

Universidade de São Paulo

**Faculdade de Filosofia, Letras e Ciências Humanas
Departamento de Ciência Política**

**FLS 5028 – Métodos Quantitativos e Técnicas de Pesquisa em Ciência Política
FLP 0406 – Métodos e Técnicas de Pesquisa em Ciência Política**

**1º Semestre de 2019
Profº. Dr. Glauco Peres da Silva**

LISTA DE EXERCÍCIOS 03

"Desenho de Pesquisa Empírica em Ciência Política"

Data de entrega: 25/03/2019 (noturno) e 27/03/2019 (vespertino)

Nome:

Período: () Vespertino, () Noturno

Exercícios Regulares

Exercício 01

Com base na bibliografia obrigatória indicada, marque “Verdadeiro” (V) ou “Falso” (F) para as afirmações a seguir e, quando falsas, justifique sua escolha em 5 linhas, no máximo.

(**V**) Um passo importante em estudos empíricos é a forma como saímos do nível teórico para a mensuração. Esse é um aspecto central seja na coleta de dados pelo pesquisador ou na utilização de dados disponíveis em diversas fontes.

Resposta: A operacionalização das variáveis, ou seja, a passagem do nível teórico (abstrato) para o nível empírico (mensuração) é o que possibilita a realização de testes de hipóteses sobre nossas teorias. Seja coletando nossos próprios dados, como em surveys, ou utilizando dados já coletados, como disponibilizados por diferentes institutos de pesquisa ou órgãos governamentais, é central identificar se eles representam de forma acurada o que desenvolvemos no nível teórico (Kellstedt e Whitten, 2015: 116-117).

(F) Dois pontos de atenção na operacionalização de uma variável são sua confiabilidade e o viés de mensuração. A primeira indica se, com a aplicação dos mesmos procedimentos, se mensurarmos os mesmos casos ou observações, os resultados produzidos seriam idênticos. Já o viés de mensuração indica se existe subestimação ou sobre-estimação de modo sistemático nos valores de uma variável. Se buscamos testar teorias, devemos priorizar a redução do viés à confiabilidade da mensuração.

Resposta: A descrição sobre o que são confiabilidade e viés de mensuração estão corretas. No entanto, o viés de mensuração é um problema maior se buscamos a aproximação do “valor real” de determinadas observações. Já se buscamos testar a teoria, é preferível que ela seja o mais confiável possível, mesmo que possuas algum viés. Em outras palavras, preferimos uma medida enviesada mas confiável a uma não viesada mas não confiável (Kellstedt e Whitten, 2015: 122-123).

(V) No processo de observar se uma medida representa teoricamente o que buscamos analisar é necessário observar sua validade, ou seja, se ela mede o que foi pretendido inicialmente. Nesse sentido, ao observamos que uma medida é confiável devemos avaliar sua validade em três aspectos: face, conteúdo e constructo. A partir disso, com medidas válidas e confiáveis podemos realizar nossas avaliações sobre as teorias causais.

Resposta: A validade de uma medida indica se ela mede o que propõe originalmente. Nesse sentido, devemos examinar sua validade de face (se ela parece ou não mensurar o que se propõe a mensurar), de conteúdo (se captura o conceito a ser medido) e de constructo (o grau em que se relaciona com medidas com as quais a teoria requer que ela esteja relacionada). No entanto, a observação da validade de uma medida ocorre somente se a medida for confiável, uma vez que se não for confiável não nos permitiria testar as teorias causais nas quais estamos interessados.

(F) A capacidade de capturar no mundo real o que espelhamos em conceitos abstratos não interfere diretamente na capacidade de avaliar o suporte empírico para uma teoria.

Resposta: Nossa análise empírica deve se basear em medidas que capturam a essência dos conceitos abstratos de nossa teoria ou não podemos confiar nos achados de nossas análises, uma vez que estão distantes de representar a teoria que queríamos testar (Kellstedt e Whitten, 2015: 131).

(V) Para distribuições assimétricas, é recomendável utilizar a mediana ao invés da média como medida de tendência central. Uma distribuição com obliquidade (*skewness*) positiva possui mais valores acima da média do que abaixo. Já uma distribuição com obliquidade negativa possui mais valores abaixo da média do que acima. A diferença entre a média e a mediana é uma das maneiras de se analisar a obliquidade de uma variável.

Resposta: Para entendermos a distribuição de nossas variáveis, é um procedimento comum começar por mensurar suas medidas de tendência central. A vantagem da mediana é que é uma medida mais conservadora do que a média, mostrando o valor da observação que está no percentil 50. Podemos utilizar outros meios para analisar a distribuição de uma variável, como o desvio-padrão, histogramas, *boxplots* e estimativas de densidade kernel (Kellstedt e Whitten, 2015: 143).

