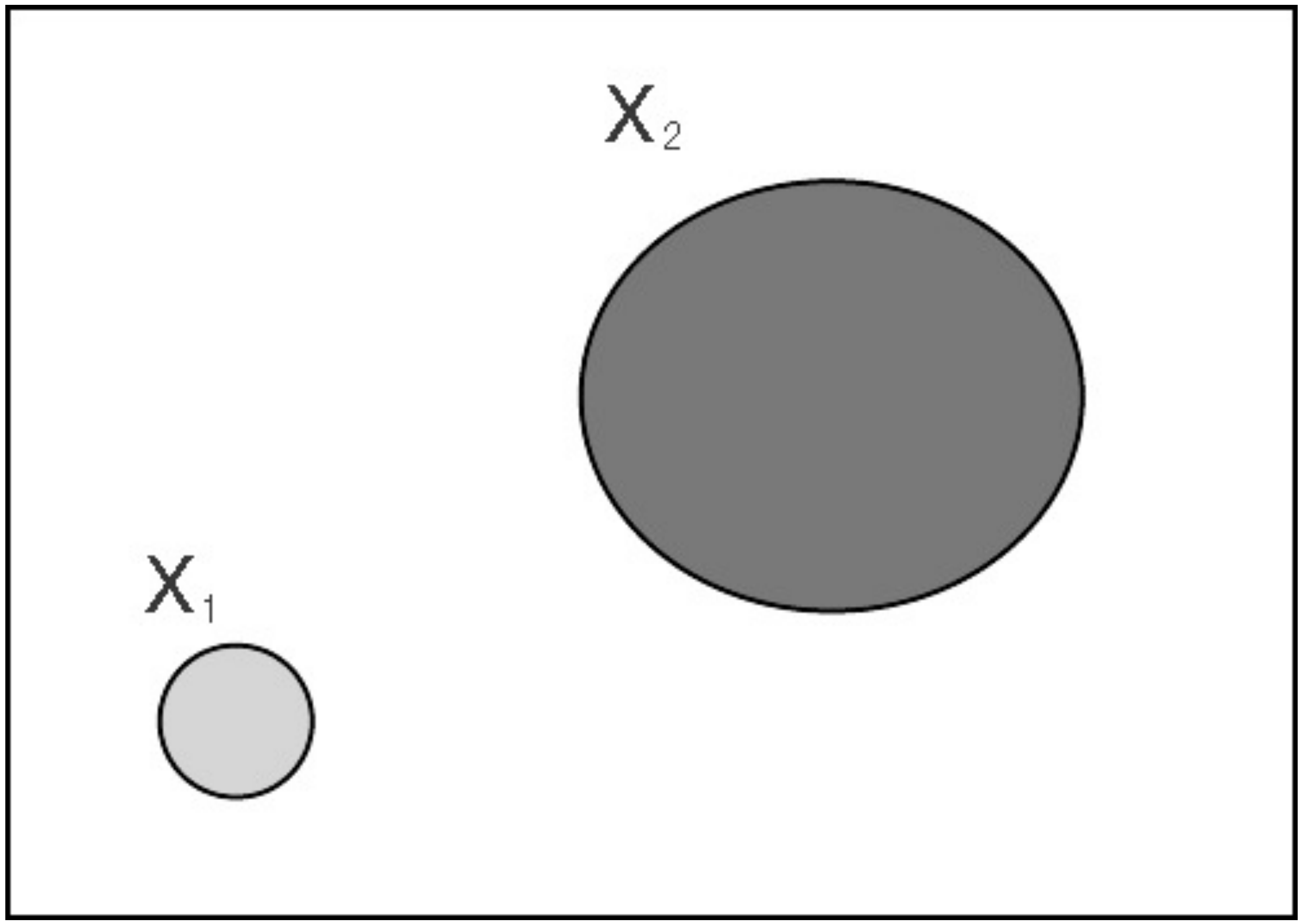
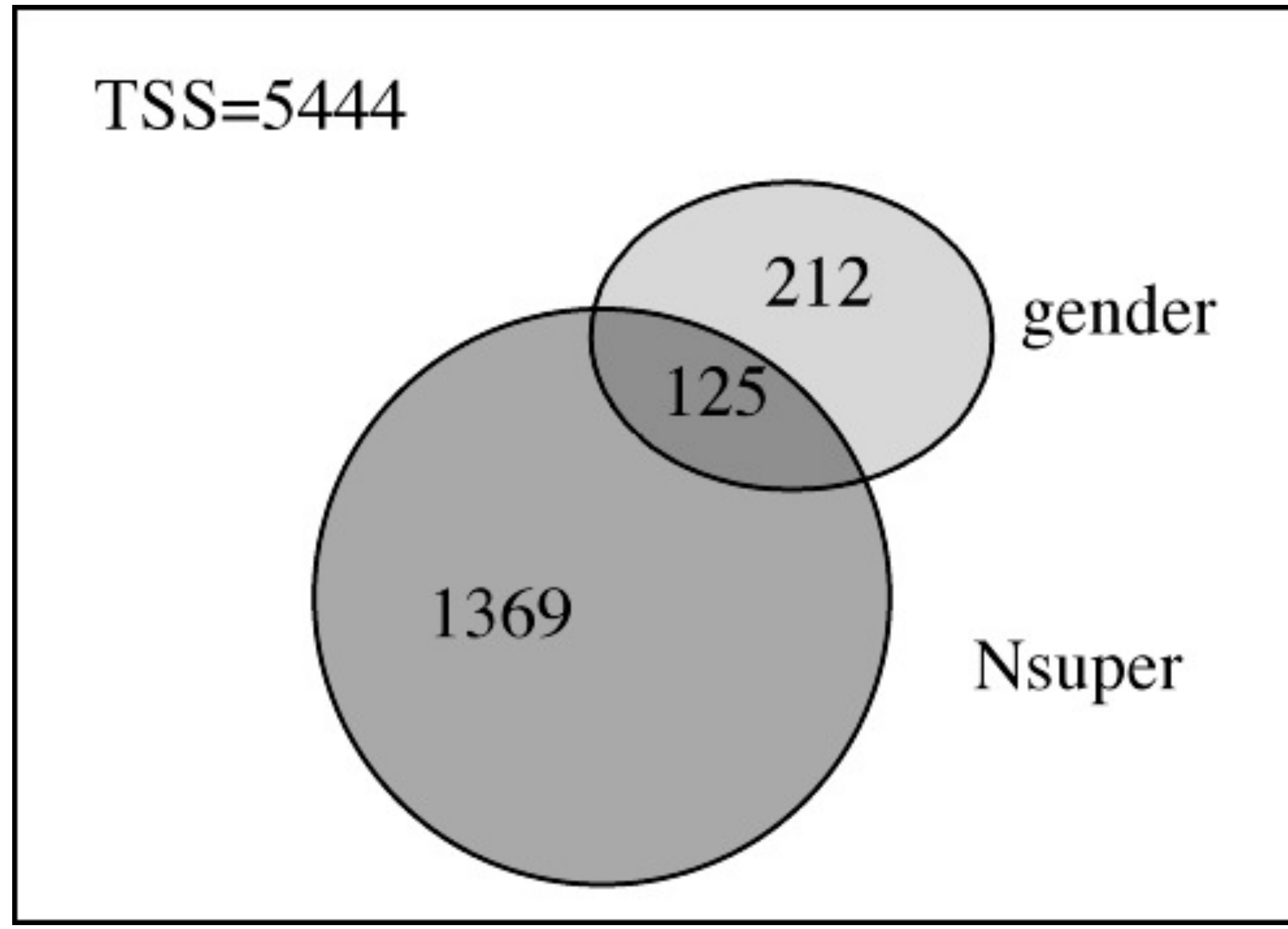
# Visualizing Multiple Regression

based on Edward H. S. Ip (2001) and
Peter E. Kennedy (2002)

