

Session 14

In the example which was presented in the previous session, suppose we have historical data.

Suppose we know the true average and standard deviation. Suppose we have 20 historical data points. It is a limited amount of data.

```
set.seed(1)      # Keeps random data from changing each time the code is run
theta <- 18      # True long-run average temp
sigma <- 3       # True std dev of temp around this mean
n <- 20          # number of simulated historical data points

y <- rnorm(n,theta,sigma)

hist(y, breaks = 10, main = "Histogram of historical temperatures in degrees C")
```

Figure 1: 49

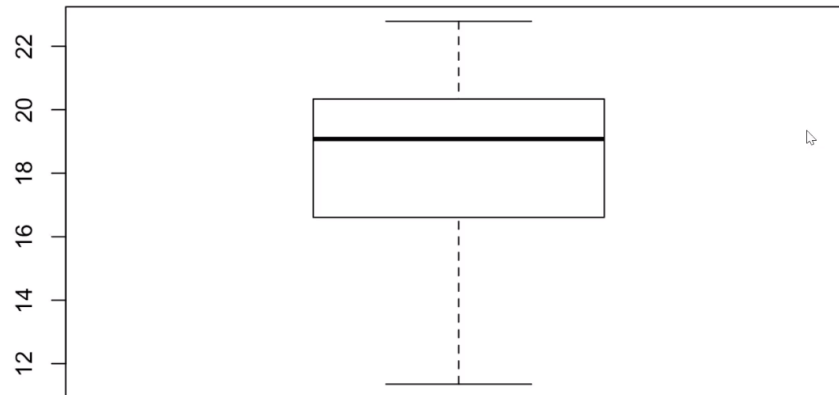


Figure 2: 50

In the above graph, you might think that the estimated average is higher than the actual average.

The estimated average and standard deviation are quite higher than the actual ones.

In this case, the decision-maker who relies on this model is overestimating the risk.

Suppose you have a forecasting system,

Instead of relying on historical data, suppose that we have a forecasting system. Also, suppose the system is biased and it tends to be low.

[Hide](#)

```

theta_hat <- mean(y)      # Sample mean
epsilon_hat <- y - theta_hat # Model residuals

ssr <- sum(epsilon_hat^2)  # Sum of squared residuals

sigma_hat <- ssr/(n-1)     # Estimated standard error
print(theta_hat)

```

[1] 18.57157

[Hide](#)

```

print(sigma_hat)

