Multiple regression I

- When one dependent variable depends on multiple independent variables values of the dependent variable can be predicted using a function of all the independent variables that are believed to affect the dependent variable.
- So, if there are k independent variables the function $f : D_{X_1} \times D_{X_2} \times \ldots \times D_{X_k} \to D_Y$ where D_{X_i} is the domain of variation for variable X_i and D_y is the domain of variation for the dependent variable Y.
- We will only look at the case where f is a linear function. So, the estimate of Y, $\hat{Y}$ is given by: $\hat{Y} = a + b_1 X_1 + \ldots + b_k X_k$.
- Once we have multiple independent variables we have two cases:
  - The independent variables are pairwise uncorrelated (also called orthogonal).

Multiple regression II

  - There is at least one pair of correlated variables (also called non-orthogonal).
- We will study the case when we have two independent variables. When more than two variables are present the same basic approach works. We also start with the orthogonal case.

Multiple regression: two variables, orthogonal I

- In this case $\hat{Y} = a + b_1 X_1 + b_2 X_2$.
- Geometrically the regression line becomes a plane and 3-points are required to completely fix the plane.
- a is the intercept; b_1 is the slope in the X_1 direction and b_2 is the slope in the X_2 direction.
- Calculating the best regression plane is exactly similar to the one variable regression case. We calculate the squared error expression:
  $\mathcal{E}(\varepsilon) = \sum_i (Y_i - (a + b_1 X_{1_i} + b_2 X_{2_i}))^2$, find $\frac{\partial \mathcal{E}}{\partial a}$, $\frac{\partial \mathcal{E}}{\partial b_1}$, $\frac{\partial \mathcal{E}}{\partial b_2}$ equate each to 0 and find the values of $a,\ b_1,\ b_2$ that minimizes ε^2. The detailed derivation is an excercise.

Exercise

Exercise 2

Using the expression for the square error above differentiate with respect to $a, \ b_1, \ b_2$*, equate each to* 0 *and derive formulae for the 3 parameters* $a, \ b_1, \ b_2$*.*

Values of $a, \ b_1, \ b_2$

The values of $a, \ b_1, \ b_2$ are (using the notation of the earlier lec.):

$$a = \mu_Y - b_1\mu_{X_1} - b_2\mu_{X_2}$$

$$b_1 = \frac{SS_{X_2}SCP_{X_1 \cdot Y} - SCP_{X_1 \cdot X_2}SCP_{X_2 \cdot Y}}{SS_{X_1}SS_{X_2} - SCP^2_{X_1 \cdot X_2}}$$

$$b_2 = \frac{SS_{X_1}SCP_{X_2 \cdot Y} - SCP_{X_1 \cdot X_2}SCP_{X_1 \cdot Y}}{SS_{X_1}SS_{X_2} - SCP^2_{X_1 \cdot X_2}}$$

Also, note that the point $(\mu_{X_1}, \mu_{X_2}, \mu_Y)$ lies on the regression plane.

Quality of the prediction

As done for simple linear regression we need to judge whether the regression can occur by chance at a chosen α-level. Also, calculate percentage of variance explained by the linear relationship and the contribution of each variable.

- Quality is measured by the squared correlation: $R^2_{Y \cdot X_1 X_2} = \frac{SCP^2_{\hat{Y} \cdot Y}}{SS_{\hat{Y}} SS_Y}$. Note the use of capital R for multi-variable correlation.
- R^2 gives the fraction of the variance in Y explained by the linear relationship.
- The F-ratio is defined as: $F_{Y \cdot X_1 X_2} = \frac{R^2_{Y \cdot X_1 X_2}}{1 - R^2_{Y \cdot X_1 X_2}} \times \frac{N-k-1}{k}$ where N is the sample size and k is the number of independent variables.
- Assuming H_0 is true the F-ratio follows a Fisher distribution with $\nu_1 = k$ and $\nu_2 = N - k - 1$.
- For a chosen α if $F \geq F_{\text{critical}}$ then H_0 can be rejected.

