

Como encontrar resposta à minha pergunta ?

O desafio da Estatística Aplicada na solução de problemas práticos.

Ana Beatriz Monteiro Fonseca

Departamento de Estatística - UFF

abmfonseca@id.uff.br

SEM
9 **EST**

9ª Semana da Estatística

Universidade Federal Fluminense

23 a 26 de outubro de 2017

Proposta:

Diferentes **personagens fictícios**, trazendo suas **situações-problema**, se apresentarão a diferentes grupos de consultoria em Estatística e solicitarão respostas a suas perguntas, levando em consideração os **dados** por eles apresentados.

Consultorias:

Colocar os participantes na posição de estatísticos encarregados de responder às perguntas dos profissionais fictícios, usando as ferramentas que já dispõem.

Dinâmica:

Participantes serão divididos em consultorias, de forma a possibilitar:

- troca de experiências,
- discussão na escolha das ferramentas,
- realização das análises e
- apresentação de resultados aos clientes

Máximo: 7 alunos por grupo

Dinâmica:

Para cada problema, cada consultoria deverá apresentar propostas de análises que respondam às perguntas apresentadas pelos personagens fictícios.

Cada proposta satisfatória dará bônus à consultoria.

Dinâmica:

Ao final, as consultorias serão comparadas quanto à performance global em cada caso apresentado, baseada em:

- Tempo de execução (minutos) – até 5 pontos
- Qualidade da proposta (originalidade)
– até 5 pontos

Primeira cliente - Maria Helena

- Engenheira de produção da indústria Zwei
- Responsável por uma linha de produção
- **Produto** : parafuso para empilhadeiras
- **Interesse principal**: comparar comprimento do parafuso com padrão desejado à produção (precisa ter comprimento igual a 4,2cm)

Primeira cliente – Questões:

Com base em uma amostra retirada da linha de produção:

1. Proponha um valor que represente bem os comprimentos dos parafusos produzidos;
2. Avalie a variabilidade e a capacidade de reprodutibilidade na linha de produção;
3. Represente graficamente a amostra produzida;
4. Avalie se existe preferência por produzir acima ou abaixo do comprimento ideal;
5. Se padrão de referência do comprimento aceitar parafusos com comprimento entre 4,1cm e 4,3cm, qual a chance de ter parafusos dentro do padrão na linha de produção?
6. Construa um intervalo de referência para o comprimento, com base nos valores observados na amostra.

Primeira cliente – Execução:

Tempo de execução máximo = 40 minutos

Podem discutir e dividir a realização das tarefas entre os membros da consultoria.

Resultados devem ser entregues à cliente, que os entregará a Diretora de Finanças, que é estatístico, para averiguação.

Primeira cliente – Solução:

1. Proponha um valor que represente bem os comprimentos dos parafusos produzidos. Com $n=20$:

Mínimo = 4,20

Média = 6,87

Moda = 4,20

Máximo = 54!!! (erro de digitação?)

Mediana = 4,40

OBS: A média se diferencia bem da mediana! Efeito dos valores discrepantes (cuidado com os valores extremos). Sem ele:

Mínimo = 4,20

Média = 4,38

Moda = 4,20

Máximo = 4,6

Mediana = 4,40

Primeira cliente – Solução:

2. Avalie a variabilidade e a capacidade de reprodutibilidade na linha de produção;

Desvio Padrão = 11,10
freq(4,2) = 5

Coef. de Variação = 161,63%
freq(outros) = 15 !

Sem o valor discrepante:

Desvio Padrão = 0,17

Coef. de Variação = 3,82%

- Variabilidade adequada (pequena) se não tivesse presença de valor discrepante
- Reprodutibilidade do experimento se verificaria ($CV \leq 10\%$) só na ausência do discrepante ➡ Rever linha de produção!

Primeira cliente – Solução:

2. Avalie a variabilidade e a capacidade de reprodutibilidade na linha de produção (cont.);

- Calcular **distâncias absolutas** para valor ideal $|x_i - 4,2|$
- Calcular **distância média** e o quanto (%) ela representa em termos do próprio valor de referência (4,2cm)

➤ É grande?

