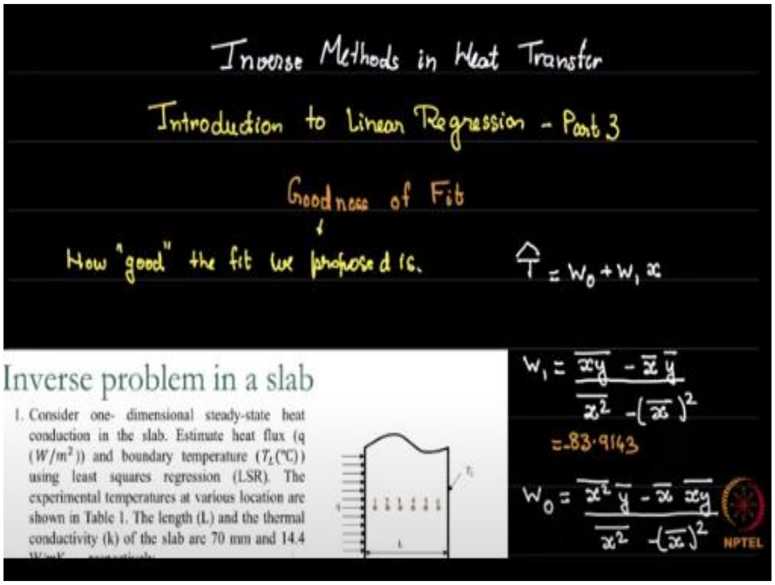
Inverse Methods in Heat Transfer
Prof. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology, Madras

Lecture No # 08
Module No # 02
Goodness of fit and coefficient of Determination

(Refer Slide Time: 00:19)



Inverse Methods in Heat Transfer

Introduction to Linear Regression - Part 3

Goodness of Fit

How "good" the fit we proposed is.

$\hat{T} = w_0 + w_1 x$

Inverse problem in a slab

1. Consider one- dimensional steady-state heat conduction in the slab. Estimate heat flux (q (W/m^2)) and boundary temperature (T_L(°C)) using least squares regression (LSR). The experimental temperatures at various location are shown in Table 1. The length (L) and the thermal conductivity (k) of the slab are 70 mm and 14.4

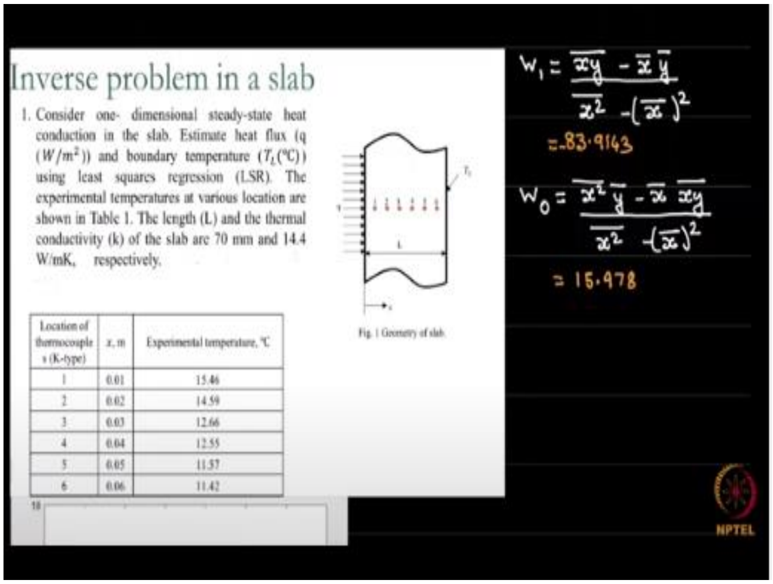
$$w_1 = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - (\bar{x})^2}$$

$= -83.9143$

$$w_0 = \frac{\overline{x^2}\,\bar{y} - \bar{x}\,\overline{xy}}{\overline{x^2} - (\bar{x})^2}$$

Welcome back. So, we will just continue our discussion from last part. In the last part we looked at how we can use linear regression to solve an inverse problem? In this part, we will look at an important quantity called the goodness of fit. What the goodness of fit determines is how good the regression fit we got was? So, what this determines is how good the fit we proposed is? So, in the last video we had seen the example of a linear model with the data that was given here in this inverse problem.

