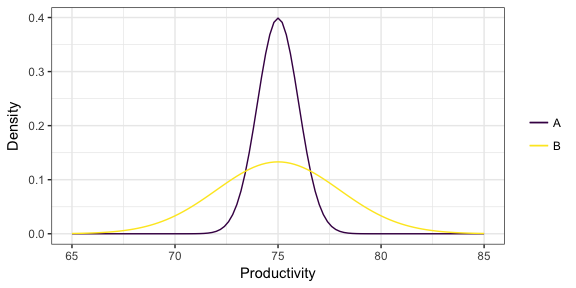
For (1) and (2) below, you're choosing between two candidates to hire. Discuss the pros and cons of choosing one candidate over the other in the following situations.

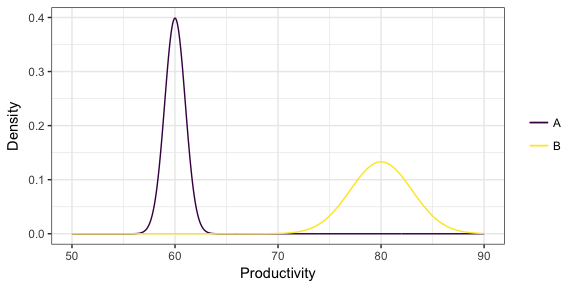
1. Both are predicted to have the same productivity score of 75, but have the following probabilistic forecasts.

[](https://github.com/vincenzocoia/BAIT509/blob/master/class_meetings/cm08-beyond_mean_mode_files/figure-html/unnamed-chunk-17-1.png)

Hiring A will give a higher probability to get a productivity between 72.5 to 77.5, and any score beyond these two stops are close to 0. The advantage of hiring A would be a much stable range for the productivity score, and there is no chances for A to get a score lower than 72.5. The disadvantage of hiring A would be no chances of getting a higher score than 77.5.

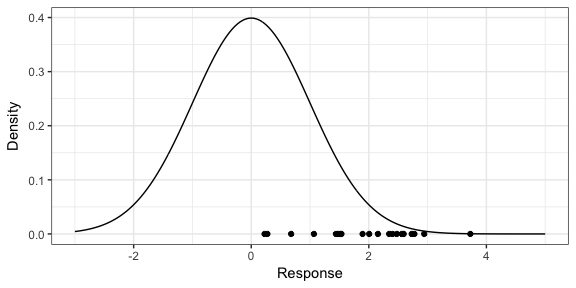
Hiring B, on the other hand, will give a wider range of productivity score from 65 to 85. The advantage for hiring B is the chances of getting higher score than 77.5 is much higher than A, but the same rule applies if B is getting a score lower than 72.5, that B might be more likely to get a lower score. The productivity of score for B is not as stable as A, giving uncertainty in this situation.

1. Two "non-overlapping" forecasts:

[](https://github.com/vincenzocoia/BAIT509/blob/master/class_meetings/cm08-beyond_mean_mode_files/figure-html/unnamed-chunk-18-1.png)

In this situation, it is more advantageous to choose B over A because the distribution of B lies between the range from 70 to 90, with a mean value of 80. Even though it has a larger variance than A, A only has a mean value of 60 and B is stochastically better than A, as the lowest value that B could give is still higher than the highest value that A could give in this situation.

1. You've formed a probabilist forecast for a particular value of the predictors, displayed below as a density. You then collect test data for that same value of the predictor, indicated as the points below the density. What is the problem with the probabilistic forecast?

[](https://github.com/vincenzocoia/BAIT509/blob/master/class_meetings/cm08-beyond_mean_mode_files/figure-html/unnamed-chunk-19-1.png)

This forecast biased since the test data points concentrate on the right side of the distribution. The highest density of the test data points is within the range of [2,3], but the shape of the probabilist forecast indicates a distribution with a mean of 0. In addition, the test points do not lie on the left side of the probabilist forecast, indicating that the forecast needs to be calibrated.