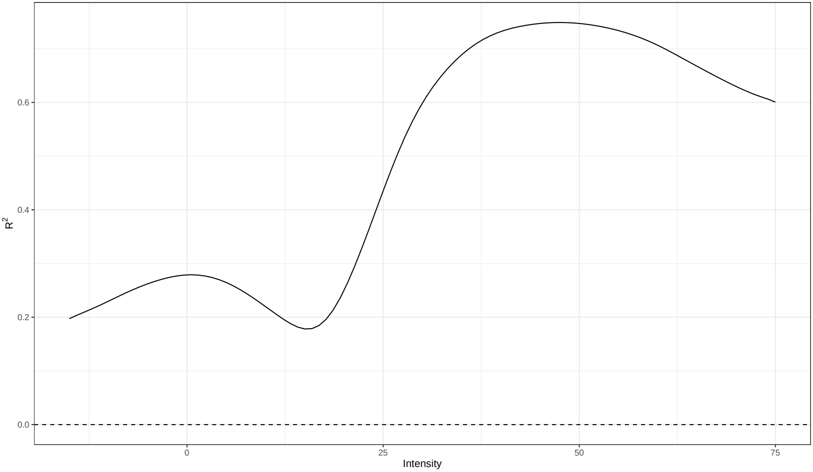


Figure 4.

Problems to solve

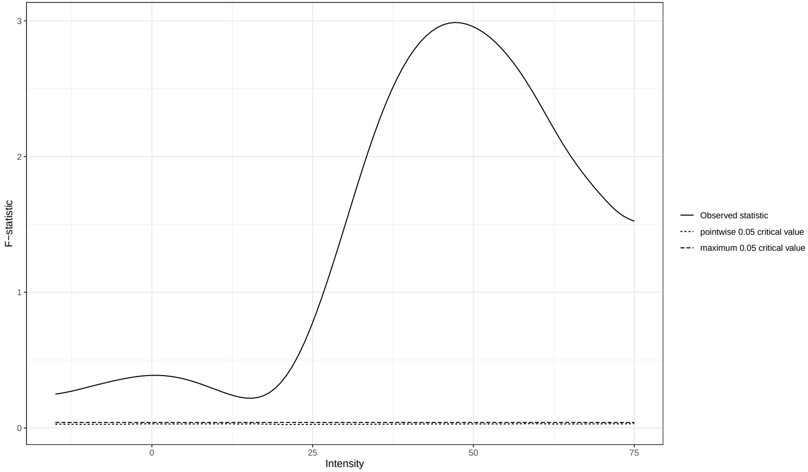


Figure 5.

Problems to solve

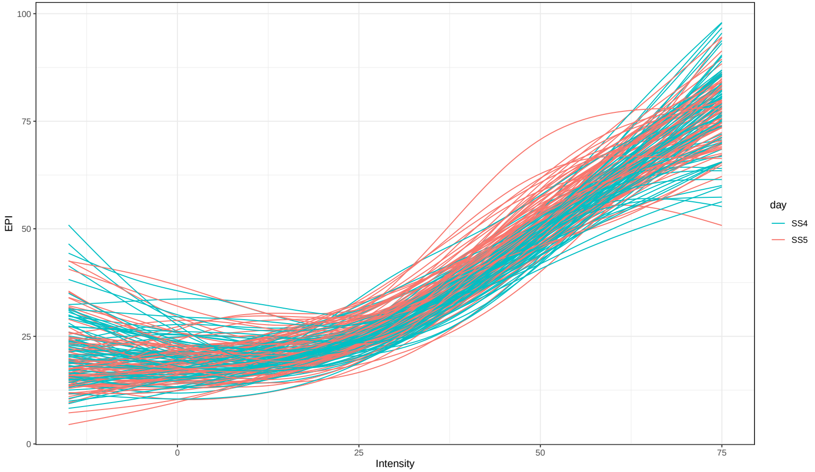


Figure 6.

Problems to solve

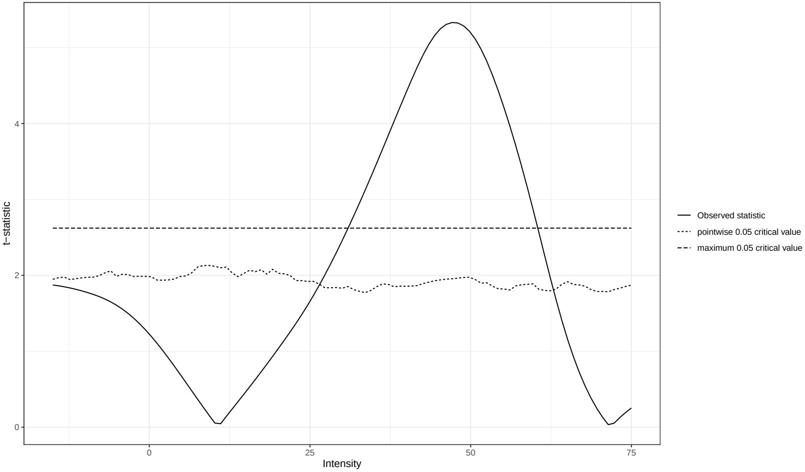


Figure 7.