**Medindo o custo não monetário do desemprego[[1]](#footnote-1)**

1. **Os objetivos da atividade**

* Trabalhar com conjuntos grandes de dados
* detectar e corrigir entradas em um conjunto de dados
* recodificar variáveis para facilitar a análise
* calcular percentis para subconjuntos dos dados
* calcular intervalos de confiança para a diferença na média entre dois grupos

1. **Contextualização**

Contratar empregados é diferente de comprar bens e serviços. A razão é que o empregador não pode redigir um contrato de trabalho executável que compreenda exatamente as tarefas que o trabalhador deve executar - quanto esforço a exercer - para receber o pagamento.

Embora o nível de esforço não possa ser especificado precisamente no contrato de trabalho, há muitas razões pelas quais as pessoas escolhem trabalhar duro e exercer esforço em seus trabalhos, mesmo que isso possa ser mental e fisicamente desgastante. Por exemplo, as pessoas podem encontrar trabalho duro e fazer um bom trabalho recompensando-as si mesmas, ou podem sentir um senso de responsabilidade por outros trabalhadores em sua equipe ou por seu empregador. Outra razão é o medo de ser demitido ou de não ser promovido a uma posição que lhe desse maior remuneração e maior segurança no emprego. Os trabalhadores não gostariam de perder seus empregos se não conseguissem imediatamente outro emprego sob as mesmas condições salariais e de trabalho. Existem, portanto, benefícios e custos para o trabalho. A diferença (benefícios menos custos) é conhecida como "renda do trabalho", o que é positivo para todos que escolhem trabalhar. Isso significa que é caro perder seu emprego.

A renda do trabalho leva em conta tanto os custos quanto os benefícios monetários e não monetários. Por exemplo, estar empregado poderia dar utilidade através de status social e conformidade com a norma social de ter um emprego; estar desempregado pode ser psicológica e financeiramente prejudicial, devido ao estresse de procurar um novo emprego e não atender às normas e expectativas sociais em relação ao trabalho. Se as sociedades têm a convicção de que pessoas capazes, em idade ativa, deveriam trabalhar, então estar desempregado pode resultar em medo de ser estigmatizado, um sentimento de vergonha por estar desempregado e de se sentir inferior a outras pessoas empregadas, tudo isso reduziria a utilidade e a satisfação com a vida de um indivíduo.

A desutilidade do desemprego é um conceito que não podemos medir diretamente, então, em vez disso, usaremos o bem-estar autodeclarado. Essa medida tem suas limitações, mas é amplamente utilizada para quantificar os custos que não podemos observar, como o efeito de se tornarem doentes crônicos ou outros eventos de mudança de vida.

Utilizaremos uma abordagem e dados semelhantes ao estudo [‘Employment status and subjective well-being’](https://tinyco.re/525262), que utilizou o European Values Study (EVS), uma pesquisa entre países, para investigar as diferenças na satisfação de vida entre pessoas de diferentes status de emprego. A hipótese era que os desempregados ficariam, em média, menos satisfeitos com a vida do que os empregados, e que essa relação entre status de emprego e bem-estar variaria de acordo com as normas sociais.

Uma explicação para uma relação entre o status de emprego e o bem-estar declarado são as normas sociais relativas a uma ética de trabalho. Se as normas sociais são um determinante importante do bem-estar, então esperamos que a diferença de bem-estar entre empregados e desempregados seja maior em países com uma ética de trabalho autodeclarada mais forte.

Embora o foco principal deste projeto seja o emprego (em tempo integral) e os desempregados, também consideraremos se o bem-estar difere para outros status de emprego, como ser aposentado. As expectativas de trabalho podem não ser tão fortes para os idosos, então a falta de emprego formal teria menos efeito sobre o bem-estar do que as pessoas em idade de trabalho que estão desempregadas.

1. **Limpar e organizar os dados**

Até agora temos trabalhado com dados formatados corretamente. No entanto, às vezes, os conjuntos de dados ficarão confusos e precisaremos "limpar" os dados antes de iniciar qualquer análise. Nos dados que usaremos, os dados do Estudo de Valores Europeus (EVS) (European Values Study) foram convertidos para o Excel de outro programa, portanto, há algumas entradas que não foram convertidas corretamente e algumas variáveis que precisam ser recodificadas, por exemplo, substituindo palavras por números ou substituindo um número por outro.

