

Análise de conglomerados em curvas de aprendizado para formação de agrupamentos homogêneos de trabalhadores

Renato Eduardo Stroeike^{a*}, Flavio Sanson Fogliatto^b, Michel Jose Anzanello^c

^{a*}renatoes@producao.ufrgs.br, UFRGS, Brasil

^bffogliatto@producao.ufrgs.br, UFRGS, Brasil

^canzanello@producao.ufrgs.br, UFRGS, Brasil

Resumo

Em diversos setores da indústria é desejado que trabalhadores reunidos em uma estação de trabalho apresentem perfil de aprendizado similar. O presente artigo apresenta um método de agrupamento de trabalhadores utilizando modelagem por curva de aprendizado e técnicas de clusterização. O método modela dados de desempenho de trabalhadores por intermédio de diversos modelos de curvas de aprendizado; os parâmetros de aprendizado dos modelos testados permitem predizer o desempenho dos trabalhadores em intervalos de tempo pré-determinados. Os valores preditos são agrupados através de ferramentas de clusterização. O maior índice de ajuste (IA), gerado a partir do Silhouette Index e do coeficiente de determinação, indica o modelo de curva mais consistente em termos de aderência aos dados e qualidade de agrupamento de perfis de aprendizado. Ao ser aplicado em dados de uma indústria de calçados, o método gerou agrupamentos consistentes de trabalhadores com base nos distintos perfis de aprendizado.

Palavras-chave

Curvas de aprendizado. Clusterização. Agrupamento de trabalhadores.

1. Introdução

Curvas de aprendizado são modelos de regressão não linear que associam o desempenho, usualmente descrito em termos de unidades produzidas por intervalo de tempo, com as características da tarefa (ANZANELLO; FOGLIATTO, 2007a). As curvas de aprendizado são baseadas no pressuposto de que o tempo de realização de uma tarefa repetitiva diminui com o passar do tempo, à medida que ocorre o fenômeno da aprendizagem.

As curvas de aprendizado são frequentemente usadas para classificar grupos de trabalhadores. Comumente designa-se por *fast learner* o trabalhador que possui uma alta taxa de aprendizado e por *slow learner* o que apresenta uma taxa menor. Nembhard e Uzumeri (2000) apontam que, tradicionalmente, grupos com baixa taxa de aprendizado tendem a alcançar um maior grau de produtividade, enquanto que grupos com alta taxa de aprendizado atingem patamares inferiores de produtividade final.

Em diversos contextos o agrupamento de trabalhadores de acordo com seu perfil de aprendizado se mostra de grande interesse. Em uma linha de montagem manual é desejado que trabalhadores reunidos em uma estação de trabalho apresentem perfis similares de aprendizado, com vistas à preservação do balanceamento da carga de trabalho na estação. Em linhas de baixa variedade isso pode não ser muito importante, mas à medida que o volume de modelos aumenta (elevada customização) o agrupamento de trabalhadores por perfis passa a ser essencial. Apesar do agrupamento de trabalhadores por características similares (por exemplo, tempo de produção ou método de trabalho) ser uma prática comum na gestão dos postos de trabalho, observa-se uma lacuna na literatura referente ao uso de curvas de aprendizado para determinação de agrupamento de trabalhadores. Tal lacuna é justificada pela dificuldade em estimar o tempo padrão de execução de uma atividade durante

o processo de aprendizado do trabalhador, visto que não se trata de uma variável pontual, mas sim de uma variável funcional (ANZANELLO; FOGLIATTO, 2007a, 2011). A adoção da média dos tempos verificados durante o processo de aprendizado como tempo padrão de operação acaba por subestimar o desempenho final do trabalhador (impactando em corridas longas de produção), ao passo que utilizar o tempo de operação após conclusão do processo de aprendizado faz com que o desempenho do trabalhador seja superestimado (gerando problemas em corridas curtas de produção, onde um número limitado de repetições normalmente é verificado).

O agrupamento de trabalhadores segundo seus perfis de aprendizado restringe-se ao estudo de Uzumeri e Nembhard (1998). Os autores utilizaram os três parâmetros da curva de aprendizado de Mazur e Hastie (1978) para descrever padrões de habilidade em uma população de trabalhadores. Ao aplicar o modelo em uma empresa de manufatura americana, os autores demonstraram existir diferentes perfis de trabalhadores, cada um mais apropriado para uma dada atividade. O principal diferencial da sistemática aqui proposta está no agrupamento dos trabalhadores com base no seu perfil de desempenho e não nos parâmetros oriundos da modelagem por curva de aprendizado, como sugerido por Uzumeri e Nembhard (1998). A utilização dos dados de desempenho permite capturar tendências específicas no perfil de aprendizado que seriam mascaradas na análise direta dos parâmetros gerados pela modelagem da curva de aprendizado.

Nesse sentido, o método para o agrupamento de trabalhadores utilizando modelagem por curva de aprendizado e técnicas de clusterização apresentado neste trabalho modela dados de desempenho de trabalhadores por meio de diversas formulações de curvas de aprendizado; os parâmetros de aprendizado permitem prever o desempenho dos trabalhadores em intervalos de tempo pré-determinados. Os valores preditos são agrupados por intermédio de ferramentas de clusterização. Uma análise hierárquica permite definir o número provável de agrupamentos a serem gerados, enquanto que a ferramenta de clusterização *k* means é efetivamente utilizada para agrupar os trabalhadores. O melhor modelo de curva, e consequente agrupamento, é escolhido por intermédio de um índice de ajuste gerado a partir de um índice de aderência de cada modelo aos dados (R^2) e de um indicador de qualidade dos agrupamentos gerados (Silhouette Index).

A principal contribuição teórica deste artigo está na integração entre curvas de aprendizado e sistemáticas de clusterização com vistas à formação de grupos de trabalhadores com perfis semelhantes de aprendizado e patamares equivalentes de desempenho

final. Destaca-se ainda a proposição de um índice que avalia conjuntamente a qualidade da aderência das curvas aos dados de desempenho (por intermédio do coeficiente de determinação, R^2) e a qualidade dos agrupamentos gerados (medida pelo Silhouette Index). Em termos práticos, o artigo contribui com uma sistemática que possibilita reduzir os impactos decorrentes da inserção de trabalhadores com perfis distintos de desempenho em uma mesma linha ou célula de produção. O agrupamento inapropriado de trabalhadores acarreta formação de gargalos produtivos, desequilíbrio nos tempos de ciclo de produção, impactos na qualidade dos produtos trabalhados e potencial descontentamento dos trabalhadores acerca do ritmo produtivo desnivelado do grupo.

O restante deste trabalho está estruturado da seguinte forma. A seção 2 traz uma breve revisão da literatura em torno dos modelos de curva de aprendizado. Na seção 3 é apresentada a metodologia proposta. Um exemplo numérico em uma indústria de calçados é apresentado na seção 4. Por fim, tem-se a conclusão do artigo na seção 5.

