

# Exercício sobre Processo de Geração de Dados (PGD)

ABJ

## Orientações gerais

- Formem grupos de 5 alunos. Todas as questões deverão ser respondidas nestes grupos;
- Os grupos terão de 40 minutos a 1 hora para responder às questões;
- Um único aluno será encarregado de anotar as respostas às questões discutidas no grupo;
- Depois das discussões em grupo, retornaremos à sala para discutir as questões conjuntamente;
- Outro aluno, que não aquele encarregado de anotar, deverá ler as respostas quando discutidas em grupo.

## Perguntas

1) Investigando o processo de geração de dados dos assuntos (e classes) processuais. Vamos usar, para esse exercício, uma [base de dados que está no Drive](#)

Essa planilha contém a contagem de todos a contagem de todos os assuntos extraídos do Esaj do TJAM, entre os anos de 2014 e 2019. Poderíamos fazer uma planilha igual para as classes, pois o raciocínio será o mesmo.

a) Os assuntos estão padronizados, ou para designar um mesmo assunto há uma série grafias distintas? Por que isso aconteceu?

Os assuntos estão padronizados. Isso acontece graças às [Tabelas Processuais Unificadas](#). As TPUs são uma iniciativa do CNJ para padronizar a classificação dos processos judiciais. Existem TPUs para classes, assuntos, movimentações e documentos processuais.

b) Na planilha, existe uma série de células cuja informação é apenas um traço “-”. O que isso significa?

Isso significa que não há classificação de nível mais específico.

c) Veja a imagem abaixo

Assunto - Nome3	Assunto - Nome4
Ministério Público	Afastamento
	Aposentadoria
	Processo Disciplinar / Si...
	Promoção
	Remoção
	Remuneração
	-

Nesta imagem, o Assunto - Nome4 é o nível de assunto mais específico que essa árvore consegue chegar. O que significa o “-” nesta coluna?

Os casos “-” são casos que são sobre o Ministério Público, mas que não receberam classificação mais específica.

**d) Com base no que foi discutido no item anterior, discuta possíveis problemas no gráfico abaixo. Podemos afirmar com certeza que “Remuneração” é o assunto mais frequente a respeito do Ministério Público? O que significam os casos de cinza?**

Nós não sabemos a real distribuição dos casos marcados com “-”. Isso dificulta a análise. Para a gente assumir alguma distribuição dos casos “-” entre as outras categorias possíveis, temos de partir de uma forte premissa de que os casos marcados com “-” foram marcados assim de forma aleatória. Podemos nos fazer algumas perguntas para nos fazer questionar essa aleatoriedade.

Será que os casos marcados com “-” são aqueles casos que se enquadram em mais de um assunto como a combinação de dois ou mais assuntos (por exemplo, remuneração + promoção) e, justamente por não se enquadrar em um único assunto, eles foram marcados com “-”? Se este for o caso, será que o fato de esses casos representarem a combinação de mais de um assunto não indica que esses casos são mais complexos que os demais, que tratam de um único tema? E por serem casos mais complexos, será que eles não demoram mais tempo para serem resolvidos, ou são casos com menos consenso entre os juízes e, portanto, são casos com mais recursos?

Uma outra série de perguntas possíveis é: será que todos os casos que estão em “-” deveriam ser classificados como “Promoção”, sendo, portanto, “Promoção” o assunto mais frequente? Será que existe algum motivo para que casos que de fato sejam de “Promoção” sejam classificados de forma mais genérica do que os outros assuntos processuais? Por exemplo, será que SEMPRE que os promotores se deparam com um caso de “Remuneração”, eles registram no sistema do