$D_m = 0,195$  Aproximadamente 4,6% de 4,2

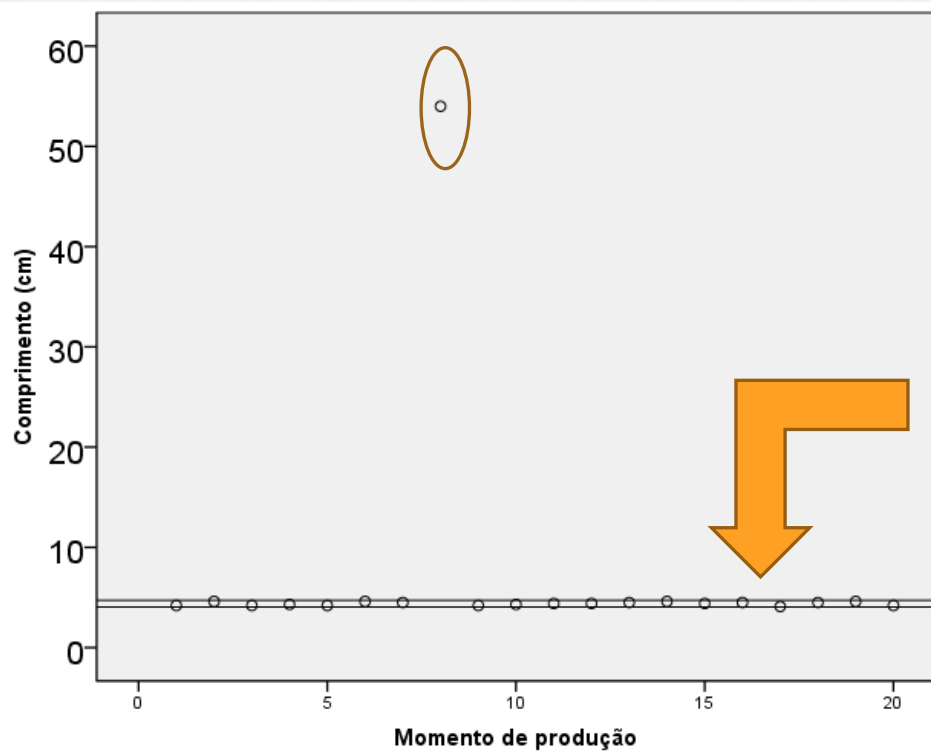
➤ **Referência é dada por quem constrói o equipamento** onde será usado o parafuso!

➤ Se **tolerância** for de 5% (ou mais), ok!

➤ Se for menor que 4%, problemas na linha de produção!

Primeira cliente – Solução:

2. Avalie a variabilidade e a capacidade de reprodutibilidade na linha de produção (cont.);

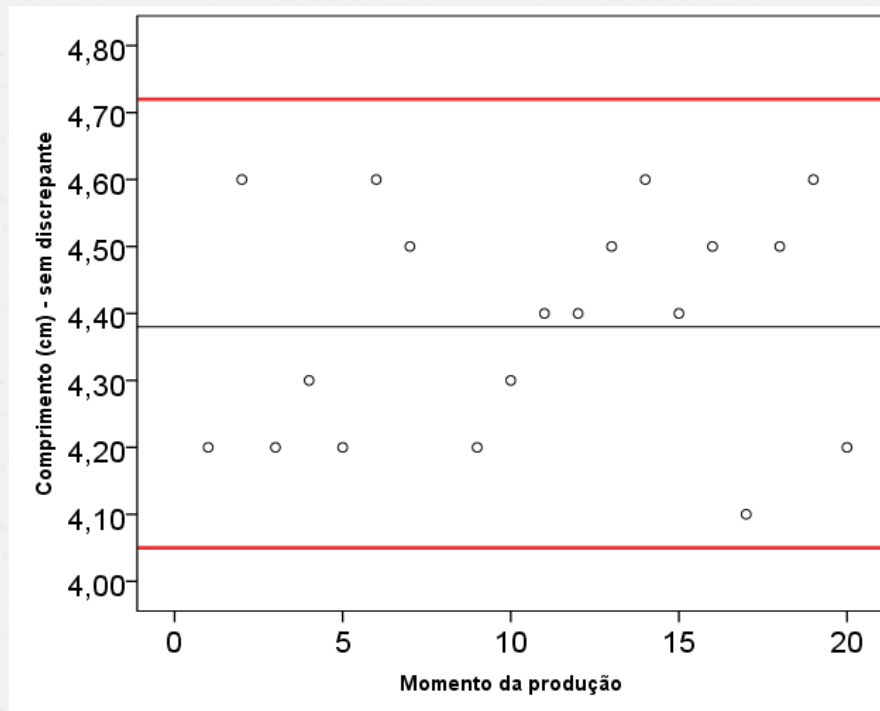


Intervalo (s/ discrepante):

(MÉDIA – 2DP; MÉDIA + 2DP)

Primeira cliente – Solução:

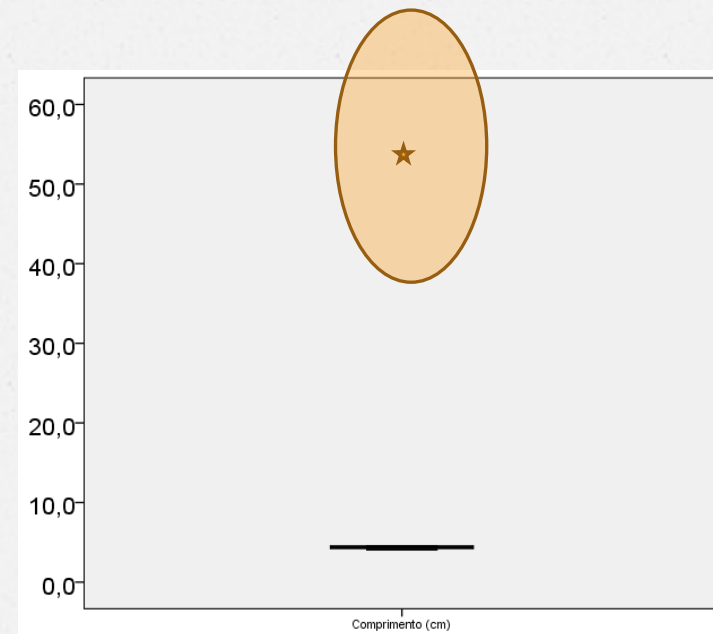
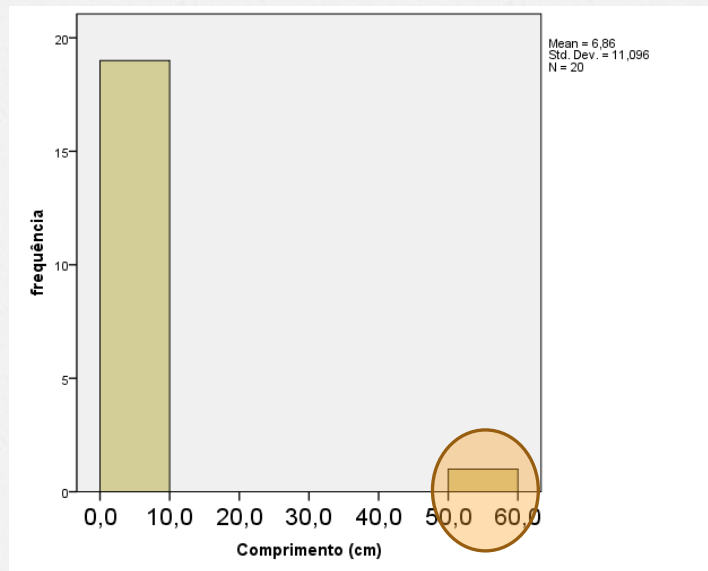
2. Avalie a variabilidade e a capacidade de reprodutibilidade na linha de produção (cont.);



Intervalo (s/ discrepante):
(MÉDIA - 2DP; MÉDIA + 2DP)

Primeira cliente – Solução:

3. Represente graficamente a amostra produzida:



Primeira cliente – Solução:

3. Represente graficamente a amostra produzida (com discrepante)

Gráfico Ramo-e-folha - Comprimento (cm)

Freq. Ramo&Folha

1,00	41 . 0
5,00	42 . 00000
2,00	43 . 00
3,00	44 . 000
4,00	45 . 0000
4,00	46 . 0000

1,00 Extremos ($\geq 54,00$)

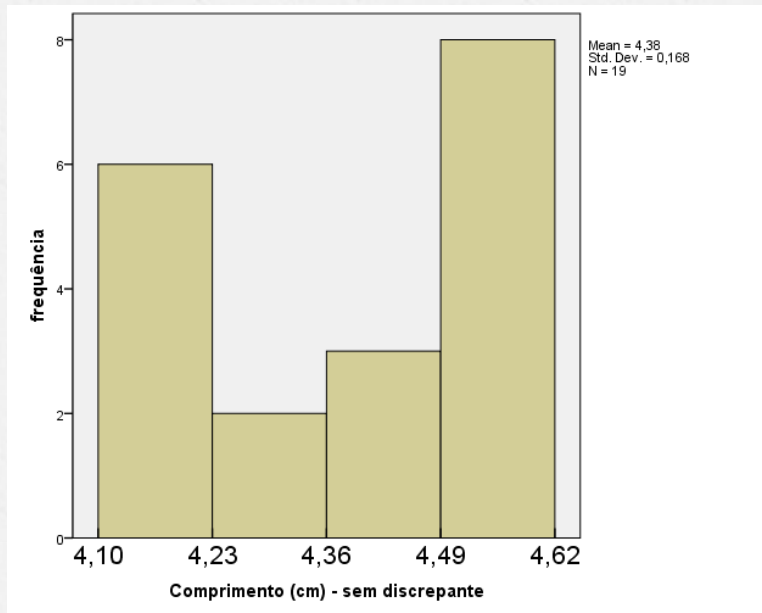
JÁ SEPARA OS
DISCREPANTES!