(Refer Slide Time: 01:20)



Inverse problem in a slab

1. Consider one- dimensional steady-state heat conduction in the slab. Estimate heat flux (q (W/m^2)) and boundary temperature (T_L(°C)) using least squares regression (LSR). The experimental temperatures at various location are shown in Table 1. The length (L) and the thermal conductivity (k) of the slab are 70 mm and 14.4 W/mK, respectively.

Location of thermocouples (K-type)	x, m	Experimental temperature, °C
1	0.01	15.46
2	0.02	14.59
3	0.03	12.66
4	0.04	12.55
5	0.05	11.57
6	0.06	11.42

Fig. 1 Geometry of slab

$$w_1 = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - (\overline{x})^2}$$

$= -83.9143$

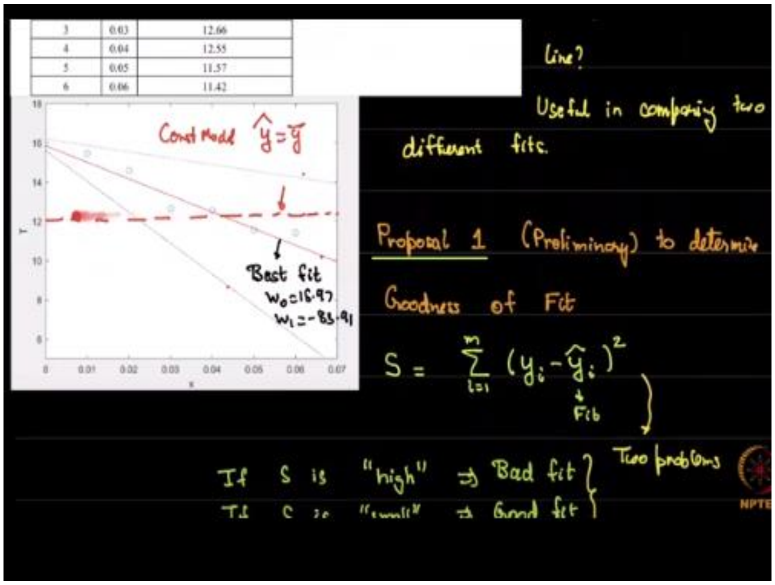
$$w_0 = \frac{\overline{x^2}\,\overline{y} - \overline{x}\,\overline{xy}}{\overline{x^2} - (\overline{x})^2}$$

$= 15.978$

NPTEL

And we had obtained both w_0 and w_1 the coefficients of the linear model and these were the values that we obtained.

(Refer Slide Time: 01:32)



3	0.03	12.66
4	0.04	12.55
5	0.05	11.57
6	0.06	11.42

line?

Useful in comparing two different fits.

Proposal 1 (Preliminary) to determine

Goodness of Fit

$$S = \sum_{i=1}^{m} (y_i - \underset{\text{Fit}}{\hat{y}_i})^2$$

If S is "high" ⇒ Bad fit
If S is "small" ⇒ Good fit

Two problems

NPTEL

Now once we obtain these values, once we plotted them remember, this was the best fit, that we had obtained this red line and the data is sort of scattered around this best fit line. So again, the question really is how good is the best fit line? Now why do we ask this question? Suppose somebody says that it is not a line which is the best fit, but let us say a quadratic curve which is the best fit how do, you compare these 2 different proposals for the hypothesis function proposals, for the function which actually fits the data.

Now in this particular case we had already used a linear fit because we set theory predicts that but suppose somebody has a new theory, they say no the theory that you used which was that you know k $\frac{dT}{dx}$ is 0 is wrong and here is experimental evidences that, Then we needed a comparison between essentially noisy data fits like our best fit line and some other fit and we need to compare 2 different theories, so this is useful in comparing 2 different fits.

