Regression
(Module 5)

Statistics (MAST20005) &
Elements of Statistics
(MAST90058)

School of Mathematics and Statistics
University of Melbourne

Semester 2, 2020

Aims of this module

- Introduce the concept of **regression**
- Show a simple model for studying the relationship between two variables
- Discuss correlation and how it relates to regression

Outline

Relationships between two variables

We have studied how to do estimation for some simple scenarios:

- iid samples from a single distribution (X_i) μ, σ^2
- comparing iid samples from two different distributions $(X_i \;\&\; Y_j)$ $\mu_X - \mu_Y$, σ_X^2 / σ_Y^2
- differences between paired measurements $(X_i - Y_i)$ $\mu_X - \mu_Y$

We now consider how to analyse bivariate data more generally, i.e. two variables, X and Y, measured at the same time, i.e. as a pair.

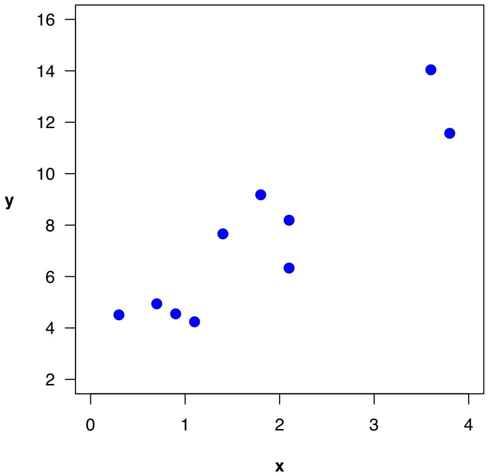
The data consist of pairs of data points, (x_i, y_i).

These can be visualised using a scatter plot.

Example data

x_i	y_i
1.80	9.18
1.40	7.66
2.10	6.33
0.30	4.51
3.60	14.04
0.70	4.94
1.10	4.24
2.10	8.19
0.90	4.55
3.80	11.57

$n = 10$



Outline

Regression

$y = f(x) + \epsilon.$

$E(y|x) = E[f(x)|x] + E(\epsilon|x)$ $\uparrow 0$

$= f(x). = \alpha + \beta x$

$\mathrm{var}(y|x) = \mathrm{var}(f(x)|x) + \mathrm{var}(\epsilon|x)$

$= 0 + \mathrm{var}(\epsilon|x)$

$= \sigma^2$

Often interested in how Y depends on X. For example, we might want to use X to predict Y.

In such a setting, we will assume that the X values are known and fixed (henceforth, x instead of X), and look at how Y varies given x.

Example: Y is a student's final mark for Statistics, and x is their mark for the prerequisite subject Probability. Does x help to predict Y?

The regression of Y on x is the conditional mean, $\mathbb{E}(Y \mid x) = \mu(x)$.

The regression can take any form. We consider simple linear regression, which has the form of a **straight line**:

$$\mathbb{E}(Y \mid x) = \alpha + \beta x \quad \text{and} \quad \boxed{\mathrm{var}(Y \mid x) = \sigma^2.}$$

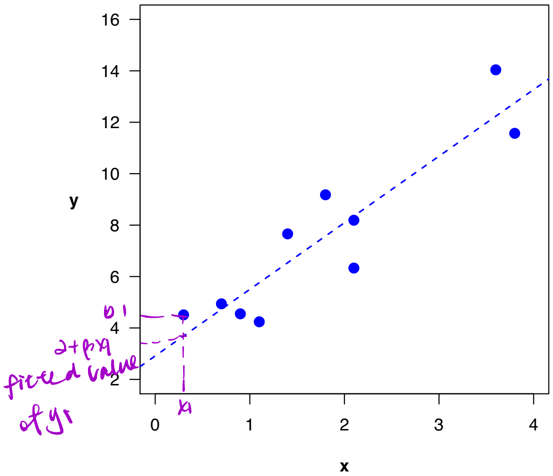
?

does not depend on x

Example: simple linear regression model

$$\mathbb{E}(Y \mid x) = \alpha + \beta x$$
$$\text{var}(Y \mid x) = \sigma^2$$

$(y_1 - (\alpha + \beta x_1))^2$



Terminology (X, Y)

- Y is called a $\boxed{\text{response}}$ variable. Can also be called an $\boxed{\text{outcome}}$ or $\boxed{\text{target variable}}$. Please do **not** call it the 'dependent' variable.
- x is called a $\boxed{\text{predictor}}$ variable. Can also be called an explanatory variable. Please do **not** call it an 'independent' variable.
- $E(Y|X)$ $\mu(x)$ is called the (linear) predictor function or sometimes the regression curve or the model equation.
- The parameters in the predictor function are called regression coefficients.

$var(Y|X) = \sigma^2$

↑

not a coefficient

since not inside the predictor function

Why ‘regression’?

It is strange terminology, but it has stuck.

Refers to the idea of ‘regression to the mean’:
if a variable is extreme on its first measurement, it will tend to be closer to the average on its second measurement, and vice versa.

First described by Sir Francis Galton when studying the inheritance of height between fathers and sons. In doing so, he invented the technique of simple linear regression.

Linearity

not necessary in x

A regression model is called linear $\boxed{\text{if it is linear in the coefficients.}}$

It doesn't have to define a straight line! of x

Complex and non-linear functions of x are allowed, as long as the resulting predictor function is a linear combination (i.e. an additive function) of them, with the coefficients 'out the front'.

coefficient and function of x

For example, the following are linear models:

$$\mu(x) = \alpha + \beta x + \gamma x^2$$

$$\mu(x) = \frac{\alpha}{x} + \frac{\beta}{x^2}$$

$$\mu(x) = \alpha \sin x + \beta \log x$$

$x_1 = \phi_1(x)$
$x_2 = \phi_2(x)$
$\vdots$
$x_p = \phi_p(x)$

$\mu(x) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$

linear combination of α & β's

The following are NOT linear models:

$$\mu(x) = \alpha \sin(\beta x)$$
$$\mu(x) = \frac{\alpha}{1 + \beta x}$$
$$\mu(x) = \alpha x^{\beta}$$

$\log(\mu(x)) = \log \alpha + \beta \log x$

$\mu^*(x) = \log(\mu(x))$
$\alpha^* = \log \alpha$

...but the last one can be re-expressed as a linear model on a log scale (by taking logs of both sides),

$$\mu^*(x) = \alpha^* + \beta \log x \qquad (\alpha^*, \beta)$$

Outline

Estimation goals

Back to our simple linear regression model:

$$\mathbb{E}(Y \mid x) = \alpha + \beta x \quad \text{and} \quad \mathrm{var}(Y \mid x) = \sigma^2.$$

- We wish to estimate the slope (β), the intercept (α), the variance of the errors (σ^2), their standard errors and construct confidence intervals for these quantities.
- Often want to use the fitted model to make predictions about future observations (i.e. predict Y for a new x).
- Note: the Y_i are not iid. They are independent but have different means, since they depend on x_i. $E(Y_i \mid x_i) = \alpha + \beta x_i$
- We have not (yet) assumed any specific distribution for Y, only a conditional mean and variance.

Reparameterisation

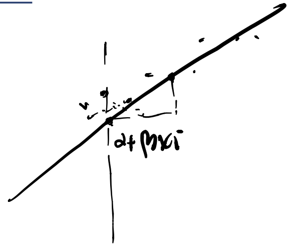
Changing our model slightly...

Let $\alpha_0 = \alpha + \beta\bar{x}$, which gives:

$$\begin{aligned}\mathbb{E}(Y \mid x) &= \alpha + \beta x \\ &= \alpha_0 + \beta(x - \bar{x})\end{aligned}$$

variable $x - \bar{x}$

(α_0, β)



Now our model is in terms of α_0 and β.

This will make calculations and proofs simpler.