Faça o download dos dados e documentação do EVS:

* Faça o download dos [dados do EVS](https://tinyco.re/9465901).
* Para a documentação, acesse o [site de download de dados](http://tinyco.re/6701388).
* Clique em "Data and Documents" no meio da página, depois clique em "Other Documetns" e faça o download do arquivo PDF chamado "ZA4804\_EVS\_VariableCorrespondence.pdf".

1. A planilha de dados contém um dicionário de dados incompleto e quatro guias com os dados coletados de cada onda da pesquisa. Os nomes das variáveis atualmente não nos dizem qual é a variável, portanto, precisamos rotulá-los novamente para evitar confusão.
   1. Use o PDF que você baixou para preencher a coluna C (Variable description) da guia Dicionário de dados.
   2. Preencha a coluna B com um novo nome apropriado para cada variável e renomeie as variáveis nas outras quatro guias de acordo.

Ao longo desta atividade, vamos nos referir às variáveis usando seus nomes originais, então a guia do dicionário de dados será útil.

1. Agora vamos dar uma olhada mais de perto em como algumas das variáveis ​​foram medidas.
   1. A variável A170 é a satisfação com a vida declarada em uma escala de 1 (insatisfeito) a 10 (satisfeito). Os entrevistados responderam à pergunta "Tudo considerado, o quanto você está satisfeito com sua vida hoje em dia?" Discuta as suposições necessárias para usar essa medida em comparações interpessoais e entre países, e se você acredita que elas são plausíveis. Você pode achar útil consultar a [Caixa 2.1 das diretrizes da OCDE para medir o bem-estar subjetivo](https://tinyco.re/1768499).
   2. O status de emprego de um indivíduo (variável X028) foi autodeclarado. Explique se a declaração incorreta do status de emprego provavelmente é um problema e forneça alguns fatores que podem afetar a probabilidade de declaração incorreta nesse contexto.
   3. As variáveis ​​C036 a C041 perguntam sobre as atitudes de um indivíduo em relação ao trabalho. Com autodeclarados, também podemos nos preocupar que os indivíduos estejam usando uma heurística (regra de ouro) para responder às perguntas. A [Tabela 2.1 das "Diretrizes da OCDE para medir o bem-estar subjetivo](https://tinyco.re/1768499) lista alguns vieses de resposta e heurísticas que os indivíduos poderiam usar. Escolha três que você acha que se aplicam particularmente as perguntas sobre satisfação com a vida ou ética profissional e descreva como podemos verificar se esse problema pode estar presente em nossos dados.
2. Vamos agora verificar e limpar o conjunto de dados para que esteja pronto para uso. Faça as seguintes alterações em todas as guias de dados relevantes:
   1. Atualmente os valores ausentes são registrados como ".a", mas gostaríamos que eles fossem células em branco. Use a ferramenta “Localizar e Substituir” do Excel para alterar o ".a" para células em branco para as variáveis ​​A009 a X047D. (Veja o passo a passo do Excel 3.1 para obter ajuda sobre como fazer isso).
   2. A variável A170 (satisfação de vida) é atualmente uma mistura de números (2 a 9) e palavras "Satisfeito" e "Insatisfeito", mas gostaríamos que fossem todos os números. Substitua a palavra "Insatisfeito" pelo número 1 e a palavra "Satisfeito" pelo número 10.
   3. Da mesma forma, a variável X011\_01 (número de filhos) não registrou filhos como uma palavra em vez de um número. Substitua "Sem filhos" pelo número 0.
   4. As variáveis ​​C036 a C041 devem ser substituídas por números que variam de 1 (Discordo totalmente) a 5 (Concordo totalmente) para que possamos calcular a média deles mais tarde. Da mesma forma, a variável A009 deve ser recodificada como 1 = "Muito pobre", 2 = "Ruim", 3 = "Boa", 4 = "Boa", 5 = "Muito boa".
   5. Gostaríamos de dividir o X025A em duas variáveis, uma para o número antes dos dois pontos e outra contendo as palavras após os dois pontos. Use as funções ESQUERDA e / ou DIREITA do Excel para criar duas variáveis ​​de acordo. (Veja o passo a passo do Excel 3.1 para obter ajuda sobre como fazer isso).
   6. **Limpar dados e dividir variáveis**

Embora o artigo que estamos acompanhando considere apenas indivíduos de 25 a 80 anos que não eram estudantes, nós manteremos essas observações em nossa análise. No entanto, precisamos remover quaisquer observações que tenham valores ausentes para A170, X028 ou qualquer uma das outras variáveis, exceto X047D.