Tamanho do Ramo: 0,1

Cada folha: 1 observação

Primeira cliente – Solução:

3. Representação gráfica a amostra produzida (sem discrepante)



Primeira cliente – Solução:

4. Avalie se existe preferência por produzir acima ou abaixo do comprimento ideal

- Assimetria?** $la = + 0,25$ (n=20) ou $la = +1,1$ (n=19)
Média = 4,38 Moda = 4,2
Média > Moda (Ass. Direita!)

Ramo-e-folha - Comprimento (cm)

Freq.	Ramo&Folha
-------	------------

1,00	41 . 0
------	--------

5,00	42 . 00000
------	------------

2,00	43 . 00
------	---------

3,00	44 . 000
------	----------

4,00	45 . 0000
------	-----------

4,00	46 . 0000
------	-----------

1,00 Extremos (>=54,00)

# Valores < 4,2	: 1
-----------------	-----

# Valores = 4,2 (ideal)	: 5
-------------------------	-----

# Valores > 4,2	: 14
-----------------	------

**Preferência por valores
acima do ideal.**

Primeira cliente – Solução:

5. Se padrão de referência do comprimento aceitar parafusos com comprimento entre 4,1cm e 4,3cm, qual a chance de ter parafusos dentro do padrão na linha de produção?

Chance \approx Proporção

Ramo-e-folha - Comprimento (cm)

Freq. Ramo&Folha

1,00	41 . 0
5,00	42 . 00000
2,00	43 . 00
3,00	44 . 000
4,00	45 . 0000
4,00	46 . 0000

1,00 Extremos ($\geq 54,00$)

Valores < 4,1 : 0

Valores no intervalo : 8

Valores > 4,3 : 12

Chance $\approx 8/20 = 0,4$


Primeira cliente – Questões:

6. Construa um intervalo de referência para o comprimento, com base nos valores observados na amostra.

Intervalos: [mínimo ; máximo]

Tirar 5% dos dados  [4,1 ; 4,6] (c/ discrepante)
[4,2 ; 4,6]

[média - k*dp; média + k*dp]

k=2:  [4,07; 4,72] (s/ discrepante)

cercas internas do box-plot:
[Q1 – 1,5 DEQ; Q3 + 1,5 DEQ]

$Q1 = 4,2; Q3 = 4,53 \rightarrow [3,71; 5,01]$

Segunda cliente - Daniela

- Médica endocrinologista
- Responsável por linha de pesquisa na UFRNB
- **Objetivo:** Verificar se existe associação entre a idade de aparecimento de Diabetes do tipo 1 e nível de TSH em mulheres
- Pacientes adultas c/ Diabetes do tipo 1 (n = 20)
- Faixas etárias: Menor que 30; 30 a 40; 40 ou mais
- Valores ref. (adultas) TSH: 0,4 a 4,3 mcUI/mL

Segunda cliente - Daniela

1. Descrever pacientes em relação à idade e o valor mínimo que engloba (acima dele) 80% dos casos;
2. Descrever pacientes em relação ao valor de TSH e valor máximo que engloba (abaixo dele) 80% dos casos e o percentual de casos dentro do intervalo de referência;
3. Verificar se o TSH varia com idade e se existe padrão de relação entre as variáveis, para usar na modelagem;
4. Verificar como se dá a distribuição das pacientes considerando a faixa de referência para o valor de TSH e faixas etárias, e representá-la graficamente.

Segunda cliente – Execução:

Tempo de execução máximo = 40 minutos

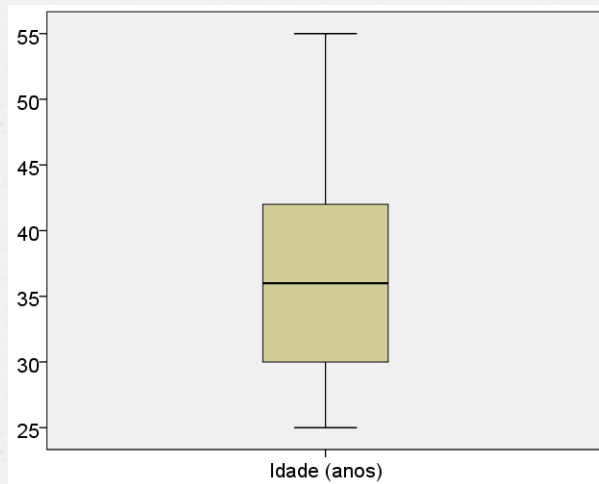
Podem discutir e dividir tarefas entre os membros da consultoria.