Least squares estimation

Choose α_0 and β to minimize the sum of squared deviations:

$$H(\alpha_0, \beta) = \sum_{i=1}^{n} (\overset{\text{true}}{y_i} - \underbrace{\alpha_0 - \beta\,(x_i - \bar{x})}_{\text{fitted value of } y_i})^2$$

$H(\alpha_0, \beta)$ is continuous in α_0 and β

Solve this by finding the partial derivatives and setting to zero:

$$0 = \frac{\partial H(\alpha_0, \beta)}{\partial \alpha_0} = 2 \sum_{i=1}^{n} [y_i - \alpha_0 - \beta(x_i - \bar{x})](-1)$$

$$0 = \frac{\partial H(\alpha_0, \beta)}{\partial \beta} = 2 \sum_{i=1}^{n} [y_i - \alpha_0 - \beta(x_i - \bar{x})](-(x_i - \bar{x}))$$

$\rightarrow \sum_{i=1}^{n} [y_i - \alpha_0 - \beta(x_i - \bar{x})] = 0$

$\Rightarrow \sum_{i=1}^{n} y_i - n\alpha_0 - \beta \sum_{i=1}^{n} (x_i - \bar{x}) = 0$

$= \beta\left(\sum_{i=1}^{n} x_i - n\bar{x}\right) = 0$

$= \sum_{i=1}^{n} y_i - n\alpha_0 = 0$

$\hat{\alpha}_0 = \bar{y}$

These are called the normal equations.

Least squares estimators

$\sum_{i=1}^{n}[y_i - \hat{\alpha}_0 - \hat{\beta}(x_i-\bar{x})][x_i-\bar{x}] = 0$

$\Rightarrow \sum_{i=1}^{n} y_i(x_i-\bar{x}) - \hat{\alpha}_0 \underbrace{\sum_{i=1}^{n}(x_i-\bar{x})}_{=0} - \hat{\beta}\sum_{i=1}^{n}(x_i-\bar{x})^2 = 0.$

Some algebra yields the least square estimators,

Y_i for r.v.

y_i for a particular value.

$$\hat{\alpha}_0 = \bar{Y}, \quad \hat{\beta} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})Y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

$\Rightarrow. \sum_{i=1}^{n} y_i(x_i-\bar{x}) = \hat{\beta}\sum_{i=1}^{n}(x_i-\bar{x})^2$

$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$

Another expression for $\hat{\beta}$ is:

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

sample covariance between X and Y

$cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(Y_i-\bar{Y})$

$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

These are equivalent, due to the following result:

$$\sum(x_i-\bar{x})(Y_i-\bar{Y}) = \boxed{\sum(x_i-\bar{x})Y_i.}$$

$= \sum(x_i-\bar{x})Y_i - \sum(x_i-\bar{x})\bar{Y}$

$= \bar{Y}\sum(x_i-\bar{x}) = 0.$

Can also then get an estimator for α:

$$\hat{\alpha} = \hat{\alpha}_0 - \hat{\beta}\bar{x}$$
$$= \bar{Y} - \hat{\beta}\bar{x}.$$

$\alpha_0 = \alpha + \beta\bar{x}$

$\Leftrightarrow \alpha = \alpha_0 - \beta\bar{x}$

plug in method

And also an estimator for the predictor function,

$$\begin{aligned}\hat{\mu}(x) &= \hat{\alpha} + \hat{\beta}x \\ &= \hat{\alpha}_0 + \hat{\beta}(x - \bar{x}) \\ &= \bar{Y} + \hat{\beta}(x - \bar{x}).\end{aligned}$$

Ordinary least squares

This method is sometimes called ordinary least squares or OLS.

Other variants of least squares estimation exist, with different names. For example, 'weighted least squares'.

$$\text{OLS}: \sum_{i=1}^{n} (y_i - \alpha_0 - \beta(x_i - \bar{x}))^2$$

$$\text{WLS}: \sum_{i=1}^{n} \underset{\substack{\uparrow \\ \text{weight}}}{w_i} (y_i - \alpha_0 - \beta(x_i - \bar{x}))^2$$

Example: least squares estimates

For our data:

$$\bar{x} = 1.78$$
$$\bar{y} = 7.52 = \hat{\alpha}_0$$
$$\hat{\alpha} = 2.91$$
$$\hat{\beta} = 2.59$$

→ easier to calculate

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \rightarrow \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

$$\downarrow$$

$$(n-1)s_x^2$$

The fitted model equation is then:

$$\hat{\mu}(x) = 2.91 + 2.59x$$

```
> rbind(y, x)
  [,1] [,2] [,3] [,4]  [,5] [,6] [,7] [,8] [,9] [,10]
y 9.18 7.66 6.33 4.51 14.04 4.94 4.24 8.19 4.55 11.57
x 1.80 1.40 2.10 0.30  3.60 0.70 1.10 2.10 0.90  3.80

> model1 <- lm(y ~ x)
> model1
```

linear model

```
Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
      2.911        2.590
```

$\hat{\alpha}$ $\hat{\beta}$

Properties of these estimators

What do we know about these estimators?

x_i fitted
Y_i random

They are all linear combinations of the Y_i,

$$\hat{\alpha}_0 = \sum_{i=1}^{n} \left(\frac{1}{n} \right) Y_i$$

$\bar{Y}$

$$\hat{\beta} = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{K} \right) Y_i$$

where $K = \sum_{i=1}^{n} (x_i - \bar{x})^2$.

This allows us to easily calculate means and variances.

Means?

$E(Y_i) = E(E(Y_i \mid x_i)) = E\{\alpha_0 + \beta(x_i - \bar{x})\} = \alpha_0 + \beta(x_i - \bar{x})$

① x_i known parameter

$$\mathbb{E}(\hat{\alpha}_0) = \mathbb{E}(\bar{Y}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i) = \frac{1}{n}\sum_{i=1}^{n}[\alpha_0 + \beta(x_i - \bar{x})] = \alpha_0$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\alpha_0}_{\alpha_0} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\beta(x_i - \bar{x})}_{=0} = \alpha_0.$$

$$\mathbb{E}(\hat{\beta}) = \sum_{i=1}^{n}\frac{(x_i - \bar{x})}{K}\mathbb{E}(Y_i) = \frac{1}{K}\sum_{i=1}^{n}(x_i - \bar{x})(\alpha_0 + (x_i - \bar{x})\beta)$$

$$= \frac{1}{K}\underbrace{\sum_{i=1}^{n}(x_i - \bar{x})}_{=0.}\alpha_0 + \frac{K}{K}\beta = \beta$$

$K = \sum_{i=1}^{n}(x_i - \bar{x})^2$

This also implies, $\mathbb{E}(\hat{\alpha}_0) = \alpha_0$ and $\mathbb{E}(\hat{\mu}(x)) = \mu(x)$, and so we have that all of the estimators are unbiased.

$E(\hat{\alpha}) = E(\hat{\alpha}_0 - \hat{\beta}\bar{x}) = E(\hat{\alpha}_0) - E(\hat{\beta})\bar{x} = \alpha_0 - \beta\bar{x} = \alpha.$

Variances?

$z = \sum_{i=1}^{n} w_i Y_i \Rightarrow Var(z) = \sum_{i=1}^{n} w_i^2 Y_i$

$$\operatorname{var}(\hat{\alpha}_0) = \operatorname{var}(\bar{Y}) = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{var}(Y_i) = \frac{\sigma^2}{n}$$

$\hat{\alpha}_0 = \sum_{i=1}^{n} \left(\frac{1}{n}\right) Y_i$ // $\quad \frac{n\sigma^2}{n^2}$ //

$$\operatorname{var}(\hat{\beta}) = \operatorname{var}\left(\sum_{i=1}^{n}\frac{(x_i-\bar{x})}{K}Y_i\right) = \sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{K}\right)^2\operatorname{var}(Y_i)$$

$$= \frac{1}{K^2}\sum_{i=1}^{n}(x_i-\bar{x})^2\operatorname{var}(Y_i) = \frac{1}{K^2}\underbrace{\boxed{\sum_{i=1}^{n}(x_i-\bar{x})^2}}_{K}\sigma^2$$

$$= \frac{1}{K^2}\sigma^2\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{K^2}\sigma^2 K$$

$$= \frac{\sigma^2}{K}$$

$= cov(X+Y, X) + cov(X+Y, Y)$.