(F) Variáveis categóricas são variáveis que não é possível de se fazer uma ordenação que distinga cada um de seus valores possíveis. Assim sendo, somente podemos usar a média e a moda para descrever suas distribuições. Já variáveis contínuas são variáveis que possuem uma diferença de unidade constante, podendo ser utilizadas tanto a média, a moda e a mediana para descrevê-las.

Resposta: Não é possível se estimar a média de variáveis categóricas. A única medida de tendência central possível de ser mensurada para tais variáveis é a moda (Kellstedt e Whitten, 2015: 136).

(V) Apesar de semelhantes, gráficos de barras e histogramas são adequados para diferentes tipos de variáveis. Os gráficos de barras são utilizados para descrever variáveis categóricas e os histogramas para variáveis contínuas.

Resposta: Apesar dessa semelhança, gráficos de barra e histogramas são usados para tipos distintos de variáveis. Gráficos de barra analisam a frequência de variáveis categóricas,

enquanto que histogramas analisam a frequência de intervalos de valores variáveis contínuas (Kellstedt e Whitten, 2015: 136; 143).

(F) A utilização das estatísticas e dados descritivos nos permitem realizar inferências causais sobre as relações teóricas que buscamos observar. Ao conhecer os dados, conseguimos identificar os padrões de relação entre as variáveis e identificar se uma teoria possui respaldo empírico.

Resposta: Estatísticas descritivas são úteis para uma primeira visão dos dados de uma variável, auxiliando a evitar erros no longo prazo. No entanto, não é possível testar teorias causais com uma única variável, uma vez que teorias causais representam uma possível relação entre duas variáveis (Kellstedt e Whitten, 2015: 145).

Exercício 02

Um conceito fundamental para a ciência política é o de democracia. Como apontado por Kellstedt e Whitten (2015: 124-129), todavia, esse é um conceito que é difícil de se operacionalizar no mundo real. Uma medida de mensuração bastante difundida por pesquisadores da área é o *Polity IV*. Os pesquisadores ligados a esse projeto criaram uma escala que vai de -10 a +10, atribuindo um *score* a cada país, operacionalizando 4 componentes de seus governos: i) regras para o recrutamento do Executivo; ii) competitividade do recrutamento do Executivo; iii) abertura do recrutamento do Executivo; e iv) limites legais ao chefe do Executivo.

Por outro lado, temos também medidas menos complexas que decorrem de uma definição minimalista de democracia, ou é uma democracia ou não é. Nesse sentido, outras medidas também são operacionalizadas como a de Cheibub, Gandhi e Vreeland (2010), o Democracy-Dictatorship (DD) Index assume a dicotomia em apenas dois valores possíveis (1 para democracia, 0 para ditadura).

(A) As duas medidas lidam com o mesmo conceito (democracia) a partir de diferentes concepções teóricas (poliarquia e definição minimalista de democracia), o que tem implicações sobre como as medidas são obtidas. Que tipo de variável (categórica, ordinal ou contínua) é cada uma dessas variáveis usadas para mensurar a democracia?

Dica: pense no significado dos zeros.

(Utilize no máximo 10 linhas para a realização dessa questão)

Resposta: A *Polity IV* é uma medida controversa por alguns motivos. No que nos interessa por agora, é a caracterização comum do índice como quantitativo. Em certo ponto, é o tratamento usual dado pela maioria dos pesquisadores e serão aceitas justificativas baseadas no fato de ser um score que varia de -10 a +10. No entanto, para que uma variável seja quantitativa existem dois aspectos importantes: que a distância entre os intervalos possua o mesmo significado, bem como o valor 0 (diferença de unidade constante). É possível estabelecer que a diferença entre regimes com *scores* +4 e +6 é a mesma que entre regimes -2 e +2? No primeiro caso, ambos são regimes no espectro democrático, já no segundo um está no autocrático e outro no democrático (vide Kellstedt & Whitten, 2015:134). É uma tarefa difícil justificar que os intervalos possuem sempre o mesmo significado. Da mesma forma, um país com score 0 é democrático ou autocrático? O que ele significa? Não é possível estabelecer teoricamente uma construção para o valor. É por isso que existe um movimento para classificação da variável como categórica ordenada e não como quantitativa.

Já para a variável *DD Index* a classificação é mais simples: é uma variável categórica com valores dicotômicos, ou seja, uma variável binária.

(B) De acordo com sua resposta no item (A), justifique que tipo de medida de tendência central você usaria para analisar o *Polity IV* e o *DD Index*.

(Utilize no máximo 5 linhas para a realização dessa questão)

Resposta: Caso considere a *Polity IV* como variável quantitativa, as medidas de tendência central aplicáveis seriam a média e a mediana. A escolha da mais adequada depende da distribuição dos valores na variável: caso a obliquidade seja muito grande, a mediana pode trazer informações bastante úteis não capturadas pela média, como a existência de valores extremos em uma das pontas da distribuição.

