

### 1ª Lista de Exercícios

#### Providencias preliminares:

1. Instale o pacote “readxl” que é necessário para importar os valores de um arquivo com extensão .xls ou .xlsx.
2. Importe o pacote através do comando “library(readxl)”.
3. Baixe, neste link, a base de dados que será utilizada. Essa base de dados apresenta a renda média por hora trabalhada (a preços de 2008) para homens (1) e mulheres (2) com diploma universitário (*College Graduates*) de 1992 a 2008 nos Estados Unidos.
4. Importe a base de dados para o Rstudio com o nome “cps”.

### QUESTÕES

1. Seja  $Z \sim N(0,1)$ , isto é, seja  $Z$  uma variável aleatória que segue uma Distribuição Normal de média 0 e variância 1. Calcule o valor da função densidade de probabilidade no ponto  $z = 3$ . Dica: pesquise sobre o comando “dnorm”.
2. Seja  $Z \sim N(0,1)$ . Calcule  $P(|Z| \leq 1.64)$ , isto é, a probabilidade de que o valor absoluto de  $Z$  seja menor ou igual a 1.64. Dica: pesquise sobre o comando “pnorm”.
3. Seja  $Y \sim N(5,25)$ . Calcule o quantil 99% dessa distribuição. Isto é, encontre o valor de  $y$  tal que os valores acima de  $y$  apresentam uma probabilidade de apenas 1% de serem observados. Dica: pesquise sobre o comando “qnorm”.
4. O que acontece se o *output* (resultado) da função “qnorm” servir de *input* (entrada) para a função “pnorm” e vice versa? O que isso demonstra sobre a relação entre as funções?
5. Seja  $X \sim t_{10000}$ , isto é, seja  $X$  uma variável aleatória que segue uma Distribuição t de Student com 10000 graus de liberdade. E seja  $Z \sim N(0,1)$ . Calcule o quantil 95% de ambas as distribuições. Dica: pesquise sobre o comando “qt”.
6. Repita o exercício anterior, mas ao invés de 10000 graus de liberdade, utilize apenas 10. O que mudou? E o que isso significa?
7. Seja  $X \sim t_1$ . Gere 1000 números aleatórios dessa distribuição e atribua esses 1000 números à variável “x”. Calcule a média amostral dos valores atribuídos a “x”. Repita o procedimento algumas vezes. Considerando que a Distribuição t de Student é simétrica em torno do zero, o que explica que as médias amostrais de “x” são tão diferentes de zero? Dica nº 1: pesquise

sobre o comando “rt”. Dica nº 2: faça esse mesmo teste usando outros valores de graus de liberdade.

8. Considerando a base de dados (“cps”) baixada preliminarmente, teste a hipótese, a um nível de significância  $\alpha=0.05$ , de que a remuneração média pela hora trabalhada é superior a 23.50 \$/h. Qual foi o resultado do teste de hipóteses? Utilize como critério para o teste de hipóteses a Estatística t (t score).
9. Repita a questão anterior, mas agora utilizando um nível de significância  $\alpha=0.01$ . Qual o resultado do teste de hipóteses?
10. Repita as duas questões anteriores, mas ao invés de utilizar a Estatística t (t score) como critério para o teste de hipóteses, utilize o p-valor. Dica: parta do valor já obtido do t score e utilize o comando “pnorm”.
11. Obtenha os valores do t score e p-valor utilizando o comando “t.test”. Repare que o p-valor obtido através do “t.test” é ligeiramente diferente do valor obtido quando se utiliza o “pnorm”. O que explica essa diferença? Qual método seria mais adequado em amostras pequenas?
12. Explique para o que serve o comando “set.seed” no R.
13. Considere a amostra aleatória  $(X_i, Y_i) \sim N(0,1)$  para  $i=1, \dots, 100$ . Gere os respectivos vetores X e Y. Utilize um *seed* igual a 123 antes de resolver essa questão.
14. Calcule a variância de X. Calcule a covariância de X consigo mesmo. Explique porque os resultados foram iguais.
15. Calcule a covariância e a correlação entre X e Y. Qual é a diferença entre essas duas medidas?