$\mathrm{cov}(X+Y, X+Y)$

$= \mathrm{cov}(X,X) + \mathrm{cov}(X,Y)$

$= \mathrm{cov}(X,X) + \mathrm{cov}(Y,Y)$

$+ \mathrm{cov}(Y,X) + \mathrm{cov}(Y,Y)$

Similarly,

$\hat{\alpha} = \hat{\alpha}_0 - \hat{\beta}\bar{x}.$

$$\mathrm{var}(\hat{\alpha}) = \left(\frac{1}{n} + \frac{\bar{x}^2}{K}\right)\sigma^2$$

$$\mathrm{cov}(\hat{\alpha}_0, \hat{\beta}) = 0$$

$$\mathrm{var}(\hat{\mu}(x)) = \left(\frac{1}{n} + \frac{(x-\bar{x})^2}{K}\right)\sigma^2$$

$\mathrm{var}(\hat{\mu}(x)) = \mathrm{var}(\hat{\alpha}_0 + \hat{\beta}(x-\bar{x}))$ since $\mathrm{cov}(\hat{\alpha}_0, \hat{\beta}) = 0$

$= \mathrm{var}(\hat{\alpha}_0) + \mathrm{var}(\hat{\beta}(\underbrace{x-\bar{x}}_{\text{known}}))$

$\frac{\sigma^2}{n}$

$(x-\bar{x})^2 \cdot \mathrm{var}(\hat{\beta})$

$= (x-\bar{x})^2 \cdot \frac{\sigma^2}{K}.$

$\Rightarrow \sigma^2\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{K}\right)$

$\mathrm{cov}(\hat{\alpha}_0, \hat{\beta})$

$= \mathrm{cov}\left(\sum_{i=1}^{n} \frac{1}{n} Y_i, \sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{K}\right) Y_i\right)$

$= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{cov}\left(\frac{1}{n} Y_i, \frac{x_j-\bar{x}}{K} Y_j\right)$ the additive law.

why?

if $i \neq j$, since Y_i's are indep

$\mathrm{cov}(Y_i, Y_j) = 0$

$\frac{1}{n^2}\sum_{i=1}^{n} \frac{(x_i-\bar{x})}{K} \mathrm{var}(Y_i)$

$i = j$

$\sum_{i=1}^{n} \mathrm{cov}\left(\frac{1}{n} Y_i, \frac{x_i-\bar{x}}{K} Y_i\right) = \sum_{i=1}^{n} \left(\frac{1}{n}\right)\left(\frac{x_i-\bar{x}}{K}\right) \mathrm{var}(Y_i)$

$= \frac{1}{n}\sigma^2 \underbrace{\sum_{i=1}^{n}(x_i-\bar{x})}_{=0} \cdot \frac{1}{K} = 0$

$\mathrm{var}(\hat{\alpha}) = \mathrm{var}(\hat{\alpha}_0 - \hat{\beta}\bar{x})$

$\bar{x}^2 \mathrm{var}(\hat{\beta})$

$= \bar{x}^2 \cdot \frac{\sigma^2}{K}.$

$= \mathrm{var}(\hat{\alpha}_0) + \mathrm{var}(\hat{\beta}\bar{x})$

$\frac{1}{n}\sigma^2$

$\underbrace{+ 2\mathrm{cov}(\hat{\alpha}_0, \hat{\beta}\bar{x})}_{=0}$. since $\mathrm{cov}(\hat{\alpha}_0, \hat{\beta}) = 0$

Can we get their standard errors?

We need an estimate of σ^2.

$\text{var}(Y \mid x_i)$

ANOVA

Analysis of variance: iid model

For $X_i \sim \mathrm{N}(\mu, \sigma^2)$ iid,

$$\underbrace{\sum_{i=1}^{n}(X_i-\mu)^2}_{\text{Total sum of squares}} = \underbrace{\sum_{i=1}^{n}(X_i-\bar{X})^2}_{\text{sample sum of square}} + n(\bar{X}-\mu)^2$$

variation of the estimator $\bar{X}$

$$\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu)^2$$
$$=\sum_{i=1}^{n}\left[(x_i-\bar{x})^2+(\bar{x}-\mu)^2+2(x_i-\bar{x})(\bar{x}-\mu)\right]$$
$$=\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{n}(\bar{x}-\mu)^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2$$
$$+2\underbrace{\sum_{i=1}^{n}x_i\bar{x}-x_i\mu-\bar{x}^2+\mu\bar{x}}$$
$$=\cancel{n\bar{x}^2}-\cancel{\mu n\bar{x}}-\cancel{n\bar{x}^2}+\cancel{\mu n\bar{x}} = 0$$

since $S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

take expectation $E\left(\sum_{i=1}^{n}(x_i-\mu)^2\right) = nE\left[(x_i-\mu)^2\right] = n\,var(x_i)^2 = n\sigma^2$

LHS $n\sigma^2$ → RHS $(n-1)E(S_x^2) + E(n(\bar{x}-\mu)^2)$

$= nE(\bar{x}-\mu)^2$

$= \sigma^2$

since $\bar{x}$ is unbiased for μ, then $E(\bar{x}-\mu)^2 = var(\bar{x})$

$\Rightarrow \cancel{(n-1)}E(S_x^2) = \cancel{(n-1)}\sigma^2$

$E(S_x^2)=\sigma^2$ unbiased

$= \frac{\sigma^2}{n}$

Analysis of variance: regression model

$E(Y_i | x_i) = \alpha + \beta x_i = \alpha_0 + \beta(x_i - \bar{x})$

$var(Y_i | x_i) = \sigma^2$

$$\sum_{i=1}^{n} (A_i + B_i + C_i)^2 = \sum_{i=1}^{n} A_i^2 + \sum_{i=1}^{n} B_i^2 + \sum_{i=1}^{n} C_i^2 + 2\sum_{i=1}^{n} A_i B_i + 2\sum_{i=1}^{n} A_i C_i + 2\sum_{i=1}^{n} B_i C_i$$

$$\begin{aligned}
&\sum_{i=1}^{n}(Y_i - \alpha_0 - \beta(x_i - \bar{x}))^2 \\
&= \sum_{i=1}^{n}(Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x}) + \underbrace{\hat{\alpha}_0 + \hat{\beta}(x_i - \bar{x})}_{\text{fitted value of } Y_i} - \alpha_0 - \beta(x_i - \bar{x}))^2 \\
&= \sum_{i=1}^{n}(Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x}) + (\hat{\alpha}_0 - \alpha_0) + (\hat{\beta} - \beta)(x_i - \bar{x}))^2 \\
&= \sum_{i=1}^{n}(Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x}))^2 + n(\hat{\alpha}_0 - \alpha_0)^2 + K(\hat{\beta} - \beta)^2
\end{aligned}$$

Note that the cross-terms disappear. Let's see...

The cross-terms...

