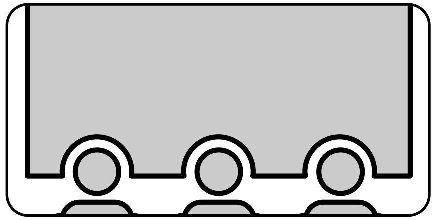
# Inference about the regression parameters

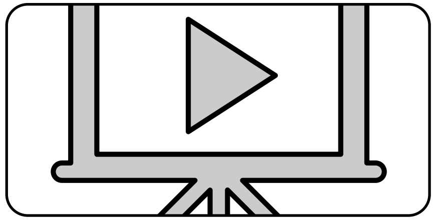
CHRIS SUTTON, FEBRUARY 2024

# Last week



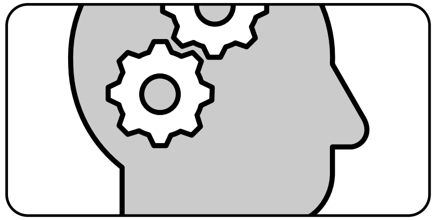
## Lectures on assessing the model

- Residuals
- ANOVA tables



## 6 more short video lectures to watch

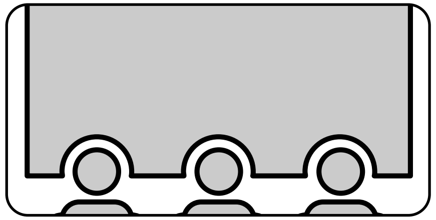
- 7 & 8 on properties of the parameters
- 9 – 12 recapping ANOVA, fitted values and residuals



## Your own data for modelling

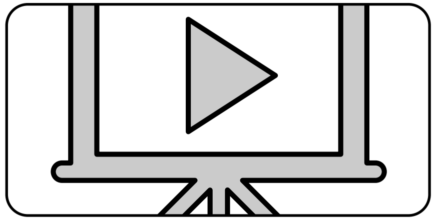
- Submitted to QM Plus with answers to the questionnaire
- You will need this for the assessed coursework coming next week

# This week



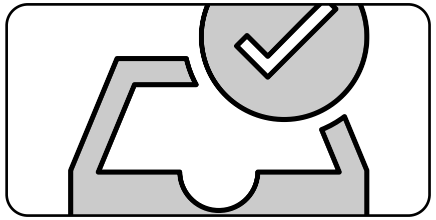
## Lectures on assessing the model

- Putting together all we have covered so far on modelling
- Confidence intervals and prediction intervals



## More short video lectures to watch

- Inference
- Using the models to make predictions



## IT Labs

- Opportunity to practice modelling in R
- Skills you will need for the two assessed courseworks

# Topics in this Statistical Modelling module

1. Principles of statistical modelling
2. The Simple Linear Regression Model
3. Least Squares estimation
4. Properties of estimators
5. Assessing the model
6. Inference about the model parameters
7. Matrix approaches to simple linear regression
8. Multiple Linear Regression Models

# Our Simple Linear Regression Model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

where the $\varepsilon_i$ are iid $\varepsilon_i \sim N(0, \sigma^2)$

# … with Least Squares estimators of the two model parameters

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\,\bar{x}$$

and

$$\widehat{\beta_1} = \frac{\sum_{i=0}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=0}^{n}(x_i - \bar{x})^2}$$

# Inference

Conclusions we would like to make:

- Confidence intervals
  - for parameters or the mean response
- Tests of significance
  - for parameters
- Prediction intervals
  - for a new observation

**inference**

**Noun: a conclusion reached on the basis of evidence and reasoning**

# Confidence intervals

For some parameter Ɵ

a 95% confidence interval for Ɵ means to find boundaries *a* and *b* such that $P(a < \theta < b) = 0.95$

More generally a $100(1 - \ \alpha)$% confidence interval for Ɵ is to find *a* and *b* such that $P(a < \theta < b) = 1 - \ \alpha$

# Confidence interval for $\beta_1$

The true value of $\beta_1$ is unknown

We have a point estimate via least squares, $\widehat{\beta_1}$

There are times when it would be more useful to have an interval within which we are confident $\beta_1$ lies

To do this we need to understand the distribution of $\widehat{\beta_1}$ and the effect of replacing $\sigma^2$ with its estimate $S^2$

