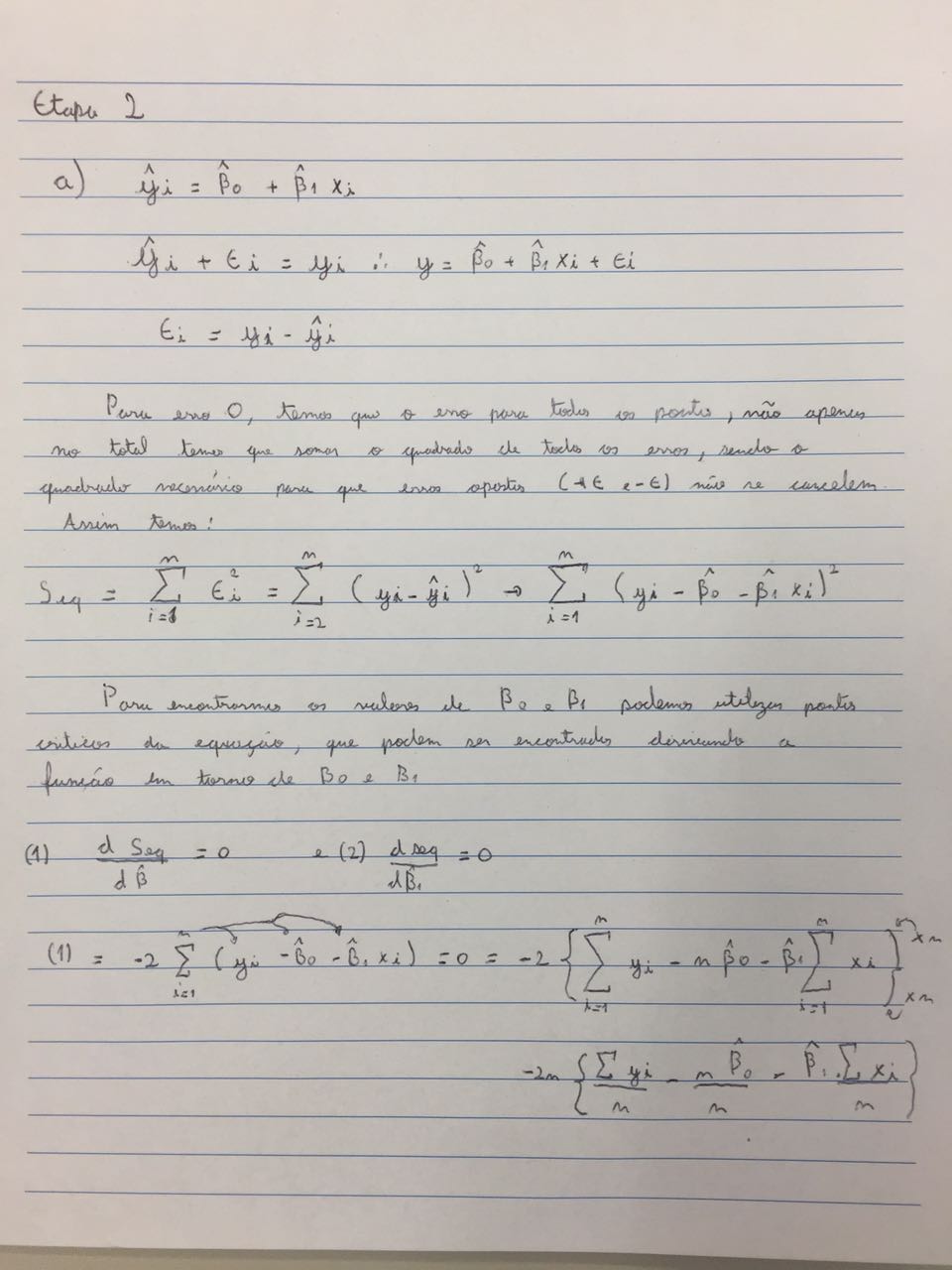
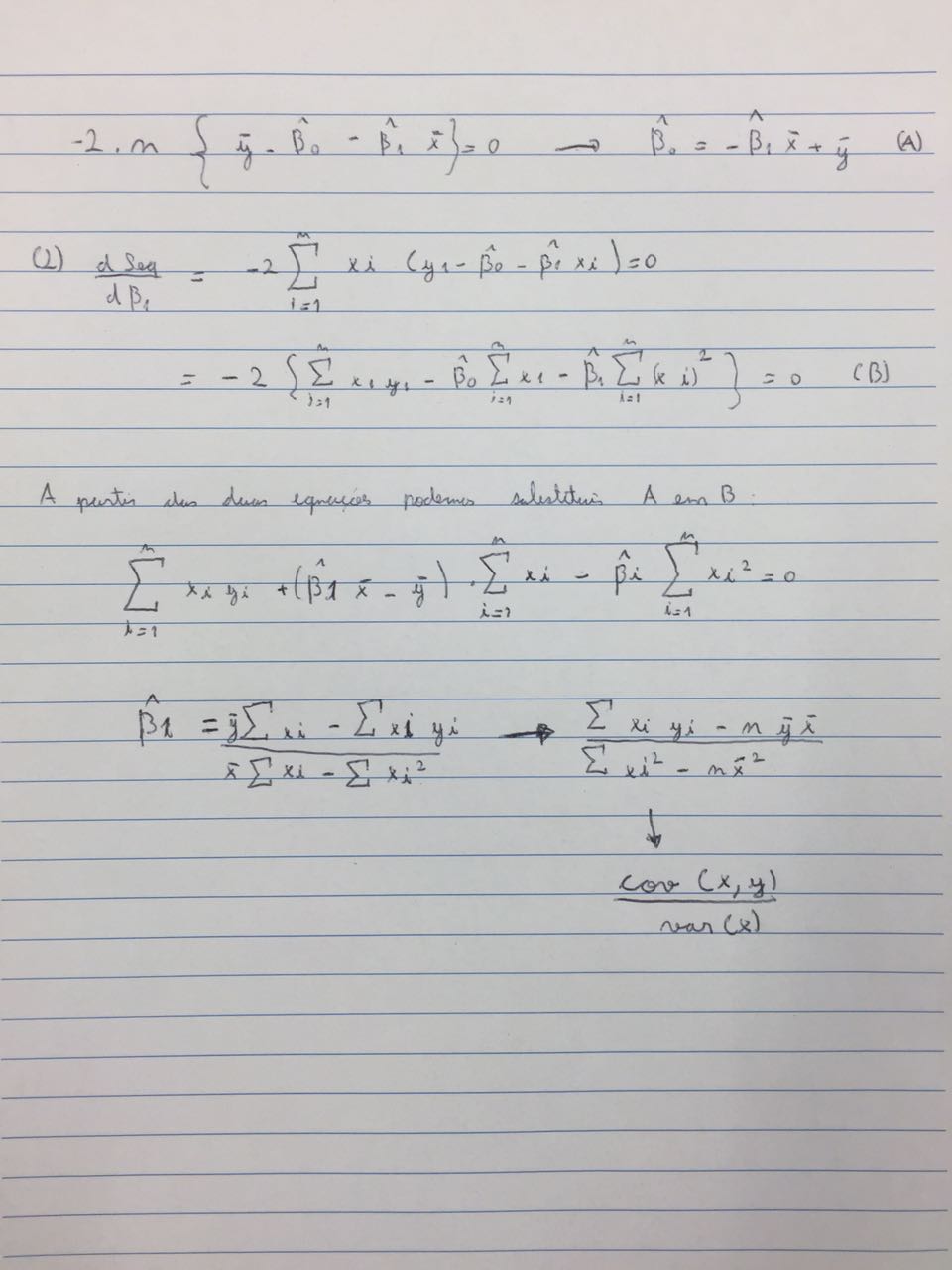
- Eduardo F. Magalhaes, Jean L. Walper -