$\hat{\alpha}_0 = \frac{1}{n}\sum Y_i$

$$t_1 = 2\sum_{i=1}^{n} Y_i(\hat{\alpha}_0 - \alpha_0) - 2\sum_{i=1}^{n} \hat{\alpha}_0(\hat{\alpha}_0 - \alpha_0) - 2\sum_{i=1}^{n} \hat{\beta}(x_i - \bar{x})(\hat{\alpha}_0 - \alpha_0)$$

$$= \cancel{2(\hat{\alpha}_0 - \alpha_0)\sum_{i=1}^{n} Y_i - 2n\hat{\alpha}_0(\hat{\alpha}_0 - \alpha_0)} - 2(\hat{\alpha}_0 - \alpha_0)\hat{\beta}\underbrace{\sum_{i=1}^{n}(x_i - \bar{x})}_{0} = 0$$

$$t_1 = 2\sum_{i=1}^{n} \overset{A_i}{\underline{(Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x}))}}\overset{B_i}{\underline{(\hat{\alpha}_0 - \alpha_0)}}$$

$$t_2 = 2\sum_{i=1}^{n} \overset{A_i}{\underline{(Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x}))}}\overset{C_i}{\underline{(\hat{\beta} - \beta)(x_i - \bar{x})}}$$

$$t_3 = 2\sum_{i=1}^{n} (x_i - \bar{x})(\hat{\beta} - \beta)(\hat{\alpha}_0 - \alpha_0) = 2(\hat{\beta} - \beta)(\hat{\alpha}_0 - \alpha_0)\underbrace{\sum_{i=1}^{n}(x_i - \bar{x})}_{0} = 0.$$

Since $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$ and $\sum_{i=1}^{n}(Y_i - \hat{\alpha}_0) = \sum_{i=1}^{n}(Y_i - \bar{Y}) = 0$, the first and third cross-terms are easily shown to be zero.

For the second term,

$$\begin{aligned}
\frac{t_2}{2(\hat\beta-\beta)} &= \sum_{i=1}^{n}(Y_i-\bar Y)(x_i-\bar x) - \hat\beta\sum_{i=1}^{n}(x_i-\bar x)^2 \\
&= \sum_{i=1}^{n}(Y_i-\bar Y)(x_i-\bar x) - \hat\beta K \\
&= \sum_{i=1}^{n}Y_i(x_i-\bar x) - \sum_{i=1}^{n}Y_i(x_i-\bar x) \\
&= 0
\end{aligned}$$

Therefore, $\boxed{\text{all the cross-terms are zero.}}$

$$\hat\beta = \frac{\sum_{i=1}^{n}(x_i-\bar x)Y_i}{\underbrace{\sum_{i=1}^{n}(x_i-\bar x)^2}_{K}}$$

$$t_2 = 2\sum_{i=1}^{n}(Y_i-\hat\alpha_0-\hat\beta(x_i-\bar x))(\hat\beta-\beta)(x_i-\bar x)$$

$$= 2(\hat\beta-\beta)\sum_{i=1}^{n}(Y_i-\hat\alpha_0-\hat\beta(x_i-\bar x))(x_i-\bar x).$$

$$= \sum_{i=1}^{n}(Y_i-\underbrace{\hat\alpha_0}_{\bar Y})(x_i-\bar x) - \hat\beta\sum_{i=1}^{n}(x_i-\bar x)^2$$

$$= \sum_{i=1}^{n}(Y_i-\bar Y)(x_i-\bar x) - \hat\beta K$$

$$= \sum_{i=1}^{n}Y_i(x_i-\bar x) - \hat\beta K.$$

since $\sum_{i=1}^{n}\bar Y(x_i-\bar x)=0$,

$$= \sum_{i=1}^{n}Y_i(x_i-\bar x) - \sum_{i=1}^{n}Y_i(x_i-\bar x)$$

$$= 0.$$

Back to the analysis of variance formula...

$E(Y_i) = \alpha_0 + \beta(x_i - \bar{x})$.

so $n\,\mathrm{var}(Y_i) = n\sigma^2$.

$$\sum_{i=1}^{n}(Y_i - \alpha_0 - \beta(x_i - \bar{x}))^2$$

known. a statistic, doesn't depend on α, β, σ^2

$$= \sum_{i=1}^{n}(Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x}))^2 + n(\hat{\alpha}_0 - \alpha_0)^2 + K(\hat{\beta} - \beta)^2$$

$\hat{\beta}$ is unbiased for β

$E(K(\hat{\beta}-\beta)^2) = KE((\hat{\beta}-\beta)^2) = K \cdot \mathrm{var}(\hat{\beta}) = K \cdot \frac{\sigma^2}{K} = \sigma^2$.

$E(n(\hat{\alpha}_0-\alpha_0)^2) = nE(\hat{\alpha}_0-\alpha_0)^2 = n\,\mathrm{var}(\hat{\alpha}_0) = \frac{n\sigma^2}{n} = \sigma^2$.

$\hat{\alpha}_0$ is unbiased for α_0

Taking expectations gives,

$$n\sigma^2 = \mathbb{E}(D^2) + \sigma^2 + \sigma^2$$

$$\Rightarrow \mathbb{E}(D^2) = (n-2)\sigma^2$$

where

$$D^2 = \sum_{i=1}^{n}(Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x}))^2.$$

Variance estimator

if let $\hat{\sigma}^2 = \sum (Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x}))^2 = D^2$.
then $E(\hat{\sigma}^2) = (n-2)\sigma^2$.

so unbiased estimator $D = \frac{1}{n-2} D^2$

Based on these results, we have $\boxed{\text{an unbiased estimator of the variance,}}$

$$\hat{\sigma}^2 = \frac{1}{n-2} D^2.$$

$E(\hat{\sigma}^2) = E(\frac{1}{n-2} D^2)$
$= E(D^2) \cdot \frac{1}{n-2} = \sigma^2$

The inferred mean for each observation is called its fitted value,
$\hat{Y}_i = \hat{\alpha}_0 + \hat{\beta}(x_i - \bar{x})$.

$\hat{\mu}(x_i)$
estimated value when $x = x_i$

The deviation from each fitted value is called a residual, $R_i = Y_i - \hat{Y}_i$.

The variance estimator is based on the sum of squared residuals,
$\boxed{D^2 = \sum_{i=1}^{n} R_i^2.}$

Example: variance estimate

For our data:

$$d^2 = 16.12$$
$$\hat{\sigma}^2 = 2.015$$
$$\hat{\sigma} = 1.42$$

$d^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

since $\hat{\sigma}^2 = \frac{1}{n-2} D^2$

then $\hat{\sigma}^2 = \frac{1}{n-2} d^2$

Standard errors

We can subsitute $\hat{\sigma}^2$ into the formulae for the standard deviation of the estimators in order to calculate standard errors.

For example,

$$\mathrm{var}(\hat{\beta}) = \frac{\sigma^2}{K}$$

$$\Rightarrow \quad \mathrm{se}(\hat{\beta}) = \frac{\hat{\sigma}}{\sqrt{K}}$$

$\mathrm{var}(\hat{\alpha}) = (\frac{1}{n} + \frac{\bar{x}^2}{K})\sigma^2$

$\mathrm{se}(\hat{\alpha}) = \hat{\sigma} \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{K}}$

Example: standard errors

For our data:

$$\mathrm{se}(\hat{\alpha}_0) = \frac{\hat{\sigma}}{\sqrt{n}} = 0.449$$

$$\mathrm{se}(\hat{\beta}) = \frac{\hat{\sigma}}{\sqrt{K}} = 0.404$$

$$\mathrm{se}(\hat{\mu}(x)) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{K}} = 1.42 \times \sqrt{\frac{1}{10} + \frac{(x-1.78)^2}{12.34}}$$

Maximum likelihood estimation

for α, β, $\mu(x)$ find sampling distribution for $\hat{\alpha}$, $\hat{\beta}$, $\hat{\mu}(x)$

Want to also construct confidence intervals. This requires further assumptions about the population distribution.

