

# Atividade 01: Nossas Perguntas

Caynan Sousa

16/11/2016

## Sobre os dados

Iremos analisar os dados do controle acadêmico da UFCG, nossos dados contemplam os períodos de 2002 a 2014 e todos os cursos da UFCG.

Iremos importar esses dados para um `dataframe`, que irems chamar de `alunosUFCG`.

```
colnames(alunosUFCG)
```

```
## [1] "Matricula"      "Cod_Curso"      "Cod_Disciplina"
## [4] "Cod_Evasao"     "Nome_Curso"     "Periodo"
## [7] "Nome_Disciplina" "Creditos"       "Departamento"
## [10] "Media_Disciplina" "Situacao"       "Centro"
## [13] "Campus"         "Creditos_Curso" "Tipo"
## [16] "Periodo_Ingresso" "Periodo_Relativo"
```

Como podemos ver todos os dados que temos dos alunos é bem auto-explicativo, a parter, talvez dos seguintes campos:

- **Cod\_Evasao:** 0 se o aluno *não* evadiu ao fim do primeiro periodo e 1 se ele evadiu.
- **Credito\_Curso:** Total de créditos necessários para o aluno concluir o curso.
- **Periodo\_Ingresso:** período que o aluno entrou no curso.

## Distribuição Notas por Curso

```
by_curso <- group_by(alunosUFCG, Cod_Curso)

grade_by_curso <- summarise(by_curso,
  name = first(Nome_Curso),
  media_disciplina = mean(Media_Disciplina, na.rm = TRUE))
```

```
# Disciplinas com Menor Média
```

```
head(arrange(grade_by_curso, media_disciplina))
```

```
## # A tibble: 6 × 3
##   Cod_Curso      name media_disciplina
##   <int>      <fctr>      <dbl>
## 1  11105110  FÍSICA (LIC) - D      2.752181
## 2  15125100 ENGENHARIA DE ALIMENTOS - D      3.901301
## 3  11134100  ESTATÍSTICA - D      4.019135
## 4  51105210  FÍSICA (LIC) - N      4.274814
## 5  21105210  FÍSICA (LIC) - N      4.343119
## 6  11105150  FÍSICA (BAC) - D      4.424567
```

```
# Disciplinas com Maior Média
```

```
head(arrange(grade_by_curso, desc(media_disciplina)))
```

```
## # A tibble: 6 × 3
##   Cod_Curso      name media_disciplina
```

```
##           <int>           <fctr>           <dbl>
## 1  31350100      DIREITO - M      8.385194
## 2  31350300      DIREITO - V      8.367123
## 3  21205100     MEDICINA - D      8.342126
## 4  31350200      DIREITO - N      8.128079
## 5  12205100     MEDICINA - D      8.076255
## 6  13311150  MÚSICA (BAC) - D      7.991890
```

So posso pensar que talvez seja pelo fato de que os cursos de Medicina e Direito, sejam um dos mais concorridos na UFCG, portanto os alunos de tais cursos provavelmente tem uma melhor base dos conteúdos do ensino médio além de provavelmente já estarem acostumados a estudar de forma regular e constante.

E o curso de Física, é um dos cursos com maior evasão na UFCG, o que pode ser justificado por diversos motivos, desde uma falta de conhecimentos básicos por parte dos alunos, como uma fraca metodologia de ensino por parte dos professores desse departamento.

## Comparação Periodos do Curso de Ciências da Computação

Vamos filtrar nossos alunos, para termos apenas os alunos do curso de computação.

Tendo em mãos os alunos de computação, podemos agrupar por disciplina e então calcular a média de cada disciplina e agrupar tais resultados por periodo; Assim obtemos um conjunto de dados que podemos ordenar baseado em notas de cada periodo.

```
alunosCC <- filter(alunosUFCG, Cod_Curso == 14102100)

by_disciplina <- group_by(alunosCC, Cod_Disciplina)

grade_by_disciplina <- summarise(by_disciplina,
  name = first(Nome_Disciplina),
  periodo = first(Periodo),
  media_disciplina = mean(Media_Disciplina, na.rm = TRUE))

by_periodo <- group_by(grade_by_disciplina, periodo)

grade_by_periodo <- summarise(by_periodo,
  media_periodo = mean(media_disciplina, na.rm = TRUE))

head(arrange(grade_by_periodo, media_periodo))
```

```
## # A tibble: 6 × 2
##   periodo media_periodo
##   <dbl>      <dbl>
## 1  2002.2      5.873579
## 2  2002.1      5.909563
## 3  2003.1      5.996020
## 4  2010.2      6.328205
## 5  2013.2      6.393889
## 6  2008.2      6.610367
```

```
head(arrange(grade_by_periodo, desc(media_periodo)))
```

```
## # A tibble: 6 × 2
##   periodo media_periodo
##   <dbl>      <dbl>
## 1  2013.1      9.109837
```

```
## 2 2008.1      8.928571
## 3 2012.2      8.585714
## 4 2012.1      8.522328
## 5 2005.1      7.918457
## 6 2014.1      7.868548
```

Algo interessante é observar que temos mais turmas da entrada no meio do ano entre as turmas com pior rendimento, e um número maior de turmas da entrada do início do ano entre as melhores turmas.

## Correlacionando Mat. Discreta e Cálculo I

Temos que agrupar nossos alunos que cursaram Matemática discreta e Cálculo I, e então podemos calcular a correlação entre a nota obtida em cada uma das disciplinas.

Note que caso o aluno tenha cursado as disciplinas mais de uma vez, iremos considerar a mediana de tais notas no nosso calculo de correlação.

```
# Filtramos alunosCC para mostrar apenas alunos que cursaram Discreta
alunosCC.disc <- filter(alunosCC, Cod_Disciplina == 1109113)
alunosCC.disc <- select(alunosCC.disc, Matricula, Media_Disciplina)
alunosCC.disc <- group_by(alunosCC.disc, Matricula)
alunosCC.disc <- summarise(alunosCC.disc,
  mediana_disc = median(Media_Disciplina, na.rm = TRUE))

# Filtramos alunosCC para mostrar apenas alunos que cursaram Calculo I
alunosCC.calc <- filter(alunosCC, Cod_Disciplina == 1109103)
alunosCC.calc <- group_by(alunosCC.calc, Matricula)
alunosCC.calc <- summarise(alunosCC.calc,
  mediana_calc = median(Media_Disciplina, na.rm = TRUE))

#Unimos os DataFrames
alunosPair <- merge(alunosCC.disc, alunosCC.calc, by = "Matricula", all = FALSE)
alunosPair <- na.omit(alunosPair)

# Calculamos a correlação
cor(alunosPair$mediana_calc, alunosPair$mediana_disc)
```

```
## [1] 0.70494
```

Temos então uma correlação fortemente positiva (+0.7).