1. Remova todas as linhas com valores ausentes da sua planilha. O passo a passo do Excel 3.2 fornece orientações sobre como fazer isso. Lembre-se que nem todas as perguntas foram feitas, portanto, consulte o PDF da documentação de dados para verificar se deve haver valores ausentes em uma coluna inteira. Nota: o Excel pode levar algum tempo para processar os comandos.
   1. **Descartando observações que satisfazem condições particulares**
2. Vamos agora criar duas variáveis, ética de trabalho e renda relativa, que usaremos em nossa comparação de bem-estar.
   1. A ética de trabalho é medida como a média de C036 a C041. Crie uma variável mostrando esta medida ética de trabalho.
   2. O estudo calculou a renda relativa de um indivíduo como um desvio da renda média no país daquele indivíduo. Explique os problemas com o uso desse método se a distribuição de renda estiver inclinada (por exemplo, uma cauda longa à direita).
   3. Em vez de usar a renda média, definiremos a renda relativa como o percentual do indivíduo na distribuição de renda. Crie uma variável mostrando essa informação. (Veja o Excel 3.3 para uma maneira de fazer isso).
   4. **Calculando percentis de valores reais**
3. Agora temos todas as variáveis que precisamos no formato que precisamos. Vamos fazer algumas tabelas para resumir os dados. Usando os dados da questão IV:
   1. Crie uma tabela mostrando a divisão da população de cada país de acordo com o status de emprego, com o país (S003) como a variável de linha e o status de emprego (X028) como a variável de coluna. Expresse os valores como porcentagens do total da linha em vez de frequências. Discuta quaisquer diferenças ou semelhanças entre os países que você acredita que seja interessante.
   2. Crie uma tabela, como mostra a Figura 1, semelhante à Tabela 1 do estudo [‘Employment status and subjective well-being’](https://tinyco.re/525262), e preencha os valores omissos. Tabelas sínteses como estas fornecem uma visão geral útil de como cada variável se parece.

Figura 1 – Estatísticas descritivas por sexo, estudo de valores europeus

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variáveis** | **Masculino (Male)** | | **Feminino (Female)** | |
| **Média** | **Desvio-padrão** | **Média** | **Desvio-padrão** |
| Wellbeing |  |  |  |  |
| Self-reported health |  |  |  |  |
| Work ethic |  |  |  |  |
| Age |  |  |  |  |
| Education |  |  |  |  |
| Number of children |  |  |  |  |

1. **Visualizando os dados**

Agora, criaremos alguns gráficos com as medidas autodeclaradas (ética do trabalho e bem-estar), começando com gráficos de colunas para mostrar as distribuições de valores. Juntamente com o status do emprego, essas são as principais variáveis de interesse; portanto, é importante analisá-las cuidadosamente antes de fazer uma análise mais aprofundada dos dados.

A distribuição da ética e do bem-estar do trabalho pode variar entre os países, mas também pode mudar ao longo do tempo dentro de um país, principalmente porque as pesquisas são realizadas cerca de uma vez por década. Para comparar distribuições para um país específico ao longo do tempo, precisamos usar o mesmo eixo horizontal; portanto, primeiro precisamos criar uma tabela de frequências para cada distribuição de interesse. Além disso, como o número de pessoas pesquisadas pode diferir, usaremos porcentagens em vez de frequências como a variável do eixo vertical.

1. Usando apenas os dados do “Wave 3” e “Wave 4”, para três países de sua escolha:
   1. Usando a medida ética do trabalho, crie uma tabela de frequências separada para a “Wave 3” e para a “Wave 4”, semelhante à Figura 2. Os valores na primeira coluna devem variar de 1 a 5, em intervalos de 0,2.