Let's assume a normal distribution:

$$Y_i \sim \mathrm{N}(\alpha + \beta x_i, \sigma^2).$$

$E(Y_i|x_i) = \alpha + \beta x_i$

$\mathrm{var}(Y_i|x_i) = \sigma^2$

Alternative notation (commonly used for regression/linear models):

$$\boxed{Y_i = \underbrace{\alpha + \beta x_i}_{\text{mean}} + \underset{\substack{\uparrow \\ \text{error term}}}{\epsilon_i}, \quad \text{where} \quad \epsilon_i \sim \mathrm{N}(0, \sigma^2).}$$

✓

Let's maximise the likelihood...

joint distribution of Yi's indep product

Since the Y_i's are independent, the likelihood is:

$$L(\alpha, \beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - \alpha - \beta x_i)^2}{2\sigma^2}\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{\sum_{i=1}^n (y_i - \alpha_0 - \beta(x_i - \bar{x}))^2}{2\sigma^2}\right\}$$

$$-\ln L(\alpha, \beta, \sigma^2) = \frac{n}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha_0 - \beta(x_i - \bar{x}))^2$$

minimize

$$= \underbrace{\frac{n}{2}\ln(2\pi\sigma^2)}_{\text{free of } \alpha, \beta} + \frac{1}{2\sigma^2}\boxed{H(\alpha_0, \beta)}$$

sum of square deviations

The α_0 and β that maximise the likelihood (minimise the log-likelihood) are the same as those that minimise the sum of squares, H.

minimise H

$\hat{\alpha}_0 = \bar{Y}$

$\hat{\beta} = \frac{\sum(x_i - \bar{x})Y_i}{\sum(x_i - \bar{x})^2}$

OLS
ordinary least squares

The OLS estimates are the same as the MLEs!

$$\psi = -\ln L(\alpha, \beta, \sigma^2) = \frac{n}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha_0 - \beta(x_i - \overline{x}))^2.$$

$$\frac{d\psi}{d\alpha_0} = \frac{-1}{\sigma^2}\sum_{i=1}^{n}(y_i - \alpha_0 - \beta(x_i - \overline{x})) = 0.$$

$$= \frac{-1}{\sigma^2}\left(\sum_{i=1}^{n} y_i - n\alpha_0\right) = 0. \quad \Rightarrow \hat{\alpha}_0 = \overline{Y}.$$

$$\frac{\partial\psi}{\partial\beta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \alpha_0 - \beta(x_i - \overline{x}))(x_i - \overline{x}) = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i\left(x_i - \overline{x}\right) - 0 - \beta\sum_{i=1}^{n}(x_i - \overline{x})^2 = 0.$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_i(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\frac{\partial\psi}{\partial\sigma^2} = \frac{n}{\cancel{2}} \cdot \frac{\cancel{2\pi\sigma}}{\cancel{2\pi}\sigma^{\cancel{2}}} + \frac{1}{2} \times (-2) \cdot \frac{1}{\sigma^3}\sum_{i=1}^{n}(y_i - \alpha_0 - \beta(x_i - \overline{x}))^2$$

$$= \frac{n}{\sigma} - \frac{1}{\sigma^3}\sum_{i=1}^{n}(y_i - \alpha_0 - \beta(x_i - \overline{x}))^2 = 0.$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \alpha_0 - \beta(x_i - \overline{x}))^2$$

$$= \frac{1}{n} \cdot D^2$$

so $\hat{\sigma}^2_{MLE} = \frac{1}{n} D^2$ biased.

$$E(\hat{\sigma}^2_{MLE}) = E\left(\frac{1}{n} D^2\right) = \frac{n-2}{n}\sigma^2$$

What about σ^2?

Differentiate by σ, set to zero, solve...

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n} D^2$$

This is biased. Prefer to use the previous, unbiased estimator,

$$\hat{\sigma}^2 = \frac{1}{n-2} D^2$$

$\frac{\partial \ln L(\alpha, \beta, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} H(\alpha_0, \beta)$

$\hat{\sigma}^2_{MLE} = \frac{1}{n} H(\hat{\alpha}_0, \hat{\beta})$

since we already get MLE for α_0, β plug in

$\hat{\sigma}^2_{MLE} = \frac{1}{n} D^2$

$D^2 = \sum_{i=1}^{n} R_i^2$ and $R_i = y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x})$

$X_1, \cdots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \Rightarrow E(S^2) = \sigma^2$

$\hat{\sigma}^2_{MLE} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \Rightarrow E(\hat{\sigma}^2_{MLE}) = \frac{n-1}{n} \sigma^2$

$E(\hat{\sigma}^2_{MLE}) = \frac{1}{n} E(D^2)$

$= \frac{n-2}{n} \sigma^2 \neq \sigma^2$

why different

Sampling distributions

The $Y_1, \cdots, Y_n$ are independent normally distributed random variables.

$\hat{\alpha} = \bar{Y}$

$\hat{\beta} = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

① assume x fixed ②

$\hat{\alpha}_0, \hat{\sigma}, \hat{\beta}, \hat{\mu}(x)$

Except for $\hat{\sigma}^2$, our estimators are linear combinations of the Y_i so will also have normal distributions, with mean and variance as previously derived.

$\hat{\alpha}_0 = \bar{Y} \sim N(\alpha_0, \frac{\sigma^2}{n})$

$cov(\hat{\alpha}_0, \hat{\beta}) = 0 \Rightarrow \hat{\alpha}_0, \hat{\beta}$ uncorrelate.

For example,

$$\hat{\beta} \sim \mathrm{N}\left(\beta, \frac{\sigma^2}{K}\right).$$

Moreover, we know $\boxed{\hat{\alpha}_0 \text{ and } \hat{\beta} \text{ are independent}}$, because they are bivariate normal rvs with zero covariance.

$X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

Using the analysis of variance decomposition (from earlier), we can show that,

$$\boxed{\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}.}$$

$\hat{\sigma}^2 = \frac{1}{n-2} D^2.$

$\sum(x_i - \mu)^2 = \sum(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$

$\hat{\sigma}^2 = \frac{1}{n-2} D^2$

$(n-2)\hat{\sigma}^2 = D^2 = \sum_{i=1}^{n}(y_i - \hat{\alpha}_0 - \hat{\beta}(x_i - \bar{x}))^2$

Therefore, we can define pivots for the various mean parameters. For example,

$$\boxed{\frac{\hat{\beta} - \beta}{\hat{\sigma}/\sqrt{K}} \sim t_{n-2}}$$

$se(\hat{\beta}) = \frac{\sigma}{\sqrt{K}}$

and

$$\frac{\hat{\mu}(x) - \mu(x)}{\underbrace{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{K}}}_{se(\hat{\mu}(x))}} \sim t_{n-2}$$

This allows us to construct confidence intervals.

$\frac{\hat{\beta} - \beta}{\sigma/\sqrt{K}} \sim N(0,1).$ $\quad \frac{\frac{\hat{\beta}-\beta}{\sigma/\sqrt{K}}}{\sqrt{\frac{(n-2)\hat{\sigma}^2}{(n-2)\sigma^2}}} = \frac{\hat{\beta} - \beta}{\hat{\sigma}/\sqrt{K}} \sim t_{n-2}$

$\hat{\beta} \pm a \cdot se(\hat{\beta}) \to \frac{\hat{\sigma}}{\sqrt{K}}$

a: quantile of t_{n-2}

Example: confidence itervals

For our data, a 95% CI for β is:

$$\hat{\beta} \pm c\,\frac{\hat{\sigma}}{\sqrt{K}} = 2.59 \pm 2.31 \times 0.404 = (1.66, 3.52)$$

where c is the 0.975 quantile of t_{n-2}.

A 95% CI for $\mu(3)$ is:

$$\hat{\mu}(3) \pm c \times \mathrm{se}(\hat{\mu}(3)) = 10.68 \pm 2.31 \times 0.667 = (9.14, 12.22)$$

for a CI, $\hat{\mu}(x)$ is used to estimate $\underbrace{E(Y|x)}_{\text{fixed value}}$ ②.