2. Referencial teórico

O primeiro modelo de curvas de aprendizado foi proposto por Wright (1936) usando dados observados em operações de montagem de aviões. O autor concluiu que a montagem de aeronaves apresentava uma redução aproximada de 20% no custo médio a cada duplicação da quantidade produzida (TEPLITZ, 1991; COOK, 1991; BADIRU, 1992; ARGOTE, 1999; ASKIN; GOLDBERG, 2001; ANZANELLO; FOGLIATTO, 2007b). Esse fenômeno de redução de custos já havia sido observado na indústria aeronáutica por Rohrbach (1927), porém não havia sido analisado o efeito de aprendizado (BISKUP, 2007).

A principal motivação no estudo de curvas de aprendizado é a redução de custos e o aumento da produtividade. Redução de custos é resultado de “aprender fazendo”, aumento da capacidade e expansão da produção (SÖDERHOLM; SUNDQVIST, 2007). Adler (1990) sugeriu que o aprendizado compartilhado no desenvolvimento de produto, na interface de manufatura, no posicionamento primário da planta e no aprendizado contínuo após o início das atividades é essencial para o aumento da produtividade.

Os fatores que proporcionam o aprendizado são complexos e ainda não plenamente compreendidos (ADLER; CLARK, 1991). Diversos trabalhos, como os de Uzumeri e Nembhard (1998) e Anzanello e Fogliatto (2007b), demonstram serem os trabalhadores que

aprendem de forma mais gradual os que alcançam um padrão de produtividade maior. Além disso, trabalhadores que aprendem mais rapidamente também tendem a esquecer mais rapidamente durante paradas na produção (NEMBHARD; UZUMERI, 2000). A modelagem de tais fenômenos deu origem a uma série de modelos distintos de curvas de aprendizado, descritas a seguir.

O modelo tido como pioneiro pela literatura é o modelo potencial, no formato proposto por Wright (1936), apresentado na Equação 1. y indica o tempo necessário para execução de uma repetição da operação em análise, enquanto que C_1 é o tempo de execução da primeira repetição. A constante de aprendizado é o parâmetro b , que varia de 0 a -1 . Quanto menor o valor do parâmetro de aprendizado, maior é a taxa de aprendizado.

$$y = C_1 x^b \quad (1)$$

Baseado no modelo de Wright, o modelo do Plateau consiste no acréscimo de uma constante C ao primeiro, sendo dado na Equação 2. A constante C impõe um limite para o processo de aprendizado, indicando o máximo desempenho que o trabalhador poderá atingir.

$$y = C + C_1 x^b \quad (2)$$

Também similar aos anteriores, o modelo de Stanford- b consiste em adicionar o parâmetro positivo B , que representa a experiência prévia do indivíduo medida em unidades previamente produzidas (ver Equação 3), ao modelo.

$$y = C_1 (x + B)^b \quad (3)$$

O modelo hiperbólico de dois parâmetros baseia-se na proposição de Mazur e Hastie (1978), sendo apresentado na Equação 4. Enfatiza-se que a notação difere dos modelos anteriormente apresentados, sendo y o número de unidades produzidas em um dado intervalo de tempo e x o intervalo de tempo considerado. Nesse modelo, o parâmetro de aprendizado é representado por r ; o parâmetro k mostra o nível máximo de aprendizado.

$$y = k \left(\frac{x}{x+r} \right) \quad (4)$$

Baseando-se em uma modificação do modelo anterior, a curva hiperbólica de três parâmetros (Equação 5) adiciona o parâmetro p a sua formulação, o que representa a experiência prévia do trabalhador na execução da tarefa. Os demais parâmetros são como na Equação 4.

$$y = k \left(\frac{x+p}{x+p+r} \right) \quad (5)$$

Outro modelo amplamente utilizado em aplicações práticas é o exponencial de três parâmetros, na Equação 6. Os parâmetros seguem definições do modelo anterior.

$$y = k \left(1 - e^{-\left(\frac{x+p}{r}\right)} \right) \quad (6)$$

Desde o seu desenvolvimento, as curvas de aprendizado têm se mostrado uma importante ferramenta em engenharia de produção. Corominas, Olivella e Pastor (2010) formalizaram o problema de atribuição e de agendamento de um conjunto de tarefas, considerando o desempenho dos trabalhadores em uma tarefa como dependente da experiência na atividade e em outras similares. Cohen, Vítner e Sarin (2008) estudaram o problema de alocação de trabalhadores em linhas de montagem, diminuindo o *makespan*. Foi o primeiro trabalho em que as curvas de aprendizado do tempo de preparação dos equipamentos foram avaliadas.

Toksari et al. (2008) elaboraram um algoritmo para balanceamento de linhas de montagem sujeitas ao efeito de aprendizado, tendo por objetivo diminuir o número de estações de trabalho e o tempo de ciclo nessas estações. Prochno (2004) relatou de forma abrangente o primeiro ano de operação na área de montagem de uma nova planta automotiva por intermédio de curvas de aprendizado. Tal análise permitiu ao autor concluir que as curvas de aprendizado são resultado de um processo de integração que reúne vários trabalhadores com aprendizado individual contínuo, em diferentes áreas da organização.

Ngwenyama, Guergachi e McLaren (2007) apresentaram uma abordagem para atualizações de *softwares* que maximiza os ganhos de produtividade da organização baseada em curvas de aprendizado. Em seu trabalho sobre aprendizado em inovação e produção, Pruett e Thomas (2008) concluíram que quanto melhor for o processo de aprendizado na área de inovação da empresa, menos dramático será o efeito do aprendizado na produção. O autor observou que o processo de aprendizado também ocorre no desenvolvimento de produtos e que possíveis deficiências no desenvolvimento só serão observadas na produção.

Nos últimos anos, a qualidade dos processos vem sendo muito valorizada e esse conceito foi acrescentado à teoria sobre curvas de aprendizado. Essa ideia pode ser vista em Jaber e Griffida (2004),

onde foi desenvolvida uma modificação na curva de Wright criando uma curva da qualidade, sendo essa o somatório de duas curvas: a primeira representa a redução de tempo para cada unidade adicional produzida, e a segunda descreve a redução no tempo em cada unidade defeituosa retrabalhada.