Figura – Tabela de frequências para ética de trabalho autodeclarada

| **Intervalo do score ética no trabalho** | **Frequência** | **Percentual de indivíduos (%)** |
| --- | --- | --- |
| 1.00 |  |  |
| 1.20 |  |  |
| … |  |  |
| 4.80 |  |  |
| 5.00 |  |  |

* 1. Plote um gráfico de colunas para cada país para mostrar a distribuição das pontuações de ética no trabalho na “Wave 3”, com a porcentagem de indivíduos no eixo vertical e a faixa de pontuações na ética no trabalho no eixo horizontal. Em cada gráfico, plote a distribuição de pontuações na “Wave 4”, em cima da distribuição da “Wave 3”.
  2. Com base em seus gráficos de I (b), parece que as atitudes em relação ao trabalho para cada país mudaram ao longo do tempo?

1. Usaremos gráficos de linhas para fazer uma comparação semelhante para o bem-estar ao longo do tempo. Usando dados para países presentes nas “Waves 1 a 4”:
   1. Crie uma tabela mostrando o bem-estar médio, por “wave” (variável de coluna) e por país (variável de linha). Dica: talvez seja mais fácil criar quatro tabelas dinâmicas separadas e copiar e colar os dados em uma nova tabela.
   2. Plote um gráfico de linhas com o número de “Wave” (1 a 4) no eixo horizontal e o bem-estar médio no eixo vertical. Certifique-se de incluir uma legenda.
   3. Dos seus resultados em II (a) e (b), parece que o bem-estar mudou ao longo do tempo? Que outras informações sobre a distribuição do bem-estar poderíamos usar para suplementar esses resultados?

Após descrever padrões em nossas principais variáveis ​​ao longo do tempo, usaremos coeficientes de correlação[[2]](#footnote-2) e gráficos de dispersão para examinar a relação entre essas variáveis ​​e as outras variáveis ​​em nosso conjunto de dados.

1. Usando os dados do “Wave 4”:
   1. Crie uma tabela como mostrado na Figura 3 e calcule os coeficientes de correlação necessários (para status de emprego e gênero, você precisará criar variáveis).

Figura – Correlação entre bem-estar, ética no trabalho e outras variáveis, Onda 4

|  |  |  |
| --- | --- | --- |
| **Variable** | **Wellbeing** | **Work ethic** |
| Age |  |  |
| Education |  |  |
| Employment status (= 1 if full-time employed, = 0 if unemployed) |  |  |
| Gender (= 0 if male, = 1 if female) |  |  |
| Self-reported health |  |  |
| Income |  |  |
| Number of children |  |  |
| Relative income |  |  |
| Wellbeing | 1 |  |
| Work ethic |  | 1 |

* 1. Interprete os coeficientes, prestando muita atenção em como as variáveis são codificadas. Explique se as relações apresentadas pelos coeficientes são o que você esperava (por exemplo, você esperaria que o bem-estar aumentasse ou diminuísse com saúde, renda etc.)

Em seguida, examinaremos a relação entre status de emprego e bem-estar e investigaremos a hipótese do artigo de que essa relação varia com a ética média do trabalho em um país.

1. Usando os dados do “Wave 4”, faça o seguinte:
   1. Crie uma tabela mostrando o bem-estar médio de acordo com o status do emprego (mostrando apenas as categorias de empregados, aposentados e desempregados em tempo integral), com o país (S003) como variável de linha e o status de emprego (X028) como variável da coluna. Comente quaisquer diferenças no bem-estar médio entre esses três grupos e se as normas sociais são uma explicação plausível para essas diferenças.
   2. Use a tabela de IV (a) para calcular a diferença no bem-estar médio (empregado de tempo integral menos desempregado e empregado de tempo integral menos aposentado).
   3. Plote gráficos de colunas mostrando essas duas diferenças no bem-estar, com o país no eixo horizontal (classificado da menor para a maior ética média de trabalho).
   4. A diferença no bem-estar entre empregados e desempregados varia de acordo com a ética de trabalho média do país?

Até o momento, descrevemos os dados usando tabelas e gráficos, mas não fizemos nenhuma declaração sobre se o que observamos provavelmente se deve ao acaso ou não. Na próxima parte, avaliaremos a significância estatística da relação entre status de emprego e bem-estar.

1. **Intervalos de confiança para diferenças na média**

O objetivo desta atividade era examinar a relação empírica entre emprego e bem-estar. Em outras palavras, se o bem-estar médio das pessoas empregadas (estatisticamente) é significativamente diferente do bem-estar médio das pessoas desempregadas (como poderíamos esperar da teoria econômica)?

