Fieno, bom dia.

Espero poder ajudar. “Outliers" são definidos como os valores numéricos de qualquer conjunto de dados aleatórios, que têm um desvio invulgarmente elevado a partir de qualquer conjunto estatístico significativo (em média), ou o valor da mediana.

Melhorando essas palavras, estes números que se distanciam necessariamente precisam ser calculados para verificarmos exatamente o grau de precisão dos desvios de nossos rols. Esses valores podem ter distâncias pequenas ou muito grandes. Detectar os valores extremos de um conjunto de dados representa um problema estatístico complexo, principalmente que isso só é visto em países como na Holanda, rssss, normalmente não aprendemos isso aqui.

Para o cálculo, podemos usar uma grande variedade de técnicas, normalmente uso as calculadoras online, por preguiça.

Em geral, a obtenção dos "outliers" em um conjunto de dados, sendo população ou amostra, é feito por meio do cálculo do desvio, para cada número, expressa como "Z-score" ou "Z-pontuação modificado", aportuguesado, isso quando necessitamos de cálculos com grau de precisão mais altos para compará-lo contra determinado limiar predefinido.

Z-score nada mais é que a quantidade de desvios padrões em relação à média estatística (em outras palavras, é medida em "Sigmas").

No Z-score modificado aplica-se a técnica de cálculo da mediana para medir o desvio e, em muitos casos, fornece detecção estatística mais robusta de outliers.

Matematicamente, o Modificado Z-score poderia ser escrito (como sugerido por Iglewicz e Hoaglin [1]), como:

Mi = 0,6745 \* (Xi-Mediana (Xi)) / MAD, onde MAD é o desvio absoluto mediano.

Qualquer número do conjunto observado com o valor absoluto do Z-score modificado superior a 3,5, de dados é considerada uma "outliers".

**Vou explicar, “tentar” uma *Solução* de planilha do Excel**

Microsoft Excel não possui um comando que calcula diretamente o "Outliers" ou a discrepância, mas pode ser construídos usando fórmula de matriz e outra função estatística: a média () [2,3].

Para se fazer isso não é necessário nenhum conhecimento em VBA, coisa que você manja, se montar um VBA, manda pra mim.

**Exemplo de cálculo**

Vamos detectar um Outliers em uma planilha do Microsoft Excel usando um conjunto de 10 números, e vamos seguir as etapas abaixo:

**Passo 1**. Abra o Excel e coloque uma amostra de 10 números sorteados na coluna A, começando com a primeira linha: 3, 1, -23, 7, 0, 12, -2, 7, 2, 1, depois de inserido, eu normalmente organizo os dados em rol.

**Passo 2**. Na primeira linha da coluna C (em outras palavras, C1) inserir a fórmula: =MED(A1:A10). O valor nesta célula corresponde à mediana calculada sobre um conjunto de dados inserido no passo 1.

**Passo 3**. Na segunda linha da coluna C (na linha 2). Coloque essa fórmula usando matriz: {=MED(ABS(MED(A1:A10)-A1:A10))} . No excel, só pra lembrar, pra inserir uma matriz, selecione as células, digite a fórmula no Excel na barra de fórmulas e, em seguida, clique na combinação: CTRL-SHIFT-ENTER (observe as chaves em torno da expressão, o que indica a fórmula de matriz) . O valor nesta célula (C2) corresponde a MAD [1].

**Passo 4**. Após isso, insira a fórmula: =SE(((ABS(C$1-A1))>3,5\*C$2);"OUTLIER";"NORMAL") na primeira linha da coluna B e estendê-lo até a 10 ª linha, em outras palavras, copiá-lo em cada célula do intervalo B1: B10.

Resultado final de "detecção de outlier" deve aparecer na coluna B, indicando os dois números "discrepantes" números (-23 e 12), conforme mostrado abaixo:



**VALEU FÍIII, ESPERO TER AJUDADO. SEI QUE AS COISAS AÍ SÃO BEM MAIS COMPLEXAS, MAS DÁ PRA DAR UMA AJUDADA NO INGLISHHHHH!!!**

**COMO DIZ UM FILÓSOFO AQUI DE PARACATU:**

***“NÃO SACO MUITO DE INGLÊS, MAS NO PORTUGUÊS EU DISTRÓIO!”***