

Métodos de Pesquisa e Elaboração de Questionários

Denise Britz do Nascimento Silva

Escola Nacional de Ciências Estatísticas - ENCE



ENCE

Instituto Brasileiro de Geografia e Estatística IBGE

 **IBGE**

Material apresentado neste seminário foi originalmente produzido em conjunto com a Prof. Maria Luiza B. Zacharias e Prof. Pedro Silva.

O que é uma Pesquisa Quantitativa ?

- “É um método sistemático para obter informações de unidades visando construir descrições quantitativas dos atributos da população da qual fazem parte.” (Groves et al., 2004)
- Envolve um processo lógico de investigação que permite a compreensão de um fenômeno e de como este afeta uma população
- É o processo de identificação, obtenção, tratamento, análise, apresentação e disseminação de informações estatísticas para atender uma certa demanda



Objetivos Típicos de Pesquisas Quantitativas

- Buscar a compreensão de algum problema econômico ou social
- Necessidade de dados para diagnósticos
- Monitorar o 'estado da nação'
- Apoiar o planejamento e a avaliação de políticas públicas
- Testar teorias/hipóteses sobre padrões de comportamento
- Obter informação útil para tomadores de decisão (nos setores público e privado)



Exemplos de Pesquisas Quantitativas

- Grandes pesquisas de institutos nacionais de estatísticas
- Pesquisas de mercado e de opinião
- Pesquisas eleitorais
- Pesquisas junto a empresas ou estabelecimentos
- Pesquisas de avaliação de impacto de políticas públicas

Métodos para Pesquisas Quantitativas

“Conjunto de conceitos, definições operacionais, procedimentos, técnicas e princípios relacionados com o planejamento, coleta, processamento e análise de pesquisas.”

Groves et al. (2004)

Groves, R.M. et al. (2004). Survey Methodology. Hoboken: John Wiley & Sons.

Tipos de pesquisa – Seleção das unidades

Censitária

coletam informações sobre **TODAS** as unidades da população

Por amostragem - seleciona parte da população de modo que seja possível estimar alguma característica sobre toda a população





Por que não fazer sempre pesquisa por amostragem e esquecer o Censo?

Por que não fazer sempre pesquisa por amostragem e esquecer o Censo?

Porque há ocasiões onde a natureza da informação faz com que o Censo seja não apenas desejável, mas essencial.

EXEMPLO



Etapas de uma pesquisa

- (Demanda pela informação)
- Definição de objetivos
- Planejamento: cadastro, desenho da amostra, forma de coleta, conceitos, plano tabular, desenho do questionário
- Coleta de dados
- Processamento: captura, codificação, crítica e imputação
- Estimação
- Análise
- Disseminação
- Documentação
- Avaliação



Processo de pesquisa

Gestão de qualidade / Metadados

Planejamento	Planejar resultados / tabulações	Descrição de variáveis	Definir método de coleta de dados	Definir cadastro de seleção e amostra	Definir processamento estatístico	Definir sistemas de produção e fluxo de trabalho		
Construção	Elaborar instrumento de coleta	Desenvolver componentes processo	Configurar fluxos de trabalho	Testar sistemas de produção	Finalizar sistemas de produção			
Coleta	Selecionar amostra	Iniciar coleta	Realizar coleta	Finalizar coleta				
Processamento	Integração de dados	Classificação e codificação	Revisar, validar & criticar	Imputação	Derivar novas variáveis & unidades estatísticas	Cálculo de pesos	Cálculo de agregados	Finalizar arquivos de dados
Análise	Esboçar saídas/resultados	Validar resultados	Examinar e explicar	Desidentificar / anonimizar	Finalizar resultados			
Disseminação	Atualizar sistemas de resultados	Elaborar produtos de disseminação	Gerenciar liberação dos produtos	Promover disseminação	Dar atendimento a usuários			
Documentação	Definir regras de arquivamento	Gerenciamento de bases de dados	Preservar dados e metadados	Disponibilizar dados e metadados				
Avaliação	Obter itens para avaliação	Realizar avaliação	Elaborar plano de ação					

Questões preliminares

Reúna o **conhecimento prévio** sobre o tema

Revisão da literatura

Discussão com especialistas

É possível obter a informação desejada?

Pesquisa amostral é melhor método de obter a informação desejada?

Haverá recursos para financiar e realizar a pesquisa?

Vale a pena fazer a pesquisa?



Definição de objetivos

- Necessidades de informação
- Usuários e usos dos dados
- Principais conceitos e definições operacionais
- Conteúdo da pesquisa
- Plano Tabular / Plano de Análise



População Alvo

População sobre a qual a pesquisa busca fornecer informações

Geralmente **finita** e pode ser contada

Seus **elementos** ou unidades são **bem definidos** e **observáveis**

Descrição deve incluir:

Onde está localizada

Quem faz parte

Quando será observada

Exemplos?



