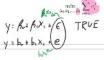
## • R-square doesn't week. • Why it doesn't week. • Why it doesn't week. • The Fix: R<sup>2</sup>

We should no longer use  $R^0$  in the multiple regression model. This is because when we add a new variable to the model,  $R^0$  must always increase (or at best stay the same). This means that we could keep adding "junk" variables to the model to arbitrarily inflate the  $R^0$ . This is not a good property for a "measure of fif" to have. Instead, we will use "adjusted Resparsel", denoted by  $R^0$ .

## 6.5.1 Why $\mathbb{R}^2$ must increase when a variable is added

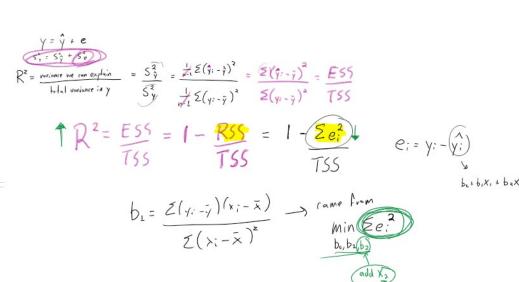
To see why  $R^2$  must always increase when a variable is added, we begin by looking again at the formula:





When we add another X variable, the dimension has been  $\sum_{i=1}^n x_i^2$  mass get sine left O(S) picks the values for the fits so that the sum of scurred vertical distances are minimized. If we give O(S) another option for minimizing those distances, the distances, the distances, the distances, the distances where O(S) can be considered as O(S). The conjugate O(S) is a variable means O(S) because a value of O(S) increases. The only way that E' stars the same is O(S) choices a value of O(S) increases.

As an example, let's try adding a nonsense variable to the house price model, pandon due rolls. Using R, 1728 die rolls are simulated (to match the house price sample size of n=1728), are recorded as a variable Dice, and added to the regression. Notice the difference in "Multiple It-squared"  $(R^2)$  and "Adjusted R-squared"  $(R^2)$  between the two regressions:



The variable Dice has no business being in the regression of house prices, and we fail to reject the null hypothesis that its effect is zero, yet the  $R^2$ increases. However.

## 6.5.2 The $\bar{R}^2$ formula

Adjusted R-squared  $(R^i)$  is a measure of fit that can either increase or decrease when a new variable is added.  $R^2$  is a slight siteration of the  $R^2$ formula. It introduces a penalty into  $R^2$  that depends or the number of Xvariables in the nuclei. (Remember that the quantum of X in the model is denoted by  $E_i$ )

• 
$$R^2$$
 always T when a variable is added  
• Why?  $R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{8e^2}{TSS} = \frac{1}{TSS} = \frac{1}{TS$ 

$$S_{\gamma}^{2} = \frac{\sum (y_{i} - \overline{y})^{2}}{n = 1}$$

$$e_{iz} y_{i} - \widehat{y}_{i}$$

$$e_{iz} y_{i} - \widehat{y}_{i}$$

$$e_{iz} y_{i} - \widehat{y}_{i}$$

$$e_{iz} y_{i} - \widehat{y}_{i}$$

$$\sum (y_{i} - \overline{y})^{2} / (n - 1)$$

$$\sum (y_{i} - \overline{y})^{2} / (n - 1)$$

$$\sum (y_{i} - \overline{y})^{2} / (n - 1)$$

$$R^2 = 1 - \frac{RSS / (n - k - 1)}{TSS / (n - 1)}$$
(6.4)

The  $R^2$  formula is such that when a variable is added to the model, k goes up, which tends to make  $R^2$  smaller. We know from the previous discussion, however, that whenever a variable is added, RSS must decrease. So, whether or and  $R^2$  increases to discussive depends on whether the new variable improves the fit of the model chough to beat the penalty incurred by k.

The justification for the (q, k-1) and (q-1) terms is from a degrees of freedom rarrow ion. How many things do we have an estimate before we can calculate RSSR k-1  $2\kappa$  must first be estimated before we can get the OLS residuals, and RSS. If you want to use RSS be resmuching else (such as a measure of fit), we recognize that we don't have a pieces of information left in the sample, we have (n-k-1). A similar argument can be made for the  $(\kappa-1)$  term in equation 6.4.