Quando calculamos diferenças entre grupos, também podemos querer saber se essas diferenças são estatisticamente significativas. Agora aprenderemos como construir intervalos de confiança para a diferença de duas maneiras, o que nos permite avaliar a significância estatística diretamente em vez de usar a regra de ouro.

Lembre-se de que a amplitude de um intervalo de confiança depende do desvio padrão e do número de observações. Ao fazer um intervalo de confiança médio para uma amostra (como o bem-estar médio dos desempregados), usamos o desvio padrão e o número de observações nessa amostra (pessoas desempregadas).

Quando observamos a diferença de médias (como o bem-estar dos empregados menos os desempregados), estamos usando dados de dois grupos (os desempregados e os empregados) para estabelecer um intervalo de confiança; portanto, o número de observações é a soma das observações nos dois grupos. Para calcular o desvio padrão para a diferença de médias, usamos os desvios padrão (DP) de cada grupo:

Essa fórmula exige que os dois grupos de dados sejam independentes, o que significa que os dados de um grupo não estão relacionados, emparelhados ou combinados com dados do outro grupo. Essa suposição é razoável para os dados de bem-estar que estamos usando. No entanto, se os dois grupos de dados não são independentes, por exemplo, se as mesmas pessoas geraram os dois grupos de dados, não podemos usar essa fórmula.

Depois de termos o novo desvio padrão e o número de observações, podemos calcular a largura do intervalo de confiança (distância da média até o final do intervalo) como antes. A diferença de médias é estatisticamente significativa se o intervalo de confiança para a diferença de médias não contiver 0.

1. Aplicaremos esse método para fazer intervalos de confiança para diferenças no bem-estar. Escolha três países: um com uma ética de trabalho média no terço superior das pontuações, um no terço médio e um no terço inferior.
   1. Crie uma Tabela Dinâmica para esses países, mostrando a pontuação média de satisfação com a vida, o desvio padrão (StdDev) da satisfação com a vida e o número de observações, com o país (S003) como variável de linha e o status do emprego (empregado em tempo integral, aposentado e apenas desempregados) como a variável da coluna.
   2. Use sua Tabela Dinâmica de I (a) para calcular a diferença de médias (empregado de tempo integral menos desempregado e empregado de tempo integral menos aposentado), o desvio padrão dessas diferenças e o número de observações.
   3. Use a função CONFIDENCE.T do Excel e os valores calculados em I (b) para determinar a amplitude do intervalo de confiança (distância entre a média e um final do intervalo) da diferença de médias (usando um nível de significância de 5%).
   4. Plote um gráfico de colunas para os países escolhidos, mostrando a diferença de bem-estar (empregado versus desempregado e empregado vs aposentado) no eixo vertical e país no eixo horizontal (classificados de acordo com a ética de trabalho baixa, média e alta). Adicione os intervalos de confiança de 1 (c) ao seu gráfico.
   5. Discuta a significância estatística de suas descobertas.
2. O método que usamos para comparar o bem-estar contava com comparações entre pessoas com diferentes status de emprego, mas o status de emprego de uma pessoa não é inteiramente aleatório. Portanto, não podemos fazer declarações causais, como 'estar desempregado causa diminuição do bem-estar'. Descreva como poderíamos usar os métodos a seguir para avaliar melhor o efeito de estar desempregado no bem-estar e fazer algumas declarações sobre causalidade:
   1. um experimento “natural”
   2. dados do painel (dados sobre os mesmos indivíduos, obtidos em diferentes momentos).

1. Preparada com base em Doing Economics: empirical projects (http://www.core-econ.org/doing-economics/index.html). [↑](#footnote-ref-1)
2. Medida numérica de quão estreitamente associadas duas variáveis ​​estão e se elas tendem a assumir valores semelhantes ou diferentes, variando de um valor de 1, indicando que as variáveis ​​assumem valores semelhantes (correlacionados positivamente) a -1, indicando que as variáveis tomem variáveis ​​diferentes (correlação negativa ou inversa). Um valor de 1 ou -1 indica que conhecer o valor de uma das variáveis ​​permitiria prever perfeitamente o valor da outra. Um valor 0 indica que conhecer uma das variáveis ​​não fornece informações sobre o valor da outra. [↑](#footnote-ref-2)