A necessidade de agrupar trabalhadores por similaridade de aprendizado é conhecida, porém ainda pouco desenvolvida na literatura sobre curvas de aprendizado (ANZANELLO; FOGLIATTO, 2007a, 2011). Pesquisadores como Cohen, Vitner e Sarin (2008) verificaram que a alocação de trabalhadores é melhor alcançada quando, para uma dada atividade em conjunto (como ocorre em uma estação de trabalho, por exemplo), são agrupados trabalhadores com perfis similares de aprendizado. A partir dessa constatação, idealmente haveriam padrões de aprendizagem similares em todas as estações de uma linha. Usando a curva hiperbólica de três parâmetros, Anzanello e Fogliatto (2007b) propuseram um método para alocar trabalhadores em tarefas específicas em linhas de montagem. Os parâmetros gerados pelas curvas de aprendizado são analisados em dois cursos distintos de ação, de acordo com a duração estimada da corrida de produção. Ferramentas de clusterização também são utilizadas, porém restringem-se ao agrupamento de modelos de produtos de acordo com sua complexidade.

Uzumeri e Nembhard (1998) foram dos poucos pesquisadores a utilizarem as curvas de aprendizado com o objetivo de identificar agrupamentos naturais de trabalhadores. Os autores observaram 3.874 casos de aprendizado em uma manufatura americana. Primeiramente foi modelada uma curva de aprendizado para cada trabalhador, sendo observados oito padrões de aprendizado distintos. Com os parâmetros dessa modelagem foi estimado o comportamento de todo o grupo por meio de um gráfico tridimensional onde cada eixo representava um parâmetro da regressão. Pôde-se observar a existência de três *clusters* de trabalhadores. O primeiro era formado pelos parâmetros de trabalhadores que saíram antecipadamente do treinamento inicial, indicando baixo desempenho. O segundo *cluster* representava o comportamento oposto, pois mostrava os dados de trabalhadores que possuíam uma experiência prévia desempenhando as atividades de forma mais eficiente. No gráfico existia ainda um terceiro *cluster*, que abrigava os parâmetros da maioria dos trabalhadores, caracterizados por um perfil médio de aprendizado.

O presente trabalho difere do de Uzumeri e Nembhard (1998) primeiramente na forma de identificar os padrões coletivos de aprendizado. Na proposta apresentada na seção 3, o agrupamento de trabalhadores ocorre de acordo com os valores

de desempenho previstos nas curvas, as quais foram geradas com base nos parâmetros obtidos na regressão. São utilizadas, também, técnicas de estatística multivariada que permitem verificar a qualidade dos agrupamentos, em oposição à identificação visual de agrupamentos proposta por Uzumeri e Nembhard (1998).

3. Metodologia proposta

O método proposto para o agrupamento de trabalhadores com perfis semelhantes de aprendizado é composto pelas etapas representadas na Figura 1. Tais etapas são detalhas na sequência.

Inicialmente, selecionam-se J ($j = 1, \dots, J$) trabalhadores que terão seus tempos de aprendizado monitorados. É desejado que cada trabalhador passe por um processo de aprendizado envolvendo uma operação distinta. No momento em que o trabalhador é apresentado a uma nova tarefa, começa a ser cronometrado o tempo demandado para execução de cada repetição. Idealmente, são coletados os tempos desde a primeira repetição (a partir do momento em que o trabalhador é apresentado à tarefa) até o momento em que não há variações significativas nos dados coletados.

Os dados de desempenho de um trabalhador j são modelados por intermédio de regressões não lineares a fim de se obter os parâmetros para os seguintes modelos de curvas de aprendizado: (i) Curva potencial, (ii) Plateau, (iii) Stanford-B, (iv) Hiperbólico de 2 parâmetros, (v) Hiperbólico de 3 parâmetros, e (vi) Exponencial de 3 parâmetros. Os modelos (i) a (iii) modelam dados do tipo “tempo demandado por repetição”, enquanto que os modelos (iv) a (vi) modelam dados do tipo “número de repetições efetivadas em determinado intervalo de tempo”. Tal diferença deve-se a aspectos estruturais dos modelos, os quais são apresentados na sequência.

A qualidade do ajuste do modelo de curva de aprendizado m ($m = 1, \dots, M$) aos dados de desempenho do trabalhador j é avaliada por meio do coeficiente de determinação R_{jm}^2 . Como um dos propósitos deste artigo é identificar o modelo que melhor se ajusta aos dados de desempenho, avalia-se a qualidade da aderência do modelo m , R_m^2 , para o conjunto de J trabalhadores conforme a Equação 7.

$$R_m^2 = \frac{\sum_{j=1}^J R_{jm}^2}{J} \quad m = 1, \dots, M \quad (7)$$

Na sequência, utilizam-se os parâmetros obtidos pelas diversas modelagens para prever valores de desempenho y para cada trabalhador com base em

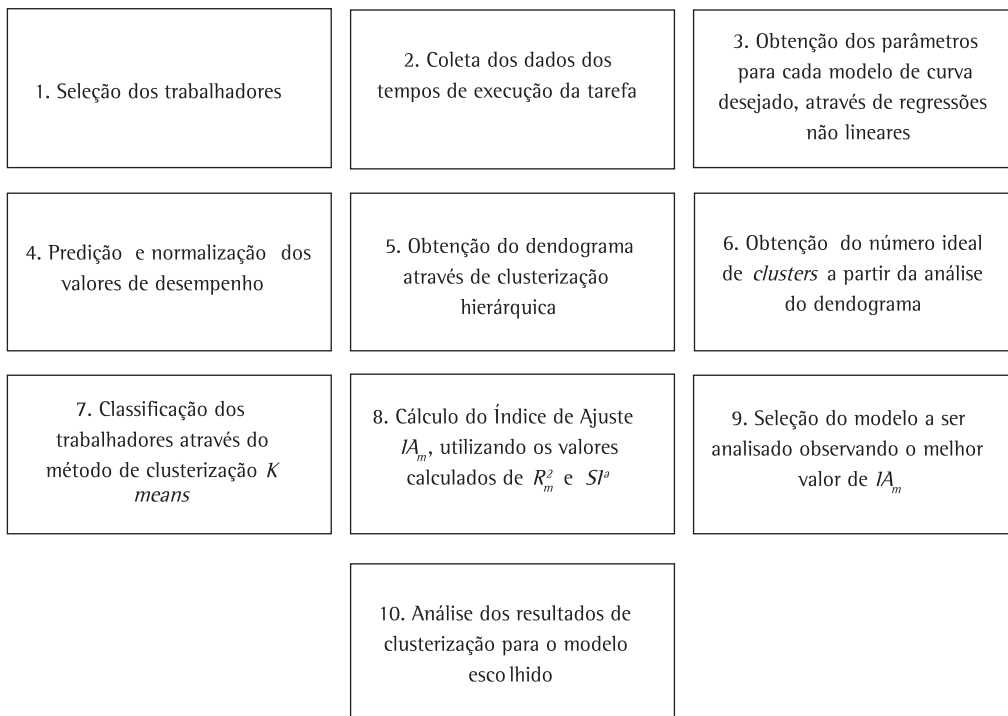


Figura 1. Estrutura do método proposto.

um intervalo de tempo ou número de repetições pré-determinado (por exemplo, predizem-se dados de desempenho equivalentes a 2 horas de operação pelo modelo exponencial). Tal procedimento faz com que exista um número comum de observações de desempenho gerado por cada modelo de curva. Os valores preditos para cada trabalhador são então normalizados, fazendo com que eles apresentem média 0 e desvio-padrão 1. As observações passam a ser consideradas variáveis de clusterização no procedimento de formação de grupos.