Caso a considere como variável categórica ordinal, Kellstedt & Whitten (2015: 136) recomendam que a moda é a única medida de tendência central adequada para avaliar tais tipos de variáveis.

Já para o *DD Index*, por ser uma variável categórica binária (também conhecida como *dummy variable*) seria recomendável utilizar a moda como medida de tendência central, pelos motivos descritos previamente quando tratamos do *Polity IV*.

(C) De acordo com sua resposta no item (A), justifique que tipo de gráfico você usaria para descrever os dados do *Polity IV* e o *DD Index*.

(Utilize no máximo 5 linhas para a realização dessa questão)

Resposta: Tanto o *Polity IV* quanto o *DD Index*, como vimos, são classificados como variáveis categóricas, o *Polity IV* sendo uma variável categórica discreta, e o *DD Index* sendo uma variável categórica binária (também conhecida como *dummy variable*). Para variáveis categóricas, é comum se utilizar gráficos como o gráfico de barras ou o gráfico de pizza. Como Kellstedt e Whitten (2015: 136) apontam, todavia, muitos pesquisadores são contrários ao uso de gráficos de pizza. O grande problema de gráficos de pizza é que costuma ser difícil para o leitor diferenciar exatamente cada seção do gráfico, não ficando claro as escalas de grandeza de cada uma dessas partes. Por esse motivo é comum que pesquisadores recomendam que seja usado um gráfico de barras para variáveis categóricas.

Caso se considere o *Polity IV* como variável contínua (vide a controvérsia desse tema no gabarito do exercício 2A), é possível outros dispositivos gráficos para analisá-lo. Um exemplo seria aplicar um histograma, que pode ser definido como uma distribuição em barras das frequências divididas em classe. Um dispositivo similar seria o gráfico de densidade kernel, que ao invés de calcular a frequência, como feito pelo histograma, calcula a função densidade de uma distribuição. Por fim, poderíamos também gerar um *boxplot*, observando graficamente o resumo dos cinco números, bem como eventuais valores desviantes (*outliers*).

Exercícios de fechamento de Bloco

Tópicos da Aula 1

A partir da leitura do seguinte excerto retirado da introdução da Democracia na América de Tocqueville responda às seguintes perguntas:

Entre os novos objetos que me chamaram a atenção durante minha permanência nos Estados Unidos nenhum me impressionou mais do que a igualdade das condições. Descobri sem custo a influência prodigiosa que exerce esse primeiro fato sobre o andamento da sociedade; ele proporciona ao espírito público certa direção, certo aspecto às leis; aos governantes, novas máximas e hábitos particulares aos governados.

Não tardei a reconhecer que esse mesmo fato estende sua influência muito além dos costumes políticos e das leis, e tem império sobre a sociedade tanto quanto sobre o governo: cria opiniões, faz nascer sentimentos, sugere usos e modifica tudo o que ele não produz. Assim, pois, à medida que eu estudava a sociedade americana, via cada vez mais, na igualdade das condições, o fato gerador de que cada fato particular parecia decorrer e deparava incessantemente com ele como um ponto central a que todas as minhas observações confluíam. Dirigi então meu pensamento para nosso hemisfério e pareceu-me que distinguia nele algo análogo ao espetáculo que o novo mundo me oferecia. (...).

Uma grande revolução democrática se realiza entre nós; todos a vêem, mas nem todos a julgam da mesma maneira. (...). Transporte-me por um momento ao que era a França há setecentos anos: encontro-a dividida entre um pequeno número de famílias que possuem a terra e governam os habitantes; o direito de comandar descende então de geração em geração com as heranças; os homens têm um único meio de agir uns sobre os outros, a força; descubro uma só origem para o poder, a propriedade da terra. Mas eis que o poder político do clero vem se estabelecer e, logo, se ampliar. O clero abre-se a todos, ao pobre e ao rico, ao plebeu e ao senhor; a igualdade começa a penetrar pela Igreja no seio do governo, e aquele que vegetará como servo numa eterna escravidão coloca-se como padre no meio dos nobres e muita vez vai sentar-se acima dos reis. Tomando-se a sociedade, com o passar do tempo, mais civilizada e mais estável, as diferentes relações entre os homens se tornam mais complicadas e mais numerosas. A necessidade das leis civis faz-se sentir vivamente. Nasce então os legistas; eles saem do recinto obscuro dos tribunais e do reduto poeirento dos cartórios e vão ocupar um lugar na corte do príncipe, ao lado dos barões feudais cobertos de arminho e de ferro. Arruinam-se os reis nos grandes empreendimentos; esgotam-se os nobres nas guerras privadas; enriquecem-se os plebeus no comércio. A influência do dinheiro começa a fazer-se sentir sobre os negócios do Estado. O negócio é uma nova fonte que se abre ao

poder, e os financistas se tornam um poder político que os demais menosprezam e adulam. Pouco a pouco, as luzes se disseminam; vê-se despertar o gosto pela literatura e pelas artes; o espírito se torna então um elemento de sucesso; a ciência é um meio de governo, a inteligência uma força social; os letrados chegam aos negócios. Entretanto, à medida que se descobrem novos caminhos para chegar ao poder, vê-se diminuir o valor do nascimento. No século XI, a nobreza tinha um preço inestimável; é comprada no século XIII; o primeiro enobrecimento ocorre em 1270, e a igualdade se introduz enfim no governo pela própria aristocracia. (TOCQUEVILLE, Alexis de. A Democracia na América, São Paulo: Martins Fontes, 2005.)