```

[1] 7.506291

Figure 3: 51

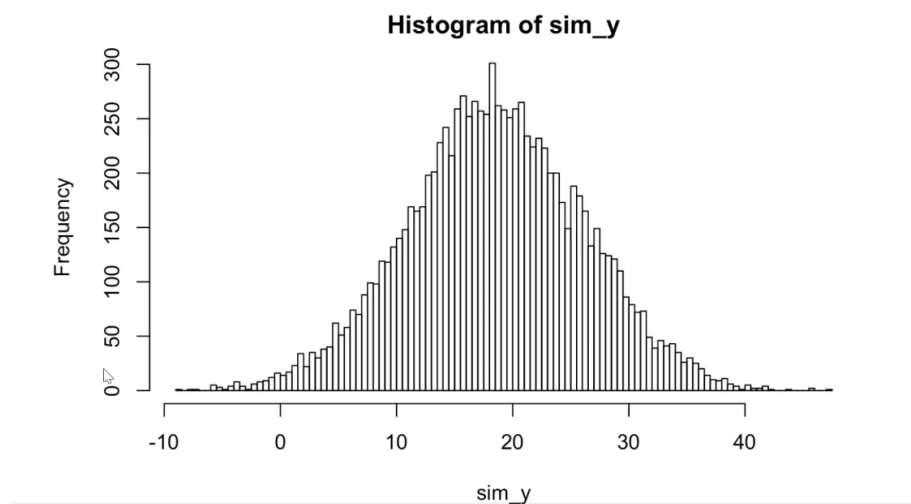


Figure 4: 52

Using a forecast

[Hide](#)

```

forecast_bias <- -0.5
forecast_std_error <- 1

forecast_errors <- rnorm(n, mean = forecast_bias, sd = forecast_std_error)

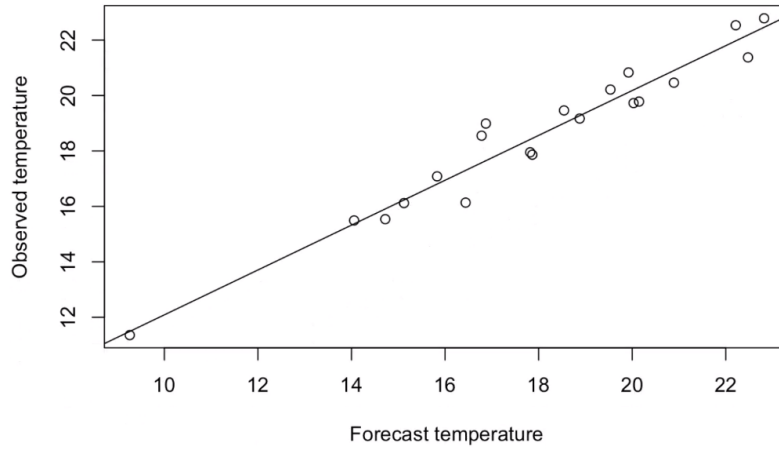
x <- y + forecast_errors

linear_model <- lm(y~x)

plot(x,y, xlab = "Forecast temperature", ylab = "Observed temperature")
abline(linear_model)

```

Figure 5: 53



It is a linear regression that looks at forecasts against the observations.

Next, we are going to figure out if you have the function system available, then what is the value of this forecasting system in providing information and delivering a better decision.

y_1, y_2, \dots, y_n historic observations

x_1, x_2, \dots, x_n historic forecasts

1- Estimate the model without the forecast

$$y_i = \theta + \varepsilon_i$$

$$\hat{\theta} = \frac{\sum y_i}{n} = \bar{y}$$

$$\hat{\varepsilon}_i = y_i - \bar{y}$$

$$s^2 = \frac{\sum (\varepsilon_i^2)}{n-1}$$

$$Y_{n+1} = N(\hat{\theta}, s^2)$$

$$E[L(Y|a = noinsurance)|\hat{\theta}, s^2] = E.L.$$

$$E.L. \geq costofinsurance$$

Here, $\hat{\theta}, s^2$ estimated with error. Therefore, if you do not handle them properly, you will get the error wrong.

If your model mis-specifies the original data generative process, you will also have the mis-specification of your actual risk and it may lead to the wrong decision.

living with some kind of misspecification might be fine, depending on the application.

2- Estimate w/ forecasts, what is the expected value of information?

Now the question is how much better your decision is going to be if you have a predictive model that helps inform the decision. It might have some signal that may be is not perfect because it is only information. Then, on average how much loss you avoid compare to if you made otherwise the same decision-making process but without the predictive model. The information is valuable if it leads you to the different decision that you would not have mad otherwise.

If you have a forecast, the forecast depends on the signal you may receive. Therefore,

$$a^* = a^*(x)$$

How to use the forecast?

Method:

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad u_i \sim N(0, \gamma^2) \text{ i.i.d}$$

Use OLS (or your preferred technique), to estimate parameters.

You can use the maximum likelihood here, too. It will give you a biased estimator that has a smaller variance.

$$\text{estimate } \hat{\beta}_0, \hat{\beta}_1, \hat{\gamma}^2$$

Generate the predictive distribution of $Y|X$.

Suppose we make the best linear fit. Then, we have now condition information for the given y's and therefore with some estimate of distribution, you can map that distribution to the distribution of the losses. The loss distribution will have a heavyweight on zero and then increasingly less and less weight on higher levels of loss. If there is a minimum threshold, all of the values before the threshold in the first graph (normal distribution) will map to a loss equal to zero, and above that they all map to higher and higher losses but with less and less probability. Here, we do not have any forecasting information.

But the idea here is instead of getting the forecast, you can have for some given x values, the distribution of the $Y|x$ which is much sharper. Then, you can map that conditional distribution. The conditional distribution will have again heavyweight on zero but much sharper distribution, in comparison to the previous graph.

