**ANÁLISE DE DADOS APLICADA AO ESTUDO DO COMPORTAMENTO (PSE5986)**

**Docentes:** Emma Otta e Pedro Trindade

**PPG:** Psicologia Experimental – Instituto de Psicologia da Universidade de São Paulo

**Encontros virtuais síncronos:** 16, 23 e 30 de setembro; 7 e 14 de outubro de 2022 das 13:30 até às 16:30 hora de Brasília

**ORIENTAÇÕES GERAIS PARA AS TAREFAS**

A disciplina está organizada em cinco blocos (1- Natureza das Variáveis, Introdução ao R e banco de Dados; 2- Estatísticas Descritivas; 3- Figuras; 4- Modelo Regressivo Linear Simples; 5- Modelo Regressivo Linear Múltiplo), sendo que cada um deles têm aproximadamente 2h de aula gravada (teoria e prática). Enquanto estiver assistindo as aulas é recomendável que comande os códigos no RStudio em sua máquina para auxiliar na consolidação do conhecimento. Só é possível aprender programar, programando. Após cada um dos blocos de aula por favor execute as tarefas descritas abaixo. As tarefas são mandatórias e serão usadas para determinação do conceito final da disciplina; elas devem ser apresentadas em formato word (.txt) e carregadas na plataforma moodle usp até 2h antes do início do encontro virtual síncrono (11:30 hora de Brasília). Além das tarefas é aconselhável que os participantes busquem explorar os códigos apresentados usando um banco de dados próprio e executem os desafios opcionais, todavia, isso não é mandatório e não serão usados para determinação do conceito final.

**LISTA DE TAREFAS**

*Aulas 1.1, 1.2 e 1.3*: Apresente pelo menos três variáveis aplicadas em estudos de sua área em cada uma das classificações da natureza das variáveis; demonstre pelo menos duas opções de conversão da classificação original das suas variáveis em uma outra classificação.

*Aulas 2.1, 2.2 e 2.3*: Apresente uma tabela com as estatísticas descritivas apresentadas na aula de todas as variáveis quantitativas do banco de dados ‘dataset.horses.csv’ somente da raça ‘Thoroughbred’ (variável ‘Breed’); em outra tabela apresente a porcentagem de ocorrência da variável ‘Stress.Score’ somente da raça ‘Quarter Mile’ (variável ‘Breed’).

*Desafio opcional*: Importe um banco de dados seu ou de seu grupo de pesquisa para o RStudio.

*Aulas 3.1 e 3.2*: Apresente um gráfico de violino e de caixas (na mesma figura) e um gráfico de dispersão juntos em uma prancha (uma figura contendo ambos os gráficos) usando as variáveis de sua preferência do banco de dados ‘dataset.horses.csv’, exceto ‘Cortisol’ e ‘IRT’; é mandatório mudar as configurações dos gráficos (estilo de tema, cor, espessura de linha, tipo de linha, tipo de ponto, etc).

*Desafio opcional*: Apresente um dos gráficos apresentados em aula usando um banco de dados seu ou do seu grupo de pesquisa.

*Aulas 4.1 e 4.2*: Apresente em uma tabela o coeficiente linear e angular e seus respectivos erro-padrão, valor de t e valor de p baseado em uma modelo regressivo linear simples usando as variáveis de sua preferência do banco de dados ‘dataset.horses.csv’, exceto ‘Cortisol’.

*Desafio opcional*: Apresente a comparação do log likelihood entre um modelo linear e um modelo quadrático usando ‘IRT’ como variável resposta e ‘Cortisol’ como variável independente.

*Aulas 5.1, 5.2 e 5.3*: Apresente em uma tabela o coeficiente linear e angular e seus respectivos erro-padrão, valor de t e valor de p baseado em uma modelo regressivo linear múltiplo usando como variável resposta o ‘Cortisol’ e conservando no modelo apenas as variáveis independentes com beta significativo.

*Desafio opcional*: Apresente o código e o output (resultado) dos testes de Cramer-Von Mises e Shapiro-Wilk, bem como do teste *t* Student comparando as raças (‘Breed’) para todas as variáveis quantitativas do banco de dados ‘dataset.horses.csv’.