População Alvo

População sobre a qual a pesquisa busca fornecer informações

Exemplos:

População residente no estado de São Paulo em setembro de 2016

Empresas formais com 5 ou mais pessoas ocupadas operando no Brasil em janeiro de 2014

Domínios registrados na internet no .br em 31 de julho de 2016

População Alvo x População de Pesquisa

- A **população alvo** é a que a pesquisa **gostaria de cobrir**.
- A **população de pesquisa** é a que a pesquisa **consegue de fato cobrir**.
- Geralmente, a **cobertura** é definida em função de **limitações do cadastro** ou cadastros que serão empregados para selecionar a amostra da pesquisa

População de Pesquisa

Para identificar a população de pesquisa, é necessário especificar :

- Unidades a serem pesquisadas
- Características definidoras das unidades
- Localização espacial das unidades
- Data ou período de referência
- Cadastro de seleção

População de Pesquisa - Exemplos

- População residente em domicílios particulares permanentes no estado de São Paulo em setembro de 2011.
 - ✓ Ficam de foram....
- Empresas registradas no CNPJ com 5 ou mais pessoas ocupadas operando no Brasil em janeiro de 2011.
 - ✓ Ficam de fora....

Conceitos e definições operacionais

Exemplo: Pesquisa Nacional por Amostra de Domicílios (PNAD/IBGE)

- **Unidade de referência**
pessoa e domicílio
- **Unidade informante**
pessoa adulta, moradora no domicílio
- **Unidade de análise**
pessoa, família, domicílio
- **Unidade de amostragem**
município, setor, domicílio

Conceitos e definições operacionais

Estabelecem **linguagem comum** entre pesquisadores, entrevistadores, respondentes, clientes e usuários - todos os envolvidos no processo

Ajudam os respondentes a entender as perguntas

Favorecem compreensão e uso adequados dos resultados da pesquisa

Participantes da pesquisa devem ter o mesmo entendimento dos termos utilizados

Cada conceito e termo utilizado deve ter uma definição (operacional) clara e precisa



Conceitos e definições operacionais

“Para conduzir uma pesquisa, devemos **traduzir os conceitos** numa **forma operacional** em que possam ser **mensurados**.” De Vaus (2002, p. 43)

- Uma definição operacional é uma tradução do conceito com a qual podemos operar na condução da pesquisa.
- Descrição de uma variável, um termo, um objeto, considerando o **processo de mensuração** que será usado para determinar sua presença, características e quantidade.



Conceitos e definições operacionais

ATENÇÃO

- Cada uma das definições (**ex. população de pesquisa, morador**) delimita um grupo de pessoas
- Definições operacionais são essenciais e devem estar **acessíveis publicamente**
- **Tarefa árdua porém imprescindível para o sucesso da pesquisa!!**



Pesquisa sobre atitude dos católicos em relação ao aborto

⇒ Quem é católico?

- é a pessoa que responde católico quando se pergunta a sua religião?
- é a pessoa que vai regularmente à missa?
- é a pessoa que afirma acreditar nos ensinamentos da Igreja Católica?

grupos diferentes

⇒ O que é oposição ao aborto?

- oposição a todo tipo de aborto?
- oposição em qualquer situação?
- oposição a todos, exceto os recomendados por razão médica?

Deve-se utilizar a definição que melhor atenda aos objetivos da pesquisa

Conceitos e definições operacionais

Não há “certo” ou “errado” em relação a conceitos e definições operacionais

O desafio é estabelecer formas úteis

Trata-se de trabalho demorado e complexo

Evite ‘reinventar a roda’ a menos que seja essencial

Muitas organizações estatísticas trabalham no assunto, convém utilizar padrões já aceitos, sempre que possível e adequado à pesquisa em questão

Exemplo: Na Inglaterra usam ‘banco de questões’
<http://discover.ukdataservice.ac.uk/variables>



Plano tabular / Plano de Análise

- Parte fundamental do **planejamento da pesquisa**, interagindo com outras decisões sobre o processo de pesquisa
- A recomendação é elaborar um **plano tabular** para a pesquisa **ANTES** de elaborar o questionário
- Isto ajudará a decidir **que perguntas incluir**, e também ajudará a **definir a forma** de algumas perguntas

Plano tabular / Plano de Análise

2.2.2 Questionnaire design

24. Once the survey objectives and tabulation plan have been determined, the relevant questionnaire can be developed. The questionnaire plays a central role in the survey process in which information is transferred from those who have it (the respondents) to those who need it (the users). It is the instrument through which the information needs of the users are expressed in operational terms as well as the main basis of input for the data processing system for the particular survey.