Contribution of each independent variable I

- $R^2_{Y \cdot X_1 X_2}$ gives the fraction of the variability in Y that is explained by the linear relationship between Y and the independent variables or equivalently $(\hat{Y})$.
- The contribution of each independent variable to the predicted variable $\hat{Y}$ is given by single variable correlations:
  $r^2_{\hat{Y} \cdot X_1} = \frac{SCP^2_{\hat{Y} \cdot X_1}}{SS_{\hat{Y}} SS_{X_1}}$ and
  $r^2_{\hat{Y} \cdot X_2} = \frac{SCP^2_{\hat{Y} \cdot X_2}}{SS_{\hat{Y}} SS_{X_2}}$.
  Also, $r^2_{\hat{Y} \cdot X_1} + r^2_{\hat{Y} \cdot X_2} = 1$

Contribution of each independent variable II

- Similarly, the contribution of each independent variable to the dependent variable Y is given by single variable correlations:
  $r^2_{Y \cdot X_1} = \frac{SCP^2_{Y \cdot X_1}}{SS_Y SS_{X_1}}$ and
  $r^2_{Y \cdot X_2} = \frac{SCP^2_{Y \cdot X_2}}{SS_Y SS_{X_2}}$.
  Also, $r^2_{Y \cdot X_1} + r^2_{Y \cdot X_2} = R^2_{Y \cdot X_1 X_2}$
- So, $r^2_{Y \cdot X_1}$, $r^2_{Y \cdot X_2}$ give the proportion of the variance of Y given by X_1 and X_2 respectively. All the above holds only when X_1 and X_2 are orthogonal.
- From the earlier equations we have the following:
  $r^2_{Y \cdot X_1} = r^2_{\hat{Y} \cdot X_1} \times R^2_{Y \cdot X_1 X_2}$
  $r^2_{Y \cdot X_2} = r^2_{\hat{Y} \cdot X_2} \times R^2_{Y \cdot X_1 X_2}$

Tests for individual variables

- To check how different from 0 the contributions of the individual variables are calculate the F-ratios:
  $F_{Y \cdot X_1} = \frac{r^2_{Y \cdot X_1}}{1 - R^2_{Y \cdot X_1 X_2}} \times (N - k - 1)$
  $F_{Y \cdot X_2} = \frac{r^2_{Y \cdot X_2}}{1 - R^2_{Y \cdot X_1 X_2}} \times (N - k - 1)$
- Each F-ratio has a Fisher distribution with $\nu_1 = 1$, $\nu_2 = N - k - 1$ assuming H_0 is true. So, calculate the critical values for the chosen α-level and if $F \geq F_{\text{critical}}$ then reject H_0.

Partitioning the variance

- Just as in the simple linear regression case the variance of Y can be partitioned into a regression part and a residual part. $SS_{\text{total}} = SS_{\text{regression}} + SS_{\text{residual}}$
- $SS_{\text{regression}}$ can be further partitioned into contributions from X_1 and X_2.
- We have: $Y = \hat{Y} + \varepsilon$ and $\hat{Y} = a + b_1 X_1 + b_2 X_2$. Substituting for a gives:

$$\begin{aligned}
\hat{Y} &= \mu_Y - b_1 \mu_{X_1} - b_2 \mu_{X_2} + b_1 X_1 + b_2 X_2 \\
&= \mu_Y + b_1 (X_1 - \mu_{X_1}) + b_2 (X_2 - \mu_{X_2}) \\
Y &= \mu_Y + b_1 (X_1 - \mu_{X_1}) + b_2 (X_2 - \mu_{X_2}) + (Y - \hat{Y})
\end{aligned}$$

So, Y is made up of μ_Y, sum of each weighted, zero mean centred independent variables and the error.

Partitioning $SS_{\text{regression}}$

$$
\begin{aligned}
\hat{Y} - \mu_Y &= b_1(X_1 - \mu_{X_1}) + b_2(X_2 - \mu_{X_2}) \\
\sum_i (\hat{Y} - \mu_Y)^2 &= \sum_i [b_1(X_1 - \mu_{X_1}) + b_2(X_2 - \mu_{X_2})]^2 \\
&= b_1^2 \sum_i (X_1 - \mu_{X_1})^2 + b_2^2 \sum_i (X_2 - \mu_{X_2})^2 + \\
&\quad 2b_1 b_2 \sum_i (X_1 - \mu_{X_1})(X_2 - \mu_{X_2}) \\
&= b_1^2 SS_{X_1} + b_2^2 SS_{X_2} + 2X_1X_2 SCP_{X_1 \cdot X_2} \\
&= b_1^2 SS_{X_1} + b_2^2 SS_{X_2} \\
&\quad SCP_{X_1 \cdot X_2} = 0 \text{ since } X_1,\ X_2 \text{ are orthogonal}
\end{aligned}
$$