# Coefficient of Determination $R^2$=SSR/SST



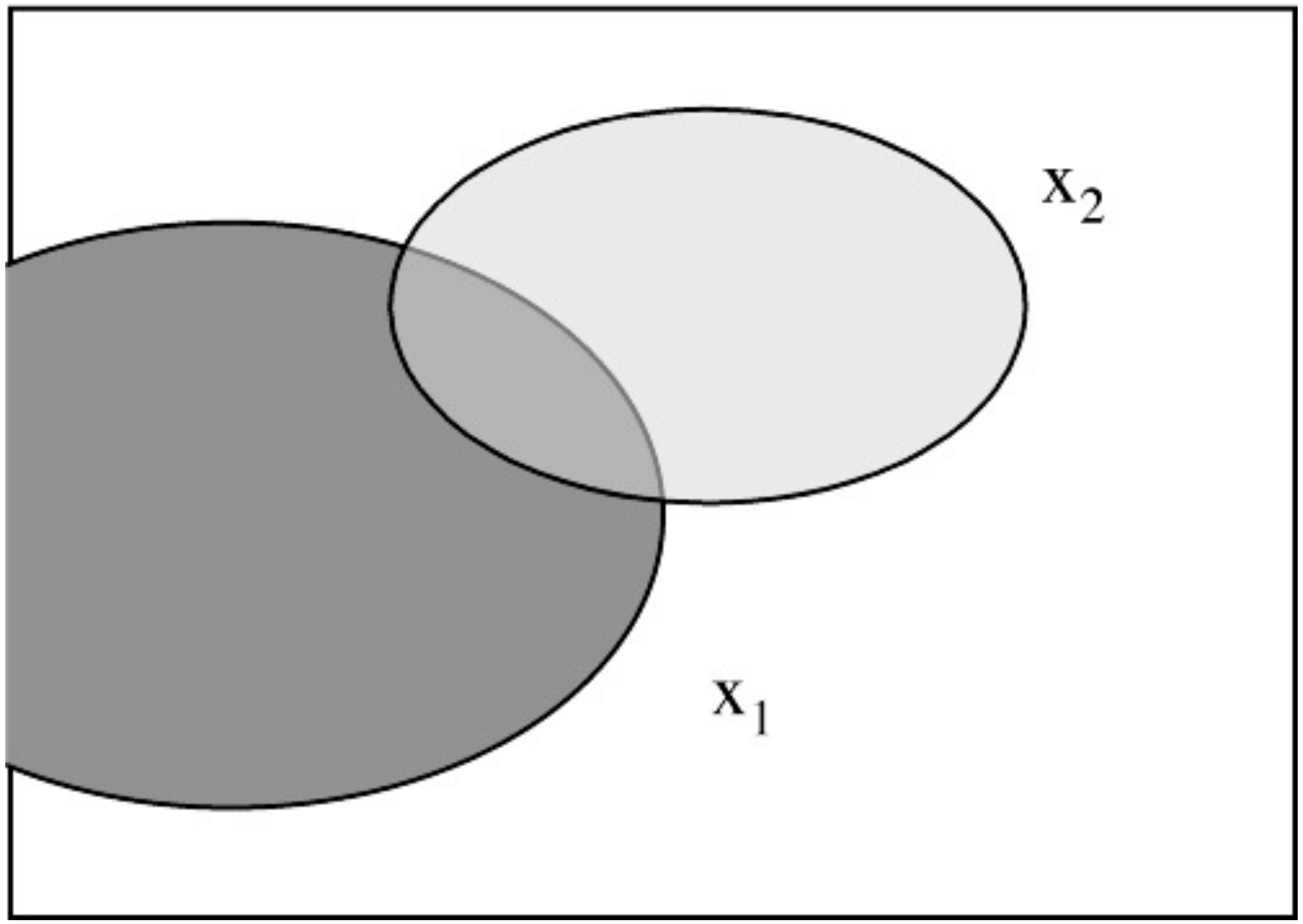

$$R^2 = r^2_{yx1} + r^2_{yx2}$$

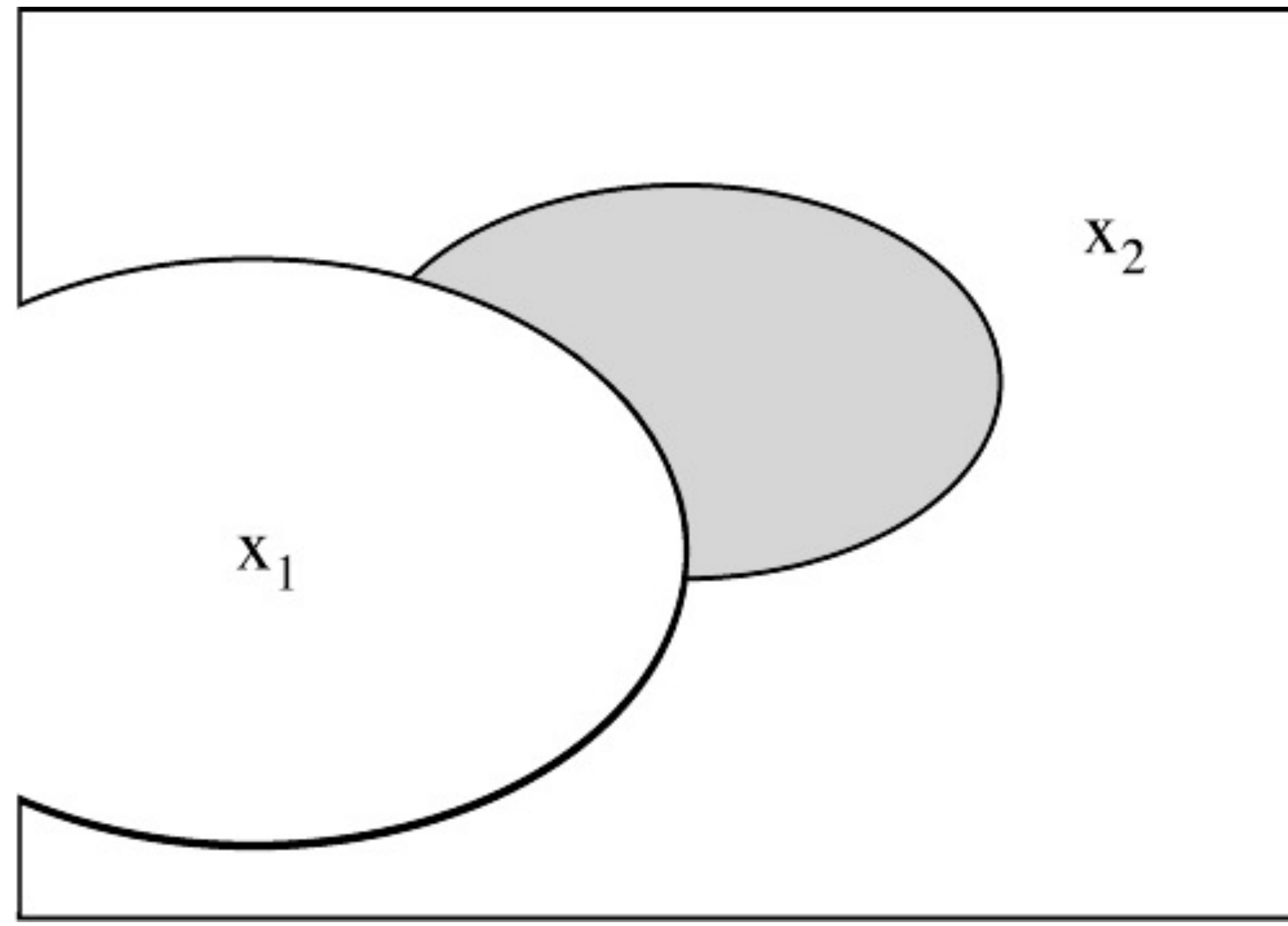
*Adding uncorrelated variables increases $R^2$*