Then given the above distribution of loss,

$$E[L(Y|X)] = E.L.(x)$$

For any given x,

$$\text{Let } a^*(x) = \operatorname{argmin} E.L.(x), C_0$$

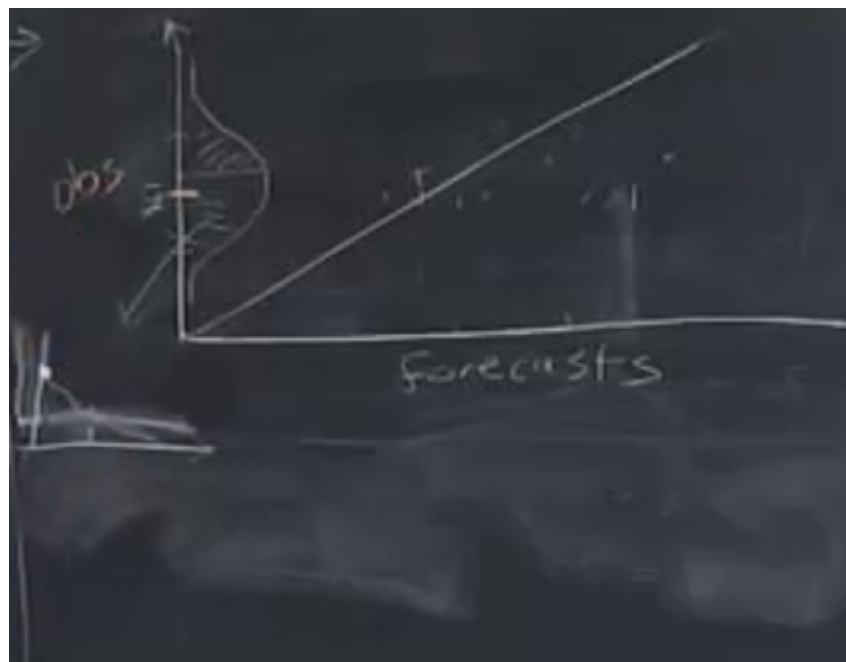


Figure 6: 55

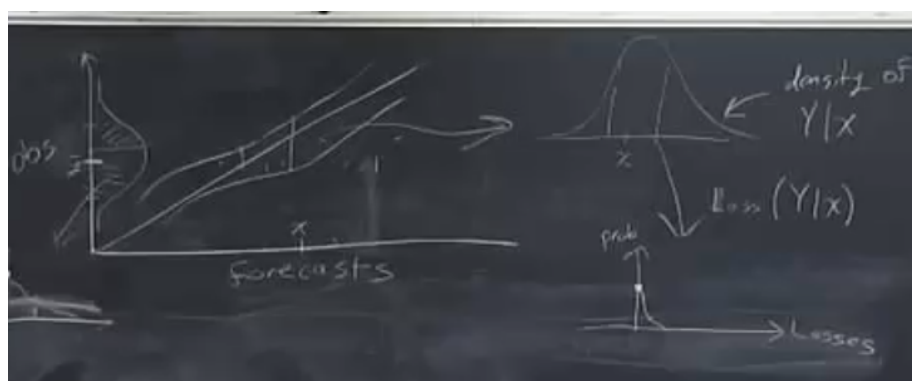


Figure 7: 56

How much value does this create by refining the prediction system in this way?
How you quantify that additional value?

Suppose we got the forecast, and when the forecast is high, you decide to buy the insurance. On the other hand, when the forecast is low, you decide not to buy the insurance. There is a critical boundary here. I know in advance, which x , I am going to map to which action. On the right part, the expected loss is higher than buying the insurance, and on the left part, the expected loss is lower than buying the insurance.

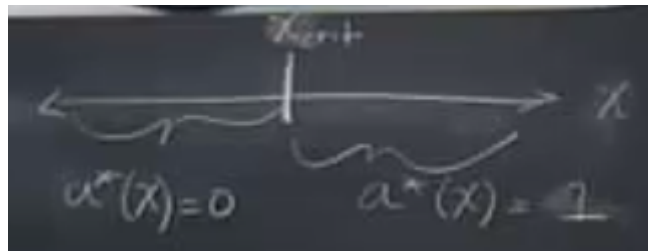


Figure 8: 57

To document how much value our method will create we can use the historic data. We can compare the action they actually did with how well we claim that could have done. That is one way to do that.

We can obtain the expected loss over the distribution of the forecast signal x and then compare it with the loss which was obtained without considering the forecasting signal. In calculating the value of the loss without considering the forecasting signal, it does not depend on x .