As observações de desempenho dos J trabalhadores gerados por um modelo m são então clusterizadas. Cada trabalhador é representado por um vetor contendo os valores preditos. Primeiramente, roda-se uma análise hierárquica para definir o número ideal de *clusters* a serem formados. Nesse tipo de análise, as observações são agrupadas segundo algum tipo de métrica de distância e as variáveis são agrupadas conforme medidas de correlação ou associação (FÁVERO et al., 2009). O número de *clusters* é então utilizado no método k means para a clusterização dos dados gerados pelo modelo m ; a qualidade da clusterização é medida através do Silhouette Index.

O Silhouette Index ($-1 < SI_j < 1$) mede o quanto uma dada observação j (trabalhador, neste caso) é similar às demais observações em seu próprio *cluster*,

comparada às observações alocadas no *cluster* mais próximo a ele. Valores de SI_j próximos a -1 indicam que o trabalhador j foi erroneamente inserido no *cluster* de destino. Valores de SI_j próximos a zero mostram que o trabalhador poderia estar tanto em seu *cluster* de destino quanto em algum outro *cluster*. Valores de SI_j próximos a 1 indicam que o trabalhador está corretamente alocado em seu *cluster*. SI_j é calculado por intermédio da Equação 8, onde $a(j)$ é a média da distância do j -ésimo trabalhador a todos os demais trabalhadores pertencentes ao *cluster* no qual o j -ésimo trabalhador foi alocado. O termo $b(j)$ é a média da distância entre o j -ésimo trabalhador e todos os trabalhadores do *cluster* vizinho mais próximo. Usualmente utiliza-se a distância euclidiana ou de Manhattan para calcular a distância entre trabalhadores.

$$SI_j = \frac{b(j) - a(j)}{\max\{b(j), a(j)\}} \quad (8)$$

A média dos valores de SI_j para os J trabalhadores oferece uma noção da qualidade geral da clusterização dos dados de desempenho gerados pelo modelo de curva m , sendo designado por SI_m . Como os valores de SI_m variam de -1 à 1 , os mesmos são reescalados para obterem-se valores entre 0 e 1, criando o Silhouette

Índice ajustado, SI_m^a . Tal procedimento faz que SI_m^a e R_m^2 operem no intervalo $[0,1]$.

Por fim, gera-se o índice de ajuste (IA_m) para avaliação do desempenho de cada modelo de curva no agrupamento de perfis, conforme Equação 9. O índice contempla dois aspectos importantes nos

esforços de agrupamento dos perfis de aprendizado: (i) a qualidade de aderência do modelo m aos dados de desempenho, e (ii) a qualidade do agrupamento gerado pelos dados de desempenho preditos por um modelo m . O agrupamento de trabalhadores gerado pelo modelo com máximo IA_m deve ser considerado.

$$IA_m = SI_m^a R_m^2 \quad m = 1, \dots, M \quad (9)$$

Tabela 1. Valores de R_m^2 , SI_m , SI_m^a e IA_m para os modelos de curva testados.

	R_m^2	SI_m	SI_m^a	IA_m
Curva Potencial	0,5283	0,8194	0,9097	0,4806
Plateau	0,5434	0,7683	0,8841	0,4804
Stanford-B	0,5339	0,8807	0,9403	0,5021
Hiperbólico 2	0,6702	0,8046	0,9023	0,6047
Hiperbólico 3	0,7279	0,9953	0,9976	0,7262
Exponencial 3	0,7315	0,9972	0,9986	0,7305

4. Resultados e discussão

Foram coletados dados de desempenho de $J = 22$ trabalhadores de uma indústria de calçados localizada no sul do país. A coleta de dados ocorreu no setor de costura, visto como o de maior complexidade em termos dos procedimentos demandados. Trabalhadores