<http://unstats.un.org/unsd/demographic/sources/surveys/Handbook23June05.pdf>

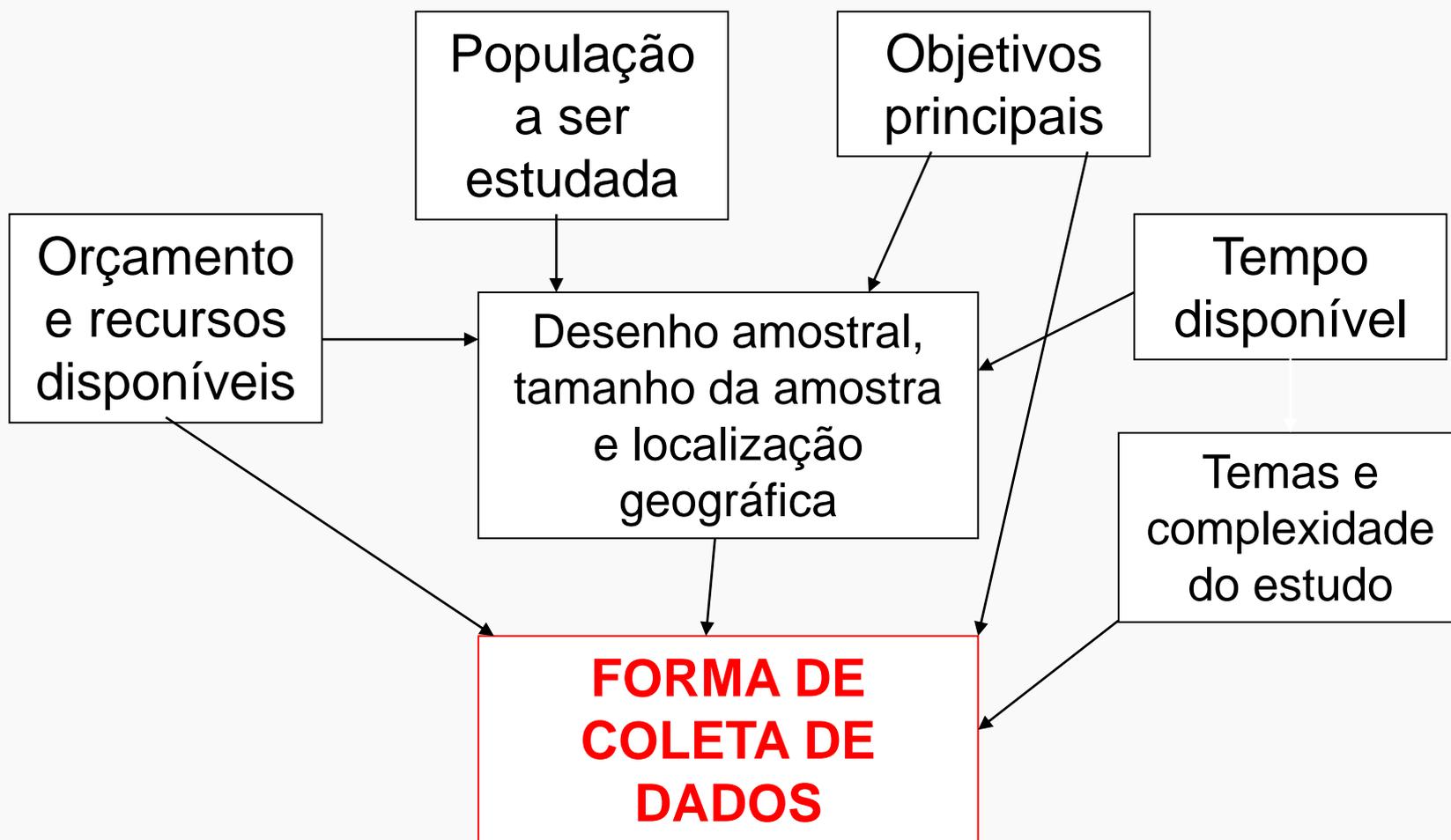


Início estratégico

Preciso planejar uma pesquisa....

Não comece a escrever as perguntas!

Quais são as questões estratégicas para considerar?



Métodos básicos de coleta

- **Autopreenchimento**
 - Correio
 - Internet
 - Computador
- **Com entrevistador**
 - Telefônica
 - Pessoal
- **Métodos combinados**

QUESTIONÁRIO

- Um questionário é um grupo ou sequência de perguntas elaboradas para obter informações sobre um tema de um respondente
- O objetivo primordial é traduzir as necessidades de informação em um conjunto de pesquisas específicas que os respondentes podem e desejam responder.

Qual estado é maior, Bahia ou Minas Gerais?



O processo de desenho de questionário

- Rever objetivos e necessidades de informação da pesquisa
- Listar conceitos e temas a serem investigados
- Buscar informações dos clientes
- Decidir quais são as informações desejadas (o que medir)
- Pesquisar a literatura / Rever questionários anteriores
- Consultar usuários de dados e respondentes / especialistas no tema
- Eventual necessidade de pesquisa qualitativa
- Operacionalizar as variáveis (como medir)
- Esboçar o questionário (Elaborar as perguntas / Definir a sequência)
- Testar e revisar o questionário
- Finalizar o questionário

O processo de desenho de questionário

- Um questionário bem desenhado deve:
 - Coletar dados eficientemente com um número mínimo de erros e inconsistências
 - Ser amigável ao respondente e ao entrevistador
 - Levar a uma redução global no custo e tempo associados com a coleta de dados



Três regras de ouro

1) O respondente pode entender a pergunta?



Três regras de ouro

2) O respondente tem condição de responder a pergunta?



Três regras de ouro

3) O respondente está disposto a responder a pergunta?



Regras básicas

- **Simplicidade:** as perguntas devem ser feitas da forma mais simples e coloquial possível
- **Clareza:** ambiguidades devem ser eliminadas

As perguntas (e alternativas de resposta) devem admitir apenas uma interpretação

- **Factibilidade:** só se deve perguntar ao entrevistado aquilo que ele sabe

Não obrigar o entrevistado a fazer contas

Não obrigar o entrevistado a lembrar de algo de um passado muito distante



Seja claro

- ***Qual marca de sabonete você usa?***

A quem se refere? Ao respondente ou a todos os moradores?

Se usar mais de uma marca, como responder?

A qual período de tempo? Esta manhã? Semana passada? Ao longo do último ano?

Onde? Em casa, na academia, em viagens?

- ***Qual marca ou quais marcas de sabonete que você pessoalmente usou em casa no último mês?***

Use palavras simples

- *Você acha que a distribuição de refrigerantes no mercado varejista é adequada?*
- *Você encontra refrigerantes com facilidade quando decide comprá-los?*

Use palavras sem ambiguidade

- *Em um mês típico, quantas vezes você faz compras em supermercados?*

Nunca

Ocasionalmente

Às vezes

Frequentemente

Regularmente

Use palavras sem ambiguidade

- *Em um mês típico, quantas vezes você faz compras em supermercados?*

Menos de 1 vez

1 ou 2 vezes

3 ou 4 vezes

Mais de 4 vezes



Evite...

- Perguntas tendenciosas ou embaraçosas

Alguma vez você já pensou em cometer suicídio?

- Perguntas com negação

Você é a favor ou contra não permitir a venda de bebidas alcoólicas em lojas de conveniência?

- Perguntas hipotéticas

Se você tivesse filhos, eles estudariam em escola pública ou privada?

Evite...

- Jargões, abreviações, palavras sofisticadas, incomuns ou ambíguas

Você sabe a localização da DPCA mais próxima?

Você já tomou a vacina influenza?

Que proporção do tempo que você vê TV à noite você passa assistindo programas novos?

Poucos respondentes são capazes de interpretar corretamente essa pergunta

“Cinquenta por cento do público não sabe realmente o que o termo 50 por cento significa”. (*Patricia Hewitt, UK Trade Secretary*)

O processo de desenho de questionário

- **Pense como deseja analisar os dados**

Evite colocar perguntas fundamentais ao final do questionário e incluir perguntas desnecessárias, apenas porque “seria interessante saber” ...

Dedique mais espaço aos itens mais importantes

- **Seja o mais específico possível com o objetivo**

Perguntas ruins se devem a objetivos não suficientemente específicos

Para alguns conceitos, a diferença entre conceito teórico e pergunta é pequena (idade, sexo)

Para outras, pode ser muito grande (renda, trabalho)



O processo de desenho de questionário

- **Elaborar as perguntas em uma sequência lógica para que:**
o questionário seja compreendido;
os entrevistados sintam-se motivados a responder
o formato dos dados possibilite a realização das análises estatísticas necessárias
- **E de modo que sejam ...**
Limitadas ao escopo da pesquisa (“caronas” alongam o questionário e comprometem sua qualidade)
Ordenadas para conquistar o entrevistado:
 - Filtros básicos no início
 - Perguntas “difíceis” no finalEvitados textos que possam induzir respostas



A importância da formulação exata

Dois monges, um Budista e um Taoísta, estavam discutindo se é pecado fumar e rezar ao mesmo tempo.

Como não chegaram a uma conclusão, decidem que cada um deve consultar seu superior a respeito.

Na semana seguinte, eles se encontram novamente.

O Budista diz “Então, o que o seu superior disse?”

O Taoísta responde: “Ele disse que tudo bem.”

“Engraçado”, o Budista responde. “Meu superior disse que era pecado”.

O Taoísta pergunta: “O que você perguntou a ele?”

O Budista diz. “Perguntei se estaria tudo bem se eu fumasse enquanto rezava”.

“Oh,”, diz o Taoísta, “Eu perguntei se estaria tudo bem se eu rezasse enquanto fumava.”

Qualidade de Perguntas – Validade

Uma pergunta é **válida** quando mede o que pensamos que mede

Validade se refere a obter resultados que reflitam o conceito que se deseja medir

Exemplo: Como está sua saúde?

- Excelente
- Boa
- Regular
- Ruim
- Péssima

O que esta pergunta mede? O estado de saúde do respondente? Ou será a percepção do respondente sobre seu estado de saúde?

Qualidade de Perguntas – Discriminação

Uma boa pergunta deve **captar variação** ou heterogeneidade na variável de interesse

Se todas as **respostas** forem **iguais**, a pergunta não permitirá discriminar, classificar ou ordenar os respondentes

Também são problemáticas perguntas que admitem **respostas vagas** sobre o fenômeno

Exemplo: Indique em qual das categorias está sua renda mensal:

- Menos de R\$15.000,00
- R\$15.000,00 ou mais

A pergunta classifica os respondentes em apenas duas categorias e em cada uma indivíduos podem ter rendas muito diferentes





Vantagens e Desvantagens de perguntas abertas e fechadas?

Perguntas fechadas com várias opções de resposta

Ler ou não as categorias de resposta?

Cartões de resposta quando se tem mais de 5 opções em entrevista pessoal

Por telefone, no máximo 4 categorias ou substituir por classificar de 0 a 10

PRIMACY EFFECT – visual

RECENCY EFFECT – oral



Orientações para redação

- Seja específico - esclareça o que se espera como uma resposta adequada
 - A quem a pergunta se aplica
 - Que informações devem ser incluídas ou excluídas da resposta
 - Em que unidades de medida a resposta deve ser fornecida (por exemplo, kg ou ton)
 - A que período de tempo a pergunta se refere (quando?)

Qual é a sua renda?

No ano passado, qual foi o rendimento total bruto de seu domicílio, ou seja, antes dos descontos e impostos? Inclua o rendimento recebido de salários, vencimentos, e todas as outras fontes.



Orientações para redação

- Forneça referências de tempo e de espaço

Com que frequência você visita sua mãe?

No último ano, com que frequência você visitou sua mãe?

Há quanto tempo você mora aqui?

Há quanto tempo você mora neste bairro?

Orientações para redação

- Não encurte as perguntas que se referem a perguntas anteriores: repita a referência

Hagar





Orientações para redação

- Evite perguntas tendenciosas

Pergunta tendenciosa:

Por favor, diga se você concorda, discorda ou não têm opinião sobre a seguinte afirmação: "O turismo é benéfico para o Rio de Janeiro e, portanto, deve ser promovido."

Pergunta neutra:

Por favor, diga se você concorda, discorda ou não têm opinião sobre a seguinte afirmação: "O turismo para o Rio de Janeiro deve ser promovido."

Orientações para redação

- Evite perguntas duplas

Com que frequência você visita seus pais?

Você geralmente compra carne e peixe congelados?

- Empregue palavras que sejam específicas e concretas (ao invés de geral e abstratas)

Na semana passada, você se exercitou?



Orientações para redação

- Evite associar a opinião a uma pessoa de prestígio

Qual é a sua opinião sobre a política do Papa em relação ao aborto?

- Evite/amenize perguntas sobre comportamentos socialmente indesejáveis

Quantas vezes você faltou aula no último mês?

Na semana passada, quantas vezes você entrou em sites de vídeos pornô na Internet?



Viés de resposta socialmente aceitável

Em questões sobre assuntos socialmente polêmicos, é esperado haver distorções nas respostas dadas pelos entrevistados.

Pessoas sendo entrevistadas naturalmente tendem responder de forma que considerem **socialmente desejável** ou esperado pelo entrevistador

Erro de memória

Recall error

Questões sobre o passado são objetos de duas principais fontes de erros.

1) Compreensão

2) Falhas de memória: são de maior importância para as questões retrospectivas do que para outros tipos de perguntas.

Nos casos em que a informação deve ser recuperada inteiramente da memória, a tarefa depende de dois fatores:

- Relevância que a própria mente dá à informação
- Tempo decorrido desde que a informação foi armazenada



Outras regras gerais

- Cuidado ao perguntar informações por terceiros

Quantas horas seu marido trabalhou semana passada em todos os empregos?

- Evite perguntas que supõem conhecimento prévio sobre o assunto

Qual o sistema eleitoral que você prefere?

voto proporcional

voto distrital

voto distrital misto

Erro de resposta

Erros de resposta causados pelo questionário podem ser devidos a:

O tipo de pergunta (aberta ou fechada)

A formulação da pergunta

O tamanho do questionário (levando ao cansaço do entrevistado)

O layout do questionário (por exemplo, saltos complicados podem causar erros, principalmente nos questionários em papel)

O processamento do questionário



Erro de resposta

Outros erros de resposta:

Erro de memória - dificuldade de lembrar

Erro telescópico – excesso de informação

Entrevistadores diferentes

Registro incorreto de respostas

Reação do entrevistador



Técnicas para reduzir erros de resposta

Tamanho da pergunta - perguntas mais longas auxiliam o entrevistado a se concentrar em um novo tópico

Curta: Que problemas de saúde você teve no ano passado?

Longa: A próxima pergunta trata sobre problemas de saúde durante o ano passado. Pedimos esta informação a todos na pesquisa. Que problemas de saúde você teve no ano passado?



Períodos de referência

Todas as perguntas sobre eventos ou comportamento precisam vir acompanhadas de períodos de referência

Longos períodos fornecem uma quantidade maior de comportamentos

Pequenos períodos amenizam dificuldade de lembrar

Períodos de referência devem ser adequados ao comportamento que está sendo estudado e ao objetivo

Questões factuais: fatos, comportamentos e eventos

Fatos: características demográficas e socioeconômicas

Idade

Estado civil

Ocupação/atividade

Sexo

Raça/Cor

Renda



Questões factuais - Idade

Qual é a sua idade?

Quantos anos você fez no seu último aniversário?

Em que ano você nasceu?

Qual é sua data de nascimento exata?

Em qual destes grupos está sua idade: menos de 30, 30 a 44, 45 a 64, 65 ou mais?



Questões factuais – Estado civil

É o estado civil ou situação conjugal?

Quem é divorciado, pode estar vivendo maritalmente (casado) com outra pessoa

Cuidado com a forma de escrever

Solteiro pode significar “não casado” e respondentes podem escolher esta resposta se são divorciados ou viúvos. A frase “nunca casou” pode ser opção melhor.

Precisa estudar melhor forma de dizer “convivendo, morando junto, coabitando...”;

Considerar ainda casais de mesmo sexo e contratos de união estável



Número de pontos em escalas de avaliação

Importante: os pontos devem cobrir todas as regiões da medição; devem ser ordinais (progredindo e não sobrepondo); cada respondente deve ter um entendimento preciso e estável de cada ponto da escala.

Escala

- 5 pontos

Escala com diferença semântica

- 7 pontos

Escalas com intervalo

- 11 pontos



Inversa →



Rotulagem dos pontos das escalas de avaliação

Após definir o tamanho da escala...

Como dispor a escala no questionário?

Usar números, palavras, ambos?





Responda, numa escala de 1 (discordo completamente) a 7 (concordo completamente), qual o seu grau de concordância com cada uma das afirmações seguintes:

Item	1 – Discordo completamente	2	3	4	5	6	7 – Concordo completamente
1 - Se não se tiver saúde não se tem nada	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 – Existem muitas coisas com que me preocupo mais do que com a minha saúde	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 – Uma boa saúde é apenas um pormenor de menor importância numa vida feliz	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4 – Não há nada mais importante do que ter uma boa saúde	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

É fácil encontrar o produto desejado navegando pelas seções do site?

- concordo plenamente
- concordo parcialmente
- não concordo nem discordo
- discordo parcialmente
- discordo totalmente

Layout de questionário

Carta de apresentação

Sobre o que é a pesquisa, por que é importante e como os resultados serão utilizados

Por que o respondente é importante para o estudo

Como o respondente foi selecionado

Garantia de sigilo/confidencialidade

Um número de telefone para tirar dúvidas

Deve ser atraente, clara, motivadora

Layout de questionário

- Introdução

fornecer o título ou assunto da pesquisa

identificar o patrocinador da pesquisa

indicar que as respostas serão confidenciais e indicar eventuais acordos para compartilhar dados com outros órgãos de estatística, departamentos de governo, clientes, etc

para pesquisas por correio, fornecer o endereço para devolução e a data para o retorno

Layout de questionário

Perguntas iniciais

- Devem se relacionar com a finalidade da pesquisa

- Dirigidas a todos os entrevistados

- Fáceis de responder

- Evitar pergunta aberta com resposta detalhada

Localização de perguntas sensíveis

- Onde o entrevistado possa se sentir confortável em responder e onde sejam mais significativas no contexto de outras perguntas

Layout de questionário

Formato

Consistência entre as fontes utilizadas, que devem ser seguidas para as perguntas, instruções, cabeçalhos e introduções de blocos

Capa

Cores

Caixas de resposta para perguntas abertas e os círculos de resposta para perguntas fechadas

Respostas múltiplas – box quadrado

Resposta única - círculo





Referências

- Bradburn, N. M. (2004). Understanding the question-answer process. *Survey Methodology*, 30(1), 5-15.
- De Leeuw, E. D., & Dillman, D. A. (2008). *International handbook of survey methodology*. Taylor & Francis.
- Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. *Handbook of survey research*, 2, 263-314.

Nova Fonte de Dados :Big Data

Definição da UNECE 2013

“Big Data are data sources that can be – generally – described as: high volume, velocity and variety of data that demand cost - effective, innovative forms of processing for enhanced insight and decision making.”

Características

- ✓ volume de dados (*massive data*)
- ✓ velocidade (capacidade de atualização de informações)
- ✓ variedade de informações
- ✓ informação não estruturada
- ✓ informação dinâmica e capturada em eventos

Big Data – Características

AAPOR Report on Big Data, 12 Fev 2015

Big Data in Survey Research AAPOR Task Force Report

Public Opin Q (2015) 79 (4): 839-880. doi: 10.1093/poq/nfv039

Variabilidade ou Volatilidade

Inconsistência/instabilidade dos dados ao longo do tempo e problemas de reprodutibilidade

Ex. algoritmos de busca – Google Flu Trend

Veracidade

Necessidade de avaliação da acurácia dos dados

Complexidade

Necessidade de integração de várias bases de dado



Big Data – Tipologia

Uma possível tipologia

- Dados gerados por pessoas
 - Media social, mensagens de texto
- Dados gerados por máquinas
 - Sensores, imagens de satélite, GPS
- Dados gerados em processos
 - Transações comerciais, cartões de crédito, registros médicos

Big Data e Pesquisas Quantitativas

Pesquisas: baseadas em dados coletados de forma estruturada e organizada, sobre uma população bem definida, com arcabouço inferencial e metodologia reprodutível

Big Data: conjunto de dados coletados de forma auxiliar e não estruturada com o objetivo de registrar um processo, sobre uma população não definida, sem arcabouço inferencial ou metodologia reprodutível

Pesquisas

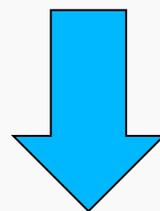
- Dados planejados
- Processo de coleta bem estruturado
- População alvo bem definida
- Métodos de Inferência Estatística disponíveis
- Boa reprodutibilidade

Big Data

- Dados orgânicos
- Sub-produto de processos
- População alvo e escopo nem sempre definidos
- Métodos de Inferência Estatística em desenvolvimento
- Problemas de reprodutibilidade

Pesquisas quantitativas no cenário atual de evolução do Big Data

A utilização de dados de fontes alternativas pode ser parte da solução para os desafios colocados para produção de informações



Novas fontes de dados devem permitir a realização de inferências e seguir padrões de qualidade das Estatísticas Públicas ou outros a serem definidos (que garantam sua qualidade/integridade técnico-científica)

Big Data e os desafios para a Produção de Informações

- Acesso às bases de dados de forma regular e *independente*
- Vieses associados ao Big Data: como utilizar o Big Data?
- Como trabalhar com os métodos estatísticos usuais de estimação em bases de dados de tamanho tão grande?
- Utilização de bases de dados coletadas por diferentes fontes de informação sem controle sobre o processo de coleta ou suas modificações
- **Proteção de dados pessoais e questões éticas**

Vieses associados ao Big Data

- **Viés de cobertura** – população do Big Data não corresponde à população de interesse
- **Viés de seleção** – população do Big Data representa a população que “informa” os dados e não o conjunto completo da população de interesse
- **Viés de representação** – Big Data contém informação sobre diferentes populações (empresas e indivíduos por exemplo)
- **Viés de medida** (erro de medida) – informações do Big Data não correspondem à medida que se deseja obter

Alguns Desafios

- Como garantir acesso regular e independente às bases de dados?
- Métodos estatísticos para lidar com viés – não há nenhuma informação sobre os dados faltantes– (*monkey-typewriter metaphor with some keys disabled* - Groves, 2011)
- Extração de sinal na presença de ruído (como obter informação útil a partir de dados sujeitos a variadas perturbações)?
- Métodos estatísticos para enormes quantidades de dados/*massive data* (presença de correlações espúrias, complicações associadas à crítica e limpeza dos dados)
- Pareamento de dados sem conhecer direito as bases originais

Big Data e os desafios para a Produção de Informações

Acesso às bases de dados de forma regular e com protocolos estabelecidos

- Dificuldades de acesso a dados de fontes privadas ou a registros administrativos
- Dados são usualmente considerados informações estratégicas para empresas (ou legalmente inacessíveis)
- Questões éticas de sigilo dos usuários/informantes
- Marco regulatório para proteção de dados pessoais - a questão do consentimento - (PL 5276/2016)
- **Novos protocolos de integração entre instituições devem ser construídos para permitir o acesso aos dados de forma independente e garantindo a neutralidade no uso e disseminação da informação**

Big Data e os desafios para a Produção de Informações Estatísticas

Quais métodos estatísticos?

With very large amounts of data, direct use of standard statistical methods, including simulation-based approaches, will tend to produce estimates of apparently very high precision, essentially because of strong explicit or implicit assumptions of at most weak dependence underlying such methods (Cox, 2015)

In fact quality of the statistics is very much dependent on how representative a particular Big Data source is of the underlying population, and on the nature of the statistical inference drawn from such data (Tam & Clarke, 2015a)

Paolo Righi - Global Conference on Big Data for Off. Statistics. 20-22 October 2015 Abu Dhabi



Big Data e os desafios para a Produção de Estatísticas Públicas

Quais métodos estatísticos?