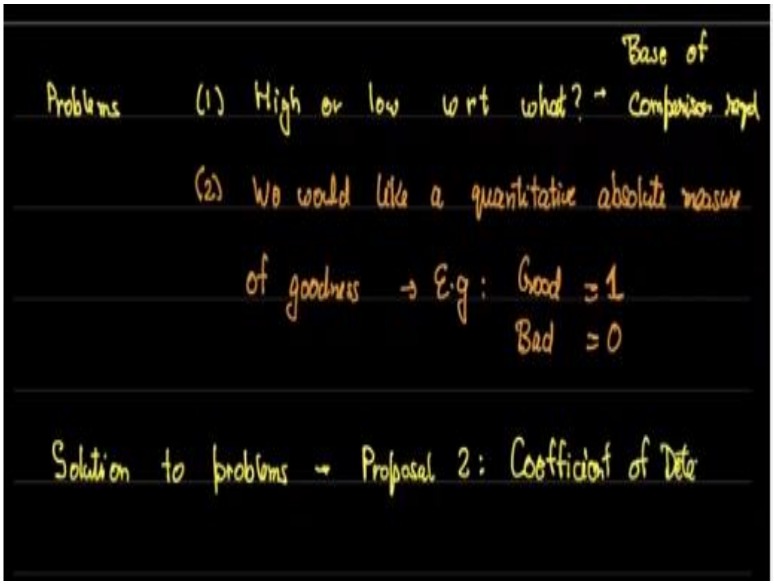
So here is an idea, we will start with the preliminary proposal. So, which I will call proposal 1. This is not quite accurate, but it will give us an idea of how to proceed to determine goodness of fit. So, the idea is very simple. we look at how good this fit is by finding out the error. So, I say we already have a measure for our error. So, we have S,

$$S = \sum (y_i - \hat{y_i})^2$$

Remember $\hat{y_i}$ is essentially our fit, And, we find this out and this is a measure of goodness of fit.

We sum this over all data points so this is our proposal remember this is not quite accurate for reasons that I will tell you. And we say that if S is high or large, then this means it is a bad fit and if S is small, then this is a good fit. So, this is our proposal on what a possible measure of goodness of fitness. So, it turns out that this has 2 problems. So, this entire proposal has 2 problems so those problems are as follows.

(Refer Slide Time: 04:55)



Problems (1) High or low wrt what? → Base of Comparison reqd

(2) We would like a quantitative absolute measure of goodness → E.g: Good = 1
Bad = 0

Solution to problems → Proposal 2: Coefficient of Dete

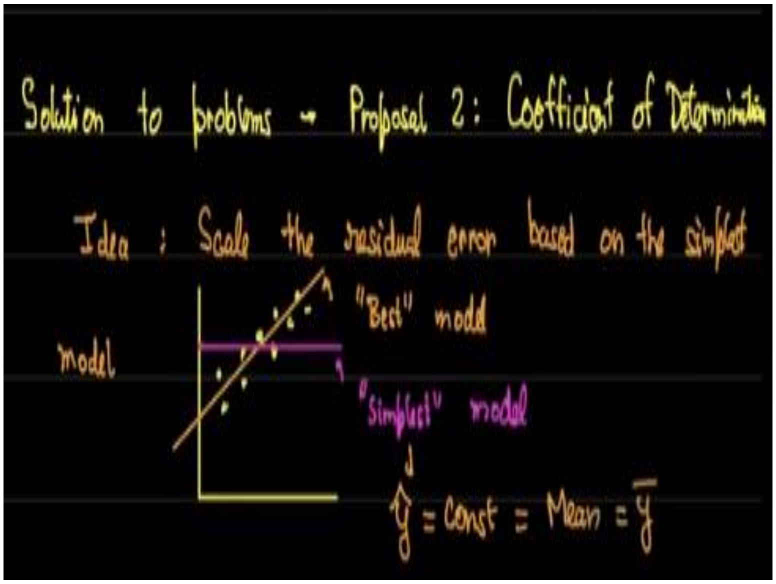
Firstly, high or low with respect to what? so this is a general lesson in science and engineering. one cannot talk about something being high or low without saying with respect to what. so there has to be a non-dimensionality or some comparison. For example, if I measure the temperature here, I have measured it in degree centigrade. Suppose I measure it in kelvin, it might look that well in this case it would not be kelvin but suppose I measure it in Frenheit it might look like the temperature gap is high.