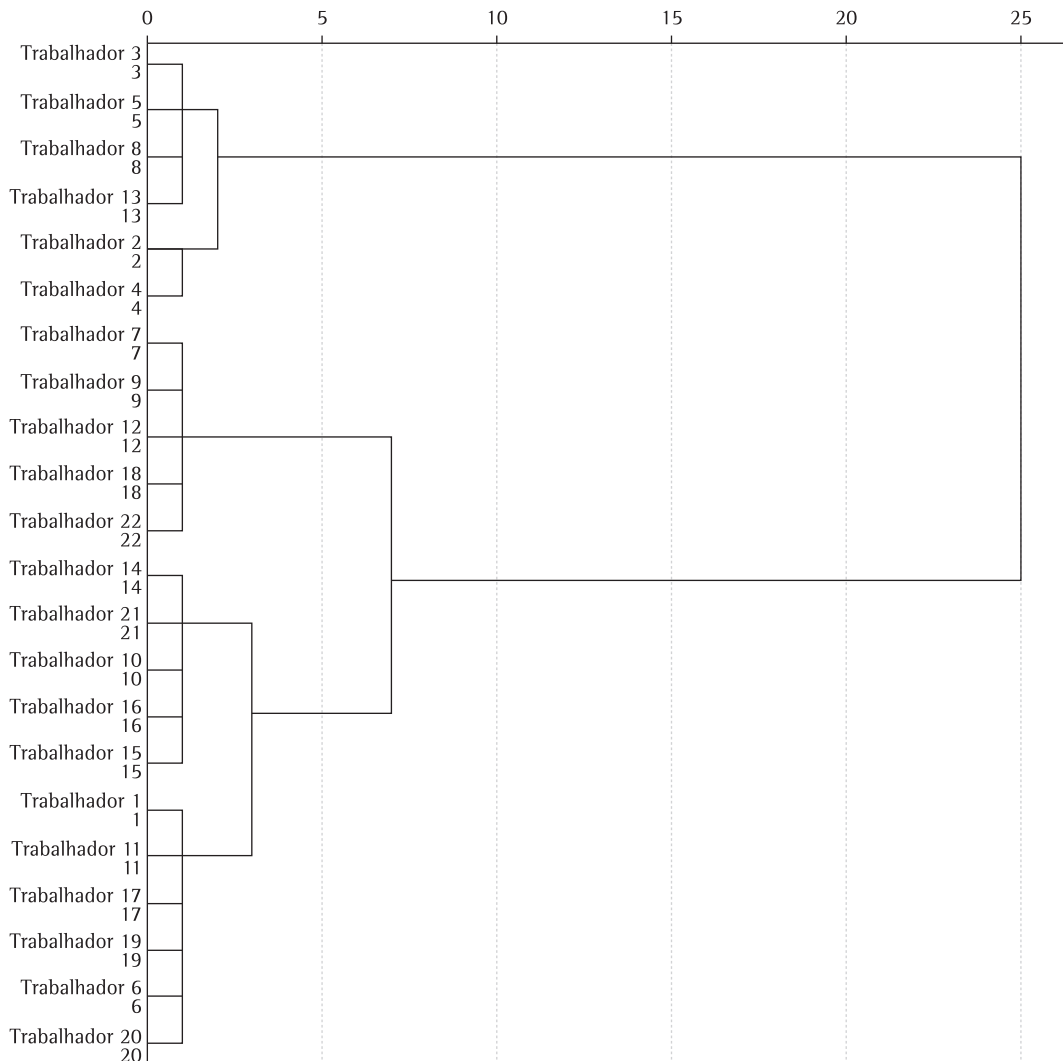


Figura 2. Dendrograma obtido pela clusterização do modelo Potencial.

com razoável experiência na operação foram considerados na coleta de dados. Os dados foram ajustados aos modelos de curvas nas Equações 1 a 6 utilizando *software* PASW Statistics 18. Os parâmetros e coeficientes R^2 obtidos para cada cruzamento de trabalhador e modelo de curva são apresentados no Anexo 1. Os valores consolidados de R_m^2 , gerados pela Equação 7, são apresentados na primeira coluna numérica da Tabela 1.

Os parâmetros dos M modelos geraram perfis distintos de desempenho para cada trabalhador. Para os modelos das Equações 1 a 3 foram preditos 140 dados de desempenho. Para os demais modelos foram preditos 22 dados.

Na sequência, rodou-se uma rotina de clusterização hierárquica sobre os dados preditos por cada modelo

de curva para a definição do melhor número de agrupamentos a serem criados. A análise dos seis dendogramas sugeriu a formação de dois clusters, como exemplificado para o modelo Potencial, na Figura 2.

A seguir, os perfis de desempenho preditos para os J trabalhadores foram clusterizados através da técnica não hierárquica k means, com $k = 2$. Os perfis foram agrupados por meio do aplicativo Matlab® versão 7.4, sendo calculado SJ_m para cada modelo de curva m (ver segunda coluna numérica da Tabela 1). O índice IA_m é gerado pela Equação 9.

A partir da análise dos valores de IA_m , selecionou-se o agrupamento de trabalhadores sugerido pelos dados preditos pelo modelo exponencial de 3 parâmetros (Exponencial 3). O referido modelo apresentou uma IA_m de 0,7305. Os *clusters* obtidos a partir dos valores preditos por esse modelo vêm apresentados na Tabela 2.

Tabela 2. Alocação dos trabalhadores aos *clusters*.

Indivíduo	Cluster alocado	Indivíduo	Cluster alocado
1	2	12	1
2	1	13	1
3	1	14	1
4	1	15	1
5	1	16	2
6	1	17	1
7	1	18	1
8	1	19	1
9	1	20	1
10	1	21	1
11	1	22	1

Os indivíduos 1 e 16, reunidos no *cluster 2*, apresentaram um comportamento similar entre si e distinto dos demais indivíduos alocados no primeiro *cluster*. Os perfis de aprendizado para tais trabalhadores, com base no modelo exponencial de 3 parâmetros (modelo escolhido para análise), são apresentadas na Figura 3. Existe nítida similaridade entre tais perfis. A título de comparação, o perfil do trabalhador 16 é confrontado com o perfil do trabalhador 19 (escolhido aleatoriamente), alocado no *cluster 2* (ver Figura 4). Existem diferenças significativas na velocidade de aprendizado e no patamar final

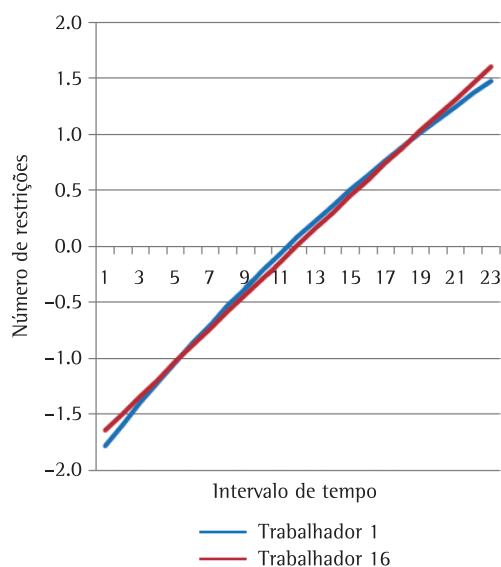


Figura 3. Perfil de aprendizado dos trabalhadores inseridos no *cluster 2*.

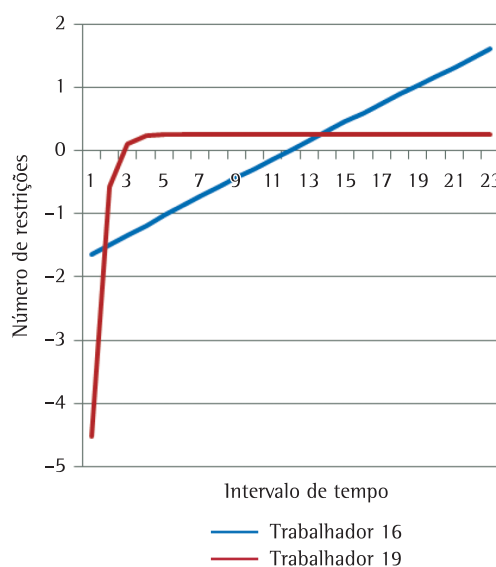


Figura 4. Comparação entre perfis de aprendizado de dois trabalhadores inseridos em *clusters* distintos.

de desempenho entre tais trabalhadores. A análise do conjunto total de curvas (não ilustrado) mostra que os trabalhadores do *cluster 2* convergem mais lentamente para o estado estacionário, indicando que ainda estão em processo de aprendizagem. Porém, uma vez atingindo o estado estacionário, tais trabalhadores apresentam produtividade mais elevada. Tal resultado corrobora os achados de Uzumeri e Nembhard (1998).

Com base nos resultados apresentados no anexo, pode-se observar ainda que os trabalhadores pertencentes ao *cluster 2* apresentam um valor médio de parâmetro p (parâmetro associado à experiência prévia) igual a 21,7. O valor médio de p dos indivíduos alocados no *cluster 1* é 33,3, sinalizando uma maior experiência prévia em relação aos trabalhadores do *cluster 1*.