# Sampling distribution for $\widehat{\beta_1}$

We showed last week that the sampling distribution is

$$\widehat{\beta_1} \sim N(\beta_1 , \frac{\sigma^2}{S_{xx}})$$

Note that even if the $y_i$ are not Normal, the $\widehat{\beta_1}$ still will be

We can standardise this

$$\frac{\widehat{\beta_1} - \beta_1}{\frac{\sigma}{\sqrt{S_{xx}}}} \sim N(0, 1)$$

# But the $\sigma^2$ here is a problem

However, the $\sigma^2$ is not known

The best we can do is replace it with our unbiased estimate from last week $S^2$

but when we do that the probability distribution changes from Normal to Students-t

From Probability & Statistics II

if $Z \sim N(0,1)$ and $U \sim \chi^2_v$ then $\frac{Z}{\sqrt{U/v}} \sim t_v$

# Student t distribution

The student t distribution applies here because we have

$$Z = \frac{\widehat{\beta_1} - \beta_1}{\frac{\sigma}{\sqrt{S_{xx}}}} \sim N(0,1) \text{ and } U = \frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}$$

[the second of these we will show formally later in the module]

therefore, $$T = \frac{\frac{\widehat{\beta_1} - \beta_1}{\frac{\sigma}{\sqrt{S_{xx}}}}}{\sqrt{\frac{(n-2)S^2}{\sigma^2(n-2)}}} = \frac{\widehat{\beta_1} - \beta_1}{\frac{S}{\sqrt{S_{xx}}}} \sim t_{n-2}$$

# Developing a confidence interval for $\beta_1$

If $\frac{\widehat{\beta_1} - \beta_1}{\frac{S}{\sqrt{S_{xx}}}} \sim t_{n-2}$ and we define $t_{\frac{\alpha}{2}}$ to be the quantity such that

$$P\left(|t_v| < t_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

then

$$P\left(\widehat{\beta_1} - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{S_{xx}}} < \beta_1 < \widehat{\beta_1} + t_{\frac{\alpha}{2}} \frac{S}{\sqrt{S_{xx}}}\right) = 1 - \alpha$$

# Comment

The confidence interval for $\beta_1$ based on $t_{\frac{\alpha}{2}}$ depends on:

- $\widehat{\beta_1}$ (which in general is a random variable) and
- $S^2$ (which depends on our observed data)

This means that it only makes sense to calculate the confidence interval given a particular set of observed data

# Confidence interval for $\beta_1$

For a particular data set

With $\widehat{\beta_1}$ and $S^2$ calculated for that data

$$[a, b] = \left[\widehat{\beta_1} - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{S_{xx}}} \, , \; \widehat{\beta_1} + t_{\frac{\alpha}{2}} \frac{S}{\sqrt{S_{xx}}}\right]$$

# Testing the significance of $\beta_1$

Last week we used the ANOVA table and F statistic to test the null hypothesis $H_0: \beta_1 = 0$

Now that we have a confidence interval for $\beta_1$ there is another way to test this same hypothesis

We have already seen $T = \frac{\widehat{\beta_1} - \beta_1}{\frac{s}{\sqrt{s_{xx}}}} \sim t_{n-2}$

# Developing the test statistic

Now under $H_0$: $\beta_1 = 0$ this test statistic becomes $T = \frac{\widehat{\beta_1}}{\frac{s}{\sqrt{S_{xx}}}} \sim t_{n-2}$

Which we can calculate for any particular data set

We then reject $H_0$ if

$$|T| > t_{n-2,\frac{\alpha}{2}}$$

This is mathematically equivalent to the F statistic test

# Estimated Standard Error of $\widehat{\beta_1}$

The estimate of the standard error is the square root of the estimated variance

$$\widehat{se(\widehat{\beta_1})} = \sqrt{\frac{S^2}{S_{xx}}}$$

We can then re-frame the confidence interval and the test statistic for $\beta_1$ in terms of this estimated standard error

$$[a, b] = \left[\widehat{\beta_1} - t_{\frac{\alpha}{2}} \widehat{se(\widehat{\beta_1})}, \ \widehat{\beta_1} + t_{\frac{\alpha}{2}} \widehat{se(\widehat{\beta_1})}\right] \text{ and } T = \frac{\widehat{\beta_1}}{\widehat{se(\widehat{\beta_1})}} \sim t_{n-2}$$