If this distance is measured in millimeter versus this distance is measurement in kilo kilometer these are just numbers. So will I say the distance of one is high or what you will say one what 1 kilometer, 1 meter, 1 millimeter and depends on the physical context. So, without context without some comparative number okay so we need some base of comparison and without that we cannot really say. now the second problem is related to the first problem.

So, we would like a quantitative measure, absolute measure of goodness. So, like I said it is related. for example, good equal to 1 on a scale of 0 to 1 or as a scale of 0 to 10, we are going to choose 0 to 1 and we can say that bad is 0. So, these 2 problems turn out to be related and if we use something of this sort this will not satisfy it. For example, good and excellent fit will be 0 but a bad fit will not be one it will become a very large number as it turns out both these problems can be addressed by what is known as the coefficient of determination.

So the solution to these problems is our proposal 2 or the, this is called the coefficient of determination.

(Refer Slide Time: 07:44)



Solution to problems → Proposal 2: Coefficient of Determination

Idea : Scale the residual error based on the simplest model

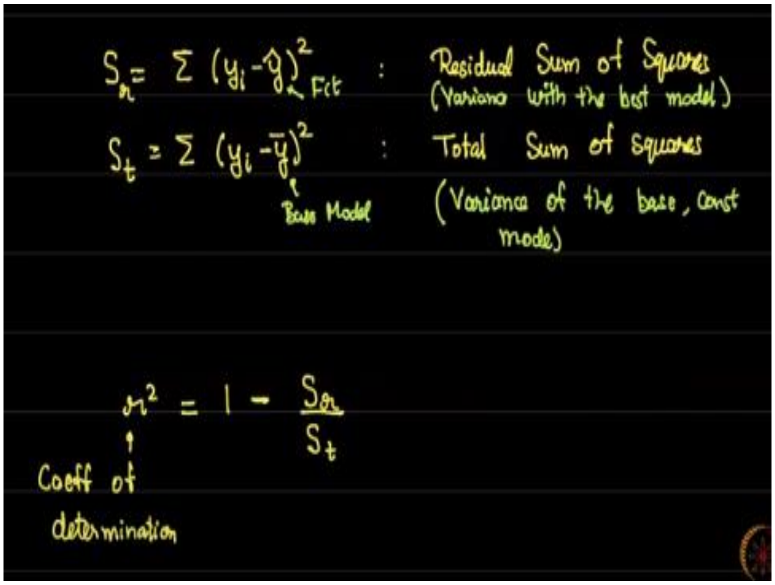
So, before we go there the idea is simple, the idea is scale, the residual error based on the simplest model. So, the idea is like this, supposes we have data, so data is scattered like this, now we have proposed the linear model so we are going to propose some model like this. So, this is the best model, but we can have the simplest model. The simplest model in this case is simply something like this.

What do I mean by the simplest model? the simplest model says $\hat{y}$ is a constant, that is regardless of what the location is. The temperature is the same and the temperature is going to be the mean that is $\bar{y}$. So, I hope this is clear. So, for example we had several model proposals here this was the best fit these 2 are bad fits, but a really bad fit could be possibly this.

So, you simply take the average of these temperatures and say that is actually what is going on physically. I have some average temperature and I have some erroneous results some errors around it. So, this is a $y = \bar{y}$ or $\hat{y} = \bar{y}$. this is a constant model or a simple model as you can see it is a very bad model. But nonetheless it is a model that we can calculate very quickly and we can say whatever it is my error base lies here.

