Hipótese estatística de correlação

Tem por objetivo, em termos estatísticos, uma correlação entre duas variáveis aleatórias. O símbolo de correlação entre duas variáveis é "r", e entre mais de duas, “R”.

**Covariância**: média dos desvios combinados.

Para cada observação, multiplicar seus desvios em X e Y ao invés de elevar ao quadrado.

Dividir pelos graus de liberdade.

A interpretação da covariância depende da escala dos dados originais.

O sinal da covariância indica proporção inversa ou direta

**Correlação**: razão geométrica – ou padronização – entre a covariância e a raiz do produto entre as variâncias de X e de Y (que equivale à multiplicação dos desvios padrões).

Razão de um produto intrínseco dos desvios e outro produto extrínseco dos desvios.

Cohen: efeitos da correlação.

Efeito fraco:

Efeito médio:

Efeito grande:

Teste de hipótese

Hipótese nula: r=0; não há correlação.

R²: coeficiente de determinação, ou coeficiente de linearidade.

Percentual de quanto uma variável explica a outra.

Regressão

Determinar a função de uma variável em relação a outras.

Tentar prever algum tipo de saída (resultado) a partir de uma ou mais variáveis previsoras.

Variável dependente em função de variáveis independentes.

Regressão simples: uma variável independente.

Regressão linear:

Método dos mínimos quadrados ordinários: encontrar a reta que melhor se ajusta aos dados. Dentre todas as infinitas possíveis linhas, a determinada pelo método dos mínimos quadrados ordinários é aquela cujos os erros quadrados serão os menores.

A curva da função afim é determinada por dois parâmetros: inclinação e intercepto/posição (onde a curva cruza com o eixo Y; ).

Erro: diferença entre o valor observado e o valor esperado de acordo com a função definida na regressão.

Teste t para o coeficiente angular

Se , então a curva determinada pelo modelo aproxima-se de um modelo meramente explicado pela média de Y (cujo ângulo é igual a zero). Logo, R² seria próximo e zero

Soma dos quadrados dos erros

Total:

; ;

Residual:

; ;

Do Modelo:

A comparação entre e equivale à comparação entre o poder de explicação do modelo e o poder de explicação da média.

: quanto da variância total é explicada pelo modelo.

Estatística F: indica quanto o seu modelo melhorou a predição em comparação com o nível de imprecisão do modelo.

Quanto maior a estatística F, maior a capacidade explicativa do modelo. Se o valor de F do modelo for maior que o valor de F crítico (a significância for menor que α)

Outliers

Uma observação que influencia o modelo sobremaneira, muito além dos demais pontos. São pontos ignoráveis.

A determinação de outliers se dá através da análise dos resíduos da observação em relação ao modelo.

**Resíduos padronizados (ou “standartizados”)**: resíduos convertidos a z-scores.

Resíduos padronizados maiores que 3,29 ou menores que -3,29 indicam outliers.

Se mais de 1% dos dados possuírem resíduos padronizados maiores que 2,58 e/ou menores que -2,58, há evidências de que o modelo possui ajuste pobre.

Se mais de 5% dos resíduos padronizados forem maiores que 1,96 e/ou menores que -1,96, então também há indícios de que o modelo não é bem adequado àquela amostra.

**Resíduos “studentizados”**: resíduo dividido por uma estimativa do seu desvio padrão que varia ponto a ponto. Fornece uma estimativa mais precisa da variância de um erro em caso específico.

**Distância de Cook**: medida de influência global de um único caso no modelo. Valores maiores que 1 são motivos de atenção.

**Pontos de alavancagem (leverage points)**: avaliam a influência de um valor observado sobre os valores previstos.

**DF Fit**: diferença entre o previsto para o caso quando o modelo é calculado incluindo este caso e removendo-o. Uma versão padronizada deste valor, geralmente é mais utilizada DF-Fit-S (Standartized DF Fit).

Se houver homoscedasticidade dos resíduos, haverá uma distribuição constante ao redor do eixo Y, sem padrões de funil ou gravata borboleta.

Homoscedasticidade é um pressuposto importante e indica que a variância dos erros (resíduos) é constante ao longo da variação da VI

Qualitativa -> categórica

Nominal

Ordinal

Quantitativa -> numérica

Discreta

Contínua

Diretório para plugin do excel

AppData>Roaming>Microsoft>Suplementos