Resultados devem ser entregues à cliente, que os entregará ao estatístico que trabalha no Comitê de Ética da Faculdade de Medicina da UFRNB, para averiguação.

Segunda cliente – Solução:

1. Descrever pacientes em relação à idade e o valor mínimo que engloba (acima dele) 80% dos casos;

- **Idade de aparecimento variou entre 25 e 55 anos**
- **Idade média = 36,9 anos**
- **Em 50% delas, diabetes apareceu a partir dos 36 anos (Me)**
- **Intervalo de referência (média \pm dp): (29,1; 44,6)**

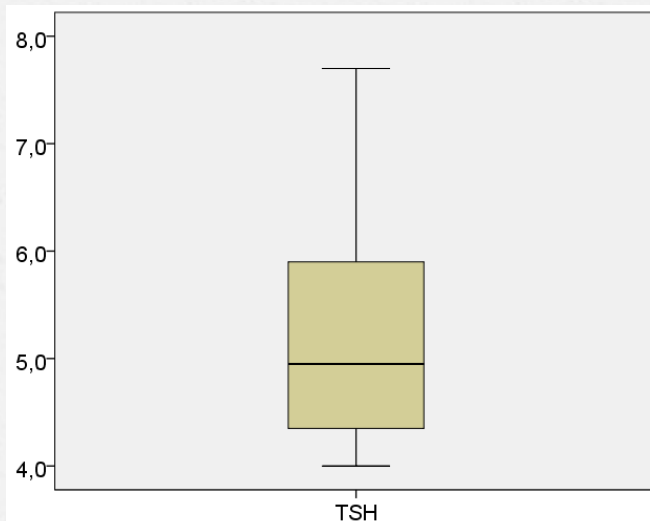


- 80% das pacientes tiveram a diabetes aparecendo a partir de 30 anos de idade (P20)

- Histograma
- Ramo-folha

Segunda cliente - Daniela

2. Descrever pacientes em relação ao valor de TSH e valor máximo que engloba (abaixo dele) 80% dos casos e o percentual de casos dentro do intervalo de referência;
- **TSH variou entre 4,0 e 7,7 mcUI/mL**
 - **TSH médio = 5,1 mcUI/mL**
 - **Em 50% das pacientes, TSH era menor que 4,95 mcUI/mL (Me)**
 - **Intervalo de referência (média \pm dp): (4,16; 6,09) mcUI/mL**

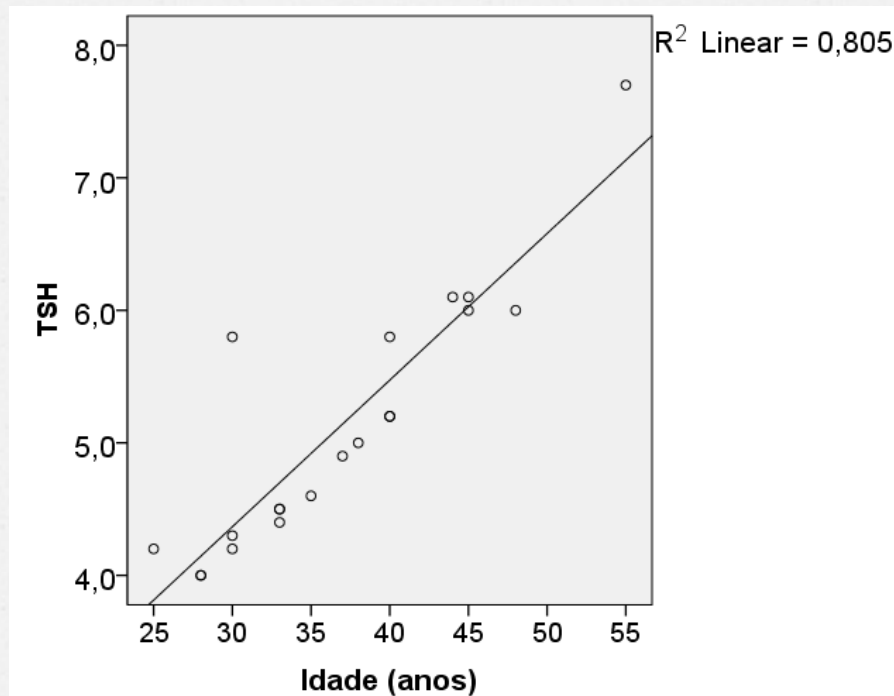


- 80% das pacientes tiveram a TSH menor que 6,0mcUI/mL (P80)

- Na faixa [0,43; 4,3] temos: 25% dos valores observados, apenas. Resto > 4,3 mcUI/mL!