No caso da indústria de calçados, tal informação é útil na alocação de tarefas aos trabalhadores. Os trabalhadores no *cluster 1*, caracterizados pelo rápido aprendizado, devem receber lotes de produção menores e ser submetidos a trocas constantes de lotes, já que adaptam-se com maior facilidade às tarefas. O contrário ocorre com os trabalhadores inseridos no *cluster 2*, aos quais devem ser alocados lotes de maior duração; tais trabalhadores convergem a um estado estacionário com uma taxa de produção mais elevada.

5. Conclusões

O presente artigo apresentou um método no qual curvas de aprendizado balizam a formação de grupos de trabalhadores com perfis similares de aprendizado. O método proposto envolve técnicas de clusterização e o Silhouette Index, juntamente com o coeficiente de determinação (R^2), para calcular o valor do índice de ajuste (IA) proposto. Esse índice permite identificar o melhor modelo de curva com base na aderência aos dados e qualidade das clusterizações geradas.

A principal contribuição teórica deste artigo reside na utilização de curvas de aprendizado e ferramentas de clusterização para a formação de grupos de trabalhadores com perfis semelhantes de aprendizado. Enfatiza-se ainda a proposição de um índice que avalia conjuntamente a qualidade dos agrupamentos gerados e a qualidade da aderência das curvas aos dados de desempenho. A contribuição prática da sistemática proposta está na redução dos impactos decorrentes da alocação de trabalhadores com perfis distintos de desempenho à mesma linha ou célula de produção.

Ao ser aplicada em dados de 22 trabalhadores de uma indústria de calçados, a sistemática proposta apontou a existência de 2 grupos de trabalhadores com

perfis distintos de aprendizado, modelados através da curva exponencial de três parâmetros. Trabalhadores alocados no *cluster 1* são caracterizados por rápido aprendizado e patamar inferior de desempenho, enquanto que trabalhadores alocados no *cluster 2* apresentam uma taxa de aprendizado menor porém atingem um patamar de desempenho maior no longo prazo.

Desdobramentos futuros incluem a utilização de técnicas de *data mining* para a classificação dos perfis de aprendizado dos trabalhadores em diversas classes de aprendizado. A utilização de medidas alternativas de aderência dos modelos aos dados também pode ser abordada com vistas ao desenvolvimento de índices de ajuste mais robustos.

Referências

- ADLER, P. S. Shared Learning. *Management Science*, v. 36, n. 8, p. 938-957, 1990. <http://dx.doi.org/10.1287/mnsc.36.8.938>
- ADLER, P. S.; CLARK, K. B. Behind the Learning Curve: A Sketch of the Learning Process. *Management Science*, v. 37, n. 3, p. 267-281, 1991. <http://dx.doi.org/10.1287/mnsc.37.3.267>
- ANZANELLO, M. J.; FOGLIATTO, F. S. Curvas de aprendizado: estado da arte e perspectivas de pesquisa. *Gestão & Produção*, v. 14, p. 109-123, 2007a. <http://dx.doi.org/10.1590/S0104-530X2007000100010>
- ANZANELLO, M. J.; FOGLIATTO, F. S. Learning curve modeling of work assignment in mass customized assembly lines. *International Journal of Production Research*, v. 45, n. 13, p. 2919-2938, 2007b. <http://dx.doi.org/10.1080/00207540600725010>
- ANZANELLO, M. J.; FOGLIATTO, F. S. Learning curve models and applications: Literature review and research directions. *International Journal of Industrial Ergonomics*, v. 41, p. 573-583, 2011. <http://dx.doi.org/10.1016/j.ergon.2011.05.001>
- ARGOTE, L. *Organizational Learning: Creating, Retaining and Transferring Knowledge*. New York: Springer, 1999.
- ASKIN, R.; GOLDBERG, J. *Design and Analysis of Lean Production Systems*. New York: John Wiley & Sons, 2001.
- BADIRU, A. B. Computational Survey of Univariate and Multivariate Learning Curve Models. *IEEE Transactions on Engineering Management*, v. 39, n. 2, p. 176-188, 1992. <http://dx.doi.org/10.1109/17.141275>
- BISKUP, D. A state-of-art review on scheduling with learning effects. *European Journal of Operational Research*, v. 188, p. 315-329, 2007. <http://dx.doi.org/10.1016/j.ejor.2007.05.040>
- COHEN, Y.; VITNER, G.; SARIN, S. Work allocation to stations with varying learning slopes and without buffers. *European Journal of Production Research*, v. 184, p. 797-801, 2008.
- COOK, J. A. Competitive Model of the Japanese Firm. *Journal of Policy Modeling*, v. 13, n. 1, p. 93-114, 1991. [http://dx.doi.org/10.1016/0161-8938\(91\)90004-1](http://dx.doi.org/10.1016/0161-8938(91)90004-1)
- COROMINAS, A.; OLIVELLA, J.; PASTOR, R. A model for the assignment of a set of tasks when work performance

- depends on experience of all tasks involved. *International Journal Production Economics*, v. 126, p. 335-340, 2010. <http://dx.doi.org/10.1016/j.ijpe.2010.04.012>
- FÁVERO, L. P. et al. *Análise de Dados - Modelagem Multivariada para Tomada de Decisões*. Rio de Janeiro: Editora Campus, 2009.
- JABER, M.; GUIFFRIDA, A. Learning Curves for Process Generating Defects Requiring Reworks. *European Journal of Production Research*, v. 159, n. 3, p. 663-672, 2004.
- MAZUR, J. E.; HASTIE, R. Learning as Accumulation: a Reexamination of the Learning Curve. *Psychological Bulletin*, v. 85, n. 6, p. 1256-1274, 1978. <http://dx.doi.org/10.1037/0033-2909.85.6.1256>
- NEMBHARD, D. A.; UZUMERI, M. V. Experimental learning and forgetting for manual and cognitive tasks. *International Journal of Industrial Ergonomics*, v. 25, n. 3, p. 315-326, 2000. [http://dx.doi.org/10.1016/S0169-8141\(99\)00021-9](http://dx.doi.org/10.1016/S0169-8141(99)00021-9)
- NGWENYAMA, O.; GUERGACHI, A.; MCLAREN, T. Using the learning curve to maximize IT productivity: A decision analysis model for timing software upgrades. *International Journal of Production Economics*, v. 105, p. 524-535, 2007. <http://dx.doi.org/10.1016/j.ijpe.2006.02.013>
- PROCHNO, P. Learning Curve? Which One? *Brazilian Administration Review*, v. 1, n. 1, p. 53-66, 2004. <http://dx.doi.org/10.1590/S1807-76922004000100005>
- PRUETT, M.; THOMAS, H. Experience-based learning in innovation and production. *R&D Management*, v. 38, n. 2, p. 141-153, 2008. <http://dx.doi.org/10.1111/j.1467-9310.2008.00504.x>
- ROHRBACH, A. Economical production of all-metal airplanes and seaplanes. *Journal of the Society of Automotive Engineers*, v. 20, p. 57-66, 1927.
- SÖDERHOLM, P.; SUNDQVIST, T. Empirical challenges in the use of learning curves for assessing the economic prospects of renewable energy technologies. *Renewable Energy*, v. 32, p. 2559-2578, 2007. <http://dx.doi.org/10.1016/j.renene.2006.12.007>
- TEPLITZ, C. J. *The Learning Curve Deskbook: A reference Guide to Theory, Calculations and Applications*. New York: Quorum Books, 1991.
- TOKSARI, M. D. et al. Simple and U-type line balancing problems with a learning effect. *Applied Mathematical Modelling*, v. 32, p. 2954-2961, 2008. <http://dx.doi.org/10.1016/j.apm.2007.10.007>
- UZUMERI, M.; NEMBHARD, D. A Population of Learners: A New Way to Measure Organizational Learning. *Journal of Operations Management*, v. 16, n. 5, p. 515-528, 1998. [http://dx.doi.org/10.1016/S0272-6963\(97\)00017-X](http://dx.doi.org/10.1016/S0272-6963(97)00017-X)
- WRIGHT, T. P. Factors Affecting the Cost of Airplanes. *Journal of the Aeronautical Sciences*, v. 3, p. 122-128, 1936.

Cluster analysis of learning curves for grouping workers with homogeneous learning profiles

Abstract

In many industrial segments, it is desirable to allocate workers with similar learning profiles in the same workstation. This paper presents a method that groups workers based on learning curve modeling and clustering techniques. Workers' performance data are modeled through several learning curve models; learning parameters allow for workers' performance prediction at intervals of predetermined time. The predicted values are then grouped by clustering techniques. The largest Adjustment Index (AI), derived from the Silhouette Index and Coefficient of Determination, indicates the model yielding superior adherence to data and better clustering of learning profiles. When applied to a shoe manufacturing process, the method generated consistent groups of workers based on their learning profiles.

Keywords

Learning curves. Clustering. Groups of workers.

Anexo 1. Parâmetros e coeficientes R^2 para os cruzamentos de trabalhadores e modelos de curvas.

	Curva Potencial	Plateau	Stanford-B	Hiperbólico de 2	Hiperbólico de 3	Exponencial de 3
				Parâmetros	Parâmetros	Parâmetros
Trabalhador 1	C = 35,970 b = -0,127 $R^2 = 0,39$	C = 24,758 b = -0,528 B = -18,370 $R^2 = 0,424$	C = 35,970 b = -0,127 B = 0 $R^2 = 0,39$	k = 32,091 r = 4,843 $R^2 = 0,55$	k = 2248,688 r = 849,729 p = 9,908 $R^2 = 0,809$	k = 47263,482 r = 22493,639 p = 9,854 $R^2 = 0,711$
Trabalhador 2	C = 184,766 b = -0,283 $R^2 = 0,822$	C = 184,766 b = -0,283 B = 0 R = 0,82 $R^2 = 0,82$	C = 220,803 b = -0,331 B = 1,126 R = 0,831 $R^2 = 0,831$	k = 15,749 r = 41,039 $R^2 = 0,88$	k = 20,382 r = 81,716 p = 7,712 R = 0,886 $R^2 = 0,886$	k = 15,303 r = 78,078 p = 11,981 R = 0,884 $R^2 = 0,884$
Trabalhador 3	C = 218,933 b = -0,237 $R^2 = 0,768$	C = 218,933 b = -0,237 B = 0 R = 0,768 $R^2 = 0,768$	C = 360,242 b = -0,365 B = 4,141 R = 0,764 $R^2 = 0,764$	k = 9,323 r = 29,525 $R^2 = 0,848$	k = 10,442 r = 49,406 p = 8,885 R = 0,858 $R^2 = 0,858$	k = 8,681 r = 67,604 p = 19,8 R = 0,848 $R^2 = 0,848$
Trabalhador 4	C = 246,666 b = -0,273 $R^2 = 0,795$	C = 227,112 b = -0,364 B = 30,454 R = 0,80 $R^2 = 0,80$	C = 258,151 b = -0,285 B = 0,203 R = 0,824 $R^2 = 0,824$	k = 9,723 r = 33,556 $R^2 = 0,89$	k = 9,615 r = 3,292 p = 0 R = 0,894 $R^2 = 0,894$	k = 7,822 r = 35,96 p = 1,056 R = 0,879 $R^2 = 0,879$
Trabalhador 5	C = 158,754 b = -0,236 $R^2 = 0,745$	C = 158,744 b = -0,236 B = 0 R = 0,745 $R^2 = 0,745$	C = 312,888 b = -0,406 B = 6,839 R = 0,796 $R^2 = 0,796$	k = 13,478 r = 27,933 $R^2 = 0,899$	k = 15,116 r = 42,577 p = 5,103 R = 0,906 $R^2 = 0,906$	k = 11,455 r = 41,624 p = 7,608 R = 0,91 $R^2 = 0,91$
Trabalhador 6	C = 62,356 b = -0,096 $R^2 = 0,477$	C = 62,356 b = -0,096 B = 0 R = 0,477 $R^2 = 0,477$	C = 67,506 b = -0,110 B = 3,955 R = 0,387 $R^2 = 0,387$	k = 16,04 r = 5,495 $R^2 = 0,681$	k = 32,161 r = 196,583 p = 98,91 R = 0,804 $R^2 = 0,804$	k = 24,102 r = 193,011 p = 114,968 R = 0,803 $R^2 = 0,803$
Trabalhador 7	C = 237,797 b = -0,153 $R^2 = 0,459$	C = 222,256 b = -0,171 B = 16,914 R = 0,459 $R^2 = 0,459$	C = 248,485 b = -0,165 B = 0,553 R = 0,451 $R^2 = 0,451$	k = 5,257 r = 15,622 $R^2 = 0,457$	k = 5,23 r = 1,543 p = 0 R = 0,458 $R^2 = 0,458$	k = 4,723 r = 28,906 p = 6,798 R = 0,448 $R^2 = 0,448$
Trabalhador 8	C = 190,752 b = -0,249 $R^2 = 0,796$	C = 190,752 b = -0,249 B = 0 R = 0,796 $R^2 = 0,796$	C = 314,890 b = -0,379 B = 4,281 R = 0,843 $R^2 = 0,843$	k = 15,46 r = 53,386 $R^2 = 0,87$	k = 31,443 r = 240,207 p = 20,262 R = 0,911 $R^2 = 0,911$	k = 18,635 r = 146,029 p = 20,413 R = 0,911 $R^2 = 0,911$
Trabalhador 9	C = 90,488 b = -0,151 $R^2 = 0,629$	C = 90,535 b = -0,156 B = 0,01 R = 0,74 $R^2 = 0,74$	C = 99,184 b = -0,174 B = 1,357 R = 0,614 $R^2 = 0,614$	k = 13,962 r = 8,401 $R^2 = 0,762$	k = 17,598 r = 36,755 p = 22,862 R = 0,804 $R^2 = 0,804$	k = 14,782 r = 55,09 p = 36,102 R = 0,802 $R^2 = 0,802$
Trabalhador 10	C = 37,317 b = -0,064 $R^2 = 0,192$	C = 14,271 b = -0,384 B = 25,857 R = 0,20 $R^2 = 0,20$	C = 37,317 b = -0,064 B = 0,064 R = 0,192 $R^2 = 0,192$	k = 22,487 r = 2,583 $R^2 = 0,896$	k = 27,083 r = 19,326 p = 29,318 R = 0,934 $R^2 = 0,934$	k = 24,437 r = 46,247 p = 53,177 R = 0,931 $R^2 = 0,931$
Trabalhador 11	C = 162,039 b = -0,134 $R^2 = 0,481$	C = 115,877 b = -0,293 B = 58,095 R = 0,487 $R^2 = 0,487$	C = 795,676 b = -0,429 B = 77,205 R = 0,375 $R^2 = 0,375$	k = 7,158 r = 10,663 $R^2 = 0,669$	k = 7,535 r = 19,366 p = 11,105 R = 0,69 $R^2 = 0,69$	k = 6,94 r = 50,783 p = 35,369 R = 0,685 $R^2 = 0,685$
Trabalhador 12	C = 375,312 b = -0,170 $R^2 = 0,635$	C = 377,526 b = -0,172 B = 0,1 R = 0,697 $R^2 = 0,697$	C = 417,531 b = -0,2 B = 1,161 R = 0,641 $R^2 = 0,641$	k = 3,567 r = 23,917 $R^2 = 0,501$	k = 3,939 r = 43,853 p = 12,225 R = 0,507 $R^2 = 0,507$	k = 3,954 r = 136,667 p = 61,7 R = 0,496 $R^2 = 0,496$
Trabalhador 13	C = 334,946 b = -0,257 $R^2 = 0,740$	C = 334,946 b = -0,257 B = 0 R = 0,74 $R^2 = 0,74$	C = 1037,856 b = -0,529 B = 11,556 R = 0,790 $R^2 = 0,790$	k = 7,631 r = 66,907 $R^2 = 0,812$	k = 13,662 r = 299,103 p = 40,165 R = 0,88 $R^2 = 0,88$	k = 8,625 r = 199,691 p = 41,208 R = 0,881 $R^2 = 0,881$
Trabalhador 14	C = 118,185 b = -0,077 $R^2 = 0,348$	C = 118,185 b = -0,077 B = 0 R = 0,348 $R^2 = 0,348$	C = 150,341 b = -0,131 B = 11,681 R = 0,270 $R^2 = 0,270$	k = 7,393 r = 5,296 $R^2 = 0,539$	k = 7,668 r = 10,322 p = 10,739 R = 0,555 $R^2 = 0,555$	k = 7,288 r = 42,263 p = 44,299 R = 0,544 $R^2 = 0,544$

Anexo 1. Continuação...

	Curva Potencial	Plateau	Stanford-B	Hiperbólico de 2	Hiperbólico de 3	Exponencial de 3
				Parâmetros	Parâmetros	Parâmetros
Trabalhador 15	C = 119,660 b = -0,052 R ² = 0,206	C = 32,416 b = -0,563 B = 95,516 R ² = 0,243	C = 119,660 b = -0,052 B = -0 R ² = 0,206	k = 6,571 r = 3,733 R ² = 0,593	k = 6,878 r = 7,316 p = 7,82 R ² = 0,613	k = 6,357 r = 20,569 p = 19,51 R ² = 0,637
Trabalhador 16	C = 62,057 b = -0,061 R ² = 0,116	C = 62,057 b = -0,061 B = 0 R ² = 0,116	C = 7566,831 b = -0,873 B = 264,602 R ² = 0,346	k = 13,109 r = 2,952 R ² = 0,201	k = 117,397 r = 278,426 p = 26,868 R ² = 0,545	k = 44655,744 r = 146055,696 p = 33,662 R ² = 0,538
Trabalhador 17	C = 59,398 b = -0,112 R ² = 0,432	C = 35,947 b = -0,278 B = 26,311 R ² = 0,441	C = 59,398 b = -0,112 B = 0 R ² = 0,432	k = 17,32 r = 4,733 R ² = 0,766	k = 17,32 r = 0,473 p = 0 R ² = 0,766	k = 73,533 r = 76,578 p = 14,443 R ² = 0,732
Trabalhador 18	C = 191,916 b = -0,165 R ² = 0,548	C = 191,916 b = -0,165 B = 0 R ² = 0,567	C = 212,044 b = -0,183 B = 1,966 R ² = 0,518	k = 6,982 r = 14,893 R ² = 0,582	k = 7,928 r = 35,434 p = 15,674 R ² = 0,595	k = 7,807 r = 108,661 p = 60,064 R ² = 0,585
Trabalhador 19	C = 202,079 b = -0,116 R ² = 0,611	C = 202,079 b = -0,116 B = 0 R ² = 0,611	C = 213,802 b = -0,123 B = 2,436 R ² = 0,594	k = 5,253 r = 12,164 R ² = 0,562	k = 8,573 r = 19,607 p = 11,217 R ² = 0,7	k = 35,493 r = 342,764 p = 33,055 R ² = 0,68
Trabalhador 20	C = 155,834 b = -0,105 R ² = 0,331	C = 76,572 b = -0,497 B = 94,302 R ² = 0,347	C = 155,834 b = -0,105 B = 0 R ² = 0,331	k = 6,095 r = 6,347 R ² = 0,445	k = 6,372 r = 10,721 p = 3,335 R ² = 0,454	k = 5,763 r = 23,149 p = 15,567 R ² = 0,747
Trabalhador 21	C = 205,844 b = -0,080 R ² = 0,315	C = 13760,454 b = -0,001 B = -13555,925 R ² = 0,326	C = 4028,995 b = -0,658 B = 105,663 R ² = 0,365	k = 4,066 r = 5,492 R ² = 0,461	k = 4,874 r = 37,109 p = 44,962 R ² = 0,562	k = 4,248 r = 68,231 p = 68,832 R ² = 0,565
Trabalhador 22	C = 101,434 b = -0,183 R ² = 0,786	C = 79,118 b = -0,353 B = 29,694 R ² = 0,802	C = 101,434 b = -0,183 B = 0 R ² = 0,786	k = 16,096 r = 14,428 R ² = 0,881	k = 16,096 r = 1,443 p = 0 R ² = 0,882	k = 13,335 r = 16,7 p = 0,473 R ² = 0,876