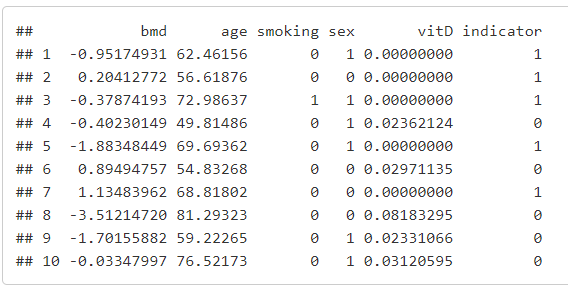
Missing indicator method

* Missing observations on one independent variable is set to fixed value
* One of the confounders (VitD) has missing values – all set to 0 and adds a missing indicator variable to VitD to indicate whether it was missing (1) or observed (0).



Linear regression model goes from

To



The linear model can be written two different ways as it differs for whether the vitD was observed or not.

**Observed** = indicator variable = 0

* It crosses out as beta5 \* indicator (0) = 0
* Only unbiased if MCAR (like CC)

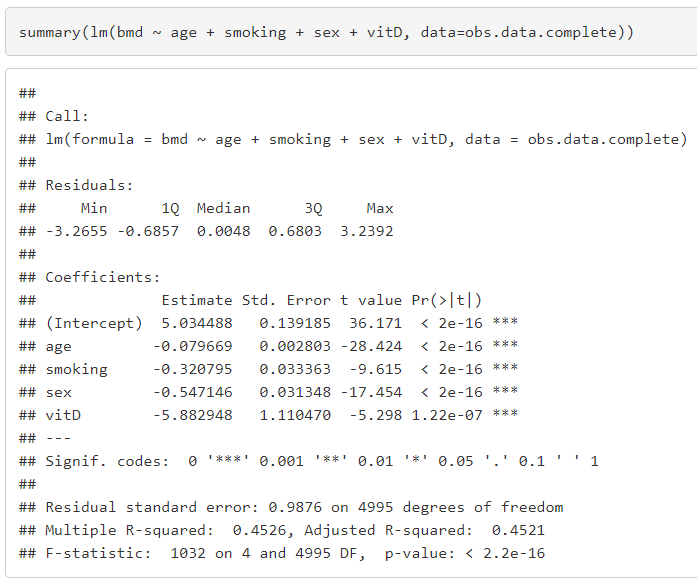
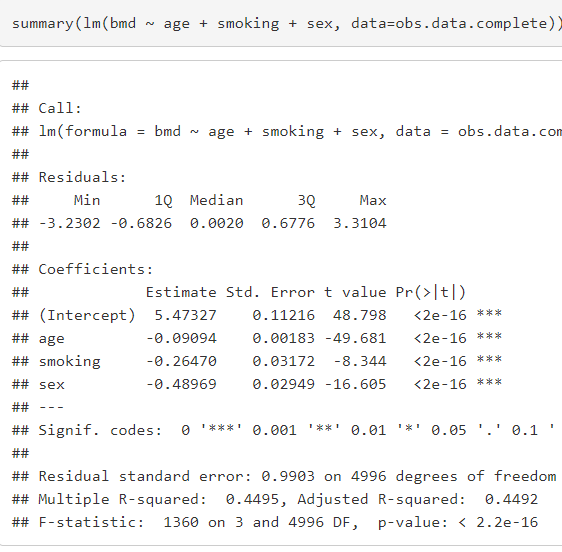
**Unobserved/Missing**= indicator variable = 1 but missing value filled with 0

* Beta4\*vitD = 0 + beta5\*indicator = Beta5
* Usually biased as part of confounder information is missing & unadjusted

The regression model estimates for the **individuals with missing VitD**, are then thus **not adjusted for vitD levels** and are therefore only unbiased if a) vitD level is not correlated to the other determinants and b) missingness of vitD level is not related to the outcome conditional on the other determinants.

Missing Indicator method is a weighted average of CCA on one part of the data (observed) and an analysis where an important cofounder is dropped (unobserved). Always biased even under MCAR.

**Verify if vitD is a confounder**

* Use linear model with all confounders and linear model without confounder in question (vitD).
* Check if estimates change drastically for other confounders

Age changes quite a bit when vitD is not included = vitD is related to the outcome

**Question 3.** Would the complete case analysis (CCA) estimate of the age effect be biased? If you compare the CCA to an analysis on the complete data (so without missings), would the standard error of the age effect accoring to CCA be larger, equal, or smaller?

*# The point estimates of the age effects are reasonably similar as they should be (the differences are random)*

*# the se is larger for the complete case analysis (as expected)*

The estimates for the effect of age would be biased after using missing indicator method. Easily tested by adding very large number instead of 0 which shows that the effect changes.

Shows that even under MCAR there is bias by using this method. Even worse than complete case analysis.

**Question 6.** What do you think of the beta estimate and p-value of the effect of the indicator variable itself? What does this finding actually tell us and what does it mean when someone wants to use this model in new individuals to make BMD predictions?

*# Significance of the indicator variable essentially means that the intercept needs some adjustment for those with missing vitD data.*

*# In other words, the prevalence of the outcome is not the same among those with and without missing vitD data*

*# From a clinician's point of view, note that it is somewhat of a strange message that missing some of the data is predictive of the outcome.*