**UNIVERSIDADE DE SÃO PAULO**

**FACULDADE DE ECONOMIA, ADMINISTRAÇÃO**

 **E CONTABILIDADE DE RIBEIRÃO PRETO**

**DEPARTAMENTO DE ECONOMIA**

# Ribeirão Preto, 1º semestre de 2012

# PROBABILIDADE E ESTATÍSTICA APLICADA II

# 1ª LISTA APLICADA – PARTE 2

# Responsável: Alexandre C. Nicolella

**6. Faça um histograma e um boxplot das seguintes bases de dados e/ou vetores: south, crime e aid [Dica: data() retornará os dados disponíveis para utilização]. Indique quais dados são distribuídos de forma assimétrica.**

**7. O vetor pi2000 contém os primeiros 2000 dígitos do π. Faça um histograma. O que você pode comentar sobre sua forma? Encontre a proporção de 1’s, 2’s e 3’s nos dados. Você pode encontrar a mesma proporção para todos os dígitos de 0 a 9?**

**8. Obtenha 100 valores oriundos de uma distribuição normal utilizando o comando rnorm(). Repita o procedimento e verifique a mudança de valores. Nomeie este objeto e calcule todas suas estatísticas descritivas (média, mediana, quartis etc.). Faça um histograma da distribuição com o eixo vertical contendo as probabilidades.**

**9. Repita todo o procedimento do exercício anterios, mas com 10.000 valores aleatórios. O que você pode verificar sobre suas estatísticas descritivas e distribuição de probabilidades? Supondo que você quisesse que os primeiros 100 valores do exercício anterior fossem os mesmo, ou seja, que os valores se originassem do mesmo processo gerador de dados (DGP), como você faria?**

**10. Análises Gráficas de Dados**

*Instale o pacote “AER” no R.*

*Nós vamos utilizar os dados “CPS1985” derivado de Berndt (1991).*

**(a)** Mostre as informações básicas da base de dados, com o nome de todas as variáveis e seu tipo (numérica, dummy)

**(b)** Mostre as primeiras 6 linhas e as últimas 6 linhas de observações

**(c)** Apresente os nomes e as estatísticas descritivas básicas de todas as variáveis da base

**(d)** Analisando a variável “salário”, apresente suas estatísticas descritivas, variância e desvio-padrão

**(e)** Apresente o histograma da variável “salário” com a frequência no eixo vertical

**(f)** Apresente o histograma da variável “log(salário)” com a probabilidade no eixo vertical. Compare com o ítem (e) e explique a diferença

**(g)** No ítem anterior, acrescente a linha de distribuição não-paramétrica Kernel no histograma, na cor laranja

**(h)** Analisando a variável “ocupação”, apresente suas estatísticas descritivas

**(i)** Mostre as mesmas estatísticas na forma de fração entre os tipos de ocupação, de tal forma que a linha some 1

**(j)** Obtenha **exatamente** o gráfico a seguir:



**(k)** Apresente um gráfico de pizza das profissões existentes, nas mesmas cores do gráfico de barras

**(l)** Apresente uma tabela de contingência das profissões por gênero masculino e feminino

**(m)** Apresente um histograma como o do ítem (j), mas com a separação em cada profissão entre homens e mulheres

**(n)** Mostre a correlação entre a variável “salário” e “educação”. Mostre também entre a variável “log(salário)” e “educação”

**(o)** Plote um gráfico de dispersão com “educação” no eixo horizontal e “log(salário)” no eixo vertical

**(p)** Apresente as estatísticas descritivas dos salários separadamente para homens e mulheres

**(q)** Apresente um boxplot dos salários separadamente por gênero na cor laranja. Há diferença considerável entre as distribuições? Há outliers?