$\mathrm{SS}(gender) = 337$, $\mathrm{SS}(gender|Nsuper) = 212$,

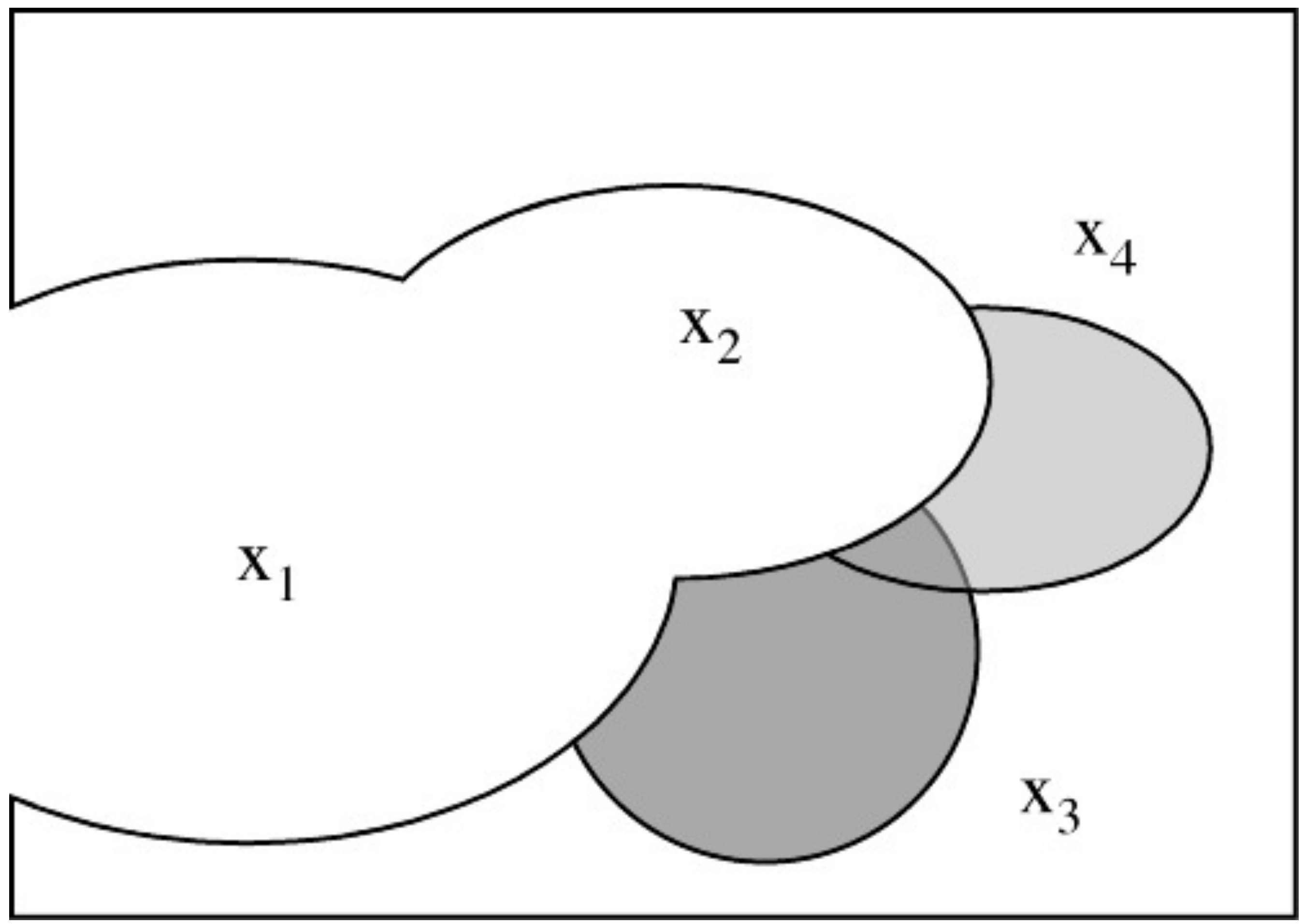
$\mathrm{SS}(Nsuper) = 1494$, $\mathrm{SS}(gender, Nsuper) = 1706$