Even if I do nothing, I will get a very simple model like this, at least any best fit model any new model that I am proposing should be more accurate than this. So, if it is only going to perform as well as a constant model, then it is a bad model. so will give you give you a bad score, if it is a really good model, I will give it a high score okay so let us use that.

(Refer Slide Time: 10:39)



$S_r = \Sigma (y_i - \hat{y})^2$ Fit : Residual Sum of Squares (Variance with the best model)

$S_t = \Sigma (y_i - \bar{y})^2$: Total Sum of squares (Variance of the base, Const mode)

Base Model

$r^2 = 1 - \frac{S_r}{S_t}$

Coeff of determination

So now remember we had this original error which was $\sum(y - \hat{y})^2$, this we are going to call S_r, which is called residual sum of squares. So, our proposal one was this itself would be the goodness of fit, but we say let us compare this with respect to something else. So, the comparative quantity that we are going to compare with is what happens with a bad model. with a bad model you simply have,

$$S_r = \sum (y_i - \hat{y})^2$$

So, this is a bad model, this is sometimes called total sum of squares, S_t, and as we see and you will know this much statistics I assume, but anyway we will come back to this.

$$S_t = \sum (y_i - \bar{y})^2$$

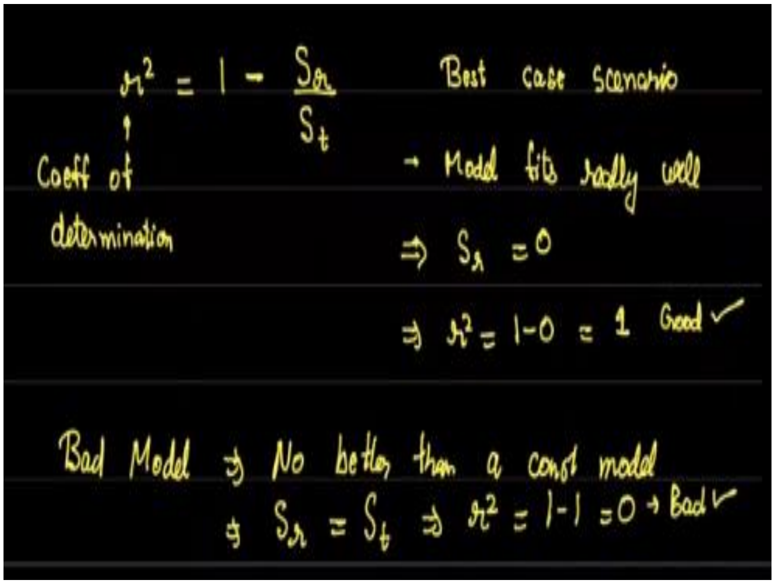
This is variance of the base model and this is variance with the proposed or the best. this is my fit; this is my base model. Now we can compare S_r and S_t. So, we can say a measure a decent measure a scaled measure of error is S_r by S_t.

But we are going to do one more thing, we will say it is,

$$r^2 = 1 - \frac{S_r}{S_t}$$

Because let me come to that quickly r^2 or let me use small r^2, because we used capital before. So, this r^2 is known as coefficient of determination let us look at its properties.

(Refer Slide Time: 13:08)