Partitioning $SS_{\text{regression}}$ contd.

$$
\begin{aligned}
SS_{\text{regression}} &= b_1^2 SS_{X_1} + b_2^2 SS_{X_2} \\
&= \frac{SCP^2_{X_1 \cdot Y}}{SS^2_{X_1}} SS_{X_1} + \frac{SCP^2_{X_2 \cdot Y}}{SS^2_{X_2}} SS_{X_2} \quad \text{substitute for } b_1,\ b_2 \\
&= r^2_{YX_1} SS_Y + r^2_{YX_2} SS_Y = SS_{Y \cdot X_1} + SS_{Y \cdot X_2} \\
R^2_{Y \cdot X_1 X_2} &= r^2_{YX_1} + r^2_{YX_2} \quad \text{after dividing above by } SS_Y
\end{aligned}
$$

Degrees of freedom and mean square values

- Three points are needed to fully define the regression plane. $(\mu_{X_1}, \mu_{X_2}, \mu_Y)$ is on the plane. So, only two points can be independently chosen so $df_{\text{regression}} = 2$.
- For df_{residual} the sample has N points, k independent variable values are known and the residual has mean 0. So, $df_{\text{residual}} = N - k - 1$.
- So, we can use mean squares(as done earlier)
  $s^2_{\text{regression}} = \frac{SS_{\text{regression}}}{df_{\text{regression}}} = \frac{SS_{\text{regression}}}{2}$
  $s^2_{\text{residual}} = \frac{SS_{\text{residual}}}{df_{\text{residual}}} = \frac{SS_{\text{residual}}}{N-k-1}$
  $s^2_{Y \cdot X_1} = \frac{SS_{Y \cdot X_1}}{df_{X1}}$
  $s^2_{Y \cdot X_2} = \frac{SS_{Y \cdot X_2}}{df_{X2}}$
- Note that $df_{X_1} = 1$, $df_{X2} = 1$.
- Mean squares, unlike sum of squares and degrees of freedom, do not add up.

Using mean squares for quality

- Similar to simple linear regression but now we get 3 F-ratios:
  $F_{Y \cdot X_1 X_2} = \frac{s^2_{\text{regression}}}{s^2_{\text{residual}}}$
  $F_{Y \cdot X_1} = \frac{s^2_{Y \cdot X_1}}{s^2_{\text{residual}}}$
  $F_{Y \cdot X_2} = \frac{s^2_{Y \cdot X_2}}{s^2_{\text{residual}}}$
- The test remains the same. Find F_{critical} in each case using the F-distribution using the relevant α, ν_1, ν_2 and reject H_0 if $F_{\text{ratio}} \geq F_{\text{critical}}$.

Problem/Exercise I

This is a pseudo replication of an experiment on retro-active interference. Retro-active interference happens when what we are learning now can interfere with what we have learnt in the past and therefore make us forget what was learnt earlier. The experiment involves learning a first list of words, then a second list of words followed by a recall of the first list of words. The controls learn only the first list and recall it. In the actual experiment a subject was presented a sentence 2, 4 or 8 times (first list - X_1) and then presented a second sentence again 2, 4 or 8 times (second list - X_2) and was then asked to recall the words in the first sentence. The number of words recalled was the dependent variable Y. The data below gives the details for 18 subjects two per condition.

	X_2		
X_1	2	4	8
2	35	21	6
	39	31	8
4	40	34	18
	52	42	26
8	61	58	46
	73	66	52

1. Plot a scatter diagram for the data.

Problem/Exercise II

2. Find the best fit minimum squared error regression plane.
3. Find $R^2_{Y \cdot X_1 X_2}$.
4. Find $F_{Y \cdot X_1 X_2}$.
5. Find the critical values for the F-distribution for $\alpha = 0.05,\ 0.01$.
6. Can H_0 be rejected at the above two alpha levels?
7. What percent of the variance can be explained by regresssion?
8. Find $r^2_{\hat{Y} \cdot X_1}$, $r^2_{\hat{Y} \cdot X_2}$, $r^2_{Y \cdot X_1}$, $r^2_{Y \cdot X_1}$
9. For the individual correlations find the F-ratios, critical values and test the corresponding H_0.
10. Compute F using $s^2_{\text{regression}}$ and s^2_{residual} and repeat the tests for H_0 at the two levels.
11. Repeat the individual variable tests using mean squares and find the fraction of the variance explained by each variable.