# Confidence interval for the mean response $\mu_i$

We can also develop confidence intervals and test hypotheses for the mean response, that is for $E[Y_i|X_i = x_i]$ which is often written $\mu_i$

Under the simple linear regression model,

$$\mu_i = E[Y_i|X_i = x_i] = \beta_0 + \beta_1 x_i$$

And $\mu_i$ is estimated by least squares at a particular value of $x_i$ as

$$\widehat{\mu}_i = \widehat{\beta_0} + \widehat{\beta_1} x_i$$

# Sampling distribution for $\mu_i$

Under the simple linear regression model, the sampling distribution of $\mu_i$ is also normal

$$\widehat{\mu_i} \sim N(\mu_i,\, \sigma^2 (\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}))$$

Which leads to a 100(1 – α)% confidence interval for $\widehat{\mu_i}$ of

$$[a, b] = \left[\widehat{\mu_i} - t_{\frac{\alpha}{2}}\, \widehat{se(\widehat{\mu_i})}\,,\ \widehat{\mu_i} + t_{\frac{\alpha}{2}}\, \widehat{se(\widehat{\mu_i})}\right]$$

# Test statistic for the mean response $\mu_i$

where, $\widehat{se(\hat{\mu}_i)} = \sqrt{S^2(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}})}$

we can test the null hypothesis, $H_0\text{: } \mu_i = M$ for some value M (which is not necessarily zero), with the test statistic

$$T = \frac{\widehat{\mu_i} - M}{\widehat{se(\widehat{\mu_i})}} \sim t_{n-2}$$

# A note of caution

For the estimation of the mean response to be valid,

The value of $x_i$ used should be within the range of observed values for $X$

The model has said nothing about the applicability of linear regression outside of this range for $x_i$

We should not use inference about $\mu_i$ as a method of extrapolation

However we can now turn to using the model to predict the response value for some new value of $x_i$ for which $y_i$ has not yet been observed

# Prediction Interval for a new observation

we can use a linear regression model to predict the response value for some new value of $x_i$ for which $y_i$ has not yet been observed

This is called a ***Prediction Interval*** sometimes just *PI* for a new observation

Let us say that we have a new value for $x_i$ which we will label $x_0$

We have yet to observe $y_0$ so we attempt to predict it

- we make this prediction as an interval rather than a single value because of the stochastic nature of the model

# Prediction interval (continued)

We seek $y_0 = \mu_0 + \varepsilon_0$

The "point prediction" would be $\widehat{y_0} = \widehat{\mu_0} = \widehat{\beta_0} + \widehat{\beta_1} x_0$

We know that $\widehat{\mu_0} \sim N(\mu_0, \sigma^2 (\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}))$

Therefore the distribution of $\widehat{\mu_0} - \mu_0$ is

$$\widehat{\mu_0} - \mu_0 \sim N(0, \sigma^2 (\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}))$$

# From $\mu_0$ to $y_0$

But rather than $\widehat{\mu_0} - \mu_0$ we would prefer the distribution of $\widehat{y_0} - y_0$

If we add and subtract $\varepsilon_0$ to the distribution equation for $\widehat{\mu_0} - \mu_0$ we have

$$\widehat{\mu_0} - \mu_0 = \widehat{\mu_0} - (\mu_0 + \varepsilon_0) + \varepsilon_0$$

$$= \widehat{y_0} - y_0 + \varepsilon_0 \sim N(0, \sigma^2(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}))$$

But we know that $\varepsilon_0 \sim N(0, \sigma^2)$ from the original model definition, so

$$\widehat{y_0} - y_0 \sim N(0, \sigma^2\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right))$$

# From distribution to PI

To get to the prediction interval we need to:

1. standardise the normal distribution
2. replace the unknown variance $\sigma^2$ with its estimator $S^2$

1. leads to $$\frac{\widehat{y_0} - y_0}{\sqrt{\sigma^2\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right)}} \sim N(0,1)$$

2. gives us $$\frac{\widehat{y_0} - y_0}{\sqrt{S^2\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right)}} \sim t_{n-2}$$

# Prediction interval for $y_0$

The $100(1-\alpha)\%$ prediction interval for $y_0$ is then

$$\widehat{y_0} \pm t_{\frac{\alpha}{2}} \sqrt{S^2\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$$

Note the prediction interval for $y_0$ is usually much wider than the confidence interval for $\mu_0$ because the random variability term $\varepsilon_0$ impacts the PI.