$r^2 = 1 - \frac{S_r}{S_t}$

↑

Coeff of determination

Best case scenario

→ Model fits really well

$\Rightarrow S_r = 0$

$\Rightarrow r^2 = 1 - 0 = 1$ Good ✓

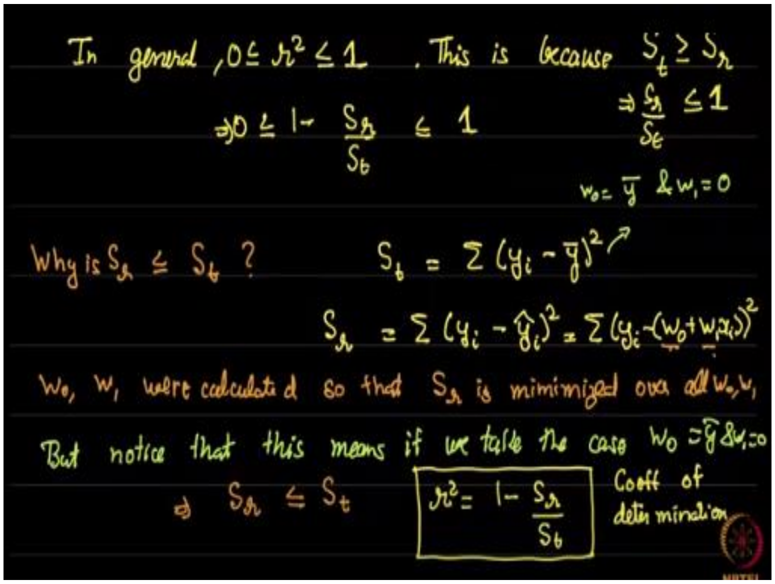
Bad Model ⇒ No better than a const model

$\Rightarrow S_r = S_t \Rightarrow r^2 = 1 - 1 = 0 \rightarrow$ Bad ✓

Now in the worst-case scenario or let us first take the best-case scenario. What is the best-case scenario? model fits really well, for example, in this case, let us say every single point on the best fit line in that case the error the error of the actual model S_r will be 0. So, this means r^2 will be 1 - 0 which is equal to 1. So, this is a good fit. So, this satisfies the condition that we had talked about, that you get a good fit in case your error is 0.

Now what happens if it is a bad model? no better than a constant model, what would this mean? This means $S_r = S_t$ so in this case $r^2 = 1 - 1$ which is 0, so this is a bad model. This condition is also satisfying.

(Refer Slide Time: 14:41)



In general, $0 \le r^2 \le 1$. This is because $S_t \ge S_r$

$\Rightarrow 0 \le 1 - \frac{S_r}{S_t} \le 1$ $\Rightarrow \frac{S_r}{S_t} \le 1$

$w_0 = \bar{y}$ & $w_1 = 0$

Why is $S_r \le S_t$? $S_t = \Sigma (y_i - \bar{y})^2$

$S_r = \Sigma (y_i - \hat{y}_i)^2 = \Sigma (y_i - (w_0 + w_1 x_i))^2$

w_0, w_1 were calculated so that S_r is mimimized over all w_0, w_1

But notice that this means if we take the case $w_0 = \bar{y}$ & $w_1 = 0$

$\Rightarrow S_r \le S_t$ $\boxed{r^2 = 1 - \frac{S_r}{S_t}}$ Coeff of determination

Now there is another certain point here. so let me add that in general r^2will always lie between 0 and 1. We have seen 2 extremes the reason is S_t is always greater than equal to S_r. Now for a moment assumes that this is true this would mean that,

$$0 \le 1 - \frac{S_r}{S_t} \le 1$$

Because this tells you $\frac{S_r}{S_t}$ is always going to be less than equal to 1. So, this number will always be positive and will always be less than equal to 1.

But why is this true? So, I will try to give you an explanation here. you will need to think a little bit more about what I am saying, but so why is $S_r \le S_t$? Now remember what these 2 quantities are,

$$S_t = \sum (y_i - \bar{y})^2$$

and

$$S_r = \sum (y_i - \hat{y}_i)^2$$

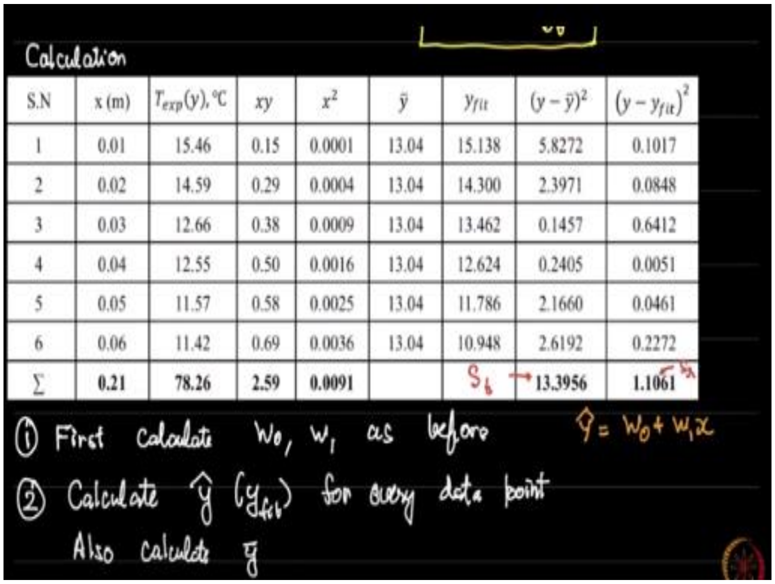
which is $(y_i - (w_0 + w_1 x_i))^2$. Now the coefficient is for the best fit, these coefficients w_0 and w_1 were calculated, so that S_r is minimized.