Example experiment - non-orthogonal variables I

This is a pseudo replication of an experiment that studies memory span in children with changing age and speech rate. In the experiment children of 3 ages 4, 7, 10 years (first independent variable X_1) were tested for immediate serial recall of 15 items. The dependent variable was the total number of words correctly recalled (Y dependent variable). The speech rate of each subject was also measured by making the subject read a passage aloud and measuring the time taken. The speech rate is then calculated as words read per unit time (X_2 second independent variable). The hypothesis is that the memory span Y will be positively correlated with age and speech rate. But it is also known that higher age usually also means higher speech rates. So, there is a correlation between age and speech rate as well. The independent variables are, therefore, non-orthogonal.

X_1(Age)	X_2(Speech rate)	Y (words recalled)
4	1	14
4	2	23
7	2	30
7	4	50
10	3	39
10	6	67

Correlated/non-orthogonal independent variables - what is same

- In some experiments (see example in previous slide) the independent variables are correlated or non-orthogonal. Some details of the regression calculations change when this is the case others remain the same.
- The bulk of the analysis comprising the items below remains unchanged.
  - i) formulae for a, b_1, b_2 ii) calculation of quality parameters $R^2_{Y \cdot X_1 X_2}$, $F_{Y \cdot X_1 X_2}$ iii) critical values from F-distribution and iv) test of the statistic to decide whether or not H_0 should be rejected.
  - Formulae for individual variable contributions $r^2_{\hat{Y} \cdot X_1}$, $r^2_{\hat{Y} \cdot X_2}$ and $r^2_{Y \cdot X_1}$, $r^2_{Y \cdot X_2}$.
  - Formulae and tests for the individual variables' statistics $F_{Y \cdot X_1 \geq F_{\text{critical}}}$ and $F_{Y \cdot X_2} \geq F_{\text{critical}}$.

Correlated/non-orthogonal independent variables - what is different

- What is different in the non-orthogonal case is:
  i) $r^2_{\hat{Y}\cdot X_1} + r^2_{\hat{Y}\cdot X_2} \neq 1$ and
  ii) $r^2_{Y\cdot X_1} + r^2_{Y\cdot X_2} \neq R^2_{Y\cdot X_1 X_2}$.
- Since X_1, X_2 are correlated there are three contributions to Y (apart from the residual) a) common part of X_1 and X_2, b) the part specific only to X_1 c) the part specfic only to X_2.
- The non-zero common part being counted twice leads to the inequalities.

Values of $a,\ b_1,\ b_2$

The values of $a,\ b_1,\ b_2$ are exactly the same as for multiple regression in the orthogonal case:

$$a = \mu_Y - b_1\mu_{X_1} - b_2\mu_{X_2}$$

$$b_1 = \frac{SS_{X_2}SCP_{X_1\cdot Y} - SCP_{X_1\cdot X_2}SCP_{X_2\cdot Y}}{SS_{X_1}SS_{X_2} - SCP^2_{X_1\cdot X_2}}$$

$$b_2 = \frac{SS_{X_1}SCP_{X_2\cdot Y} - SCP_{X_1\cdot X_2}SCP_{X_1\cdot Y}}{SS_{X_1}SS_{X_2} - SCP^2_{X_1\cdot X_2}}$$

Also, note that the point $(\mu_{X_1}, \mu_{X_2}, \mu_Y)$ lies on the regression plane.

The formulae for the statistics and tests for H_0 rejection remain unchanged. So, look at the slides for the non-orthogonal case for the details.

