properties of OLS Matrix

Sum of Squared Residuals	$(y - X\hat{\beta})'(y - X\hat{\beta})$
	$y'y - \hat{\beta}'X'y - y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$
	$y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$
Minimise the SSR	$\partial(SSR)/\partial\hat{\beta} = -2X'y + 2X'X\hat{\beta} = 0$
from the minimum we get: "normal equation"	$(X'X)\hat{\beta} = X'y$
Solve for OLS estimator $\hat{\beta}$; by pre multiplying both sides by (X′X)	$(X'X)-1(X'X)\hat{\beta} = (X'X)-1X'y$
by definition, $(X'X)-1(X'X) = I$	$I\hat{\beta} = (X'X)-1X'y$
	$\hat{\beta} = (X'X)-1X'y$
Properties	
The observed values of X are uncorrelated with the residuals.	$X'e = 0$ implies that for every column xk of X, $x'ke = 0$.
substitute in $y = X\hat{\beta} + e$ into normal equation	$(X'X)\hat{\beta} = X'(X\hat{\beta} + e)$
	$(X'X)\hat{\beta} = (X'X)\hat{\beta} + X'e$
	$X'e = 0$
The sum of the residuals is zero.	If there is a constant, then the first column in X (i.e. X1) will be a column of ones. This means that for the first element in the X′e vector (i.e. X11 ×e1 +X12 ×e2 +...+X1n ×en) to be zero, it must be the case that ei = 0.
The sample mean of the residuals is zero.	$e = \sum e\, i/n = 0$.

properties of OLS Matrix (cont)

The regression hyperplane passes through the means of the observed values (X and y).	This follows from the fact that e = 0. Recall that $e = y - X\hat{\beta}$. Dividing by the number of observations, we get $e = y - x\hat{\beta} = 0$. This implies that $y = x\hat{\beta}$. This shows that the regression hyperplane goes through the point of means of the data.
The predicted values of y are uncorrelated with the residuals.	$\hat{}'e = (X\hat{\beta})'e = b'X'e = 0$
The mean of the predicted Y's for the sample will equal the mean of the observed Y's : y^-=y-	
The Gauss--Markov Theorem: Proof that $\hat{\beta}$ is an unbiased estimator of β	$\hat{\beta} = (X'X)^{-1}X'y=(X'X)^{-1}X'(X\beta + \varepsilon)$
	$\beta + (X'X)^{-1}X'\varepsilon$
given $(X'X)^{-1}X'X = I$	$E[\hat{\beta}] = E[\beta] + E[(X'X)^{-1}X'\varepsilon] = \beta + (X'X)^{-1}X'E[\varepsilon]$
where $E[X'\varepsilon]=0$	$E[\hat{\beta}]=\beta$
Proof that $\hat{\beta}$ is a linear estimator of β.	$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$; where $(X'X)^{-1}X'= A$
	$\hat{\beta} = \beta + A\varepsilon$ => linear equation

Hetroskedasticity

consequence:	the statistics used to test hypotheses under Gauss--Markov assumptions are not valid in the presence of hetroskedasticity.
Valid estimator (any form)	$\sum[(x1- x-)^2\, u\hat{}i^2]/[SST^2x]$
	$SSTx=\sum(x1- x-)^2$
Robust Standard error	$Var\hat{}(\hat{\beta}j)=\sum[r\hat{}ij^2\hat{u}^2i]/[SSR^2j]$

Inference

Normality Assumption:	zero mean and Variance
	Var(u)= σ^2
T-test:	($\hat{\beta}$ j- β j)/se($\hat{\beta}$ j)~ t n-k-1 =t df
H0 : βj = 0	used in testing hypotheses about a single population parameter as in .
Test statistic	t $\hat{\beta}$ j=($\hat{\beta}$ j)/se($\hat{\beta}$ j)~ t n-k-1
	t = (estimate − hypothesised value)/ standard error
Alternative Hypothesis/one sided	
H1: βj > 0	t $\hat{\beta}$j > c [c @5%]
H1: βj < 0	t $\hat{\beta}$j <- c [c @5%]
Two sided	
H1: βj =/= 0	|t$\hat{\beta}$j | > c [c @2.5%]
If H0, rejected	x j is statistically significant, (significantly different from zero), @ the 5% level
if H0, not rejected	x j is statistically insignificant @the 5% level
P-value	smallest significant level at which the null hypotheses would be rejected
Confidence Interval	$\hat{\beta}$j ±c·se($\hat{\beta}$j)
	where c is 97.5 percentile in a t n-k-1 distribution
CI given; @ 5% significant level	H0 :βj =aj is rejected against H1:βj ≠ =aj ; if aj is not in the 95% confidence interval
H0:β1<β2 ⇔ β1−β2<0	t= ($\hat{\beta}$1−$\hat{\beta}$2) /se($\hat{\beta}$1 − $\hat{\beta}$2)
se($\hat{\beta}$1 − $\hat{\beta}$2) = √Var($\hat{\beta}$1 − $\hat{\beta}$2)	
	Var($\hat{\beta}$1 − $\hat{\beta}$2) = Var($\hat{\beta}$1) + Var($\hat{\beta}$2) − 2Cov($\hat{\beta}$1, $\hat{\beta}$2)
alternative to calculating se($\hat{\beta}$1 − $\hat{\beta}$2)	Let θ = $\hat{\beta}$1 − $\hat{\beta}$2; β1 = θ + $\hat{\beta}$2

Inference (cont)

H0: θ=0, H1: θ<0	Substituting β1 = θ + $\hat{\beta}$2 into the model we obtain
β0 +θ x1 +β2(x1 +x2)+β3 x3 +u	
F Test	F =[(SSRr-SSRur)/q] / [SSRur/(n-k-1)]
q	=number of restrictions
n-k-1= df ur	= df r- df ur
R^2 F stat	SSR= SST(1 - R^2)
	F= [(R^2ur-R^2r)/q] / [1-R^2ur)/(df ur)]
remember to not square the R value thats already been done	
Overall significance of the regression	
Testing joint exclusion	[R^2/R]/[(1-R^2)/(n-k-1)]

Data Scaling

Changes:	
if Xj is * by c	Its coefficient is / by c
If dependant variable is * by c	ALL OLS coefficients are * by c
neither t nor F statistics are affected	
Beta coefficients	obtained from an OLS regression after the dependant and independent variables have been transformed into z-scores

Dummy Variables

Dummy/Binary Variables	= yes/no variables
	= take on the values 0 and 1 to identify the mutually exclusive classes of the explanatory variables.
	= leads to regression models where the parameters have very natural interpretations
Given: wage= β0+ ∂0 female + β1 edu + u	
	To solve for ∂0:
	∂0=E(wage|female,edu)-E(wage|male,edu)
	where level of education is the same
Graphically ∂0 =	an intercept shift
	male intercept= β0

Dummy Variables (cont)	
	female intercept= β0+∂0
dummy variable trap=	when both dummy variables (male & female) are included; resulting in perfect collinearity
If a qualitative variable has m levels;	then (m−1) dummy variables are required and each of them takes value 0 and 1.
Hypothesis test	
Test whether the two regression models are identical:	
	H0 :β2 =β3 0
	H1 :β2 ≠0 and/or β3 ≠0.
Acceptance of H0 indicates that only single model is necessary to explain the relationship.	
Test is two models differ with respect to intercepts only and they have same slopes	
	H0 :β3 =0
	H1:β3 ≠0.
Treating a **quantitative variable** as qualitative variable increases the complexity of the model.	
	The degrees of freedom for error are reduced.
	Can effect the inferences if data set is small