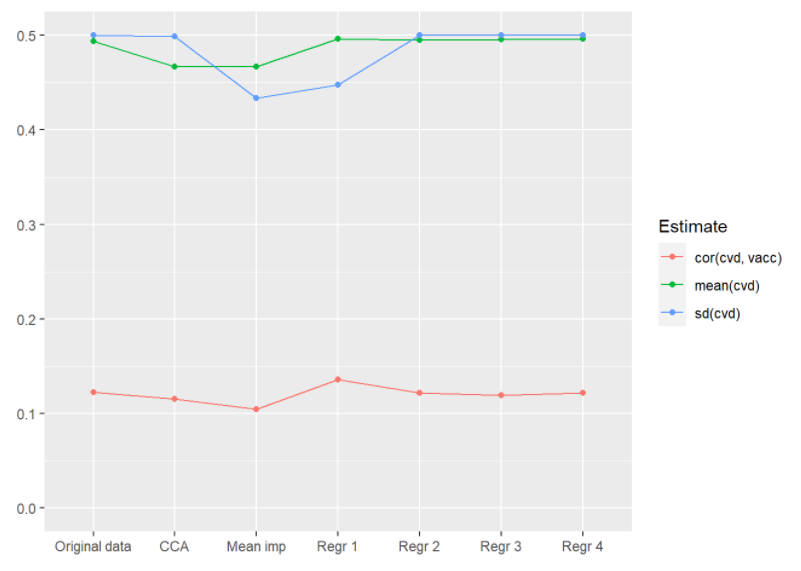
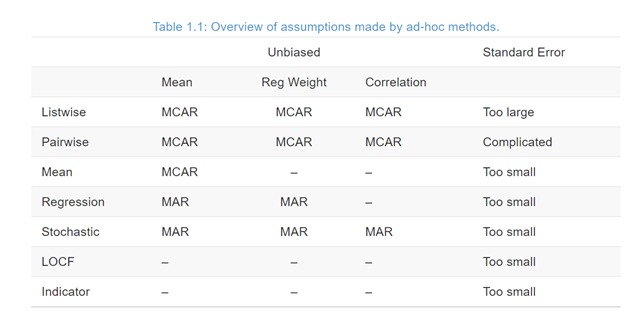
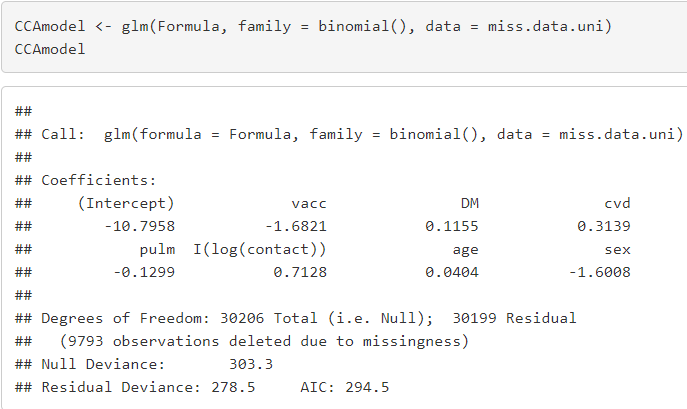
**Missing Data practical II**



****

**Simple Methods**

1. Complete Case Analysis
2. Dropping predictors with missing values
3. Mean imputation
4. **Complete Case Analysis**

* Just omit which is default for a lot of packages

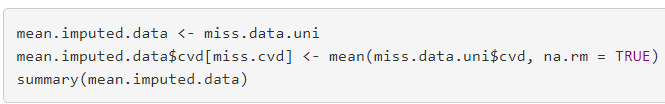
**Question 2.** How many participants were used for estimating the parameters of CCAmodel?

30206 +1 = 30207 patients OR 8 parameters = residual degrees of freedom (30199)

1. **Drop Covariates**

* Instead of deleting all cases with missing values we can delete the covariate with missing values (in this case only cvd)
* **Key Problem**: we don’t adjust for the confounder anymore which might introduce bias in the odds ratio

1. **Mean Imputation (equal to intercept only)**

* Replace missing value by the mean of their observed values (=imputed dataset)
* Mean stays the same
* Standard deviation is smaller in mean imputation
* Correlation weaker because imputed values do not covary with other variables

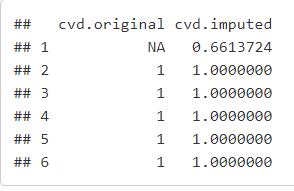
**Question 10.** Estimate the adjusted odds ratio for influenza vaccination in the imputed data.

Associations are distorted and standard error doesn’t matter because it’s biased

**Regression**

1. **Predict**

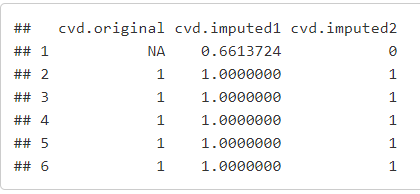
Instead of imputing same value for all NAs develop prediction model

* Cvd = binary = logistic regression analysis
* values can be between 0 and 1
* Sd still too small (no uncertainty in imputed values)
* Correlation increased

If there was only values missing for one variable this could gibe good estimates for regression coefficients but no uncertainty in imputed values = may lead to small sd

1. **Predict + noise**

* Improve predictor model by adding noise to predicted value
* For binary outcomes model-based noise by generating the value from a binomial distribution of one ‘trial’
* Imputed values are now 0s or 1s again
* Problematic when samples too small

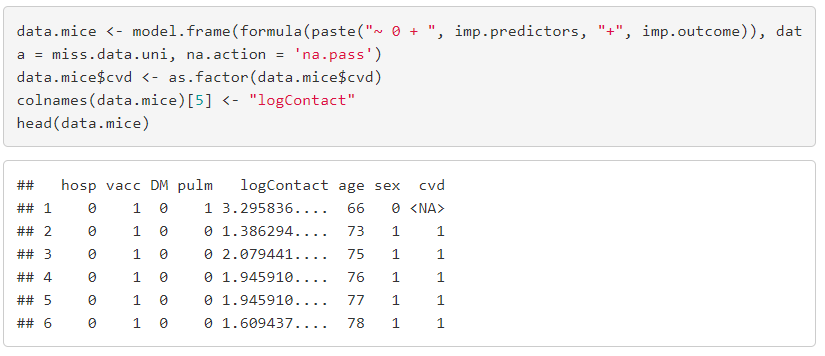


1. **Predict + noise + parameter uncertainty**

* The values of the parameters (intercept & slope) used to make a prediction for the missing values are estimates themselves
* Rather than directly using the estimated regression coefficients to create predicted values we add variability by relating to the standard errors of the coefficients
* Add random draw of regression coefficients rmt()

1. **Multiple Imputation**

* We’ve been treating all data (including imputed) as though it was observed but didn’t add uncertainty from imputation (the model)
* To preserve the uncertainty from imputation we need to repeat the sampling procedures
* Good values will not vary much across imputations
* Bad values will vary a lot across imputations

1. **Multiple Imputation in mice()**