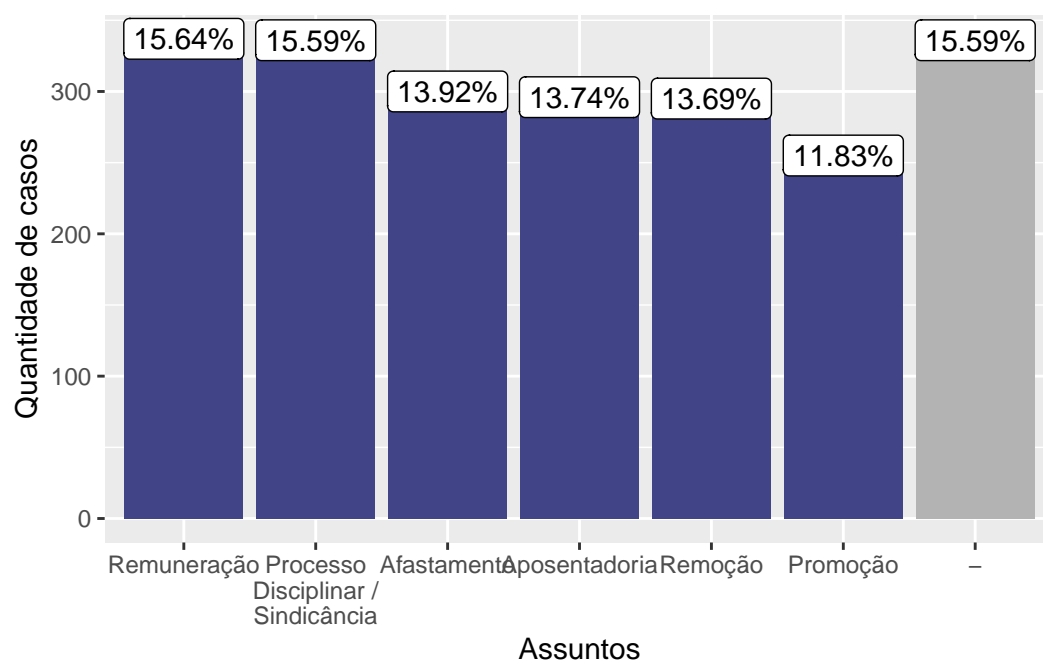


Figure 1: Distribuição de sub-assuntos dentro de DIREITO ADMINISTRATIVO E OUTRAS MATÉRIA DE DIREITO PÚBLICO > Agentes Políticos > Ministério Público

tribunal como o assunto “Remuneração”, mas quando o caso é de “Promoção”, eles registram no sistema do tribunal ou com o assunto “Promoção” ou com o assunto “Ministério Público”?

**e) Um pesquisador deseja listar *todos* os processos relacionados à remuneração dos promotores públicos. Com base na planilha deste exercício e nas discussões dos itens anteriores, quais assuntos você recomendaria a ele utilizar para filtrar a sua busca?**

Uma primeira resposta seria dizer que basta procurar pelo assunto “Remuneração”. Entretanto, o enunciado diz que o pesquisador deseja listar *todos* os processos relacionados à remuneração dos promotores. Nesse caso, o que acontece com os processos que foram mal classificados? O pesquisador estaria deliberadamente excluindo esses processos da análise. E teria problema excluir de forma sistemática os processos mal classificados? Novamente, aqui há um problema do que podemos assumir sobre o caso. Será que os casos mal classificados possuem características específicas, ou eles são aleatórios? Se for o caso de aleatoriedade, não tem problema o pesquisador deixar de fora esses casos (apesar de ser da sua vontade pessoal listar *todos* os processos). Mas se não tiver como o pesquisador defender a aleatoriedade da má-classificação, então talvez seja melhor ele tentar pegar os processos mal classificados. Neste caso, ele precisaria procurar pelos assuntos “Remuneração” e “Ministério Público”.

Se ele quiser ser mais abrangente ainda, ele deveria pegar também os assuntos “Agentes Políticos” e “DIREITO ADMINISTRATIVO E OUTRAS MATÉRIAS DE DIREITO PÚBLICO”. Mas se ele pegar esses assuntos tão genéricos, ele vai acabar pegando muito lixo na pesquisa dele, fazendo com que muitos processos que estão fora do escopo dele entrem na análise. Aqui, então, entra uma discussão importante a respeito da parcimônia x precisão. O que está em jogo aqui é saber sopesar a simplicidade da pesquisa com a sua precisão. Escolher assuntos tão genéricos vai fazer com que tenha muito lixo na lista de processos. Se isso não for retirado da base, então a análise não vai mais ser sobre processos de Remuneração do Ministério Público, mas sim, uma análise sobre qualquer tema sobre o Ministério Público, ou sobre qualquer outro Agente Político, ou ainda sobre qualquer assunto de Direito Administrativo. Neste caso, talvez seja melhor sacrificar alguns processos que foram mal classificados, a fim de se preservar o escopo da base de dados.

## 2) Investigando o processo de geração de dados do valor da causa

**O valor da ação é uma informação que pode ser obtida nas informações de capa dos processos. No Esaj, essa informação aparece disposta da seguinte maneira:**

A imagem mostra a interface do sistema Esaj com os seguintes dados:

1021742-63.2020.8.26.0196		Extinto	Tramitação prioritária
Classe Procedimento Comum Cível	Assunto Interpretação / Revisão de Contrato	Foro Foro de Franca	Vara 1ª Vara Cível
Juiz João Sartori Pires			
Distribuição 14/08/2020 às 11:03 - Livre	Controle 2020/001493	Área Cível	Valor da ação R\$ 3.373,20

Botões: Visualizar autos, Peticionar, Recolher

Sabendo que essa informação estava disponível nas capas dos processos, uma pesquisadora extraiu essa informação de 100 processos escolhidos aleatoriamente. A partir disso, foi encontrado um valor médio da ação de R\$ 17654.1, conforme a figura abaixo.