So, remember we were trying to get the best fit line, so what we were trying to do was to find out what w_0 and w_1 will minimize S_r over all w_0 and w_1. But notice, that this means, if we take the special case, where w_0 is $\bar{y}$ and w_1 is 0. So, if we take one specific line, for example coming here, when we try to find out the best fit line, what we are saying is I take this line or this line, or this line, or indeed even this line the error of this should be lower than every possible line, because that's how we optimize w_0 and w_1.

So, one such line over which this has to be better is the straight line which is horizontal. so same thing mathematically this has to be better S_r has to be better than this line which is $w_0 = \bar{y}$ and $w_1 = 0$, which is this case. This is the same as choosing $w_0 = \bar{y}$ and $w_1 = 0$. So, this would mean that S_r definitely has to be less than equal to S_t and these are all positive quantities anyway.

So, I hope this is clear, the formula for $r^2 = 1 - \frac{S_r}{S_t}$ and this is the coefficient of determination. this tells you how good or bad the fit is and we will be reusing this over the next few weeks.

(Refer Slide Time: 19:25)



Calculation

S.N	x (m)	$T_{exp}(y)$, °C	xy	x^2	$\bar{y}$	y_{fit}	$(y-\bar{y})^2$	$(y-y_{fit})^2$
1	0.01	15.46	0.15	0.0001	13.04	15.138	5.8272	0.1017
2	0.02	14.59	0.29	0.0004	13.04	14.300	2.3971	0.0848
3	0.03	12.66	0.38	0.0009	13.04	13.462	0.1457	0.6412
4	0.04	12.55	0.50	0.0016	13.04	12.624	0.2405	0.0051
5	0.05	11.57	0.58	0.0025	13.04	11.786	2.1660	0.0461
6	0.06	11.42	0.69	0.0036	13.04	10.948	2.6192	0.2272
Σ	**0.21**	**78.26**	**2.59**	**0.0091**		S_t →	**13.3956**	**1.1061** ← S_r

① First calculate W_0, W_1 as before $\hat{y} = W_0 + W_1 x$

② Calculate $\hat{y}$ (y_{fit}) for every data point

Also calculate $\bar{y}$

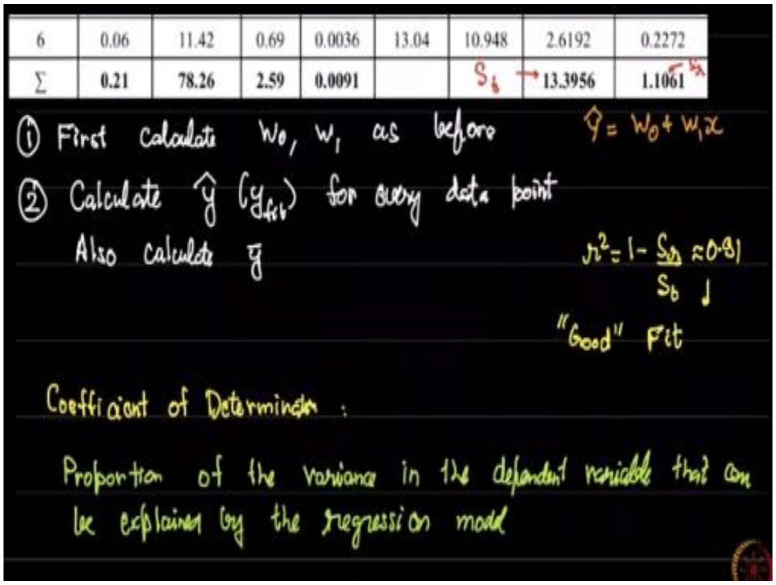
So here is this calculation for our example. so just a simple calculation. Again, I will leave the calculation to you. I will just slowly show you these are the results of some calculation and cutting and pasting it from a excel sheet. Very similar to what we did before, but the calculation works like this. you need to first calculate w_0 and w_1 as before. so if you recall from the previous video we first cataloged x, this is a summation of x then we catalog y, summation of y, we also needed x y and x square okay.