# Coefficient of Partial Determination

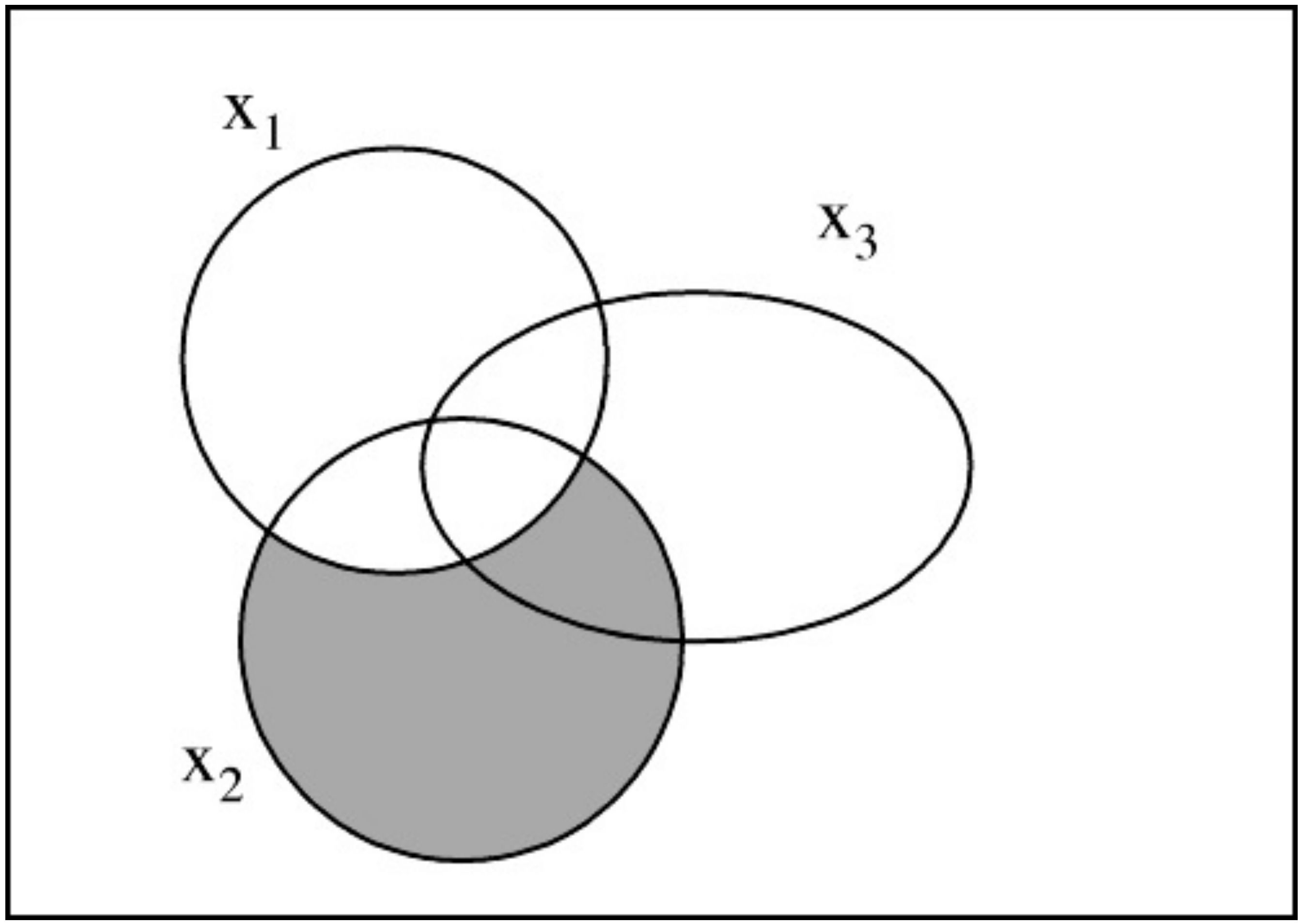


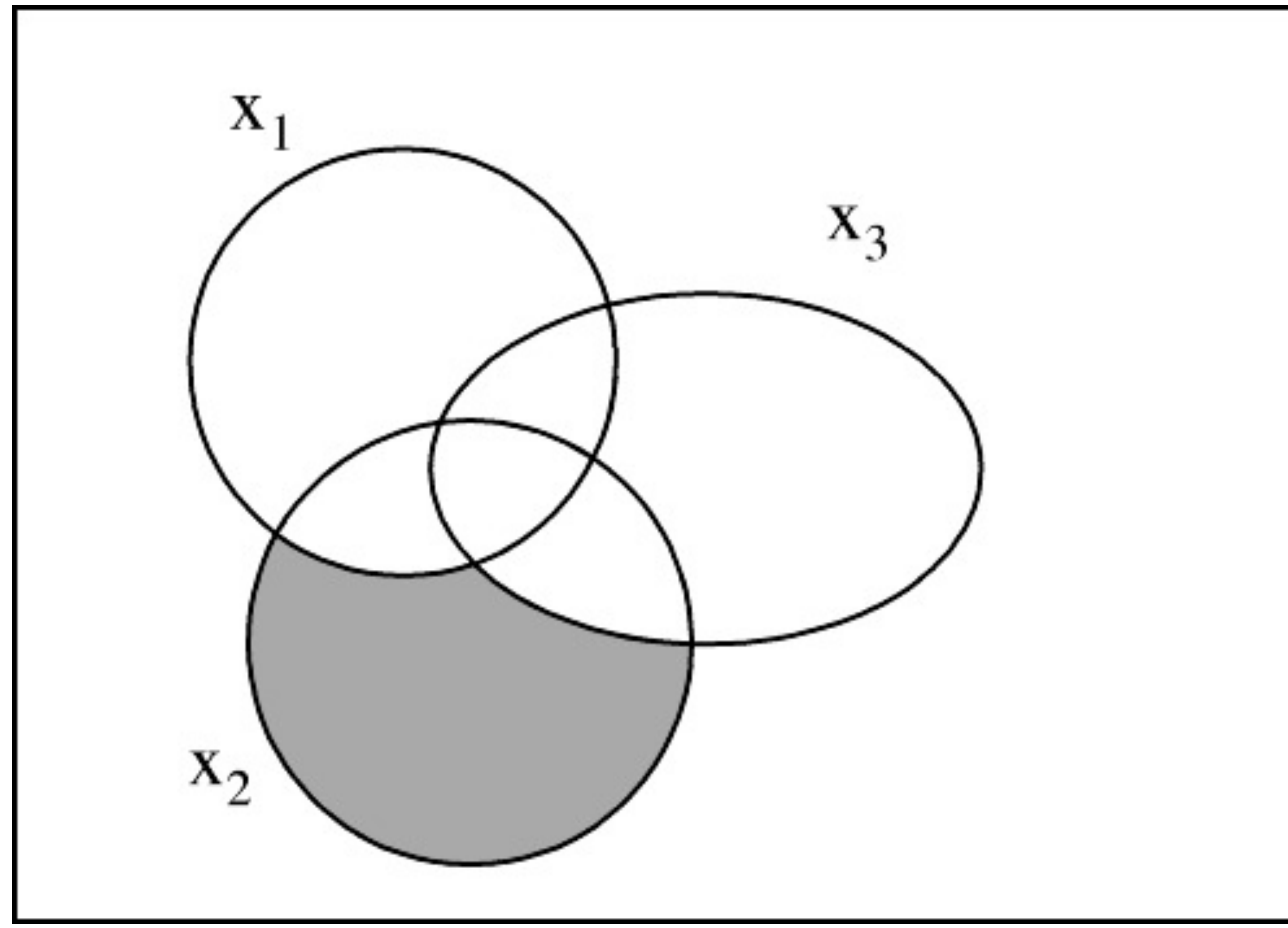
# Multiple Partial Correlation



# Sequence Matters

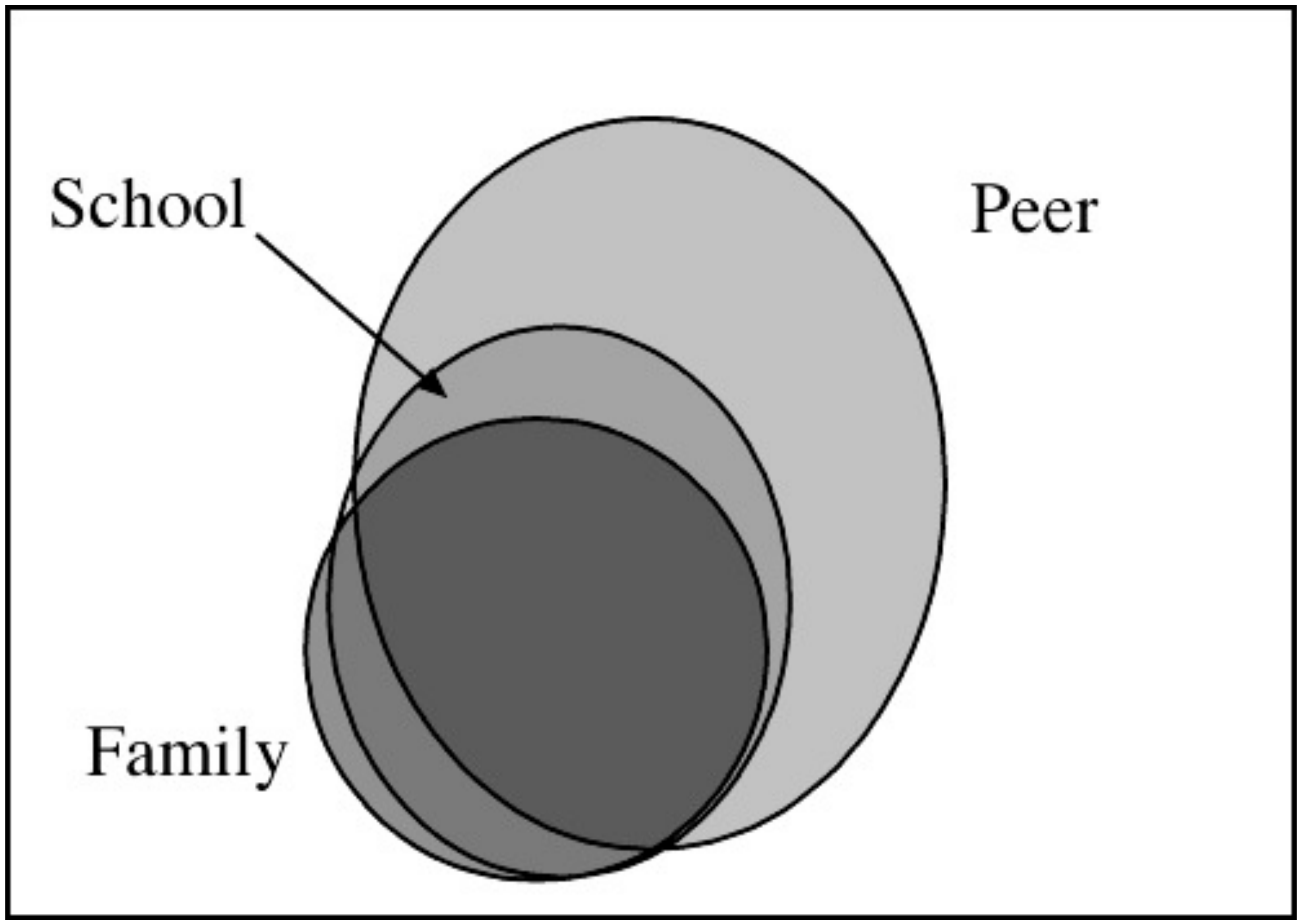



*$x_2$ enters second*




*$x_2$ enters last*

# Multicollinearity




*F-test =*

*area covered / area not covered*

*= significant*