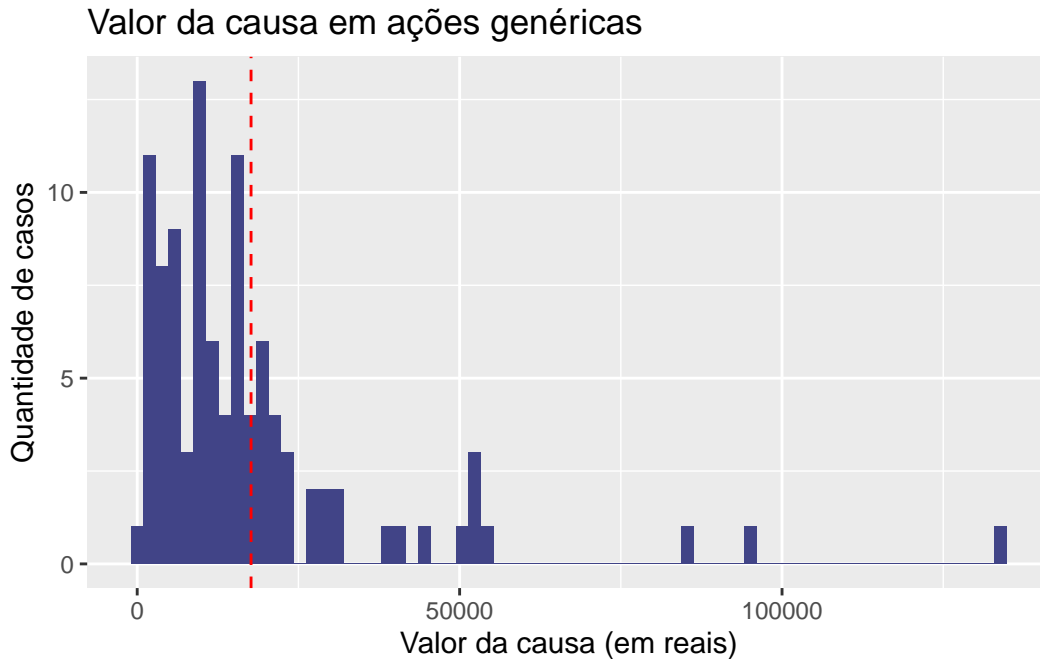


Figure 2: ?(caption)

A partir disso, responda:

a) O valor da ação corresponde ao valor efetivamente pedido? E se o valor pedido na inicial for impugnado, a informação da capa irá refletir o valor originalmente pedido, o valor contestado, ou novo valor arbitrado pelo juízo?

O valor da ação corresponde ao pedido inicial, e não ao valor contestado, tampouco ao valor arbitrado pelo juízo.

b) O valor da ação é corrigido? Ou ele pode estar desatualizado?

O valor da ação pode não estar sempre corrigido. Não existe um sistema dos TJs que garanta a atualização dos valores.

c) Se há vários pedidos alternativos, o valor da causa refletirá o primeiro deles, a soma deles ou o maior deles?

Pelo art. 292, VII, o valor da causa refletirá o pedido de maior valor.

**d) Se houver pedidos subsidiários, o valor da causa refletirá apenas o valor do pedido principal ou o valor da soma de todos os pedidos?**

Pelo art. 292, VIII, o valor da ação refletirá o valor do pedido principal.

**e) Em caso de prestações vincendas, o valor da causa corresponde à soma total de todas as prestações que serão pagas, à soma de algum período específico (como o período de um ano), ou apenas a uma única prestação?**

Em cada caso, o valor da ação pode corresponder a alguma prestação diferente. O valor pode corresponder tanto à soma de todas as prestações, bem como à soma de apenas de um ano das prestações, ou apenas a uma única prestação. Por exemplo, no caso de prestações alimentícias, pelo art. 292, III, o valor da ação corresponde à soma de 12 (doze) prestações mensais pedidas pelo autor.

**f) Se o pedido contiver um erro, o valor da ação vai refletir esse erro?**

Sim, vai refletir o erro.

**g) Assumindo que a amostragem realizada pela pesquisadora não está enviesada, e que a quantidade de casos é grande o suficiente para gerar resultados consistentes, como você avalia a utilização da variável valor da causa para representar o valor real do pedido? Discuta se podemos falar que o valor médio da ação encontrado reflete o valor na realidade? Esse valor pode estar subestimado ou sobrestimado?**

O valor da causa é um valor complicado de ser utilizado para representar o valor real do pedido. A depender do caso, ele pode ser um valor mais ou menos consistente. Por exemplo, em ações de alimentos, por causa da lei, sabe-se que os valores da causa refletem a soma de 12 meses da prestação. Neste caso, então, é possível ter mais certeza sobre o que significa o valor da ação.

