

Segundo Charnet, Freire, Charnet e Bonvino (1999, p. 11), dado Y uma variável resposta e X uma variável preditora:

O modelo de regressão linear simples descreve a variável Y como uma soma de uma quantidade determinística e uma quantidade aleatória. A parte determinística, uma reta em função de X , representa a informação sobre Y que já poderíamos “esperar”, apenas com o conhecimento da variável X .

Tal quantidade determinística não se define exclusivamente por uma medida quantitativa. Pode-se utilizar uma variável preditora com caráter indicador, ou seja, ela assume valor 1 quando se atende uma categoria de interesse e 0 caso contrário. Esse tipo de variável indicadora denomina-se *Dummy*.

Utilizam-se modelos com as características apresentadas em diversas áreas de pesquisa, incluindo genética. Por exemplo, para se estudar proteínas com diferentes abundâncias entre pacientes, que sofrem de distúrbio bipolar, e pacientes controle.

Para esse estudo, dados foram coletados de cada indivíduo em duas réplicas. Dessa forma se torna interessante estudar a influência da réplica sobre a predição da abundância para cada proteína. Sendo assim, parâmetros foram estimados para cada uma das variáveis, baseados nos dados dos indivíduos de cada grupo. Logo modelos foram ajustados para cada proteína do estudo.

Uma medida qualitativa influencia a predição de uma variável resposta se o parâmetro de sua respectiva preditora se faz indispensável ao modelo. Um teste de hipótese é aplicado para determinar se, de fato, a variável preditora para pacientes afetados e controles deve permanecer no modelo. Procedimento no qual se repete para todas as proteínas.