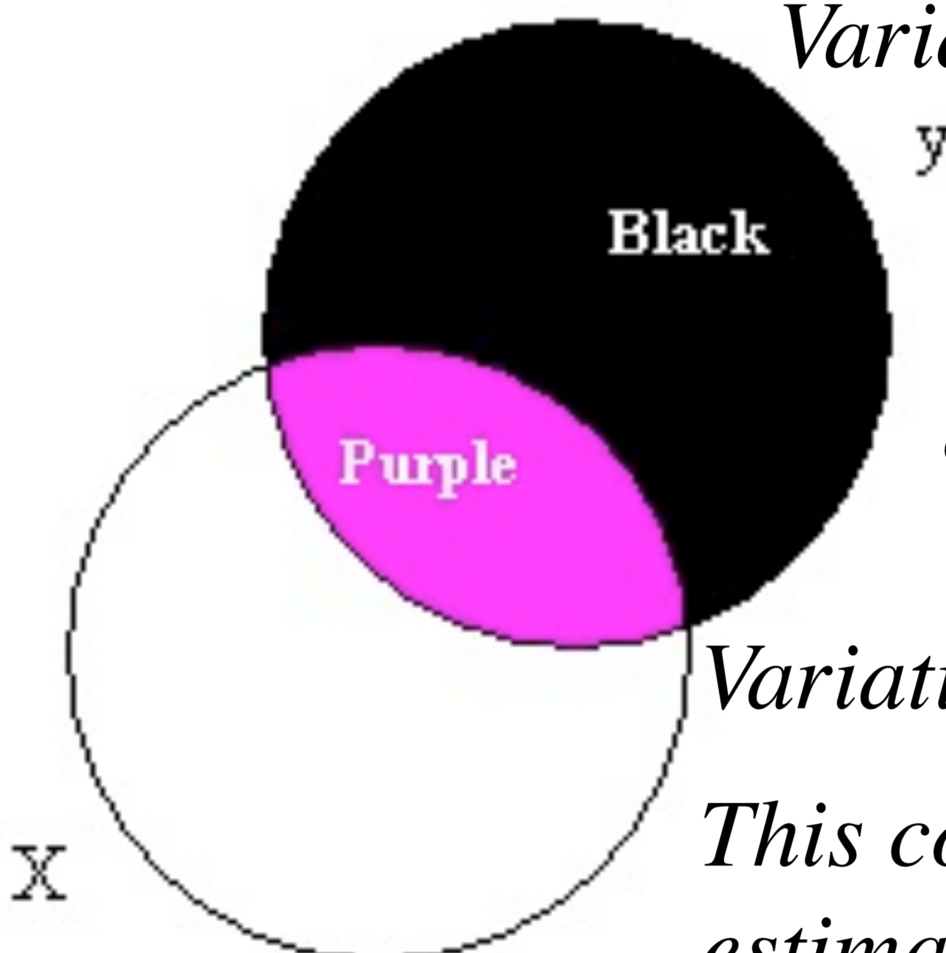
*t-test =*

*additional area covered by last variable / area not covered*

*= not significant for any variable*

# Alternative Explanation: Estimation

*Variation in*




*Variation in y that cannot be explained by X (error variance)*

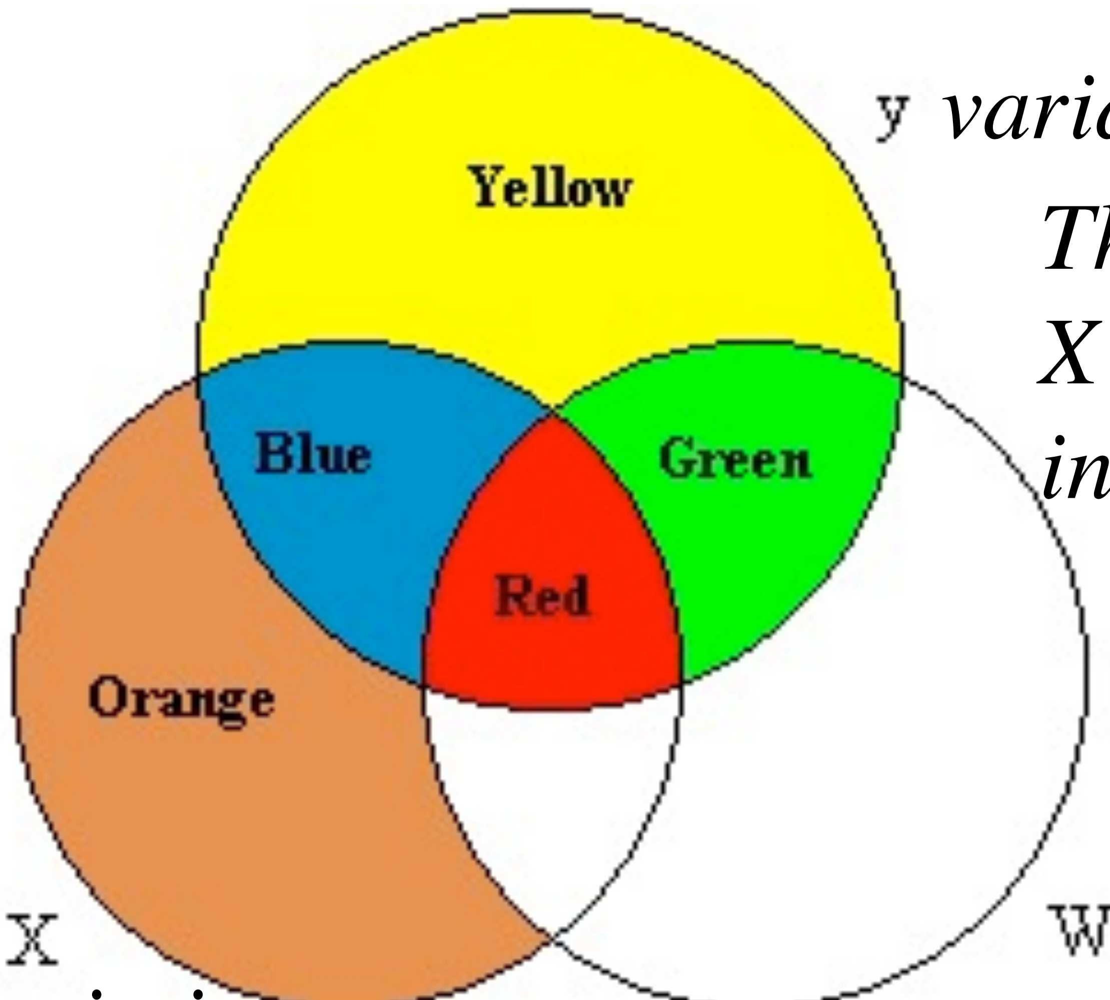
*Variation that y and X have in common.*