O valor da ação poderá estar subestimado caso existam pedidos subsidiários, ou no caso de o período para se pagar as prestações for superior a 1 ano.

E o valor da ação poderá estar sobrestimado caso existam pedidos alternativos, e o juiz conceder o pedido de menos valor. Mesmo neste caso, o valor que irá aparecer na capa do processo é o valor do pedido do maior pedido.

### **3) Investigando o processo de geração de dados da data de distribuição**

**Um pesquisador pegou a data de distribuição que está no sistema, conforme a imagem abaixo.**

**A respeito dessa data, responda:**

**a) A data de distribuição que aparece no sistema corresponde sempre à data de peticionamento? E se o processo for redistribuído, a data irá corresponder à data de distribuição inicial, ou à data de redistribuição?**

1027243-24.2022.8.26.0100				
Classe	Assunto	Foro	Vara	Juiz
Procedimento Comum Cível	Indenização por Dano Material	Foro Regional II - Santo Amaro	9ª Vara Cível	ANDERSON CORTEZ MENDES
Distribuição	Controle	Área	Valor da ação	<a href="#">Recolher</a>
29/03/2022 às 12:43 - Livre	2022/000986	Cível	R\$ 7.600,60	

Em geral sim, mas isso não será verdade sempre, pois essa data representa a data mais recente de distribuição. Se houver redistribuição do processo, então a data do sistema irá representar a data de redistribuição, e não a data de distribuição.

**b) Uma pesquisadora buscou calcular o tempo médio de duração de alguns processos. Para tanto, ela utilizou a data de distribuição registrado no sistema e a data das sentenças, que ela obteve manualmente. Dado o que foi respondido na questão anterior, discuta se o tempo que ela encontrou é confiável, ou se ele poderá estar subestimado ou sobrestimado.**

Como o sistema registra a data de redistribuição, então o tempo calculado poderá estar subestimado. Em muitos casos, a redistribuição ocorre apenas alguns dias depois da distribuição, então, em geral, essa subestimação é mínima. Entretanto, pode haver casos em que a distribuição e a redistribuição é separada por meses, podendo levar a subestimações do tempo maiores.

#### 4) Investigando o processo de geração de dados da data de liminares

Ao se estudar a Justiça Federal do Ceará, um pesquisador criou um robô para extrair todas as movimentações processuais de todos os processos de seu interesse. Essas movimentações são padronizadas pela **TPU de movimentações**. Ele queria calcular o tempo até as decisões de tutela antecipada dos processos. Para tanto, ele precisava de duas datas: (a) a distribuição do processo e (b) a data da decisão liminar.

Para pegar a data de distribuição, ele pegou a data da primeira movimentação registrada no sistema da JFCE, independente de qual fosse a descrição da movimentação. E para pegar a data da decisão de tutela antecipada, o pesquisador buscou pegar as datas das movimentações que continham em sua descrição as palavras “liminar” ou “antecipação de tutela”, porque essas eram as descrições das movimentações correspondentes a decisões de tutela antecipada que ele observou na TPU de movimentações.

A partir disso, o pesquisador encontrou um problema: em alguns processos havia duas ou mais movimentações identificadas por seu algoritmo como uma decisão de antecipação da tutela. A dúvida que surgiu foi qual data ele deveria pegar? A data da primeira movimentação identificada ou da última? Ajude o pesquisador

a decidir qual data escolher, indicando as limitações ou os potenciais de sua decisão.

Existem alguns problemas que o pesquisador deve pensar:

1. Se a petição inicial trouxer mais de um pedido liminar, esses pedidos podem ser julgados em tempos diferentes. A primeira decisão liminar pode julgar apenas uma parte dos pedidos, enquanto a última decisão liminar pode julgar o resto dos pedidos. Neste caso, escolher a primeira data, significaria escolher o tempo para se ter *alguma* resposta judicial para o problema; enquanto escolher a última data significaria escolher o tempo para se ter uma satisfação completa dos pedidos liminares.
2. Se a petição inicial trouxe apenas um pedido liminar e, depois, em outra peça, a parte autora demanda um novo pedido liminar, podemos presumir que a primeira decisão liminar tenha julgado os pedidos da petição inicial, enquanto a última decisão liminar tenha julgado o último pedido. Neste caso, escolher a última data seria um problema muito grande, pois a pesquisa estaria calculando o tempo entre a distribuição do pedido (que normalmente corresponde à petição inicial) e a decisão que julga um pedido que NÃO foi realizado neste momento, sobrestimando demais o tempo de decisão.

Dessa forma, parece ser mais acertado escolher sempre a data da primeira liminar.