Segunda cliente - Daniela

3. Verificar se o TSH varia com idade e se existe padrão de relação entre as variáveis, para usar na modelagem;



- Padrão parece linear
- $r \approx + 0,90$
(fortemente linear)

Segunda cliente - Daniela

4. Verificar como se dá a distribuição das pacientes considerando a faixa de referência para o valor de TSH e faixas etárias, e representá-la graficamente.

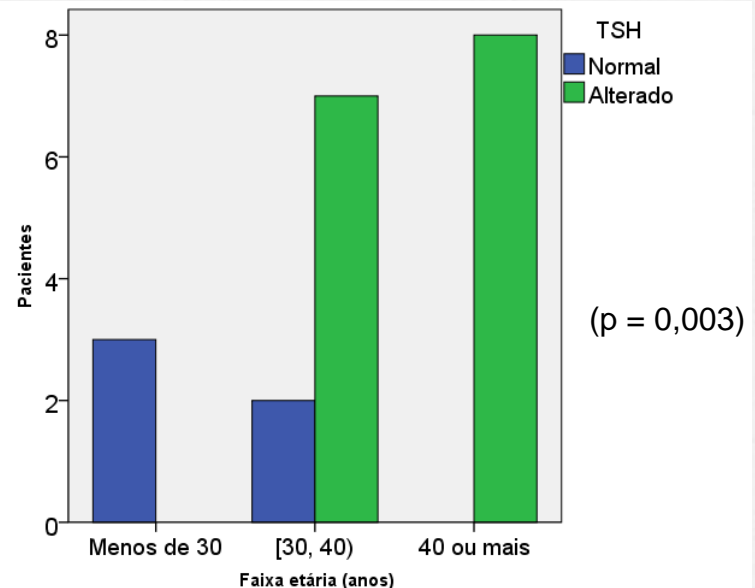
- **Faixas etárias:**

Menor que 30; 30 a 40; 40 ou mais

- **Valores ref. (adultas) TSH:**

0,43 a 4,3 mcUI/mL

Faixa etária	TSH	
	Normal	Alterado
Menos de 30 anos	3	0
[30, 40)	2	7
40 ou mais	0	8



Terceira cliente – Ana Rosa

- Química da indústria Morzk
- Chefe do laboratório de desenvolvimento de kits de reação química
- **Interesses:**
 - comparar se dois kits produzem as mesmas respostas (presença de reação) em 1 mesmo experimento
 - verificar se avaliação de cada kit depende da faixa de temperatura do local (refrigerada/ambiente)

Terceira cliente – Questões:

1. Avaliar mesmas amostras quanto à positividade da reação em cada kit;
2. Avaliar distribuição da positividade da reação em função da faixa de temperatura;
3. Comparar desempenho dos dois kits em cada faixa de temperatura;
4. Mensurar proporção de discordância entre kits em cada caso.

Terceira cliente – Execução:

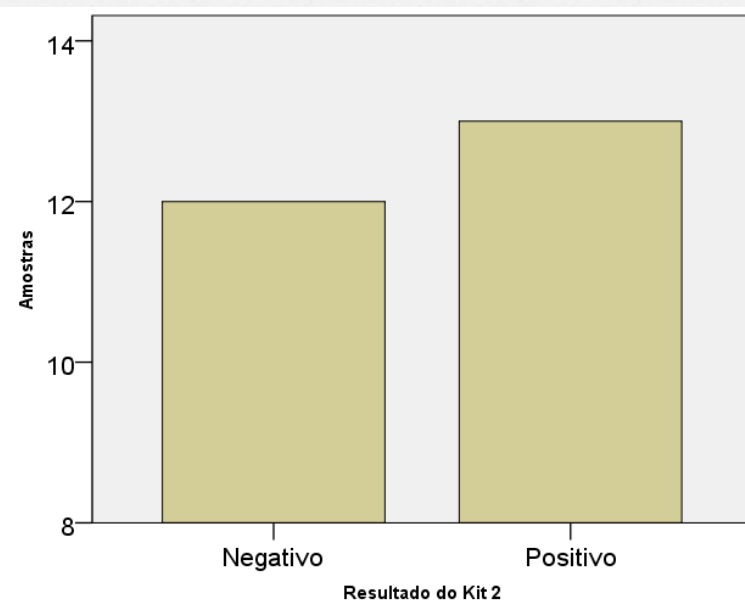
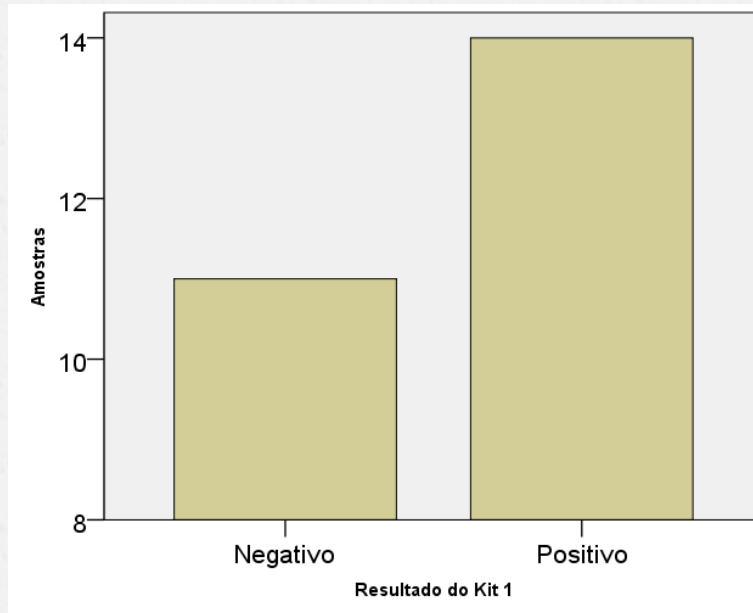
Tempo de execução máximo = 40 minutos

Podem discutir e dividir tarefas entre os membros da consultoria.

Resultados devem ser entregues à cliente, que os entregará ao estatístico que trabalha no setor de controle de qualidade da empresa.

Terceira cliente – Questões:

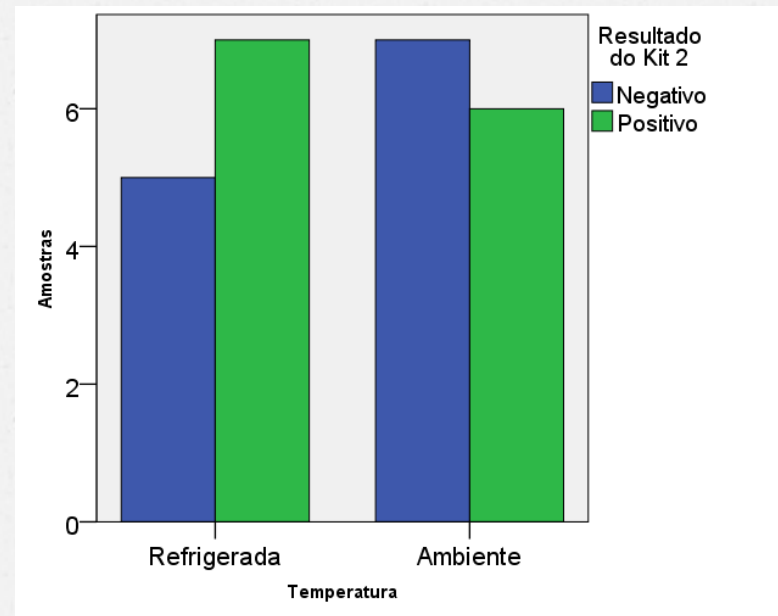
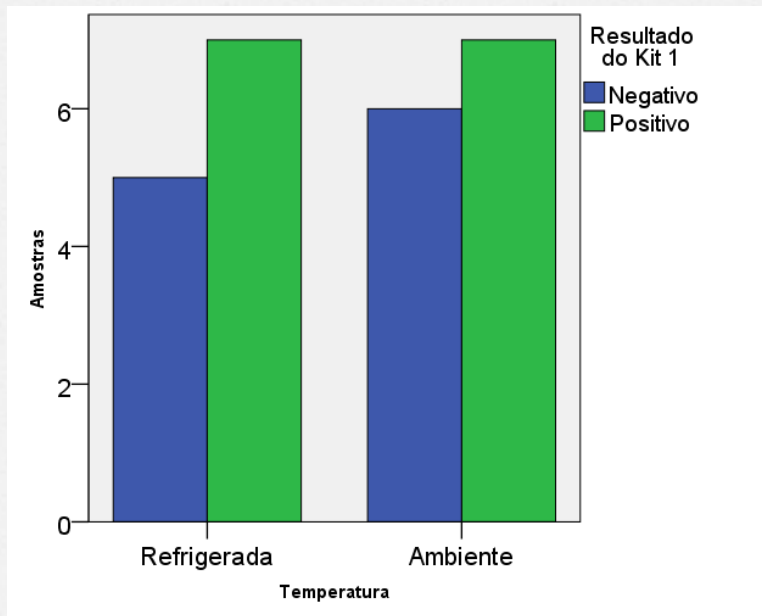
1. Avaliar mesmas amostras quanto à positividade da reação em cada kit;