(A) Tocqueville está tratando do surgimento de uma era democrática tanto no que diz respeito da forma de governo e também na forma como a sociedade se organiza. Com base no excerto apresentado acima apresente o tipo de causalidade que o autor utiliza, justifique sua resposta em no máximo 10 linhas.

Resposta: De acordo com o excerto apresentado, Tocqueville sustenta que a igualdade de condições exercia influência importante sobre os rumos da sociedade estadunidense da época. O tipo de causalidade apresentada pelo autor é aquele descrito por Silva (2018) como “perspectiva via mecanismos e capacidades” (pp. 34 e 38) uma vez que o autor procura identificar elementos de causalidade da transição à “grande revolução democrática [francesa da época]” através da apresentação de “etapas em que uma causa ou um conjunto de causas gera um efeito” (SILVA, 2018, pp. 38). Quando Tocqueville descreve algumas das diversas transições sociais que ocorreram na França entre o século XII e o século XIX ele busca apresentar partes de um processo que ao se conectarem apresentam um mecanismo causal que possibilita a ascensão da república na França e possivelmente (segundo o autor) o estabelecimento da democracia de acordo com aquilo que fora encontrado nos Estados Unidos.

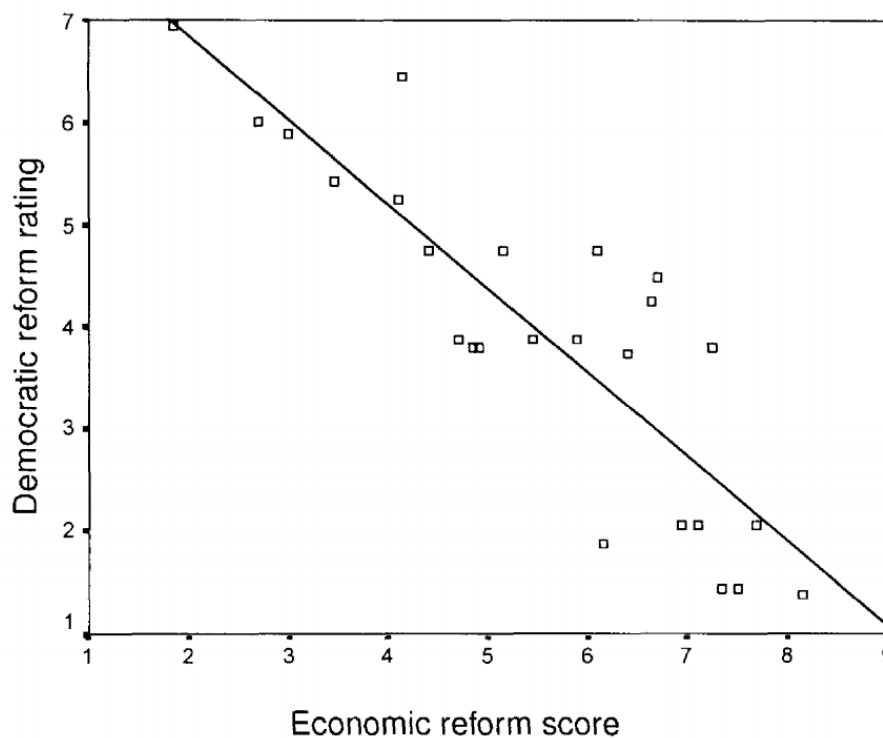
(B) Apresente um desenho de pesquisa que busque explicar o mesmo fenômeno, ou seja o surgimento da democracia. Crie uma teoria que contenha todos os elementos que foram explicitados na capítulo Panorama geral nas Ciências Sociais. Ou seja, a sua explicação deve conter elementos ontológicos, epistemológicos, o tipo de explicação e o método que será usado para mensurar suas variáveis. Lembre-se que sua nova teoria deve estar pautada em variáveis independentes diferentes daquelas usadas por Tocqueville. Não

esqueça de explicar suas escolhas, apresentando os mecanismos causais, em no máximo 15 linhas.

Resposta livre.

