1. In class, we have seen a way of constructing valid predictive coverage of the form

\[ \hat{C}(X) = [\hat{\mu}(X) - Q, \hat{\mu}(X) + Q], \]

where \( Q \) is the \( \lceil (1 - \alpha)(n+1) \rceil \) smallest value of the calibration residuals \( R_i = |Y_i - \hat{\mu}(X_i)|, \ i = 1, \ldots, n \) (see the lecture notes). Here we propose studying a different strategy.

(a) Suppose you hold two functions \( \hat{q}(X; \alpha/2) \) and \( \hat{q}(X; 1 - \alpha/2) \) which are estimates of the \( \alpha/2 \) and \( 1 - \alpha/2 \) conditional quantiles of \( Y \mid X \). We assume these are fixed (perhaps they have been fitted on an independent data set with a very fancy technique). For each data point \( (X_i, Y_i) \), \( i = 1, \ldots, n \) compute a score as follows:

\[ V_i = \max \{ \hat{q}(X_i; \alpha/2) - Y_i, Y_i - \hat{q}(X_i; 1 - \alpha/2) \}. \]

Interpret this score. What does it geometrically correspond to?

(b) Let \( Q \) be the \( \lceil (1 - \alpha)(n+1) \rceil \) smallest value of the \( V_i \)'s. Show that the interval

\[ \hat{C}(X) = [\hat{q}(X, \alpha/2) - Q, \hat{q}(X, 1 - \alpha/2) + Q] \]

has at least \( 1 - \alpha \) predictive coverage.

(c) Compare this approach with (1). Which one do you prefer and why?

(d) Suppose the estimated quantiles are very close to the estimated quantiles. How does the prediction interval (2) look like? Same question for (1) if the estimated regression function is very close to the true regression function.

(e) Bonus problem [You do not need to answer this to get a perfect homework score. This will however give you bonus points.] Implement (1) and (2) and implement it on a low-dimensional dataset of your choice (real or simulated) exhibiting heteroscedasticity; that is, the conditional variance should vary quite a bit. Comment on performance and differences.

2. In multiple testing, weighted procedures are used for various reasons; for instance, (1) to increase power, (2) to incorporate previous knowledge about individual hypotheses, (3) to represent difference in importance (some hypotheses may be more important than others) or (4) to incorporate other measures of loss. The purpose of this question is to examine weighted procedures. Here, we assume that we have \( m \)-hypotheses \( H_1, \ldots, H_m \) with \( p \)-values \( p_1, \ldots, p_m \). As usual, the \( p \)-values corresponding to true nulls are assumed to be uniformly distributed.

(a) Assign each hypothesis a weight \( w_i \geq 0 \) so that the family of weights obeys

\[ \frac{1}{m} \sum_{1 \leq i \leq m} w_i = 1. \]

Now consider the procedure that rejects if \( p_i/w_i \leq \alpha/m \), and fail to reject if \( w_i = 0 \). Show that this procedure controls the FWER (strongly) at level \( \alpha \).
(b) Define $p_i^* = p_i/w_i$ and sort the $p_i^*$’s. Consider the step-down procedure which keeps on rejecting $H(i)$ (according to the sorting above) until for the first time,

$$p(i) > \frac{\alpha}{\sum_{j=1}^{m} w(j)}.$$ 

Show that this procedure also controls the FWER at level $\alpha$. Show that this procedure rejects more than that in 1. What happens when all the $w_i$’s are the same? When $w_1 = m$ and all the others vanish?

(c) A more conceptual approach to a weighted controlling procedure is as follows: assign a weight $w_i$ to each hypothesis as above and define

$$wFDR = E \left[ \frac{\sum_{i=1}^{m} w_i V_i}{\sum_{i=1}^{m} w_i R_i} \right].$$

As usual $V_i = 1$ if $H_i$ is falsely rejected whereas $R_i = 1$ if $H_i$ is rejected. Explain in words why this may be a good criterion to use. Propose a step-up procedure (based on the sorted original $p$-values) à la Benjamini Hochberg to control the wFDR at level $q$. 

2