

**Escolha do conjunto de dados** - os conjuntos de dados devem ser escolhidos seguindo os seguintes critérios:

- Escolher por volta de 10 variáveis quantitativas CONTÍNUAS e que elas não meçam as mesmas coisas (por ex.: número de pessoas com ensino médio completo e porcentagem de pessoas com ensino médio completo - escolher só uma delas)
- IMPORTANTE: lembrar que  $n \gg p$  (muito mais observações do que variáveis)
- Para definir o conjunto de dados:
  - 1 - Ir no site <http://imrs.fjp.mg.gov.br/> e > gerar consultas > selecionar período (ano)
  - 2 - escolher o ano de 2010 (como foi o ano do Censo, há muitas variáveis disponíveis)
  - 3 - definir o estado de MG ou alguma mesorregião de MG e selecionar todos os municípios
  - 4 - definir que tipo de indicadores estudar (há dados sobre saúde, educação, segurança pública, meio ambiente etc.) e assuntos diferentes podem ser estudados ao mesmo tempo
  - 5 - selecionar o tipo de indicador e selecionar as variáveis de interesse
  - 6 - ir em concluir e depois em Download csv
  - 7 - abrir o arquivo usando o Calc (programa de planilhas do LibreOffice) ou o Excel mesmo
  - 8 - mudar os nomes das variáveis (na primeira linha, a de cabeçalho) para nomes curtos, sem acentos ou espaços (ver as planilhas que montei no exemplo - “seg\_sul” e “educ\_sul”)  
Obs.: antes de mudar os nomes das variáveis anotar em algum lugar quais as variáveis escolhidas no site
  - 9 - as duas primeiras colunas devem conter o código do IBGE do município e o nome do município (e as variáveis devem se chamar “ibge7” e “mun”)
  - 10 - salvar as planilhas como csv com nomes também curtos e sem acentos ou espaços

Parte 1 - Análise preliminar dos dados e discussão sobre o que foi observado:

- a) Mesorregião escolhida (ou o estado de MG), variáveis utilizadas e explicação das variáveis
- b) Resumo estatístico das variáveis
- c) Vetor de médias, matriz de covariâncias e matriz de correlações entre as variáveis
- d) Diagramas de dispersão entre os pares de variáveis
- e) Análise da normalidade multivariada do conjunto de dados