-Projeto 3 : Ciência dos dados : Etapa 2-

A)

(Feito totalmente em papel)



B)

Os erros são variáveis aleatórias e independentes entre si, isso implica que a distribuição deles terá o formato de uma distribuição normal. O valor esperado do erro, como mostramos nas nossas contas no item A, deverá ser próximo a zero, portanto a média dessa distribuição normal será zero.

A variância, como sabemos, é o quadrado do desvio padrão, portanto tendo o desvio padrão das amostras, ou melhor, da diferença dos valores amostrais para o valor aproximado pela nossa reta, teremos os valores para montar essa normal e formamos a distribuição dos erros.

Montando uma normal com media zero e usando o desvio padrão dos erros, como dito acima, teríamos a representação visual da distribuição. Na pratica, se os erros minimizados método dos mínimos quadrados tiverem uma variação baixa, e a normal for estreita, teremos um bom modelo, com um erro consistente.

C)

Com um teste de hipóteses, considerando que o erro será zero, podemos considerar a hipótese de que βo = 0. Assim se a hipótese for aprovada, não haveria uma relação entre as amostras, portanto Ho: βo = 0, e Ha: βo ≠ 0.

Então faríamos o teste de hipótese, e caso Ho fosse aprovado, teríamos amostras que não se relacionam uma com a outra, e a reação entre as duas será nula.

D)

É possível ter um modelo de regressão múltiplo, só que isso adicionara mais dimensões a nossa análise. Por exemplo, para um modelo de regressão duplo, ou seja, com duas variáveis explicativas, teríamos além de βo e β1, um β2, essa incógnita representa outro coeficiente angular, só que dessa vez, girando a reta em outro eixo, então podemos dizer que em um gráfico, o β2 seria o coeficiente angular da reta no espaço, em relação ao eixo Z.

As suposições continuam as mesmas, mesmo com a dimensão adicionada, e o teste de hipótese também é semelhante, com a mesma hipótese nula, só que agora, será necessário fazer vários testes de hipóteses, dependendo do número de variáveis analisadas. Então se fossemos analisar 4 variáveis por exemplo, fazer um teste de hipótese não comprova a relação entre as outras variáveis, por exemplo, se uma variável não tiver nenhuma correlação com as outras, não podemos afirmar que nenhuma das variáveis analisadas está relacionada.

O modelo para a regressão múltipla (para duas variáveis explicativas) fica assim:

