

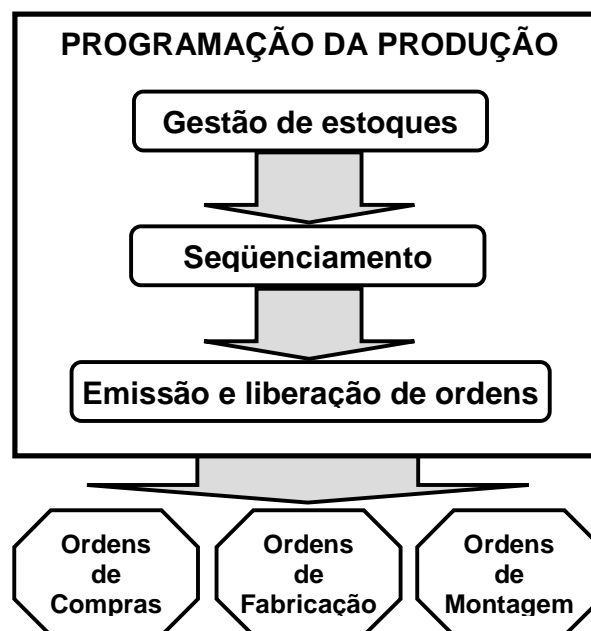
Cap. IV – SEQÜENCIAMENTO E EMISSÃO DE ORDENS

Como já vimos, as atividades de programação da produção, realizadas pelo PCP, buscam implementar um programa de produção que atenda ao PMP (Planejamento-Mestre da Produção).

Estas atividades podem ser divididas em três grupos hierarquicamente relacionados:

- A gestão dos estoques
- O seqüenciamento das ordens
- A emissão e liberação das ordens

Hierarquia das funções da programação da produção



A administração de estoques é encarregada de planejar e controlar os estoques, definindo:

- Tamanho de lotes
- Modelos de reposição
- Estoque de segurança

Escolhida uma sistemática de gestão dos estoques, são geradas – de forma direta ou indireta – as necessidades de compra, fabricação e montagem dos itens para atender o PMP.

Geralmente, as **ordens de compra**, uma vez geradas, vão para o setor encarregado das compras, fora da esfera do PCP.

Já as **necessidades de fabricação e montagem** precisam passar por um sistema produtivo com limitações de capacidade.

A adequação do programa gerado aos recursos (instalações, máquinas, homens, etc.) é **função do seqüenciamento**.

Uma vez estabelecidas todas as informações necessárias à execução do programa de produção, ou seja, definição para cada ordem da:

- a especificação do item;
- o tamanho do lote;
- a data de início e de conclusão das atividades; e,
- a seqüência e o local onde serão executadas,

o PCP pode partir para a emissão e liberação do programa de produção.

Emitido e liberado, este programa de produção passará para a esfera do acompanhamento (controle) da produção, última etapa das funções do PCP.

SEQÜENCIAMENTO E EMISSÃO DE ORDENS

As atividades da programação da produção são dependentes de como o sistema produtivo está projetado: **para empurrar ou puxar a produção**.

Nos sistemas de puxar a produção, normalmente implementados com o *kanban*, as atividades de programação da produção são deixadas a cargo dos setores de produção.

Já nos sistemas convencionais de empurrar a produção, há a necessidade de definir a cada programa de produção sua seqüência, baseada em critérios predeterminados, e emitir ordens autorizando a compra, fabricação e montagem dos itens.

Trataremos do seqüenciamento e emissão de ordens sob a ótica convencional, segundo TUBINO (2000).

Os problemas da programação da produção são apresentados, convencionalmente, em função do tipo do sistema (ou do subsistema) de produção, ou seja:

- Processos contínuos
- Processos repetitivos em massa
- Processos repetitivos em lote
- Processos por projeto

1. SEQÜENCIAMENTO NOS PROCESSOS CONTÍNUOS

Os processos contínuos de produção são empregados para produtos que não podem ser identificados individualmente, com altíssima uniformidade na produção e demanda, onde os produtos e os processos são totalmente dependentes.

Assim, é economicamente viável estruturar um sistema produtivo em grande escala, direcionado para o tipo de produto que se pretende produzir, permitindo sua automatização, por ex. refinaria de petróleo.

Como os processos contínuos se destinam à produção de poucos itens, normalmente um por instalação, não existem problemas de seqüenciamento quanto à ordem de execução das atividades.

Os problemas de programação resumem-se à definição da velocidade que será dada ao sistema produtivo para atender determinada demanda estabelecida do PMP.

Caso mais de um produto seja produzido na mesma instalação, procura-se atender o PMP com lotes únicos de cada item.

2. SEQÜENCIAMENTO NOS PROCESSOS REPETITIVOS EM MASSA

Os processos repetitivos em massa, à semelhança dos processos contínuos, são empregados na produção em grande escala de produtos altamente padronizados, porém identificáveis individualmente, por ex. eletrodomésticos.

Geralmente a montagem de processos repetitivos em massa exige produtos com demandas grandes e estáveis, com poucas alterações de curto prazo nos seus projetos, permitindo instalações altamente especializadas e pouco flexíveis.

Nestes sistemas procura-se trabalhar com o máximo de padronizações dos itens componentes, sendo diferenciados os produtos na composição da montagem final, garantindo alta taxa de produção.

O trabalho da programação da produção consiste em buscar um ritmo equilibrado entre os vários postos de trabalho, principalmente nas linhas de montagem, conhecido como **balanceamento de linha**, de forma a atender economicamente uma taxa de demanda, expressa em termos de tempo de ciclo de trabalho.

Exemplo:

Um produto é montado em uma linha que trabalha 480 minutos por dia (8 horas) a partir de 6 operações seqüenciais, com os seguintes tempos unitários:

Operação 1	Operação 2	Operação 3	Operação 4	Operação 5	Operação 6
0,8 min	1,0 min	0,5 min	1,0 min	0,5 min	0,7 min

Para se balancear uma linha, a primeira informação importante é a definição dos limites técnicos da capacidade de produção, em termos de tempo de ciclo, desta linha.

O limite superior da capacidade de produção é obtido empregando-se como tempo de ciclo o maior tempo unitário de operação:

$$\text{Limite superior} = 1,0 \text{ min por unidade.}$$

O limite inferior é obtido empregando-se a soma dos tempos das 6 operações:

$$\text{Limite inferior} = 4,5 \text{ min por unidade.}$$

Fórmula para o cálculo da capacidade de produção é:

$$\mathbf{CP = TP / TC}$$

Onde: CP = Capacidade de produção por dia

TP = Tempo disponível para a produção por dia

TC = Tempo de ciclo por unidade

Logo:

$$CP_{\text{superior}} = 480 \text{ min por dia} / 1,0 \text{ min por dia} = 480 \text{ unidades por dia}$$

$$CP_{\text{inferior}} = 480 \text{ min por dia} / 4,5 \text{ min por unidade} = 106,6 \approx 106 \text{ unidades por dia}$$

Assim, no exemplo, temos que a capacidade de produção pode variar entre:

- 106 unidades por dia, com apenas um posto de trabalho executando todas as atividades; e,
- 480 unidades por dia, com 6 postos de trabalho, cada um executando uma atividade.

No entanto, o tempo de ciclo no qual devemos operar é função do tempo disponível para a produção por dia dividido pela taxa de demanda (**D**) esperada por dia (proveniente do PMP):

$$\mathbf{TC = TP / D}$$

Admitindo-se que a taxa de demanda esperada seja de 240 unidades por dia, teremos:

$$TC = 480 \text{ min por dia} / 240 \text{ unidades por dia} = 2,0 \text{ min por unidade}$$

O número de postos de trabalho necessários para suportar uma demanda de 240 unidades por dia, com ritmo de 2,0 min por unidade, será função da forma como combinaremos as atividades individuais em grupos de no máximo 2 min de tempo.

Teoricamente, podemos calcular o número mínimo de postos de trabalho para atender uma determinada taxa de demanda, ou tempo de ciclo, da seguinte forma:

$$\mathbf{N_{\text{mínimo}} = \sum t / TC}$$

Onde: $N_{\text{mínimo}}$ = Número mínimo de postos de trabalho

t = Tempo de cada operação

No exemplo:

$$N_{\text{mínimo}} = 4,5 \text{ min por unidade} / 2,0 \text{ min por unidade} = 2,25 \text{ postos}$$

Como não existe posto de trabalho fracionado, este número deve ser arredondado para 3 postos.

Uma alternativa de composição destes 3 postos seria:

- Posto 1 = operação 1 + operação 2 = 0,8 + 1,0 = 1,8 min
- Posto 2 = operação 3 + operação 4 = 0,5 + 1,0 = 1,5 min
- Posto 3 = operação 5 + operação 6 = 0,5 + 0,7 = 1,2 min

Existem outras alternativas de composição. No exemplo, que é simples, poderíamos listar todas as alternativas e escolher a melhor.

Na prática, a situação pode ser bem mais complexa, existindo uma gama bem maior de atividades nas linhas de montagem, com limitações físicas associadas ao *layout* e equipamentos, incompatibilidades entre operações e fatores humanos.

Apesar de, teoricamente, este ser um problema que pode ser resolvido com um pacote computacional de programação dinâmica, na prática, normalmente, as empresas não se dispõem a implementar programas nesse sentido, sendo preferido empregar heurísticas que busquem boas e rápidas soluções.

Uma das regras mais simples empregadas consiste em, definido o tempo do ciclo, iniciarmos com a 1ª operação e irmos associando a ela tantas operações quanto possível, até o limite do tempo de ciclo. Uma vez fechada esta estação passamos à seguinte e assim por diante.

Logicamente, quanto maior a possibilidade de combinar tarefas distintas, mais perto do tempo de ciclo chegaremos.

A eficiência da alternativa adotada é avaliada em função ao tempo livre gerado, assim calculado:

$$I_{\text{eficiência}} = 1 - \frac{\sum \text{tempo livre}}{N \times TC}$$

Onde: $I_{\text{eficiência}}$ = Índice de eficiência da alternativa

Tempo livre = tempo de ciclo menos o tempo de cada posto

N = número de postos de trabalho

No exemplo:

$$I_{\text{eficiência}} = 1 - \frac{(2,0 - 1,8) + (2,0 - 1,5) + (2,0 - 1,2)}{3 \times 2,0} = 0,75 \text{ ou } 75\%$$

Ou seja, em média os postos de trabalho estarão ocupados 75% do tempo.

3. SEQÜENCIAMENTO NOS PROCESSOS REPETITIVOS EM LOTE

Os processos repetitivos em lote caracterizam-se pela produção de um volume médio de itens padronizados em lotes, onde cada lote segue uma série de operações que necessita ser programada à medida que as operações anteriores sejam concluídas.

Estes sistemas produtivos são relativamente flexíveis, empregando equipamentos menos especializados, que permitem, em conjunto com funcionários polivalentes, atender a diferentes volumes e variedades de pedidos dos clientes.

Os processos repetitivos em lotes situam-se entre os extremos da produção em massa e da produção sob projeto.

A quantidade de produtos é insuficiente para justificar a massificação da produção e a especialização das instalações, porém justifica a montagem de lotes repetitivos e a manutenção de estoques para absorver os custos de preparação dos equipamentos.

Desta forma, o sistema de gestão de estoques define a quantidade e o momento em que os itens são necessários, **cabendo ao seqüenciamento definir as prioridades na alocação dos recursos.**

A questão do seqüenciamento em processos repetitivos em lotes pode ser analisada sob dois aspectos:

- A escolha da ordem a ser processada dentre uma lista de ordens (decisão 1); e,
- A escolha do recurso a ser usado dentre uma lista de recursos disponíveis (decisão 2).

A **1ª decisão** (escolha da ordem a ser processada dentre uma fila de espera de ordens a processar) se resume ao estabelecimento de prioridades entre os diversos lotes de fabricação concorrentes por um mesmo grupo de recursos, no sentido de atender a determinados objetivos.

Conforme os objetivos que se pretendem atingir, regras de decisão diferentes podem ser utilizadas.

Nos processos repetitivos em lotes esta decisão é crítica para o desempenho do sistema produtivo, pois, via de regra, a maior parcela do lead time de um produto fabricado em lotes compreende o tempo em que o lote espera nas filas dos recursos para ser trabalhado.

Desta forma, ganhos resultantes de um bom seqüenciamento têm fator multiplicador no desempenho do sistema, pois teremos *lead times* padrões previstos mais perto dos *lead times* reais, reduzindo a margem de erro do programa executado em relação ao planejado.

A **2ª decisão** diz respeito à escolha do recurso a ser utilizado dentre os recursos disponíveis.

Na prática esta fica restrita às situações onde existem variações significativas no desempenho dos equipamentos, seja no tempo de processamento ou de *setup*.

Geralmente, nos processos repetitivos em lotes, quanto maior o volume de produção e, conseqüentemente, a repetição na programação dos lotes, a decisão quanto a que recurso usar é estabelecida na etapa do projeto do sistema produtivo.

À medida que o sistema de produção se aproxima da produção sob projeto, onde o grau de repetição de lotes de um mesmo item diminui, fazendo com que a seqüência de produção se altere a cada novo pedido, a decisão sobre que recurso escolher é mais premente.

GRÁFICO DE GANTT

Instrumento de visualização para a programação da produção, chamado de gráfico de Gantt em homenagem a Henry Gantt, pioneiro na sua utilização.

Auxilia na análise de diferentes alternativas de seqüenciamento.

O gráfico de Gantt pode ser empregado de diferentes formas, sendo que uma das mais comuns consiste em listar as ordens programadas no eixo vertical e o tempo no eixo horizontal.

Vários *softwares* comerciais geram o gráfico de Gantt.

3.1. Regras de seqüenciamento

As regras de seqüenciamento são heurísticas usadas para selecionar, a partir de informações sobre os lotes e/ou sobre o estado do sistema produtivo, qual dos lotes esperando na fila de um grupo de recursos terá prioridade de processamento.

Geralmente as informações mais importantes são:

- Tempo de processamento (*lead time*); e,
- Data de entrega

Estas regras normalmente assumem que os tempos e custos dos *setups* são independentes da seqüência escolhida, e são adicionados ao tempo de processamento do lote.

Geralmente a eficiência de um seqüenciamento é medida em termos de três fatores:

- *Lead time* médio;
- Atraso médio; e,
- Estoque em processo médio.

Porém nada substitui um bom PMP e a utilização equilibrada dos recursos produtivos. Se não tivermos capacidade de produção suficiente, não será uma boa regra de decisão que resolverá o problema de atendimento ao cliente.

As regras mais empregadas na prática são apresentadas na tabela a seguir:

Sigla	Especificação	Definição
PEPS	Primeira que entra primeira que sai	Os lotes serão processados de acordo com sua chegada no recurso
MTP	Menor tempo de processamento	Os lotes serão processados de acordo com o menor tempo de processamento no recurso
MDE	Menor data de entrega	Os lotes serão processados de acordo com as menores datas de entrega.
IPI	Índice de prioridade	Os lotes serão processados de acordo com o valor da prioridade atribuída ao cliente ou ao produto.
ICR	Índice crítico	Os lotes serão processados de acordo com o menor valor de: (data de entrega – data atual) / tempo de processamento
IFO	Índice de folga	Os lotes serão processados de acordo com o menor valor de: $\frac{\text{(data de entrega - } \sum \text{ tempo de processamento restante)}}{\text{número de operações restante}}$
IFA	Índice de falta	Os lotes serão processados de acordo com o menor valor de: Quantidade em estoque / taxa de demanda

Outra regra de seqüenciamento é a heurística conhecida como “**Regra de Johnson**”, que minimiza o *lead time* total de um conjunto de ordens processadas em dois recursos sucessivos (máquina A e máquina B) desde que as seguintes condições sejam satisfeitas:

- Os tempos de processamento das ordens (incluindo os *setups*) devem ser conhecidos e constantes, bem como independentes da seqüência de processamento escolhida;
- Todas as ordens são processadas na mesma direção, da máquina A para a B;
- Não existem prioridades; e,
- As ordens são transferidas de uma máquina para outra apenas quando completadas.

Uma vez cumpridas essas condições, a determinação da seqüência pela regra de Johnson segue os seguintes passos:

1º. Selecione o menor tempo entre todos os tempos de processamento da lista de ordens a serem processadas nas máquinas A e B (no caso de empate escolha qualquer uma).

2º. Se o tempo escolhido for na máquina A, programe esta ordem no início. Se o tempo escolhido for na B, programe esta ordem para o final.

3º. Elimine a ordem escolhida da lista de ordens a serem processadas e retorne ao 1º passo até programar todas as ordens.

Exemplo:

Cinco ordens de fabricação precisam ser estampadas na máquina A e, em seguida, usinadas na máquina B. Os tempos de processamento, incluindo *setups*, e as prioridades atribuídas são apresentadas na tabela. Para aplicação da regra PEPS, consideraremos que as ordens deram entradas em carteira na seqüência de sua numeração.

Ordens de Fabricação	Processamento (horas)		Entrega (horas)	Prioridade
	Máquina A	Máquina B		
OF 1	5	5	15	4
OF 2	8	6	20	1
OF 3	4	5	13	3
OF 4	2	4	10	2
OF 5	4	3	9	5

A tabela a seguir mostra as seqüências obtidas pelas regras apresentadas, exceto a regra do índice de falta (IFA) por falta de dados de estoque.

Regra	Seqüências
PEPS	OF1 – OF2 – OF3 – OF4 – OF5
MTP	OF4 – OF5 – OF3 – OF1 – OF2
MDE	OF5 – OF4 – OF3 – OF1 – OF2
IPI	OF2 – OF4 – OF3 – OF1 – OF5
ICR	OF5 – OF2 – OF3 – OF1 – OF4
IFO	OF5 – OF3 – OF4 – OF1 – OF2
Johnson	OF4 – OF3 – OF1 – OF2 – OF5

Os gráficos de Gantt para as estas sete seqüências são mostrados a seguir:

Gráfico de Gantt para a Regra PEPS

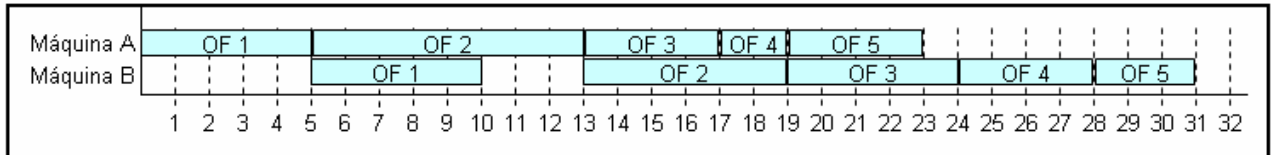


Gráfico de Gantt para a Regra MTP

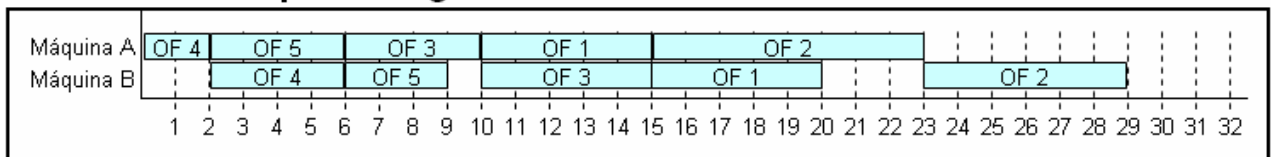


Gráfico de Gantt para a Regra MDE

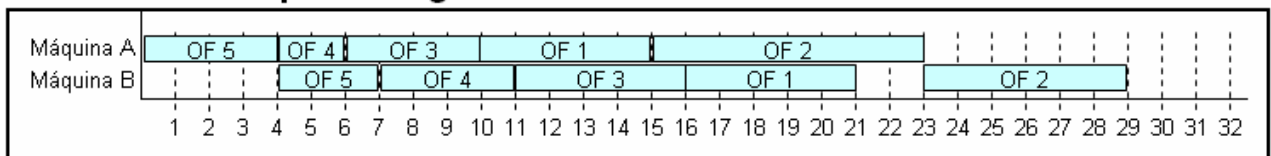


Gráfico de Gantt para a Regra IPI

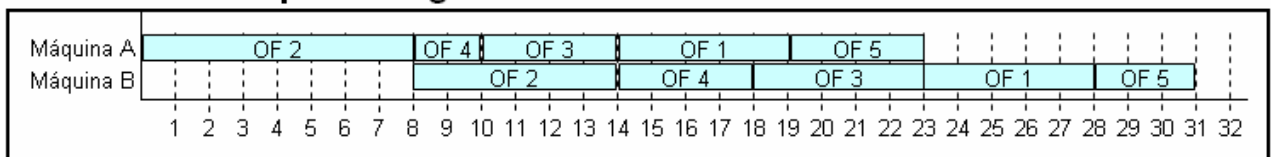


Gráfico de Gantt para a Regra ICR

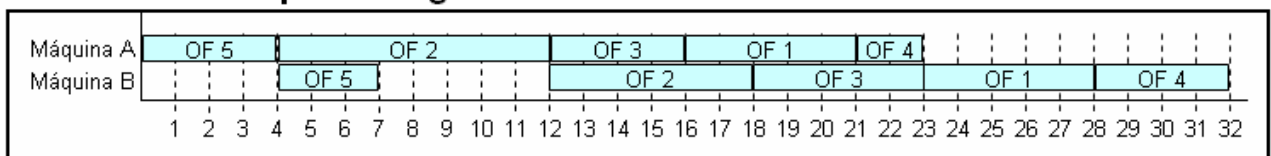


Gráfico de Gantt para a Regra IFO

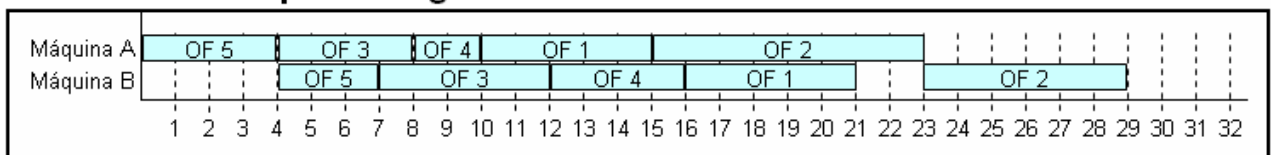
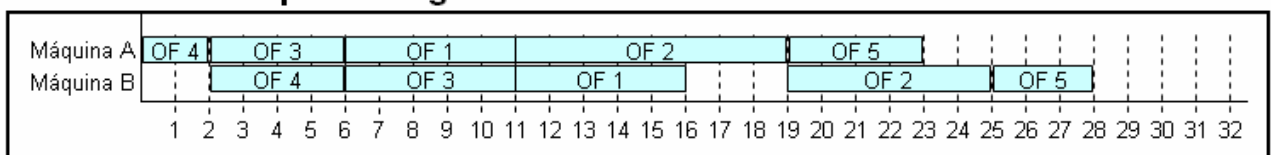


Gráfico de Gantt para a Regra de Johnson



A tabela abaixo resume o desempenho de cada regra no exemplo:

Regra	Lead Time Total (h)	Lead Time Médio (h)	Atraso Médio (h)	Tempo de Espera Médio (h)
PEPS	31	$31 / 5 = 6,2$	$(0+0+11+18+22) / 5 = 10,2$	$(0+0+2+5+5) / 5 = 2,4$
MTP	29	$29 / 5 = 5,8$	$(0+0+2+5+9) / 5 = 3,2$	$(0+0+0+0+0) / 5 = 0$
MDE	29	$29 / 5 = 5,8$	$(0+1+3+6+9) / 5 = 3,8$	$(0+1+1+1+0) / 5 = 0,6$
IPI	31	$31 / 5 = 6,2$	$(0+8+10+13+22) / 5 = 10,6$	$(0+4+4+4+5) / 5 = 4,2$
ICR	32	$32 / 5 = 6,4$	$(0+0+10+13+22) / 5 = 9,0$	$(0+0+2+2+5) / 5 = 1,8$
IFO	29	$29 / 5 = 5,8$	$(0+0+7+7+9) / 5 = 4,6$	$(0+0+3+2+0) / 5 = 1,0$
Johnson	28	$28 / 5 = 5,6$	$(0+0+1+5+19) / 5 = 5,0$	$(0+0+0+0+2) / 5 = 0,4$

COMENTÁRIOS:

Regra PEPS:

- É a mais simples, sendo pouco eficiente.
- É muito empregada quando o cliente está presente.
- Faz com que lotes com tempo grande retardem toda a seqüência de produção, gerando tempo ocioso nos processos à frente, fazendo com que o tempo de espera médio seja elevado (2,4 h).

Regra MTP

- Obtém um índice de *lead time* médio baixo, reduzindo os estoques em processo, agilizando o carregamento das máquinas à frente e melhorando o nível de atendimento ao cliente.
- No exemplo foi a regra com o melhor desempenho global, perdendo apenas para a regra de Johnson no que se refere ao *lead time*.
- Como ponto negativo, a regra MTP faz com que as ordens com tempos de processamento longos sejam sempre preteridas, principalmente se for grande a dinâmica de chegada de novas ordens com tempos menores.
- Uma solução para este problema seria associarmos uma regra complementar que possibilitasse à uma ordem que fosse preterida um determinado número de vezes, ou após um determinado tempo, avançar para o topo da lista.

Regra MDE

- Como prioriza as datas de entrega dos lotes, faz com que os atrasos se reduzam, o que é conveniente nos processos sob encomenda.
- Porém, como não leva em consideração o tempo de processamento, pode fazer com que lotes com potencial de conclusão rápido fiquem aguardando.
- Nos processos repetitivos em lotes, onde trabalhamos com estoques, as vantagens em priorizar apenas as datas de entrega não são muito claras.

Regra IPI

- Baseada em atribuirmos um índice de prioridade para cada ordem, esta regra teve o pior desempenho entre as sete regras testadas quanto ao atraso e tempo de espera médios.
- Seu uso é mais conveniente apenas como critério de desempate para outra regra.

Regras ICR, IFO e ITA

- Baseadas em cálculos de índices, são normalmente empregadas em sistemas MRPII, dentro de um módulo chamado “controle de fábrica”, que se encarrega de gerar prioridades para as ordens liberadas pelo módulo MRP.
- As regras ICR e IFO são baseadas no conceito de folga entre a data de entrega do lote e o tempo de processamento, sendo que a regra IFO considera não só a operação imediata, como todas as demais à frente.
- As regras ICR e IFO privilegiam o atendimento ao cliente, porém, devido a simplicidade do exemplo, a regra ICR obteve o pior *lead time* (32 h) e um atraso médio alto.
- A regra IFA relaciona os estoques atuais com a demanda, buscando evitar que os estoques se esgotem, o que causa prejuízo ao fluxo produtivo, sendo mais empregada para os itens intermediários que compõem os produtos acabados.

Regra de Johnson

- Apresentou o menor *lead time* (28 h) e um baixo tempo de espera para processamento na 2ª máquina, garantido pela sua heurística de seqüenciar tempos rápidos de início para o 1º recurso e tempos rápidos de conclusão para o 2º.
- Infelizmente as restrições desta regra são muito fortes, fazendo com que ela seja de aplicação limitada.

Atualmente, dentro do conceito de “PLANEJAMENTO FINO DA PRODUÇÃO”, foram gerados *softwares* que procuram, dentre outras coisas, fornecer uma ferramenta para que os diversos setores produtivos possam seqüenciar dinamicamente um programa de produção dentro de um horizonte limitado pelo PMP (normalmente com periodicidade semanal), conforme as ordens sejam concluídas e problemas surjam no dia-a-dia.

Estes *softwares*, dentro de uma nova visão de programação, não buscam otimizar a seqüência de produção, mas sim buscar boas soluções através de programações com recursos finitos e a consideração de fatores, como os tempos de *setup*, paradas para manutenção dos equipamentos etc., que convencionalmente não são consideradas nas soluções matemáticas de otimização.

De um modo geral, as seguintes características são importantes para as regras a serem empregadas para a definição do seqüenciamento de um programa de produção:

- Simplicidade: as regras devem ser simples e rápidas de entender e aplicar.
- Transparência: a lógica deve estar clara, caso contrário o usuário não verá sentido em aplica-la.
- Interatividade: como os problemas de programação afetam os programadores, superiores e operadores, as regras devem facilitar a comunicação entre estes agentes do processo produtivo.
- Gerar prioridades palpáveis: as regras aplicadas devem gerar prioridades de fácil interpretação. Os usuários entendem melhor uma regra baseada na data de entrega do que, por exemplo, em um índice muito elaborado.
- Facilitar o processo de avaliação: as regras de seqüenciamento devem promover, simultaneamente à programação, a avaliação de desempenho de utilização dos recursos produtivos. Por exemplo: qual o nível de atendimento das datas de entrega?

3.2. Teoria das restrições (TOC – Theory of Constraints)

A teoria das restrições tem sua origem no final da década de 70, quando o físico israelense Eliyahu M. Goldratt desenvolveu um sistema (*software*) de programação da produção com capacidade finita, comercialmente conhecido como OPT (*Optimized Production Technology*).

A disponibilidade de recursos computacionais mais potentes permitiu que o OPT, ao contrário dos *softwares* baseados na lógica do MRP (da década de 60), fosse desenvolvido em cima de uma base de dados que considerava a estrutura do produto (lista de materiais) e a estrutura do processo (rotina de operações) simultaneamente, tornando viável a análise, em paralelo, da capacidade de produção e do seqüenciamento do programa.

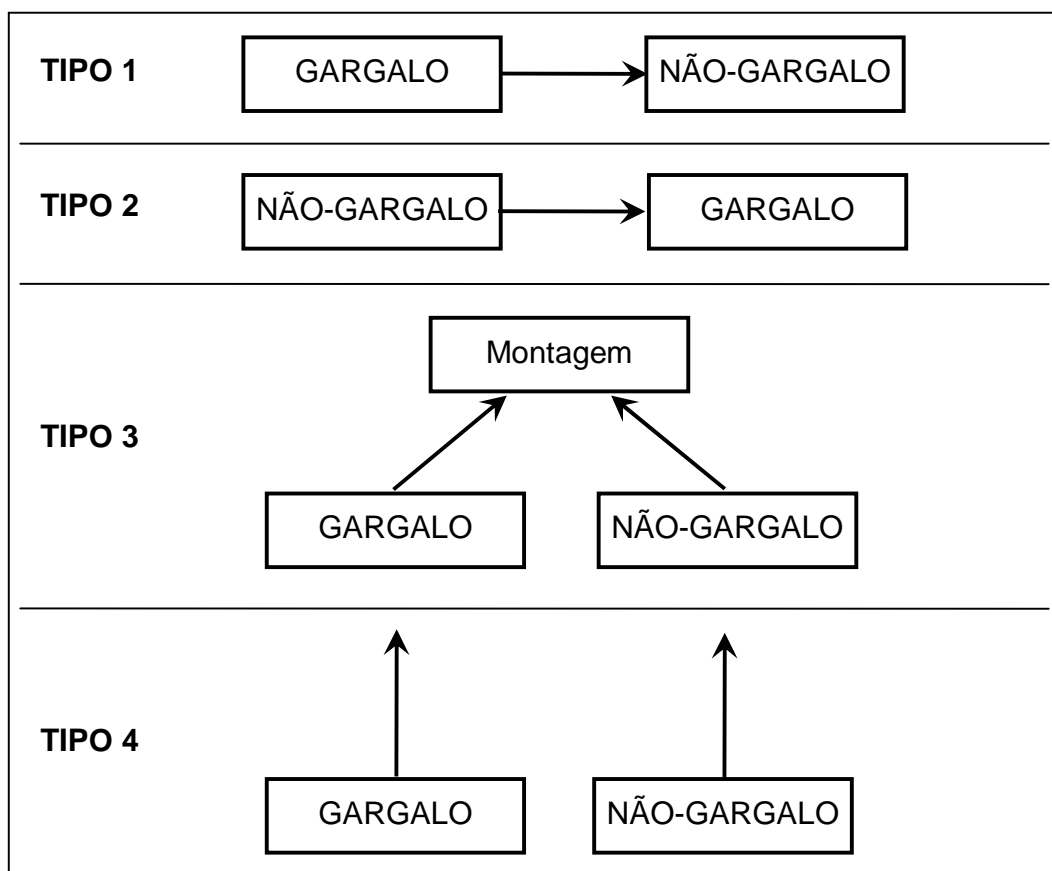
Inicialmente, é importante esclarecer que, apesar do nome pelo qual a técnica ficou conhecida sugerir que se trata de um sistema otimizante (a tradução de OPT seria algo como “Tecnologia de Produção Otimizada”), o OPT não é uma técnica otimizante no sentido científico do termo.

Nada garante que, pela sua aplicação, atinjam-se soluções ótimas, já que a técnica é baseada em uma série de procedimentos heurísticos, muitos dos quais os proprietários dos direitos de exploração do sistema nem mesmo tornaram públicos até o presente.

O OPT não é uma técnica que já tenha caído no domínio público como o MRP II, assim qualquer empresa que decida adota-la deverá fazê-lo por meio das empresas (uma americana e outra inglesa) que detêm os direitos de explorá-los. Talvez por essa razão, sua utilização é bastante restrita.

A nível acadêmico, contudo, as questões levantadas com relação a programação finita da rede de atividades em um sistema de produção convencional, foram estruturadas em um conjunto de regras ou conceitos conhecido como “TEORIA DAS RESTRIÇÕES”, que tem por base o princípio do GARGALO.

GARGALO é um ponto do sistema produtivo (máquina, transporte, espaço, homens, demanda, etc.) que limita o fluxo de itens no sistema. Identificam-se quatro tipos básicos de relacionamento entre recursos gargalos e não-gargalos.



- **TIPO 1:** o fluxo produtivo flui do recurso gargalo para o não-gargalo. Neste caso, o recurso não-gargalo fica limitado a trabalhar apenas na velocidade do fornecimento de itens pelo recurso gargalo.
- **TIPO 2:** a situação se inverte, recurso não-gargalo abastece recurso gargalo. Nesta situação, caso o recurso não-gargalo opere 100% de seu tempo, apenas parte desta produção passará pelo gargalo (fluxo) e parte formará estoques em processo no sistema.
- **TIPO 3:** caracteriza uma situação de montagem, onde um recurso gargalo e um não-gargalo abastecem uma linha de montagem. Neste caso, o fluxo produtivo da linha de montagem fica limitado pela produção do recurso gargalo e, se acionado mais do que nesta velocidade, o recurso não-gargalo estará apenas gerando estoques em processo antes da montagem.
- **TIPO 4:** tanto o recurso gargalo como o não-gargalo atendem diretamente à demanda do mercado. Esta situação é colocada apenas para caracterizar que os gargalos podem ser tanto internos quanto externos ao sistema produtivo. Neste caso, o recurso gargalo trabalhará de acordo com sua limitação de capacidade, sendo toda ela absorvida pelo mercado, enquanto que o recurso não-gargalo deverá ser acionado apenas para atender o fluxo gerado por sua demanda (que neste caso é o gargalo deste fluxo).

Da constatação de que os recursos produtivos podem ser divididos em gargalos e não-gargalos, e de que a forma como eles se relacionam definem o fluxo produtivo, os custos com estoques e as despesas operacionais, um conjunto de 10 regras ou princípios (alguns autores falam em 9, excluindo a 10ª) é usado para direcionar as questões relativas ao seqüenciamento de um programa de produção.

Regra 1: *A taxa de utilização de um recurso não-gargalo não é determinada por sua capacidade de produção, mas sim por alguma outra restrição do sistema.*

O sistema produtivo é sempre limitado por um recurso gargalo, de nada adiantando programar um recurso não-gargalo para produzir 100% de sua capacidade, pois estaremos apenas gerando estoques intermediários e despesas operacionais.

Neste sentido, a Teoria das Restrições procura deixar claro a diferença entre utilizar um recurso e ativar um recurso, gerando a regra 2:

Regra 2: *Utilização e ativação de um recurso não são sinônimos.*

Convencionalmente, os recursos são utilizados 100% de seu tempo. Um recurso parado é visto como perda de eficiência.

A Teoria das Restrições advoga que os recursos devem ser *ativados* apenas na medida em que incrementem o fluxo produtivo, ficando parados sempre que atingirem as limitações dos gargalos.

Regra 3: *Uma hora perdida num recurso gargalo é uma hora perdida em todo o sistema produtivo.*

Como os recursos gargalos não possuem tempos ociosos, caso algum problema venha a ocorrer com estes recursos, a perda de produção se repercutirá em todo o sistema, reduzindo o fluxo.

Da mesma forma, ao transformarmos tempo improdutivo (como parada para *setup* ou manutenção corretiva) em tempos produtivos nos recursos gargalos, todo o sistema estará ganhando. Por outro lado:

Regra 4: *Uma hora ganha num recurso não gargalo não representa nada.*

Como os recursos não-gargalos, por definição, possuem tempos ociosos, qualquer ação que venha apenas acelerar o tempo produtivo destes recursos estará transformando tempo produtivo em mais tempo ocioso. Neste sentido, uma melhora nos tempos de *setup* nos recursos não-gargalo, por si só, não incrementa o fluxo produtivo.

Porém, uma diminuição no tamanho dos lotes que passam por recursos não-gargalos, visando agilizar a chegada dos mesmos aos recursos gargalos, é bem vista, pois estará agilizando o fluxo apenas pela transformação dos tempos ociosos em tempos de *setup*.

Regra 5: *Os lotes de processamento devem ser variáveis e não fixos.*

Como consequência das regras 3 e 4, o tamanho dos lotes de processamento deve variar conforme o tipo de recurso pelo qual estão passando.

Em um recurso gargalo os lotes devem ser grandes para diminuir os tempos de preparação (*setup*), transformando-os em tempos produtivos.

Já nos recursos não-gargalos, os lotes devem ser pequenos para reduzir os custos dos estoques em processo e agilizar o fluxo de produção dos gargalos.

Regra 6: *Os lotes de processamento e de transferência não necessitam ser iguais.*

Convencionalmente, os lotes de produção só são movimentados quando totalmente concluídos. Isto simplifica o fluxo de informações dentro do sistema, mas gera um aumento de *lead time* médio dos itens (pois o primeiro item terá que esperar o último para ser transferido) e nos estoques em processo dentro do sistema.

Segundo a Teoria das Restrições, para evitar estes problemas, os lotes de transferência devem ser considerados segundo a ótica do fluxo, enquanto os lotes de processamento segundo a ótica do recurso no qual será trabalhado.

Regra 7: *Os gargalos governam tanto o fluxo como os estoques do sistema.*

No sentido de garantir a máxima utilização dos recursos gargalos, devemos não só seqüenciar o programa de produção de acordo com suas restrições de capacidade, como também projetar estoques de segurança na frente dos mesmos, buscando evitar interrupções nos fluxos.

Os estoques de segurança, dentro da Teoria das Restrições, são conhecidos como *time buffer*, pois procura-se antecipar no tempo a entrega dos lotes que irão abastecer os gargalos, dando-se tempo para corrigir eventuais problemas antes que os mesmos afetem o fluxo dos gargalos.

Regra 8: *A capacidade do sistema e a programação das ordens devem ser consideradas simultaneamente, e não seqüencialmente.*

Nos sistemas baseados na lógica MRP, o seqüenciamento das ordens é realizado tendo por base índices (ICR, IFO, IFA) que empregam *lead times* padrões predeterminados.

Já na Teoria das Restrições, como trabalha olhando a lista de materiais e a rotina de operações simultaneamente, considera que os *lead times* não são fixos, mas sim resultado da seqüência escolhida para o programa de produção.

Desta forma, para cada alternativa de seqüenciamento analisada, diferentes *lead times* serão obtidos.

Regra 9: *Balanceie o fluxo e não a capacidade.*

Assim como na filosofia JIT/TQC, a Teoria das restrições considera que o importante em um sistema produtivo em lotes, sujeito a passar por recursos gargalos, é buscar um fluxo contínuo destes lotes, acelerando a transformação de matérias-primas em produtos acabados.

A utilização dos recursos, aqui chamada de ativação, deve se dar no sentido de maximizar o fluxo, justificando todas as decisões que convencionalmente são consideradas improdutivas (movimentar pequenos lotes, duplicar *setups*, deixar recursos parados, etc.).

Regra 10: *A soma dos ótimos locais não é igual ao ótimo global.*

Esta regra sintetiza todas as demais, ao considerar que em um sistema produtivo as soluções devem ser pensadas de forma global (em relação ao fluxo), pois um conjunto de soluções otimizadoras individuais para cada recurso ou grupos de recursos (departamentos) geralmente não leva ao ótimo global.

Na prática, sem dispormos do *software* OPT, é muito difícil implementar estes conceitos dentro de um sistema produtivo convencional, principalmente em função da mudança constante dos gargalos.

Porém, no livro “A Meta”, Goldratt propõe o emprego de 5 passos como forma de direcionar as ações da programação da produção dentro destas regras, qual seja:

1º. IDENTIFICAR a(s) restrição(ões) do sistema.

Goldratt faz uma analogia com uma corrente: “Pense na corrente e no fato de que sua resistência é determinada pelo seu elo mais fraco.” Portanto, para aumentar a sua resistência o primeiro passo é identificar qual é o elo mais fraco da corrente.

2º. Decidir como EXPLORAR a(s) restrição(ões) do sistema.

O termo explorar a restrição é, na analogia com a corrente, ‘reforçar o elo mais fraco’. Há duas maneiras distintas de reforçar o elo mais fraco da corrente (o gargalo). Uma é adicionar mais capacidade, aumentando os recursos humanos ou adquirindo mais recursos de capital. Outra maneira de explorar a restrição é tirar o máximo da capacidade do recurso.

3º. SUBORDINAR tudo o mais à decisão anterior.

Deve subordinar todas as ações à decisão anterior, pois é o elo mais fraco da corrente que determina a sua resistência. Ou seja, não adianta melhorar as condições de outros elos se não tratar da restrição

4º. ELEVAR a(s) restrição(ões) do sistema.

Elevar a restrição é o mesmo que reforçar o elo mais fraco da corrente. Ou seja, aumentar a capacidade do recurso restritivo – ainda que isso signifique comprar mais máquinas ou contratar mais pessoas.

5º. Se em um passo anterior uma restrição tiver sido quebrada, volte ao passo 1.

Mas, CUIDADO, não deixe que a INÉRCIA cause uma restrição no sistema.

Ao reforçar o elo mais fraco a corrente torna-se mais forte. Ao repetir o processo e nada acontecer, ou seja, não fortalecer mais a corrente é sinal que este não é o elo mais fraco. Isto significa que a restrição foi quebrada. Portanto reinicia todo o processo voltando ao passo um, ou seja, identificando qual o recurso que restringe o sistema.

4. SEQÜENCIAMENTO NOS PROCESSOS POR PROJETOS

Os processos por projeto buscam atender a demanda específica de um determinado cliente, que muito provavelmente não se repetirá nos próximos pedidos. Desta forma, os recursos produtivos são temporariamente alocados a este produto, e uma vez concluído, passam para uma próxima tarefa. Sendo assim, os produtos são projetados em estreita ligação com as necessidades dos clientes, dificultando a padronização das operações e das instalações e equipamentos. Geralmente o projeto a ser executado exige a criação de uma estrutura própria de PCP que, ao final do mesmo, se desfaz ou se desloca para um próximo projeto.

Nos processos por projeto, a principal questão a ser resolvida pelo PCP, em particular pelo seqüenciamento das tarefas, está ligada a alocação dos recursos disponíveis no sentido de garantir a data de conclusão do projeto.

Desta forma, o PCP de processos por projeto busca seqüenciar as diferentes atividades do projeto, de forma que cada uma delas tenha seu início e conclusão encadeados com as demais atividades que estarão ocorrendo em seqüência e/ou paralelo com a mesma.

A técnica mais empregada para planejar, seqüenciar e acompanhar projetos é a técnica conhecida como PERT/CPM.

O PERT (*Program Evaluation and Review Technique*) e o CPM (*Critical Path Method*) são duas técnicas, desenvolvidas independentemente na década de 50, que buscaram solucionar problemas de PCP em projetos de grande porte. Devido às soluções semelhantes encontradas, atualmente são conhecidas, simplesmente, como técnica PERT/CPM.

Esta técnica permite que os gerentes do projeto, em particular o PCP, tenham:

- Visão gráfica das atividades que compõem o projeto;
- Estimativa de quanto tempo o projeto consumirá;
- Visão de quais atividades são críticas para o atendimento do prazo de conclusão do projeto;
- Visão de quanto tempo de folga dispomos nas atividades não-críticas, o qual pode ser negociado no sentido de reduzir a aplicação de recursos e, conseqüentemente, custos.

4.1. REDE PERT/CPM

A primeira providência para utilizar a técnica PERT/CPM consiste em elaborar um diagrama de rede que represente as dependências entre todas as atividades que compõem o projeto. A partir da montagem da rede, podemos trabalhar com os tempos e a distribuição de recursos necessários para atingirmos a previsão de conclusão. Neste sentido, especial atenção deve ser dada a esta etapa, pois a validade das conclusões obtidas dependerá da correta montagem deste diagrama de precedência.

DIAGRAMA DE REDE

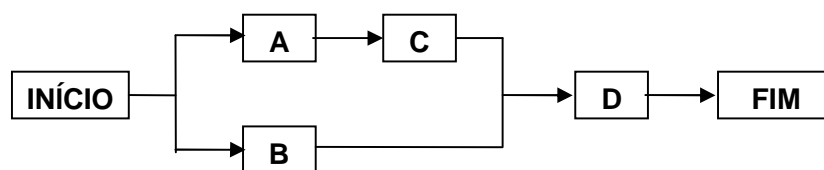
Existem dois tipos de diagramas de rede:

- **Método de Diagrama de Precedência (ou de Atividade no Nó)**

Nós ou caixas representam atividades e setas mostram as dependências entre as atividades.

É o método utilizado pela maioria dos softwares de gerenciamento de tempo.

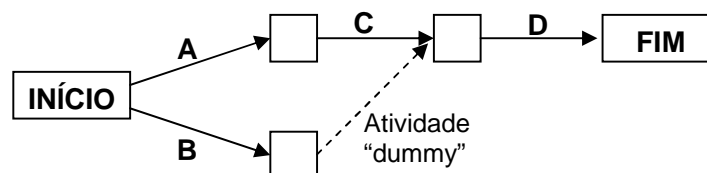
Não utiliza atividades “dummy” (atividades fantasmas).



- **Método de Diagramação de Flechas (ou de Atividade na Seta)**

Setas representam atividades.

Utilizam atividades “dummy”

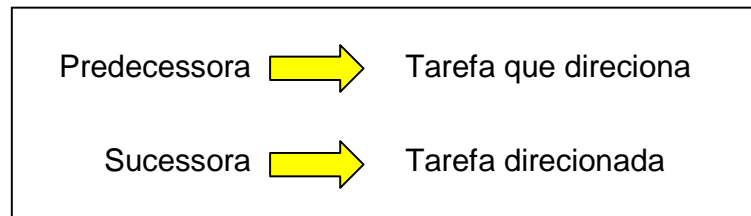


Vamos estudar o Método de Diagrama de Precedência (ou de atividade no nó)

TIPOS DE DEPENDÊNCIA

- Dependências Obrigatórias ou Mandatórias
- Dependências Arbitrárias ou Preferenciais
- Dependências Externas

Dependências não estão relacionadas com CRONOLOGIA, mas com CAUSA e EFEITO!

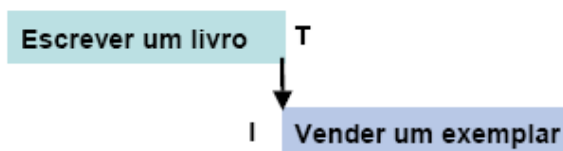


TIPOS DE RELACIONAMENTO ENTRE TAREFAS

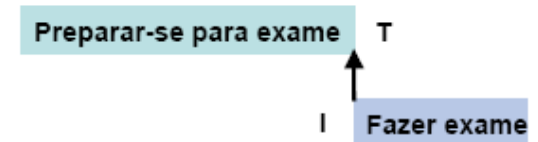
Término-Início (TI)	Uma tarefa deve finalizar antes que outra inicie
Início-Término (IT)	Quando uma tarefa inicia, outra deve terminar
Início-Início (II)	Duas tarefas devem iniciar ao mesmo tempo
Término-Término (TT)	Duas tarefas devem terminar ao mesmo tempo

RELAÇÕES DE PRECEDÊNCIA

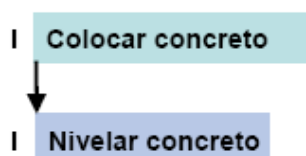
Término-Início



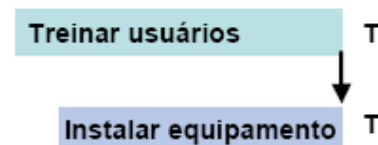
Início-Término



Início-Início



Término-Término



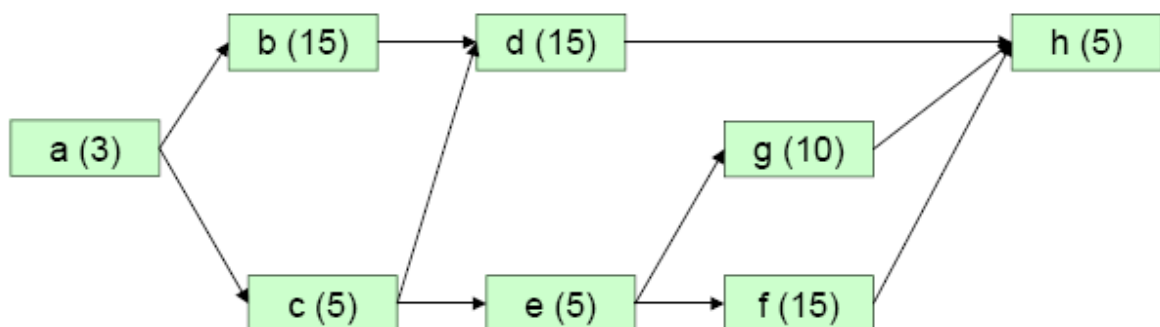
CÁLCULO DOS TEMPOS DE UMA REDE PERT/COM

INÍCIO MAIS CEDO <i>Early Start – ES</i>	A data mais cedo na qual a atividade pode iniciar
TÉRMINO MAIS CEDO <i>Early Finish – EF</i>	A data mais cedo na qual a atividade pode terminar
INÍCIO MAIS TARDE <i>Late Start – LS</i>	A data mais tarde na qual a atividade pode iniciar
TÉRMINO MAIS TARDE <i>Late Finish – LF</i>	A data mais tarde na qual a atividade pode terminar

- Para calcular as DATAS MAIS CEDO (*early dates*) utiliza-se o processo *FORWARD PASS*, caminhando a partir do início para o fim, somando as durações das atividades de acordo com as dependências lógicas.
- Para calcular as DATAS MAIS TARDE (*late dates*) utiliza-se o processo *BACKWARD PASS*, caminhando a partir do fim para o início, subtraindo as durações das atividades de acordo com as dependências lógicas.

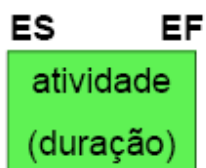
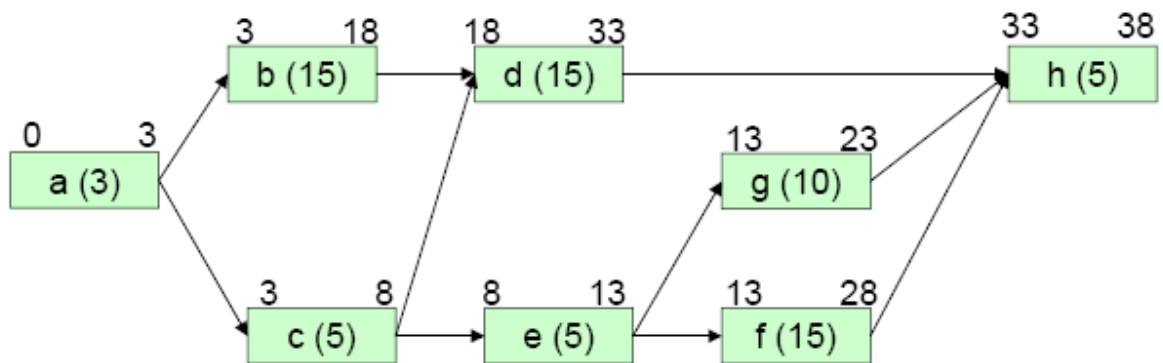
Para o diagrama de rede abaixo, podemos calcular:

- 1) “Early dates” (forward pass)
- 2) “Late dates” (backward pass)
- 3) Caminho Crítico

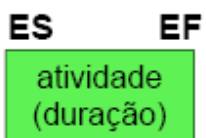
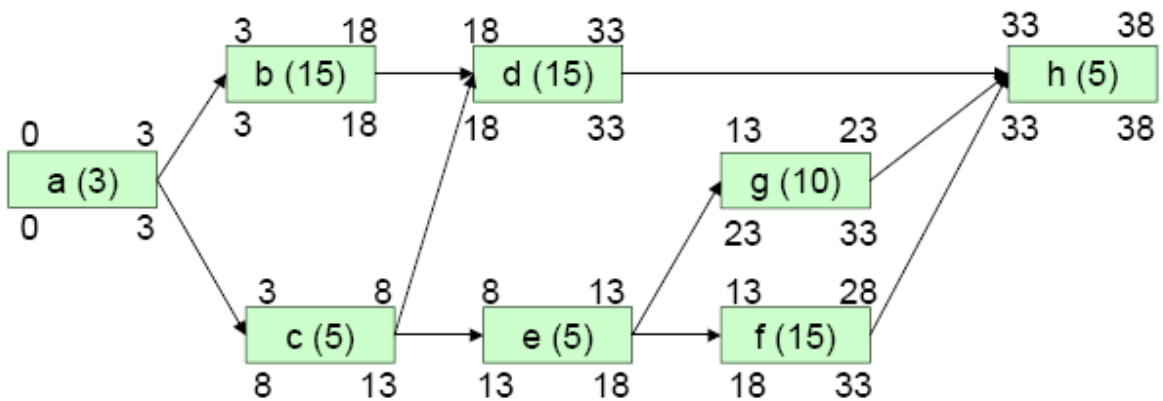


Os número dentro dos () são as durações das atividades em dias

Forward Pass



Backward Pass



Conceitos importantes:

MARCOS (Milestone):

- Compromisso específico dentro do plano do projeto
- Momento específico que não consome tempo ou recursos
- Expressa um estado: concluído, testado, recebido, etc.

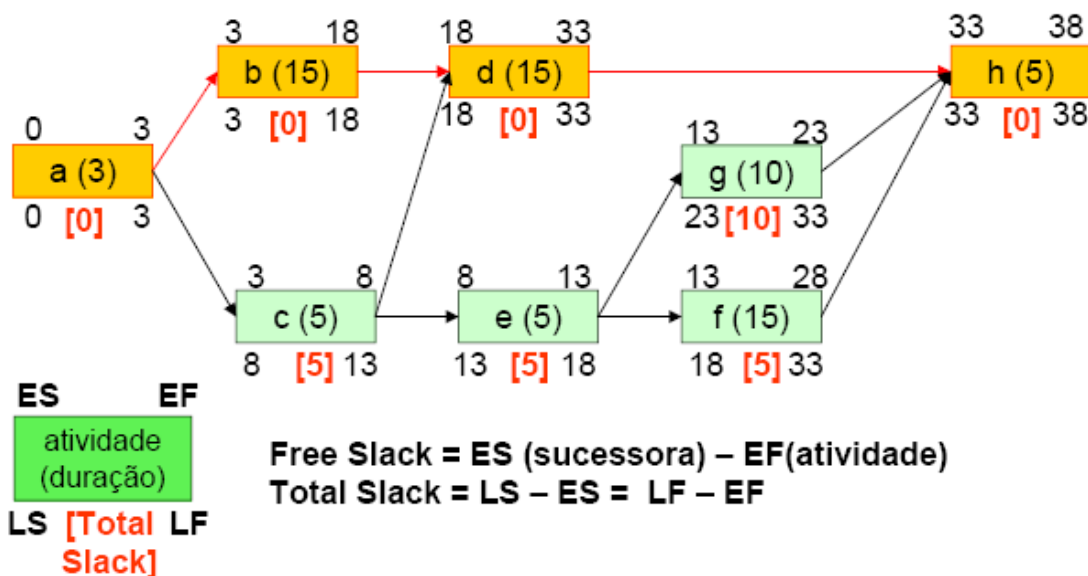
FOLGA (slack)

- **Folga Livre (free slack)** = quantidade de tempo em que a atividade poderá atrasar, sem atrasar o início da sua atividade sucessora
- **Folga Total (total slack)** = quantidade de tempo em que a atividade poderá atrasar sem atrasar o fim do projeto

CAMINHO CRÍTICO (Critical Path)

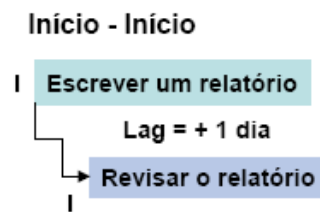
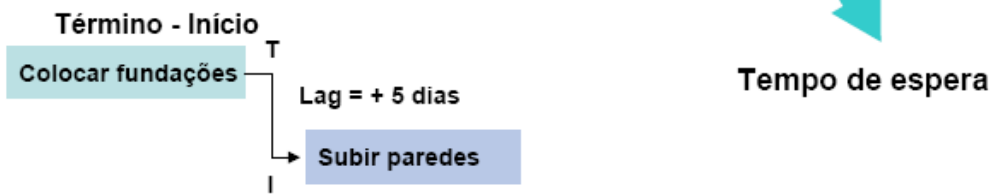
- Caminho mais longo na rede de atividades (o de maior duração)
- Determina qual a data mais cedo em que o projeto pode ser completado
- Não possui folgas
- Indica as atividades que merecem maior atenção e monitoramento
- Podem existir vários caminhos críticos em um projeto
- Pode mudar ao longo do projeto

Achando a Folga (slack) e o Caminho Crítico



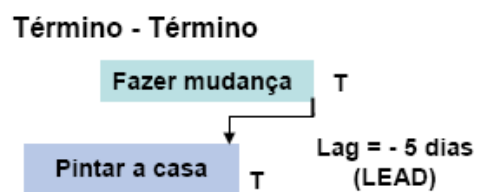
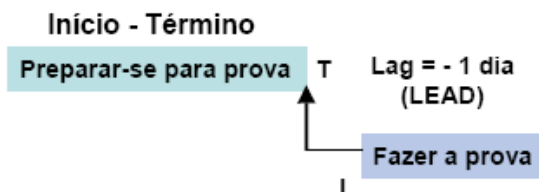
► **LAG – Folga de programação ou defasagem imposta entre atividades**

Lag – Defasagem positiva entre atividades



► **LEAD – Antecipação da programação (negative lag)**

Lead – Defasagem negativa entre atividades



ESTIMATIVAS DE DURAÇÃO DAS ATIVIDADES

Estimativa Análoga

- Baseia-se em informações históricas de projetos similares
- Também conhecida como estimativa *top-down*
- Pouco precisa

Estimativa Paramétrica

- Fórmulas matemáticas para estimar o esforço necessário para a atividade
- Mais precisas, quando disponíveis.

Estimativa de Três Pontos

- Método utilizado no PERT
- Baseia-se em três tipos de estimativas:
 - Pessimista (P) (considera o pior cenário)
 - Mais Provável (M) (considera expectativas realistas, utilização dos recursos previstos, dependências e interrupções)
 - Otimista (O) (considera o melhor cenário possível)

FÓRMULA PERT: $(P + 4 \times M + O) / 6$

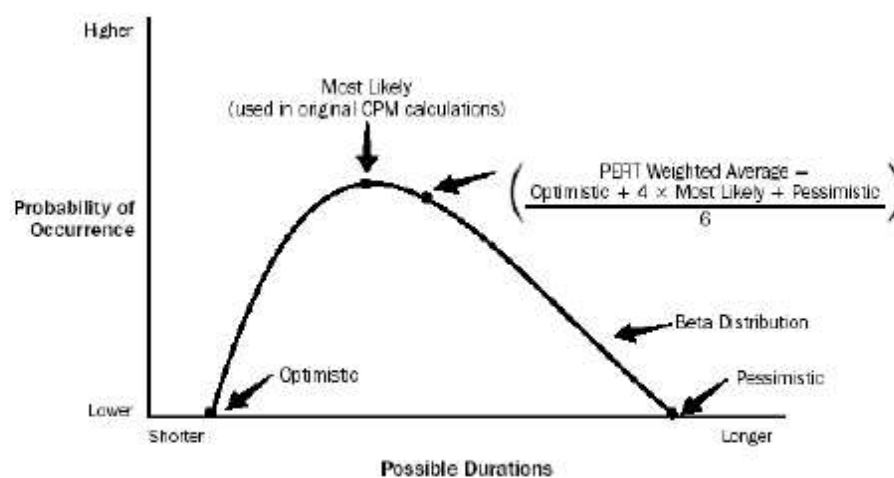


Figure 6-4. PERT Duration Calculation for a Single Activity

Estimativas Três Pontos

Fórmulas - PERT

Duração da atividade	Valor esperado da duração da atividade	$(P+4M+O)/6$
Desvio padrão da duração da atividade	Utilizado para se saber os limites mínimo e máximo da variação da duração	$(P-O)/6$

Legenda: P = Pessimista M= Mais Provável O= Otimista

+ - Um DP => 68,3%

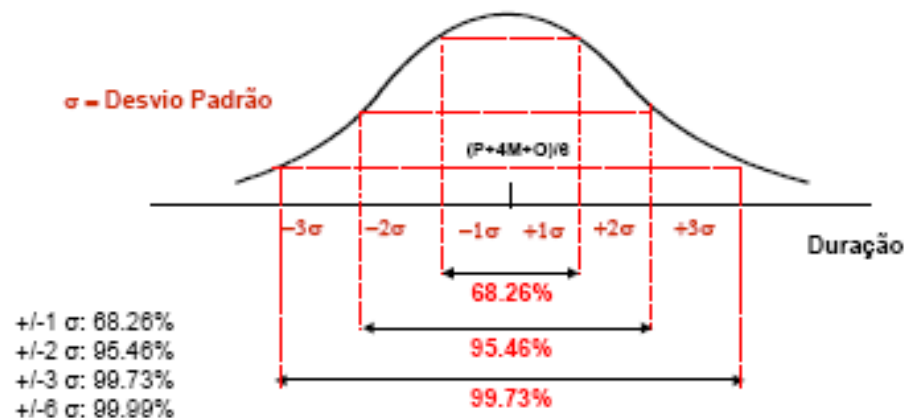
+ - Dois DP => 95,4%

+ - Três DP => 99,7%



Probabilidade de finalizar a tarefa

Estimativas Três Pontos – Desvio Padrão



ACELERAÇÃO DE UMA REDE

Compressão da Duração – “Fast Tracking”

- **Características:**
 - Técnica usada para encurtar o cronograma
 - Atividades que normalmente seriam executadas em seqüência passam a ser executadas em paralelo ou com alguma sobreposição
- **Desvantagem:**
 - Aumenta o risco e pode causar re-trabalho no projeto
- **Alternativas:**
 - Decompor o trabalho e executar atividades em paralelo
 - Alterar dependências do tipo Término-Início
 - Colocar mais recursos

Compressão da Duração – Crashing

- **Características:**
 - Análise das perdas e ganhos de custo e tempo para determinar como obter a maior compressão com o menor custo
- **Desvantagem:**
 - Geralmente resultam em aumentos nos custos diretos e em recursos adicionais
 - Normalmente menos viável que “fast-tracking” devido aos custos adicionais envolvidos
- **Regras:**
 - Comprimir sempre as atividades do caminho crítico
 - Escolher as atividades críticas de menor custo para comprimir

5. EMISSÃO E LIBERAÇÃO DAS ORDENS

A última atividade do PCP antes do início da produção propriamente dita, consiste na emissão e liberação das ordens de fabricação, montagem e compras, que permitirão aos diversos setores da empresa executarem suas atividades de forma coordenada, no sentido de atender ao PMP projetado para o período em questão.

Uma ordem de fabricação, montagem ou compras deve conter as informações necessárias para que os setores responsáveis possam executar suas atividades.

Basicamente, estas ordens contêm a especificação do item, o tamanho do lote, a data de início e de conclusão das atividades. Dependendo do tipo de produção, junto com as ordens de fabricação e montagem, devem seguir também os desenhos e instruções técnicas para produção.

É conveniente que o PCP, antes de formalizar uma programação de produção através da liberação da ordem, verifique se todos os recursos necessários estão disponíveis para o atendimento destas ordens, evitando que ordens sejam emitidas e, por falta de recursos, não sejam atendidas.

Emitidas e liberadas as ordens, o sistema produtivo passará a etapa de execução do programa, e o PCP iniciará suas atividades de acompanhamento e controle da produção.