Deriving prediction intervals

Use the same trick as we used for the simple model,

for a PI, $\hat{\mu}(x)$ is used to estimate Y $\rightarrow$ r.v much more variation

$$Y^* \sim \mathrm{N}\left(\mu(x^*), \sigma^2\right)$$

$$\hat{\mu}(x^*) \sim \mathrm{N}\left(\mu(x^*), \left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{K}\right)\sigma^2\right)$$

$$Y^* - \hat{\mu}(x^*) \sim \mathrm{N}\left(0, \left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{K}\right)\sigma^2\right)$$

$$\frac{Y^* - \hat{\mu}(x^*)}{\sigma\sqrt{1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{K}}} \sim t_{n-2}$$

A 95% PI for Y^* is given by:

$$\hat{\mu}(x^*) \pm c\,\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{K}}$$

Example: prediction interval

A 95% PI for Y^* corresponding to $x^* = 3$ is:

$$10.68 \pm 2.31 \times 1.42 \times \sqrt{1 + \frac{1}{10} + \frac{(3 - 1.78)^2}{12.34}} = (7.06, 14.30)$$

Much wider than the corresponding CI, as we've seen previously.

```
> model1 <- lm(y ~ x)
> summary(model1)

Call:
lm(formula = y ~ x)
```

$R_i = y_i - (\hat{\alpha} + \hat{\beta} x_i)$

```
Residuals:
     Min       1Q   Median       3Q      Max
-2.01970 -1.05963  0.02808  1.04774  1.80580

Coefficients:
```

testing

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.9114     0.8479   3.434 0.008908 **
x             2.5897     0.4041   6.408 0.000207 ***
```

$\hat{\alpha}$ $se(\hat{\alpha})$ $\hat{\beta}$ $se(\hat{\beta})$

```
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

$\hat{\sigma}$ $n-2$

```
Residual standard error: 1.419 on 8 degrees of freedom
Multiple R-squared:  0.8369, Adjusted R-squared:  0.8166
F-statistic: 41.06 on 1 and 8 DF,  p-value: 0.0002074
```

```
> # Confidence intervals for mean parameters
> confint(model1)
                2.5 %   97.5 %
(Intercept) 0.9560629 4.866703
x           1.6577220 3.521623

> # Data to use for prediction.
> data2 <- data.frame(x = 3)

> # Confidence interval for mu(3).
> predict(model1, newdata = data2, interval = "confidence")
      fit      lwr      upr
1 10.6804 9.142823 12.21798

> # Prediction interval for y when x = 3.
> predict(model1, newdata = data2, interval = "prediction")
      fit   lwr     upr
1 10.6804 7.064 14.2968
```

α β

95% CIs for α and β

$\mu(x)$

$\hat{\mu}(x^*)$

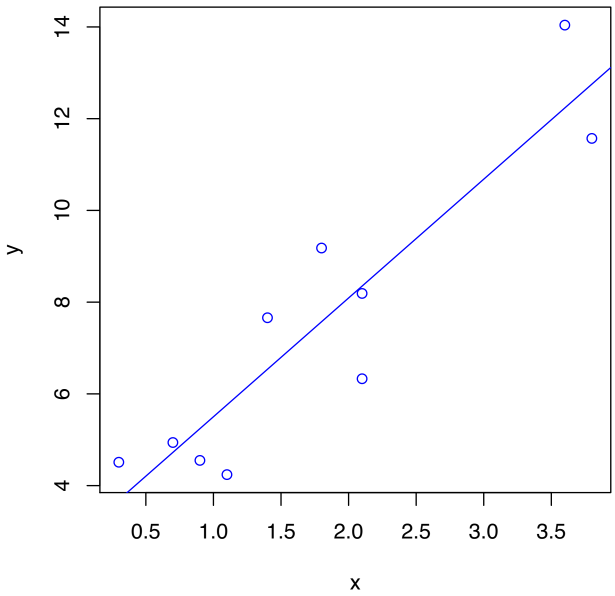
R example explained

- The `lm` (linear model) command fits the model.
- `model1` is an object that contains all the results of the regression needed for later calculations.
- `summary(model1)` acts on `model1` and summarizes the regression.
- `predict` can calculate CIs and PIs.
- R provides more detail than we need at the moment. Much of the output relates to hypothesis testing that we will get to later.

Plot data and fitted model

```
> plot(x, y, col = "blue")
> abline(model1, col = "blue")
```

The command `abline(model1)` adds the fitted line to a plot.



14
12
10
8
6
4
y
0.5
1.0
1.5
2.0
2.5
3.0
3.5
x

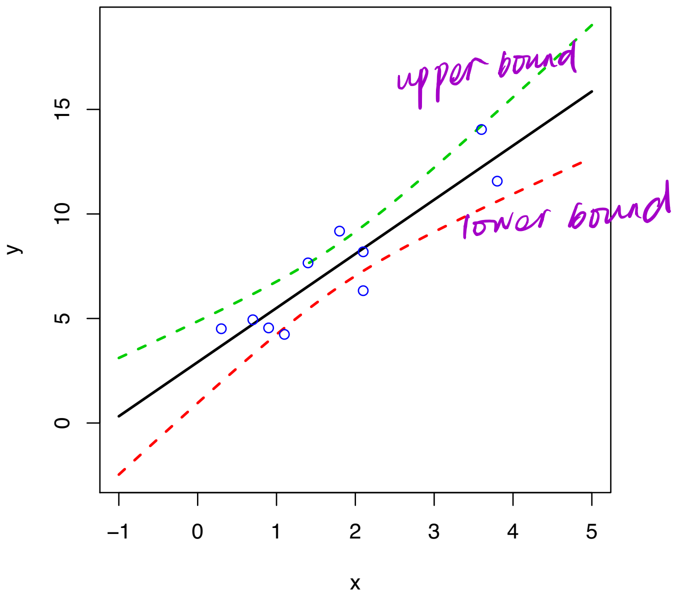
Fitted values and CIs for their means

$\hat{\mu}(x)$

```
> predict(model1, interval = "confidence")
         fit       lwr       upr
1   7.572793  6.537531  8.608056
2   6.536924  5.442924  7.630925
3   8.349695  7.272496  9.426895
4   3.688285  1.963799  5.412771
5  12.234204 10.247160 14.221248
6   4.724154  3.280382  6.167925
7   5.760023  4.546338  6.973707
8   8.349695  7.272496  9.426895
9   5.242088  3.921478  6.562699
10 12.752138 10.603796 14.900481
```

Confidence band for the mean

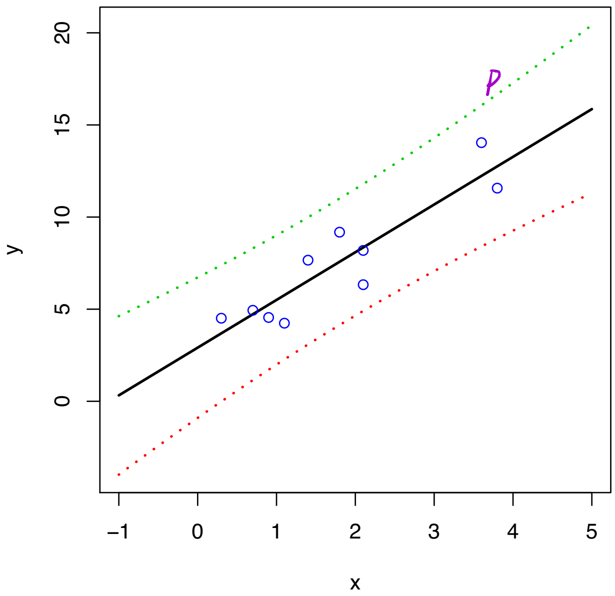
```
> data3 <- data.frame(x = seq(-1, 5, 0.05))
> y.conf <- predict(model1, data3, interval = "confidence")
> head(cbind(data3, y.conf))
      x       fit       lwr      upr
1 -1.00 0.3217104 -2.468232 3.111653
2 -0.95 0.4511941 -2.295531 3.197919
3 -0.90 0.5806777 -2.122943 3.284298
4 -0.85 0.7101613 -1.950472 3.370794
5 -0.80 0.8396449 -1.778124 3.457414
6 -0.75 0.9691286 -1.605906 3.544164

> matplot(data3$x, y.conf, type = "l", lty = c(1, 2, 2),
+         lwd = 2, xlab = "x", ylab = "y")
> points(x, y, col = "blue")
```