Exercise/Problem

Exercise 3

For the experimental data given in the first slide calculate the parameters a, b_1, b_2. *Then compute* $R^2_{Y\cdot X_1X_2}$, $F_{Y\cdot X_1X_2}$ *and test whether or not* H_0 *can be rejected.*
Further calculate $r^2_{\hat{Y}\cdot X_1}$, $r^2_{\hat{Y}\cdot X_2}$ *and* $r^2_{Y\cdot X_1}$, $r^2_{Y\cdot X_2}$ *and verify that* $r^2_{\hat{Y}\cdot X_1} + r^2_{\hat{Y}\cdot X_2} \neq 1$ *and* $r^2_{Y\cdot X_1} + r^2_{Y\cdot X_2} \neq R^2_{Y\cdot X_1X_2}$. *This is unlike the orthogonal case.*

Contribution of each variable

- To isolate the specific contribution of each independent variable compute how much X_1 is predicted by X_2 and vice versa. The residual of each prediction (that is $(\varepsilon_{X_1} = X_1 - \hat{X}_1)$ and $(\varepsilon_{X_2} = X_2 - \hat{X}_2)$) is clearly uncorrelated with the predictor and gives the specific contribution of the corresponding independent variable.
- The equation for dependence of X_1 on X_2 is:
  $\hat{X}_1 = a_{X2} + b_{X2}X_2$
  and for the dependence of X_2 on X_1 is:
  $\hat{X}_2 = a_{X1} + b_{X1}X_1$
- The formulae for the parameters in the two equations are (similar to the simple linear regression formulae):
  $a_{X2} = \mu_{X_1} - b_{X2}\mu_{X_2}, \qquad b_{X2} = \frac{SCP_{X_1 \cdot X_2}}{SS_{X_2}}$
  $a_{X1} = \mu_{X_2} - b_{X1}\mu_{X_1}, \qquad b_{X1} = \frac{SCP_{X_1 \cdot X_2}}{SS_{X_1}}$
- The specific contributions of the two independent variables are: ε_{X_1} and ε_{X_2}.

Contribution of each variable contd.

- The *part correlation* (so called because correlation is between a part of the independent variable and dependent variable Y) $r^2_{Y\cdot\varepsilon_{X_1}}$ and $r^2_{Y\cdot\varepsilon_{X_2}}$ is given by:

$$r^2_{Y\cdot\varepsilon_{X_1}} = \frac{SCP^2_{Y\cdot\varepsilon_{X_1}}}{SS_Y SS_{\varepsilon_{X_1}}}$$

$$r^2_{Y\cdot\varepsilon_{X_2}} = \frac{SCP^2_{Y\cdot\varepsilon_{X_2}}}{SS_Y SS_{\varepsilon_{X_2}}}$$

Exercise 4

Calculate the regression between X_1 (dependent) and X_2(independent) and vice versa by finding a_{X2}, b_{X2} and a_{X1} and b_{X1}.
Calculate the part correlation of ε_{X_1} and ε_{X_2} with Y - that is $r^2_{Y\cdot\varepsilon_{X_1}}$ and $r^2_{Y\cdot\varepsilon_{X_2}}$.

Individual variable contributions as increments

- Another way to calculate specific individual variable contributions is by looking at the increment that is obtained when the k^{th} variable is added. Assume, $R^2_{Y \cdot X_1 \dots X_{k-1}}$ then add the k^{th} variable to get $R^2_{Y \cdot X_1 \dots X_k}$. The specific contribution of the k^{th} variable will be

$$r^2_{Y \cdot \varepsilon_{X_k}} = R^2_{Y \cdot X_1 \dots X_k} - R^2_{Y \cdot X_1 \dots X_{k-1}}$$

- For $k = 2$, the individual variable contributions are given by $r^2_{Y \cdot \varepsilon_{X_1}} = R^2_{Y \cdot X_1 X_2} - r^2_{Y \cdot X_2}$ and $r^2_{Y \cdot \varepsilon_{X_2}} = R^2_{Y \cdot X_1 X_2} - r^2_{Y \cdot X_1}$.
- Let $R^2_{Y \cdot X_1 \cap X_2}$ denote the part of the variance explained by the common portion of X_1 and X_2. Then, $R^2_{Y \cdot X_1 \cap X_2} = R^2_{Y \cdot X_1 X_2} - (r^2_{Y \cdot \varepsilon_{X_1}} + r^2_{Y \cdot \varepsilon_{X_2}})$