Kit 1 parece marcar mais positivos que o Kit 2 !

Terceira cliente – Questões:

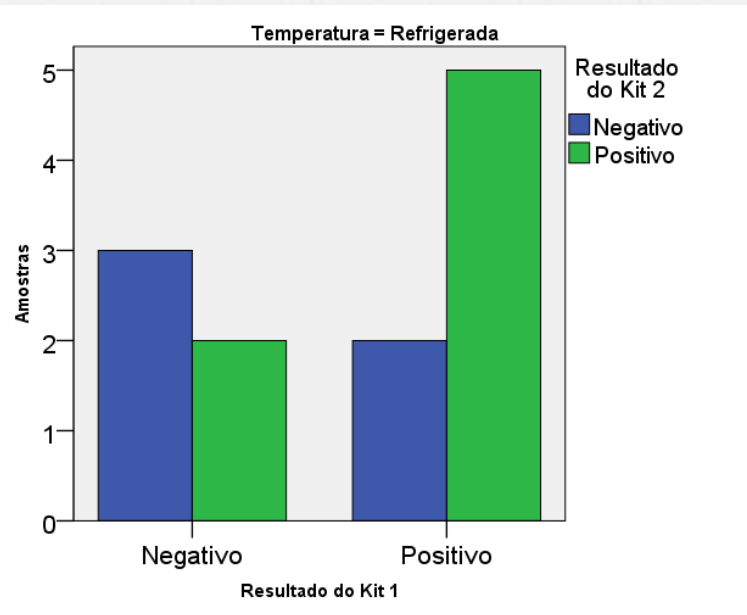
2. Avaliar distribuição da positividade da reação em função da faixa de temperatura;



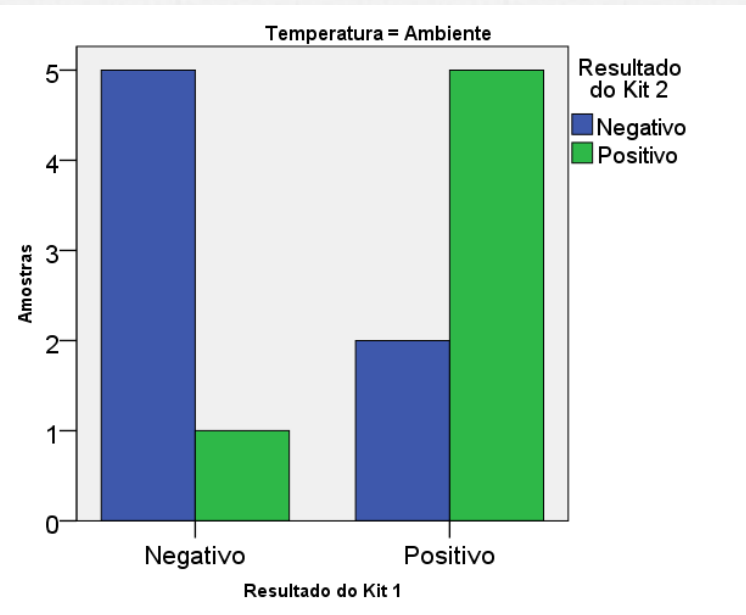
TEMPERATURA PARECE IMPACTAR MAIS O KIT 2 !

Terceira cliente – Questões:

3. Comparar desempenho dos dois kits em cada faixa de temperatura;



Discordância entre kits !



Concordância maior entre kits !

Terceira cliente – Questões:

4. Mensurar proporção de discordância entre kits em cada caso.

Temp. Refrigerada	Kit2		
Kit1		Neg	Pos
	Neg	3	2
	Pos	2	5

$$Pd = 4/12 \cong 0,33$$

Temp. Ambiente		Kit2	
		Neg	Pos
Kit1	Neg	5	1
	Pos	2	5

$$Pd = 3/13 \cong 0,23$$