upper bound
lower bound
y
x
-1
0
1
2
3
4
5
0
5
10
15

Prediction bands for new observations

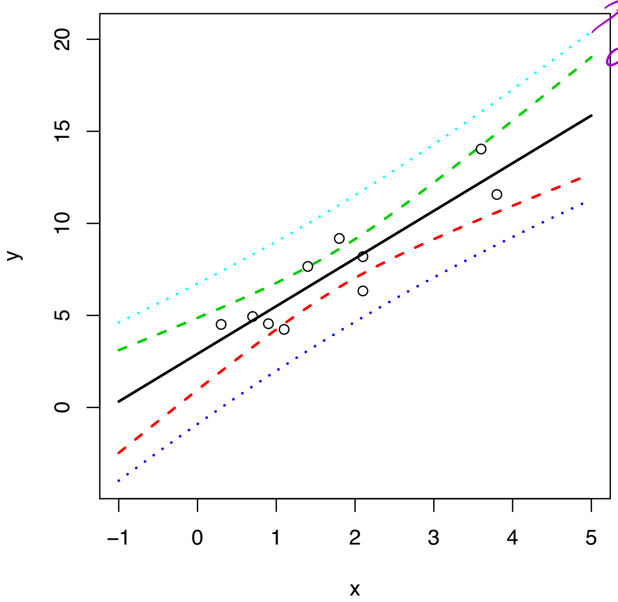
```
> y.pred <- predict(model1, data3, interval = "prediction")
> head(cbind(data3, y.pred))
      x       fit       lwr      upr
1 -1.00 0.3217104 -3.979218 4.622639
2 -0.95 0.4511941 -3.821827 4.724215
3 -0.90 0.5806777 -3.664763 4.826119
4 -0.85 0.7101613 -3.508034 4.928357
5 -0.80 0.8396449 -3.351646 5.030936
6 -0.75 0.9691286 -3.195606 5.133863

> matplot(data3$x, y.pred, type = "l", lty = c(1, 3, 3),
+         lwd = 2, xlab = "x", ylab = "y")
> points(x, y, col = "blue")
```



P
20
15
10
5
0
y
−1
0
1
2
3
4
5
x

Both bands plotted together

```
> matplot(data3$x, y.pred, type = "l", lty = c(1, 2, 2, 3, 3),
+         lwd = 2, xlab = "x", ylab = "y")
> points(x, y, col = "blue")
```



prediction band

confidence band

Checking our assumptions

What modelling assumptions have we made?

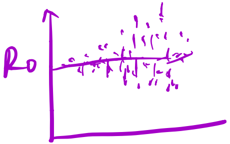
- Linear model for the mean $E(Y|X) = \alpha + \beta x$
- Equal variances for all observations (homoscedasticity) σ^2
- Normally distributed residuals

Ways to check these:

linear model

- Plot the data and fitted model together (done!)
- Plot residuals vs fitted values $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$
- QQ plot of the residuals $Y_i = \alpha + \beta x_i + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$

check residual



In R, the last two of these are very easy to do:

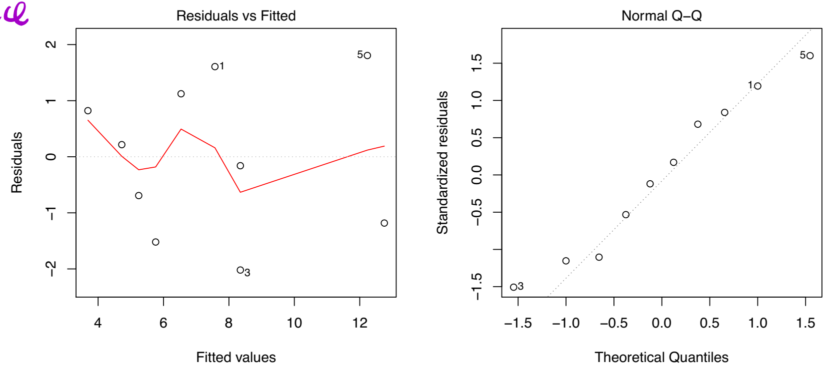
```
> plot(model1, 1:2)
```

FV

FV↑ var(R0)↑

contradict constance variance

$\varepsilon_i \sim N(0, \sigma^2(x))$



Outline

Multiple regression

- What if we have more than one predictor?
- Observe $x_{i1}, \dots, x_{ik}$ as well as y_i (for each i)
- Can fit a multiple regression model:

multiple predictor

k+1 coefficient

$$\mathbb{E}(Y \mid x_1, \dots, x_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

- This is linear in the coefficients, so is still a linear model
- Fit by method of least squares by minimising:

$$H = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik})^2$$

- Take partial derivatives, etc., and solve for $\beta_0, \dots, \beta_k$.
- The subject Linear Statistical Models (MAST30025) looks into these types of models in much more detail.

Two-sample problem

- The two-sample problem can be expressed as a linear model!
- Sample $Y_1, \ldots, Y_n \sim \mathrm{N}(\mu_1, \sigma^2)$ and $Y_{n+1}, \ldots, Y_{n+m} \sim \mathrm{N}(\mu_2, \sigma^2)$.
- Define indicator variables (x_{i1}, x_{i2}) where $(x_{i1}, x_{i2}) = (1, 0)$ for $i = 1, \ldots, n$ and $(x_{i1}, x_{i2}) = (0, 1)$ for $i = n+1, \ldots, n+m$.

from first sample

from second sample

- Observed data: $\boxed{(y_i, x_{i1}, x_{i2})}$
- Then $Y_1, \ldots, Y_n$ each have mean $1 \times \beta_1 + 0 \times \beta_2 = \mu_1$ and $Y_{n+1}, \ldots, Y_{n+m}$ each have mean $0 \times \beta_1 + 1 \times \beta_2 = \mu_2$.
- This is in the form a multiple regression model.
- The general linear model unifies many different types of models together into a common framework. The subject MAST30025 covers this in more detail.

Outline

Correlation coefficient

(Revision) for two rvs X and Y, the correlation coefficient, or simply the correlation, is defined as:

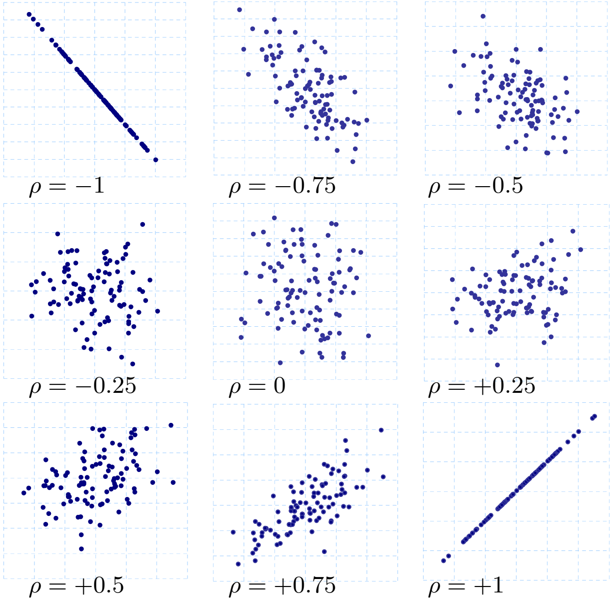
$$\rho = \rho_{XY} = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var} X \operatorname{var} Y}} = \boxed{\frac{\sigma_{XY}}{\sigma_X \sigma_Y}}$$

$|\rho| \le 1$

This is a quantitative measure of the strength of relationship, or association, between X and Y.