- Amostragem
- Métodos de replicação para análise e estimação de medidas
 - Bag of Little Bootstraps
(http://web.cs.ucla.edu/~ameet/blb_icml2012_final.pdf)
- Modelos de Superpopulação
- Métodos para estimação na presença de dados faltantes (hipóteses MAR)

Iniciativas na América Latina e Brasil

“Opportunities and Requirements for Leveraging Big Data for Official Statistics and the Sustainable Development Goals in Latin America.”
Data-Pop Alliance (Harvard Humanitarian Initiative, MIT Media Lab and Overseas Development Institute). November 2015.

IBGE/Brasil

Mapa de cobertura e uso da terra (uso de imagens de satélites)

Uso de webscraping para o componente de preços de passagens aéreas do SNIPC



**Escola Nacional de
Ciências Estatísticas**

Programa de Pós-Graduação

**Mestrado e Doutorado
em População, Território
e Estatísticas Públicas**

20ANOS

**PROGRAMA DE PÓS-GRADUAÇÃO EM POPULAÇÃO,
TERRITÓRIO E ESTATÍSTICAS PÚBLICAS**

Objetivos do Programa

Os cursos de Mestrado e Doutorado em População, Território e Estatísticas Públicas da ENCE têm por objetivo promover a formação de pesquisadores e profissionais para o desempenho de atividades de pesquisa e ensino, capazes de desenvolver e aplicar abordagens teóricas, modelos e métodos para a análise de processos sócio-espaciais bem como para a produção de informações e estatísticas públicas que ampliem o conhecimento da realidade populacional e territorial e de sua dinâmica.

<http://www.ence.ibge.gov.br/>



Algumas referências

- Buelens, B., Daas, P., Burger, J., Puts, M., and Van den Brakel, J. (2014), “Selectivity of Big Data”, Discussion Paper 201411, Statistics Netherlands, The Hague/Heerlen, The Netherlands.
- Citro, C. F. (2014). From Multiple Modes for Surveys to Multiple Data Sources for Estimates. *Survey Methodology*, 40, 137–161.
- Couper, M. (2013). Is the sky falling? New technology, changing media, and the future of surveys. *Survey Research Methods* 7, 145 – 156
- Daas, P. and Puts, M. (2014) Big data as a source of statistical information. *The Survey Statistician* 69, 22- 31. <http://isi.cbs.nl/iass/N69.pdf>
- Daas, P. J. H., Roos, M., Van de Ven, M., and Neroni, J. (2012). Twitter as a potential data source for statistics. Discussion Paper 201221, Statistics Netherlands, The Hague/Heerlen, The Netherlands.
- Data-POP Alliance. Resources and Publications. <http://datapopalliance.org/resources/>



Algumas referências

- Groves, R. M. (2011). Three Eras of Survey Research. *Public Opinion Quarterly* (Special Issue)75 (5): 861-871.doi: 10.1093/poq/nfr057
- Holt D. (2007). The Official Statistics Olympic Challenge: Wider, Deeper, Quicker, Better, Cheaper (with discussion). *American Statistician*, 61, 1–15.
- Pferffermann, D. (2015). Methodological Issues and Challenges in the Production of Official Statistics 24th Annual Morris Hansen Lecture. *Journal of Survey Statistics and Methodology* 3 (4):425-483.doi: 10.1093/jssam/smv035
- Tam, S. and Clarke, F. (2015). Big Data, Official Statistics and Some Initiatives by the Australian Bureau of Statistics. *International Statistical Review*. Volume 83, Issue 3, pages 436–448, December 2015.
- Tam, S. and Clarke, F. (2015). Big data, statistical inference and official statistics. Research Paper. Australian Bureau of Statistics.
- United Nations (2014). Report of the Global Working Group on Big Data for Official Statistics. Available at <http://unstats.un.org/unsd/statcom/doc15/2015-4-BigData-E.pdf>





Referências

Backstrom, C. H. e Hursch-César, G. (1981). *Survey Research, 2nd edition*. Nova Iorque: Macmillan Publishing Company.

Czaja, R. e Blair, J. (2005). *Designing surveys: a guide to decisions and procedures, 2nd edition*. Londres: Sage. Capítulos 1 e 2.

De Vaus, D. (2002). *Surveys in social research, 5th edition*. Londres: Routledge. Capítulos 1 a 5.

Groves, R.M. et al. (2004). *Survey Methodology*. Hoboken: John Wiley & Sons.

Salant, P. e Dillman, D. A. (1994). *How to conduct your own survey*. Nova Iorque: John Wiley & Sons.

Referências

UNECE (2009a). *Making data meaningful: Part 1 - a guide to writing stories about numbers*. Geneva, United Nations Economic Commission for Europe. 1: 28.

UNECE (2009b). *Making Data Meaningful: part 2 - a guide to presenting statistics*. Geneva, United Nations Economic Commission for Europe: 52.

Wild, C. J. and G. A. F. Seber (2004). *Encontros com o acaso: um primeiro curso de análise de dados e inferência*. Rio de Janeiro, LTC - Livros Técnicos e Científicos Editora S.A.