Tópicos da Aula 2

Em um estudo sobre democratização em 28 ex-países socialistas, Steven Fish (1998)¹ identificou uma alta correlação entre nível de reformas econômicas e nível de democracia. Para o autor, a realização de reformas econômicas logo após o fim do regime socialista (privatização, liberalização de preços, menos intervenção do Estado na economia) possibilitou um maior nível de democratização. Logo, países que realizaram poucas reformas econômicas em direção a uma economia capitalista, democratizaram-se menos. A relação causal é explicada por meio do seguinte mecanismo: reformas econômicas possibilitam o surgimento de classes médias e elites econômicas que, por dependerem menos de empregos estatais como fonte de renda, são mais livres para demandarem maior participação e abertura política.



¹ Fonte: FISH, Steven M. *Democratization's Requisites: The Postcommunist Experience*, *Post-Soviet Affairs*, 14:3,1998: 212-247.

Com base apenas no enunciado, no gráfico e nas quatro etapas de análise da causalidade, responda:

(A) Há mecanismo causal crível que conecta X a Y? Se não há, justifique (máximo de 3 linhas).

Resposta: Sim, mas aberto à contestação

(B) Podemos eliminar a possibilidade de que X causa Y? Se não podemos, justifique (máximo de 3 linhas).

Resposta: Não. A relação pode ser inversa, ou seja, a democratização pode estar causando a abertura econômica. A causalidade pode, ainda, ter direção dupla.

(C) Existe covariação entre X e Y? Se não há, justifique (máximo de 3 linhas).

Resposta: Sim. O próprio enunciado afirma.

(D) Controlou-se por todas as variáveis colineares Z que podem tornar a associação entre X e Y espúria? Se não controlou-se, justifique (máximo de 5 linhas).

Resposta: Com base apenas no enunciado, não. A relação pode ser afetada por fatores como renda per capita, crescimento econômico, desigualdade de renda, desigualdade de gênero, diversidade étnica, presença de recursos naturais, mobilização da sociedade civil, dentre outros.

(E) Trata-se de um experimento ou de um estudo observacional? Justifique (máximo de 3 linhas).

Resposta: Estudo observacional. Não é possível atribuir diferentes valores da variável independente (reformas econômicas) aleatoriamente a grupos de tratamento e controle.

Tópicos da Aula 3

Para responder os exercícios aplicados da Aula 3, baixe a base de dados (*CPDS_2016.xlsx*) e seu *Codebook* (*Codebook.pdf*), disponíveis no Moodle. Ela é composta por algumas variáveis da “Comparative Political Data Set” para o ano de 2016.

Antes de iniciar a resolução, dê uma olhada no banco de dados e no codebook para identificar o os conceitos que cada uma das variáveis representa e como foram operacionalizadas para a análise empírica. Esse processo é muito importante para saber se as medidas são confiáveis e válidas, além de saber como proceder nossa análise desde as estatísticas e dados descritivos. Feito isso, responda às seguintes questões:

(A) As variáveis “bic” e “judreview” são variáveis que operacionalizam características institucionais. Existe alguma diferença na operacionalização das variáveis? Justifique.

Resposta: As duas variáveis operacionalizam categorizações para as observações, mas possuem uma diferença. A variável “judreview” indica se o país possui um corpo judiciário independente que julga a constitucionalidade da legislação, operacionalizada em “Sim” ou “Não”, sendo uma variável categórica dicotômica (Kellstedt e Whitten, 2015: 133;136-137). Já a variável “bic”, ainda que indique se o país é ou não bicameralista, possui uma ordem entre as quatro categorizações, indo de “unicameralismo” a “bicameralismo forte”, sendo uma variável categórica ordinal (Kellstedt e Whitten, 2015: 133-134;138-141).

(B) Qual a diferença da variável “womenpar”, da participação de mulheres no parlamento, com as duas variáveis do exercício A?

Resposta: Diferentemente das variáveis do exercício A, que representam categorias que não possuem intervalos iguais entre as unidades, a variável “womenpar” possui intervalos iguais entre as unidades de medida, característica das variáveis quantitativas (Kellstedt e Whitten, 2015: 134-135;138-142).

(C) Faça a representação gráfica mais adequada para representar as variáveis “bic” e “judreview”. Justifique suas escolhas.

Resposta: Sendo variáveis categóricas, os gráficos de setores e de barras são as opções disponíveis. No entanto, como salientam Kellstedt & Whitten (2015, 136) muitos estatísticos são contrários ao gráfico de setores e recomendam fortemente o uso do gráfico de barras .

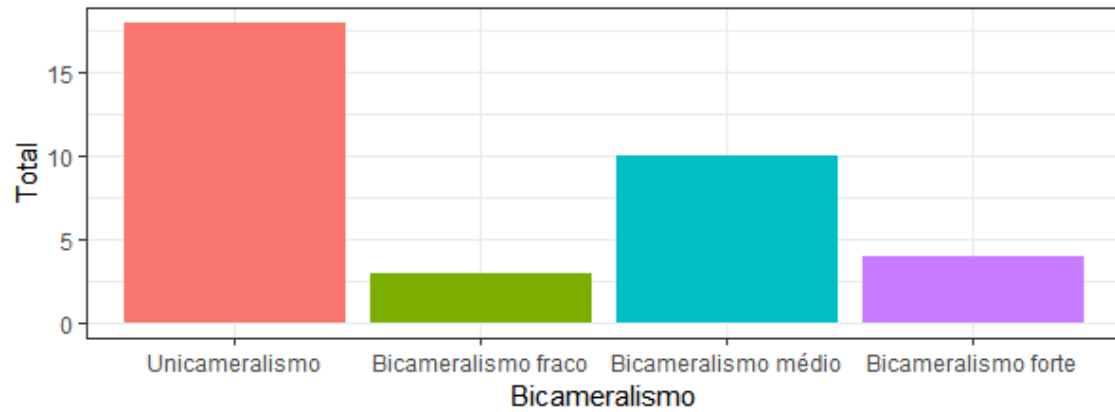


Gráfico de barras para a variável "bic"

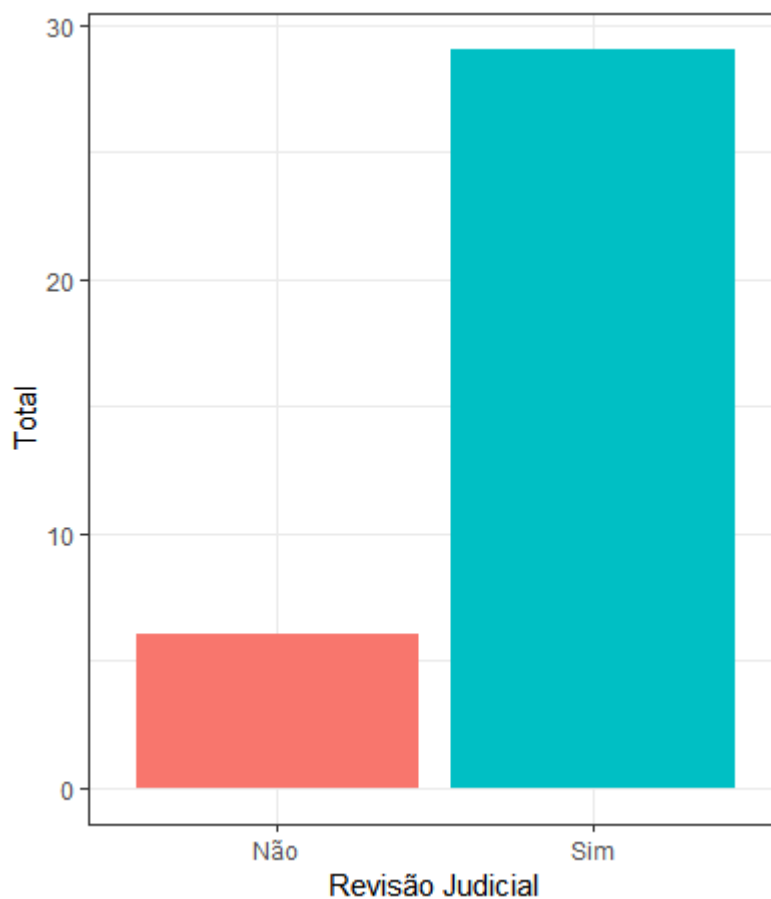


Gráfico de barras para a variável "revjud"

No entanto, ainda que menos usual e com as diversas críticas, o gráfico de setores pode ser útil se uma das categorias é muito dominante sobre as outras. Algo como o que

acontece com a variável “revjud”, mas com maior destaque se o número de categorias for elevado.

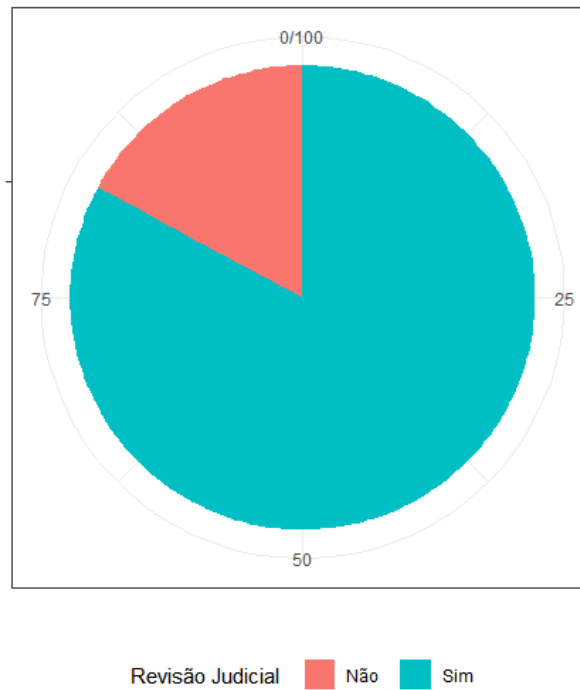
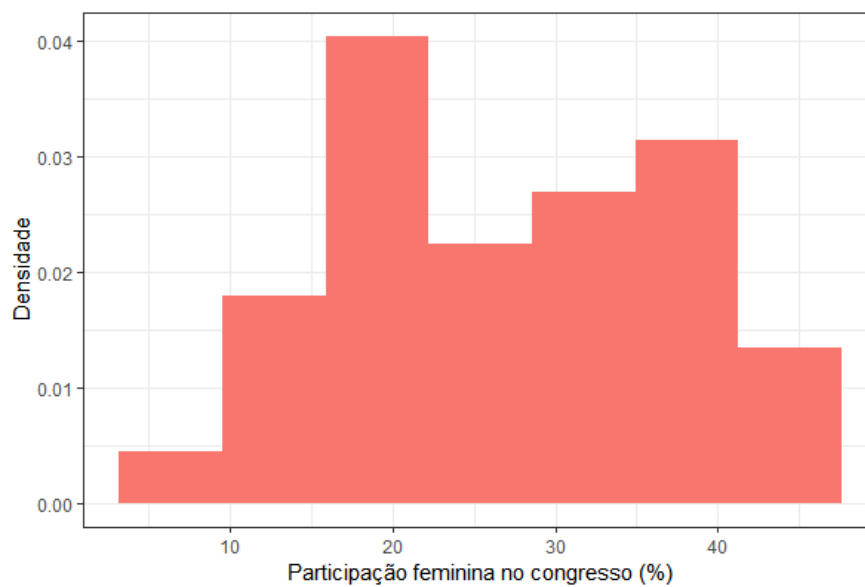


Gráfico de setores para a variável “revjud”

(D) Faça a representação gráfica mais indicada para apresentar a distribuição da variável “womenpar”. Justifique a escolha.

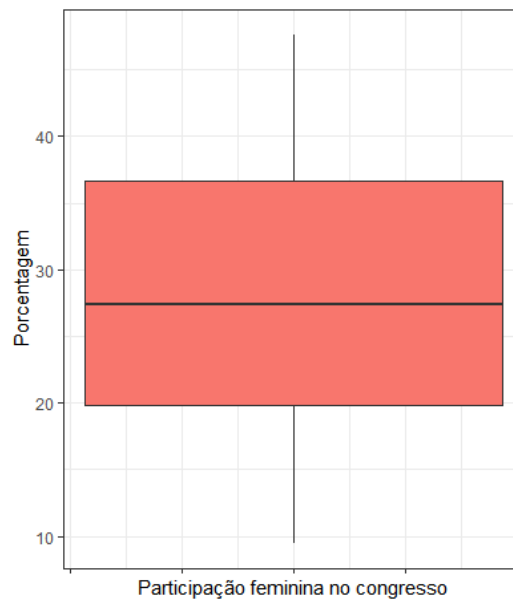


Resposta: Diferentemente do que ocorreria com as variáveis “judrev” e “bic”, a variável “womenpar” apresenta continuidade entre os valores, sendo uma variável quantitativa

Dessa forma, o histograma é o gráfico mais adequado para a representar a distribuição da variável (Kellstedt & Whitten, 2015: 143-144).

(E) A variável “womenpar” possui *outliers* (valores atípicos)? Justifique sua resposta graficamente.

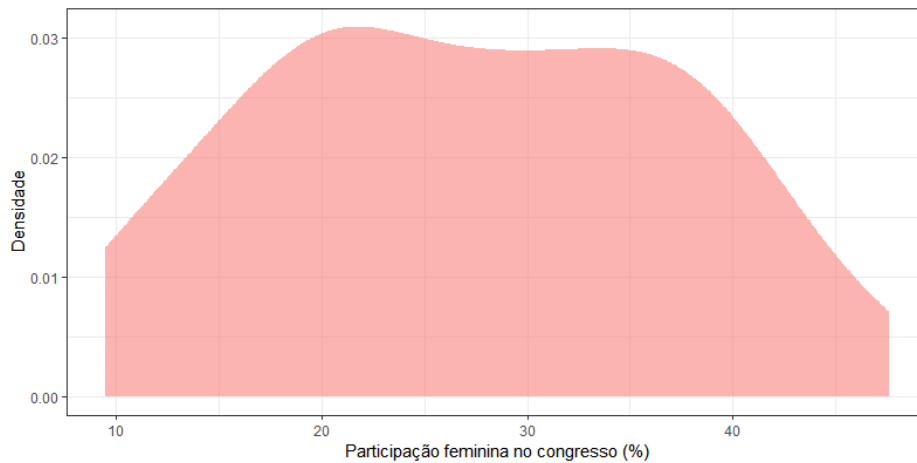
O boxplot representa graficamente a dispersão, identificando a mediana, os quartis e possíveis *outliers*. Observando o gráfico não são encontrados *outliers* (Kellstedt & Whitten, 2015: 139-141).



Boxplot para a variável “womenpar”

(F) A distribuição das observações na variável “womenpar” é simétrica ou assimétrica? Justifique graficamente sua resposta.

Resposta: O histograma apresentava uma leve concentração e observações à esquerda do gráfico, o que é ressaltado pelo gráfico de densidade. No entanto, não existe uma assimetria em destaque quando observamos o gráfico de densidade, apesar das pequenas diferenças (Kellstedt & Whitten, 2015: 143-144).



(G) Qual medida de tendência central é mais adequada para cada uma das três variáveis? Por que? Apresente as medidas e como foram obtidas.

Resposta: A variável “*judrev*” sendo categórica, não possuindo uma ordem natural ou intervalos iguais entre os valores que assume. Dessa forma, a moda é a medida de tendência central mais adequada e indica qual categoria possui o maior número de observações. A observação dos gráficos descritivos no exercício C nos mostra o “Sim” como moda (Kellstedt e Whitten, 2015: 136).

Já a variável “*bic*” possui uma ordenação, ainda que não possua intervalos com valores iguais entre as categorias. Ainda assim é possível a utilização da mediana, mostrando em que posição no espectro unicameral-bicameral está a observação que divide em dois grupos iguais. Como são 35 observações, a observação na posição 18 indica a mediana e, se observarmos o gráfico de barras, temos que ela ainda está na categoria “Unicameralismo”, que é também a moda (Kellstedt e Whitten, 2015: 136;138-139).

Por fim, a variável “*womenpar*” pode ser representada pela média ou pela mediana (Kellstedt e Whitten, 2015: 138-139;142-143).

No caso da média, seu valor é de 27,3 e é calculada como:

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{x} = \frac{28,7 + 31,1 + (\dots) + 32 + 19,2}{35}$$

$$\bar{x} = \frac{955,5}{35}$$

$$\bar{x} = 27,3$$

Mediana = 54,89

Como o número de observações é ímpar, a mediana é obtida com a observação situada exatamente no centro da distribuição.

$$mediana = \frac{35}{2} = 17,5$$

Como $n = 35$, a 18ª observação representa a mediana e assume o valor de 27,4.

Observadas as duas medidas e associando com os gráficos descritivos, a média é uma medida mais indicada uma vez que a distribuição não apresenta assimetria destacada ou *outliers*, com a pequena diferença entre média e mediana referendando essa percepção.

(H) Calcule a variância e o desvio-padrão e explique o que eles indicam.

Resposta:

Variância = 104,01

$$s^2 = \frac{\Sigma(x - \bar{\mu})^2}{n - 1}$$

$$s^2 = \frac{(x_1 - \bar{\mu})^2 + (x_2 - \bar{\mu})^2 + (x_3 - \bar{\mu})^2 + (\dots) + (x_{n-1} - \bar{\mu})^2 + (x_n - \bar{\mu})^2}{n - 1}$$

$$s^2 = \frac{(28,7 - 27,3)^2 + (31,1 - 27,3)^2 + (\dots) + (32 - 27,3)^2 + (19,2 - 27,3)^2}{35 - 1}$$

$$s^2 = 104,01$$

Desvio-padrão = 10,2

$$s = \sqrt{\frac{\Sigma(x - \bar{\mu})^2}{n - 1}}$$

$$s = \sqrt{\frac{(28,7 - 27,3)^2 + (31,1 - 27,3)^2 + (\dots) + (32 - 27,3)^2 + (19,2 - 27,3)^2}{35 - 1}}$$

$$s = \sqrt{104,01}$$

$$s = 10,2$$

Deve-se ressaltar que a variância e o desvio-padrão são medidas de dispersão dos dados em torno da média. Por indicar um valor fornecido pela dos quadrados dos desvios, a variância possui difícil interpretação, o que é solucionado pelo desvio-padrão que indica a distância típica das observações em relação à média (Kellstedt e Whitten, 2015: 142-143).

Informação Extra:

O desvio-padrão de 10,2 não é alto a ponto de indicar uma grande dispersão dos dados em torno da média ou muito baixo para indicar uma maior proximidade da média.

A Regra Empírica é útil para interpretação da dispersão dos dados quando considerada toda a amostra, com 64% dos dados em intervalo de um desvio-padrão da média, 95% a dois desvios-padrão, e quase todos os dados a três desvios-padrão. No entanto, esses valores dependem da distribuição e pode não funcionar tão bem para distribuições assimétricas ou com poucos valores observados.

Quando observado o intervalo das observações, pode-se pensar na influência de possíveis valores atípicos na média e, conseqüentemente, nas medidas de dispersão, ponto que será retomado. Desvios-padrões podem ser muito afetados por valores atípicos, especialmente em pequenos conjuntos de dados como os do nosso exercício.