Partitioning the variance

- The total variance can be partitioned into four parts:

$$\begin{aligned}\text{total variance} \quad = \quad & \text{variance due to common part of } X_1, X_2 + \\ & \text{variance specifically due to } X_1 + \\ & \text{variance specifically due to } X_2 + \\ & \text{residual variance (unexplained)}\end{aligned}$$

- As an equation:

$$R^2_{Y \cdot X_1 \cap X_2} + r^2_{Y \cdot \varepsilon_{X_1}} + r^2_{Y \cdot \varepsilon_{X_2}} + \text{residual variance} = 1$$

Exercise 5

Calculate $r^2_{Y \cdot \varepsilon_{X_1}}$, $r^2_{Y \cdot \varepsilon_{X_2}}$ *and* $R^2_{Y \cdot X_1 \cap X_2}$ *using the 'contribution as increment' method and the residual variance. Compare with the values calculated in the earlier exercise.*

Alternate formulae for part correlation coefficients

- The part correlation coefficients can be directly calculated using other correlation coefficients:

$$r^2_{Y\cdot\varepsilon_{X_1}} = \frac{(r_{Y\cdot X_1} - r_{Y\cdot X_2} r_{X_1\cdot X_2})^2}{1 - r^2_{X_1\cdot X_2}}$$

$$r^2_{Y\cdot\varepsilon_{X_2}} = \frac{(r_{Y\cdot X_2} - r_{Y\cdot X_1} r_{X_1\cdot X_2})^2}{1 - r^2_{X_1\cdot X_2}}$$

Exercise 6

Compute the part correlation coefficients with alternate formulae and confirm that you get the same values.

F test for part correlation coefficients

- The formulae for the F-ratio for the part correlation coefficients is similar to the earlier formulae:

$$F_{Y \cdot \varepsilon_{X_1}} = \frac{r^2_{Y \cdot \varepsilon_{X_1}}}{1 - R^2_{Y \cdot X_1 X_2}} (N - K - 1)$$

$$F_{Y \cdot \varepsilon_{X_2}} = \frac{r^2_{Y \cdot \varepsilon_{X_2}}}{1 - R^2_{Y \cdot X_1 X_2}} (N - K - 1)$$

- The sampling distribution is the Fisher F-distribution with $\nu_1 = 1$ and $\nu_2 = N - k - 1$. The critical values can be found for the chosen α and then reject H_0 if $F - \text{ratio} \geq F_{\text{critical}}$.

Exercise 7

Calculate the F-ratio for both part correlation coefficients and determine if H_0 can be rejected $\alpha = 0.5$ and $\alpha = 0.01$.

More than two independent variables

- When $k > 2$ then the formulae for the parameters in the regression equation become very complicated and it is best to use libraries to do the necessary calculations.
- The incremental contribution of a variable approach can be used to calculate the part correlation coefficients. The alternate formulae (extended versions) can also be used to calculate the same coefficients.

Multi-collinearity

- If one or more independent variable is completely determined by another independent variable - that is the correlation coefficient between the two variables is 1 then the denominatior $1 - r^2_{X_1 \cdot X_2}$ is 0.
- This situation is akin to having dependent variables in systems of linear equations.
- Calculations can also break down if two independent variables are highly correlated (that is a correlation coefficient very close to 1). Then $1 - r^2_{X_1 \cdot X_2}$ can be very small leading to overflows.

More general use of part correlation coefficient

- To calculate the dependencies in a set of dependent variables after the effect of one variable in the set has been removed part correlation methods can be used. For example, if Y, W and Z are a set of dependent variables then to eliminate the effect of Z from Y and W find the residual $\varepsilon_{Y_Z} = Y - \hat{Y}_Z$ and $\varepsilon_{W_Z} = W - \hat{W}_Z$ then find the correlation coefficient between ε_{Y_Z} and ε_{W_Z}. One can write the formula in terms of other correlation coefficients:

$$r^2_{\varepsilon_{Y_Z} \cdot \varepsilon_{W_Z}} = \frac{(r_{Y \cdot W} - r_{Y \cdot Z} r_{W \cdot Z})^2}{(1 - r^2_{Y \cdot Z})(1 - r^2_{W \cdot Z})}$$