*This co-movement is used in OLS estimation of the slope coefficient $b_X$*

*Larger overlap => smaller $b_X$ variance*

*Variation in*

# Multiple Regression




y *variation*

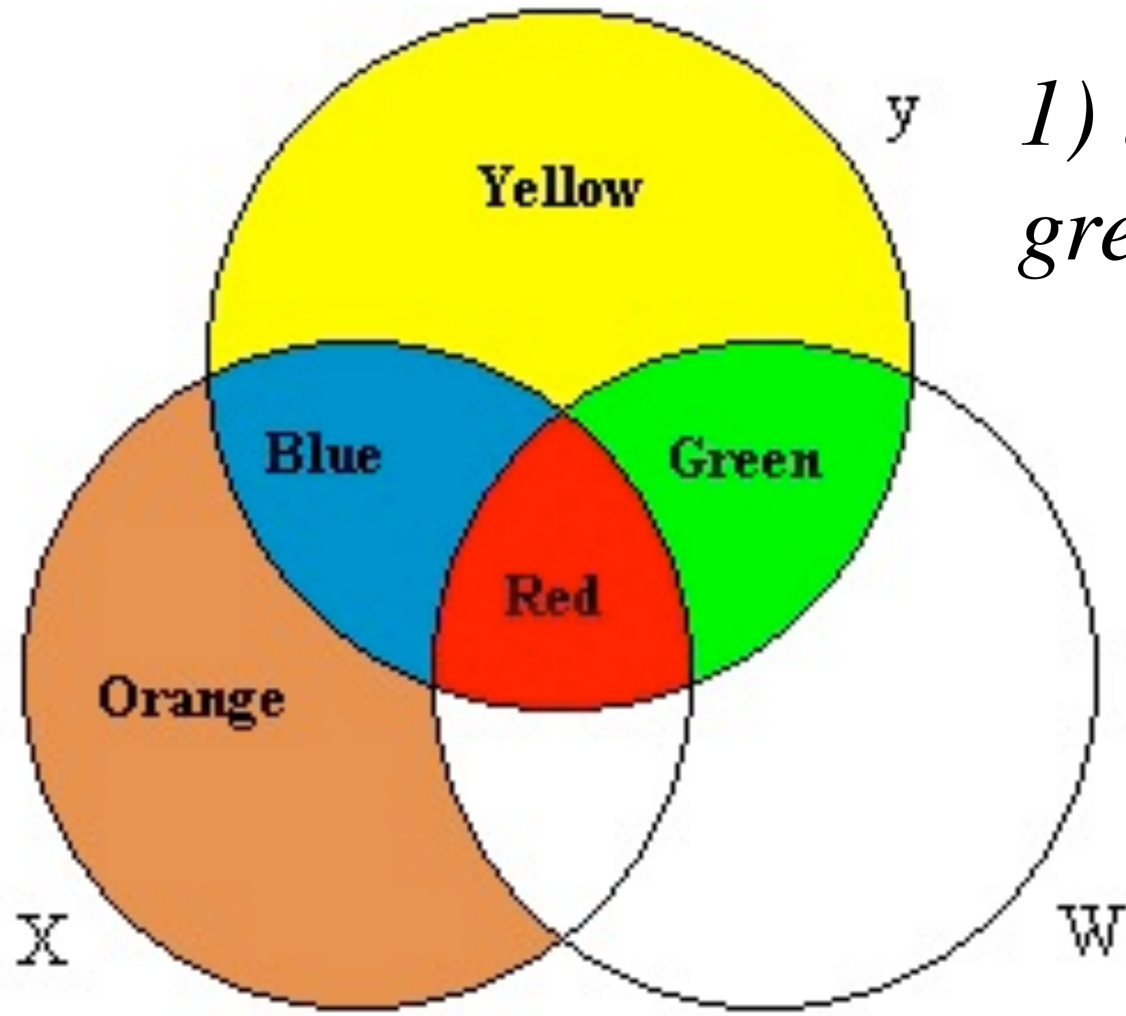
*The overlap between the y circle and the X and W circles represents the variation in y explained by variation in X and in W.*

W *variation*

X *variation*

*The ratio of the overlap (the blue + red + green area) to the y circle is interpreted as the $R^2$ from regressing y on X and W.*