So, these quantities were needed so in order to calculate w_0 and w_1 and that is what we did right at the beginning. Now once you do that you have some expressions. So, you actually have numbers for w_0 and w_1, it is only after this. So, after this step you then calculate $\hat{y}$ which I have called y-fit here for every data point, also calculate $\bar{y}$. So once summation of y is here this divided by 6 which is 13.04 approximately is $\bar{y}$, we have calculated it.

Now we calculate y fit what is y-fit? y-fit basically is you go everywhere and calculate $\hat{y}$ which is y-fit as $w_0 + w_1x$. Since we also already know w_0 and w_1 you can now see. So, you have now recalculated $w_0 + w_1x$, so $w_0 + w_1x$ is y fit. So, you calculate this for these 6 quantities. Once you do that you square it $(y - \text{y-fit})^2$, this is the gap between your prediction and the actual data you get this.

And this is the gap between your data and the average data here so you get that too. This quantity now is what we called S_t and this quantity here is what we call S_r.

(Refer Slide Time: 22:23)



6	0.06	11.42	0.69	0.0036	13.04	10.948	2.6192	0.2272
Σ	**0.21**	**78.26**	**2.59**	**0.0091**		S_t →	**13.3956**	**1.1061** ← S_r

① First calculate w_0, w_1 as before

$\hat{y} = w_0 + w_1x$

② Calculate $\hat{y}$ (y_{fit}) for every data point

Also calculate $\bar{y}$

$r^2 = 1 - \frac{S_r}{S_t} \approx 0.91$ ↓

"Good" Fit

Coefficient of Determination :

Proportion of the variance in the dependent variable that can be explained by the regression model

Now once you calculate that r^2 is easy to calculate now, $r^2 = 1 - \frac{S_r}{S_t}$ in this case this comes to approximately 0.91 with the numbers here and you can see this is what we would call a good fit. Some people prefer above 0.95 but 0.9 and above generally is a decent fit for the data. So, we would have to rethink in case let us say it is 0.6, 0.5 or something of that, so we would have to think maybe I assumed it is a linear model and it is not really a linear model we will come to such discussions later on in the course.

So, what we have done so far is talk about the goodness of fit and how we can calculate it in this simple case. there is sort of a formal definition for the coefficient of determination, I will write that down. So, this is the proportion of the variance in the dependent variable that can be explained by the regression model. So, what I mean by that is, see this original figure there is some noise, there is some noise in the data or there are some variants in the data. This data also had some variance, when there was really nothing just a constant.

So, if we just put a constant fit here which was just the mean then there is some variance there. Now when you put a model, you should invariably see that some proportion of this variance is actually reduced. so that will call an explanation. So basically, we are able to explain a decent proportion. if all of it was explained 100% of explained by a model, that means we have got a really good model but at least we are able to explain 90% of what is happening in this data, which we could not have explained if we had just assumed the mean.

So that is what this means physically the proportional or the proportion of the variance in the data, that we can basically explain simply by the regression model. Now there are some subtleties here about coefficient of determination versus a correlation coefficient we will come to this, when we come to the probabilistic portions of this course. For now, I will stop the video here. in the next video we will look at quadratic models instead of just linear models. Thank you.