We will now consider inference on ρ,
based on an iid sample of pairs (X_i, Y_i).

Note: unlike in regression, X is now considered as a random variable.


$\rho = -1$
$\rho = -0.75$
$\rho = -0.5$
$\rho = -0.25$
$\rho = 0$
$\rho = +0.25$
$\rho = +0.5$
$\rho = +0.75$
$\rho = +1$

$\hat{\beta} = \dfrac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$

Sample covariance

To estimate $\mathrm{cov}(X, Y)$ we use the sample covariance:

$$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}\right)$$

$\sigma_{XY} = E(XY) - E(X)E(Y)$

random X

You can check that this is unbiased, $\mathbb{E}(S_{XY}) = \sigma_{XY} = \mathrm{cov}(X, Y)$.

?

$(n-1)E(S_{xy}) = E\left(\sum_{i=1}^{n} x_i y_i\right) - \frac{1}{n} E\left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right).$

$= n\mu_{xy} - E\left(\frac{1}{n}\left[\sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} x_i y_j\right]\right)$

$= n\mu_{xy} - \frac{1}{n}\left[n\mu_{xy} + n(n-1)\mu_x\mu_y\right]$

Sample correlation coefficient

$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$

To estimate ρ we use the sample correlation coefficient (also known as Pearson's correlation coefficient):

$$R = R_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

$R^2 = \frac{S_{XY}^2}{S_X^2 S_Y^2}$

You can check that $|R| \leqslant 1$, just like $|\rho| \leqslant 1$.

This gives a $\boxed{\text{point estimate of } \rho.}$

For further results, we make some more assumptions...

Bivariate normal

Assume X and Y have correlation ρ and follow a bivariate normal distribution,

mean vector   covariance matrix

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathrm{N}_2 \left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix} \right) \rightarrow \text{cov}(X,Y)$$

In this case, the regressions are linear,

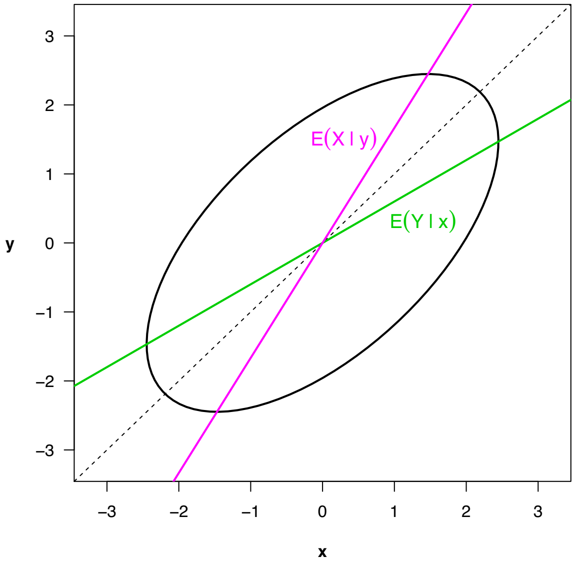
E(X)

?

$$\mathbb{E}(X \mid Y = y) = \mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y - \mu_Y) = \alpha' + \beta' y \quad \text{linear in } y$$

$$\mathbb{E}(Y \mid X = x) = \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X) = \alpha + \beta x \quad \text{linear in } x$$

Note: $\beta' \neq 1/\beta$


E(X | y)
E(Y | x)
y
x
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

Variance explained

An alternative analysis of variance decomposition:

$$\sum (Y_i - \bar{Y})^2 = \sum (Y_i - \hat{\alpha} - \hat{\beta}\boxed{x_i})^2 + \hat{\beta}^2 \sum (X_i - \bar{X})^2$$

$$= (1 - R^2) \sum (Y_i - \bar{Y})^2 + R^2 \sum (Y_i - \bar{Y})^2$$

This implies that R^2 is the proportion of the variation in Y 'explained' by x.

In this usage, R^2 is called the coefficient of determination.

in regression

$E(Y|x) = \alpha + \beta x$

$var(Y|x) = \sigma^2$

$\hat{\alpha}_0 = \bar{Y} = \hat{\alpha} + \hat{\beta}\bar{x} \Rightarrow \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}$

$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (Y_i - \hat{\alpha}_0)^2 = \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}\boxed{\bar{x}})^2$

to transform to x_i, do -1 trick

$= \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}x_i + \hat{\beta}x_i - \hat{\beta}\bar{x})^2$

$= \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}x_i)^2 + \sum_{i=1}^n (\hat{\beta}x_i - \hat{\beta}\bar{x})^2 = \sum_{i=1}^n R_i^2 + (\hat{\beta})^2 \sum_{i=1}^n (x_i - \bar{x})^2$

cross term

$= 2\sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}x_i)(\hat{\beta}x_i - \hat{\beta}\bar{x})$

?

$\Rightarrow \sum_{i=1}^n R_i^2 = (1 - R^2)\sum (Y_i - \bar{Y})^2$

$\hat{\beta} = \frac{\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2} \Rightarrow \hat{\beta}^2 = \frac{s_{xy}^2}{s_x^2 \cdot s_x^2}$

$\Rightarrow (\hat{\beta})^2 \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{s_{xy}^2}{s_x^2 \cdot \cancel{s_x^2}} \cdot (n-1)\cancel{s_x^2} = \frac{s_{xy}^2}{s_x^2 s_y^2}(n-1)s_y^2 = R^2 \sum_{i=1}^n (Y_i - \bar{Y})^2$

Remarks

- For simple linear regression, the coefficient of determination is the same as the square of the sample correlation, with both being denoted by R^2.
- Also, the proportion of Y explained by x is the same as the proportion of X explained by y. Both are equal to R^2, which is a symmetric expression of both X and Y.
- For more complex models, the coefficient of determination is more complicated: it needs to be calculated using all predictor variables together.

do regression of Y on X
or do regression of X on Y.
same coefficient of determinaion
R^2.

CI for ρ

Approximate sampling distribution

Define:

$$g(r) = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right)$$

This function has a standard name, $g(r) = \text{artanh}(r)$, and so does it's inverse, $g^{-1}(r) = \tanh(r)$. The function $g(r)$ is also known as the Fisher transformation.

The following is a widely used approximation:

R is used to estimate ρ

$$g(R) \approx \text{N}\left(g(\rho), \boxed{\frac{1}{n-3}} \right)$$

$g(R)$ is used to estimate $g(\rho)$

We can use this to construct approximate confidence intervals.

free of ρ

Example: correlation

For our data:

$$r = 0.91$$
$$r^2 = 0.84$$

An approximate 95% CI for $g(\rho)$ is: $se(g(r)) = \sqrt{\frac{1}{n-3}}$

$$g(r) \pm \frac{c}{\sqrt{n-3}} = 1.56 \pm 1.96 \times 0.378 = (0.819, 2.30)$$

where $c = \Phi^{-1}(1 - \alpha/2)$. Transforming this to an approximate 95% CI for ρ:

$$(\tanh(0.819), \tanh(2.30)) = (0.67, 0.98)$$

```
> cor(x, y)
[1] 0.9148421

> cor(x, y)^2
[1] 0.836936

> cor.test(x, y)

Pearson's product-moment correlation

data:  x and y
t = 6.4078, df = 8, p-value = 0.0002074
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6726924 0.9799873
sample estimates:
      cor
0.9148421
```

n-2

$R = \hat{\rho}$

```
> model1 <- lm(y ~ x)
> summary(model1)

Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max
-2.01970 -1.05963  0.02808  1.04774  1.80580

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.9114     0.8479   3.434 0.008908 **
x             2.5897     0.4041   6.408 0.000207 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.419 on 8 degrees of freedom
Multiple R-squared:  0.8369, Adjusted R-squared:  0.8166
F-statistic: 41.06 on 1 and 8 DF,  p-value: 0.0002074
```

R^2

test $\rho=0$