# Estimating $b_X$ and $b_W$

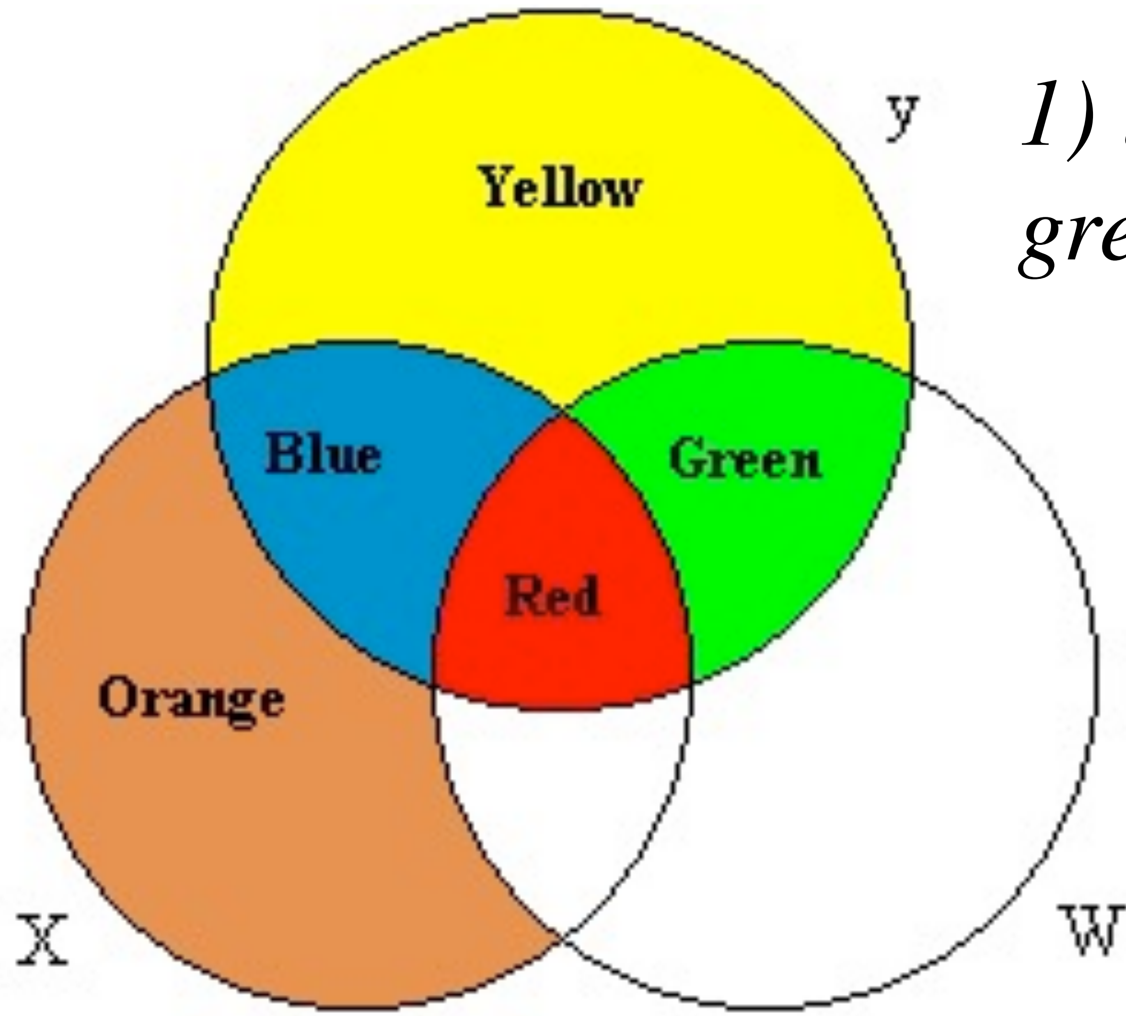



*1) use blue + red to estimate* $b_X$ *and green + red to estimate* $b_W$

*2) throw away red, use blue to estimate* $b_X$ *and green to estimate* $b_W$

*3) divide red somehow*

# Estimating $b_X$ and $b_W$




*1) use blue + red to estimate $b_X$ and green + red to estimate $b_W$*

*2) throw away red, use blue to estimate $b_X$ and green to estimate $b_W$*

*3) divide red somehow*

*Excluding the red area will result in unbiased $b_X$ and $b_W$ estimates*

$$b_X = (X^{*\prime}X^*)^{-1}X^{*\prime}y^* \text{ where } y^* = M_w y \text{ and } X^* = M_w X$$

# Multicollinearity

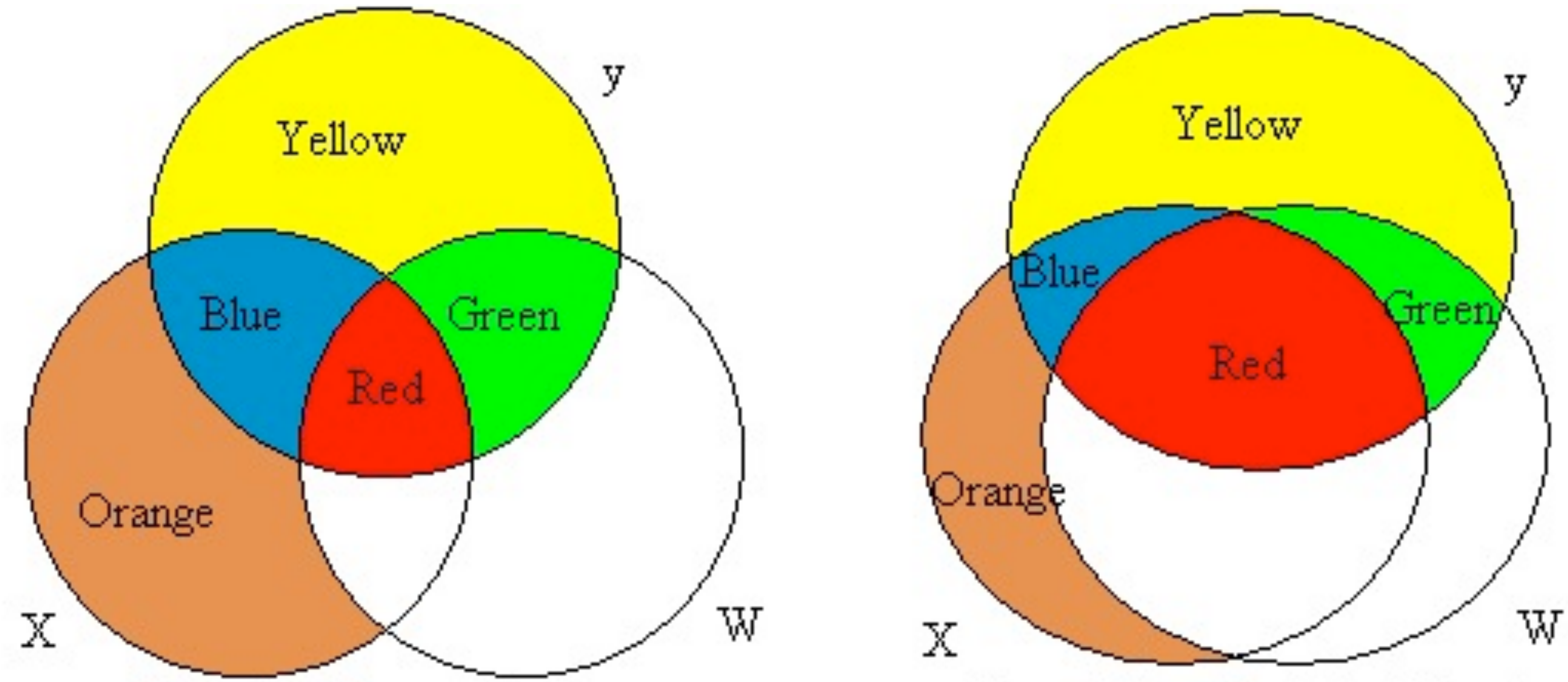


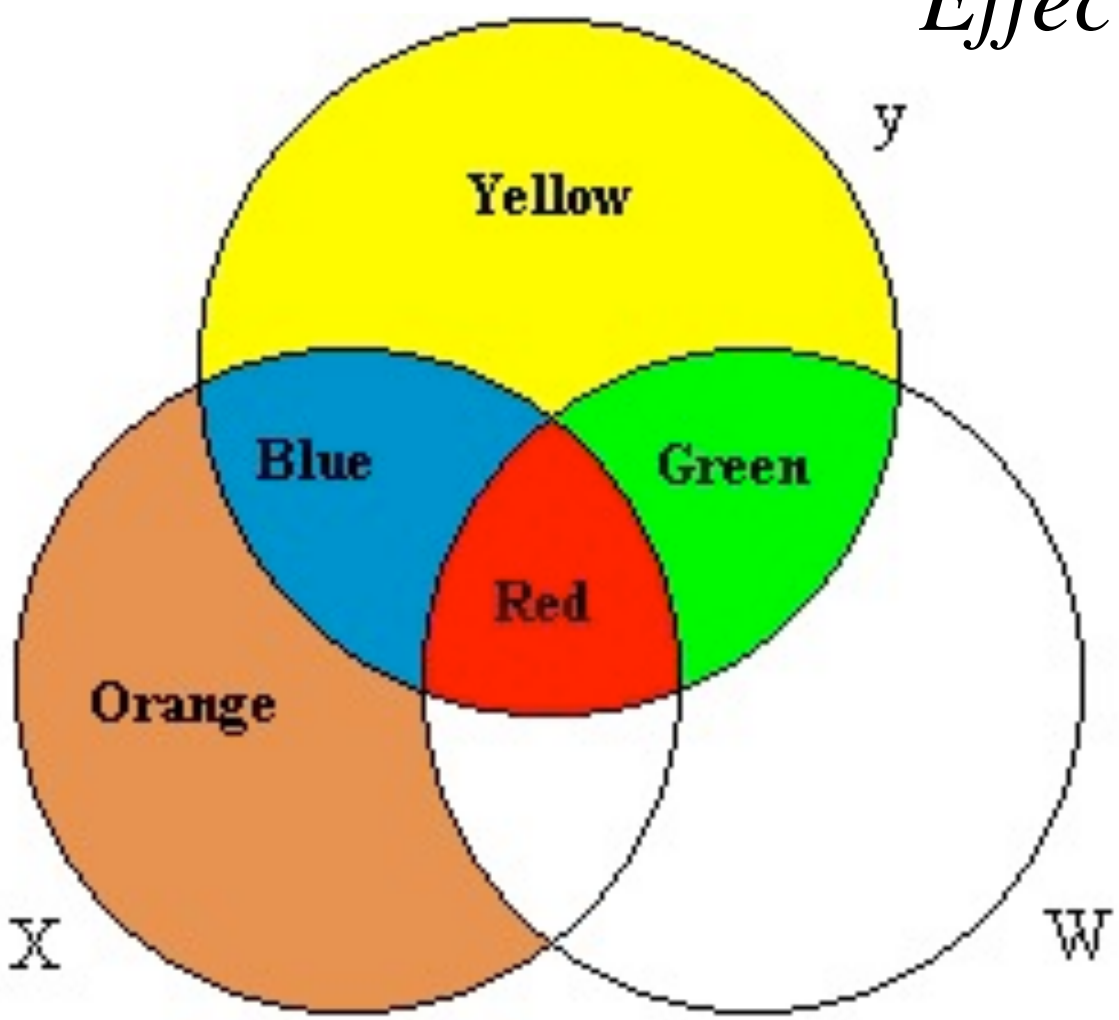

Figure 3a Modest collinearity

Figure 3b Considerable collinearity

*Effect on: bias and variance of $b_X$ and $b_W$.*

*What is the effect of perfect collinearity?*

# Omitted Variable




*Effect on: bias and variance of $b_X$ and $b_W$.*

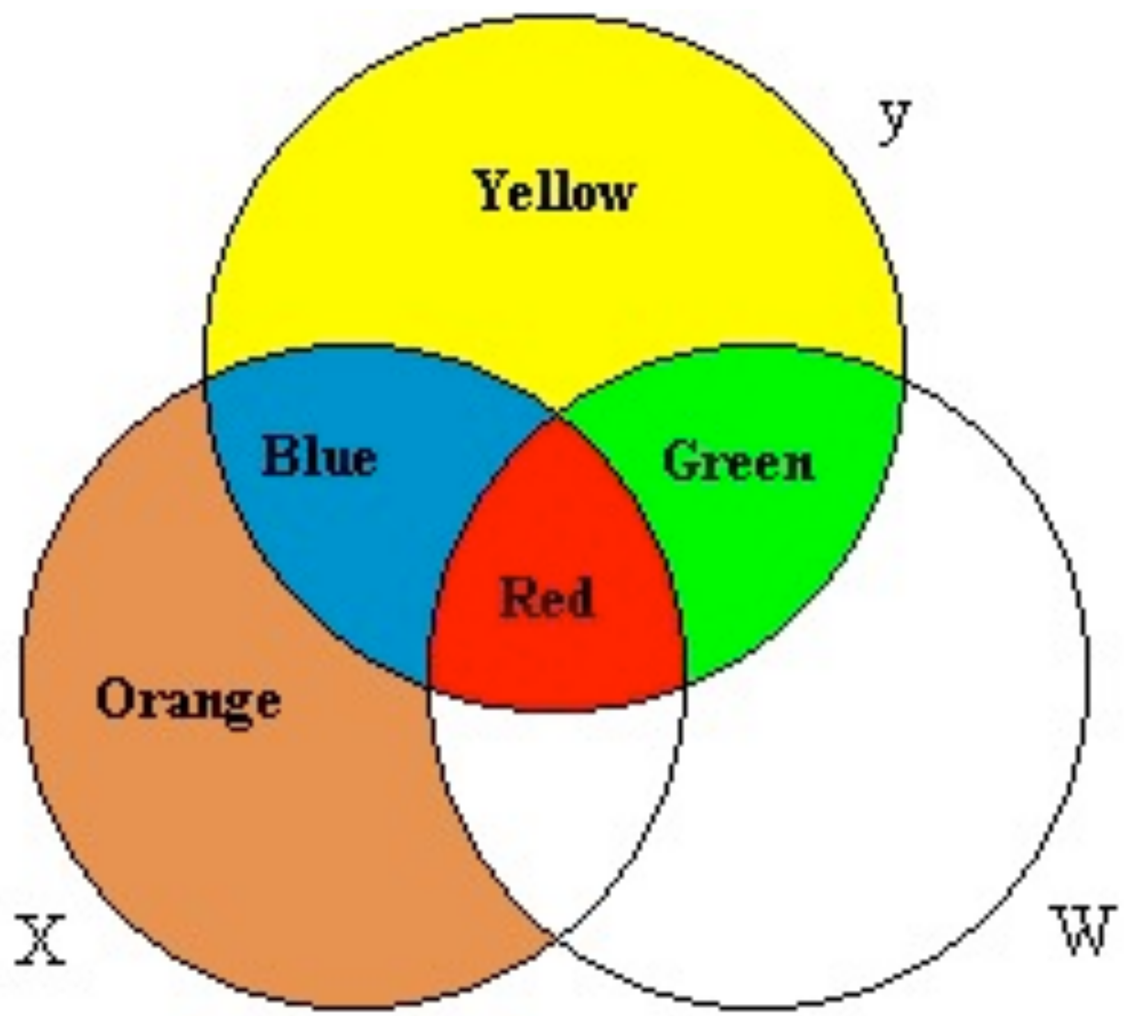
*MSE = (bias)$^2$ + variance*

*=> drop highly collinear variable*

*Effect on: bias of error variance*

*What if W is orthogonal to X?*

# Application: Detrending Data




*1) regress y on X and W*

*2a) regress y on W, save residuals $r_y$*

*2b) regress X on W, save residuals $r_X$*

*2c) regress $r_y$ on $r_X$*

*Compare results from 1) and 2)*