6.

Amostragem - Introdução - Tipos - Estimativas dos parâmetros - Grandes e pequenas amostras.

Ponto 18 – Amostragem.

Ponto 19 – Testes de significância.

AMOSTRAGEM.

1. Introdução. Definições.

Amostra tem no conhecimento popular o sentido de pequena porção de um material qualquer a partir da qual se procura determinar as qualidades inerentes àquele material.

A necessidade do emprêgo de amostra, surge da impossibilidade de, em certos casos examinar à totalidade de elementos de uma população afim de pesquizar alguma qualidade na mesma.

Exº: altura de homens brasileiros, qualidade de uma barra de aço, de um cabo, maior ou menor confiança que se pode ter em um método de análise ou num analista, pesquisas de mercado.

<u>Amostra</u> – é um selecionado número de indivíduos de uma população, cada um dos quais é um membro dela.

Objetivo da teoria: fixar os métodos de extrair uma boa amostra.

Teorema de Tchebycheff – $P(|x-m| \le \alpha) > 1-\frac{1}{\alpha^2}$

Lei dos grandes números – $P(|\bar{x}-a| \le \eta) > 1-\frac{1}{\varepsilon}\cdot\frac{\sigma^2}{n}$

Vantagens de teoria.
- i. Custo reduzido _ parte do agregado original.
- ii. Maior rapidês - pode ser a mais importante consideração mesma amostragem.
- iii. Maior amplitude e flexibilidade - pessoal especializado.
- iv. Maior precisão - pessoal especializado.

x

2. Classificação das populações e das amostras.

Populações.
- Q.to á extensão: infinita; muito grande; finita
- Q.to á continuidade: continua; discontinua ou discreta
- Q.to á existencia: abstrata (hipotetica); concreta.
- Q.to á homogeneidade: homogêneas; heterogêneas.

Explicar.

Populações de populações.

Sub-populações.

Classif. amostras.
- Qto. à extensão
  - grandes
  - pequenas
- Qto. ao método de seleção
  - randomicas
    - pp. ditas
    - sistemáticas
    - duplas
    - sequenciais
    - geográficas
  - intencionais
  - randomico-intencionais.

Explicar.
Esquemas - Simpson e Kafka.

---

3. Planejamento de amostragem.

O planejamento depende do material que se pretende amostrar, bem como das características que apresenta, o meio em que está imerso, etc.

Principais medidas:

1. - Fixação dos objetivos do planejamento.

2. - Definição da população a ser amostrada.
   1. Definição simples: caso da tecnologia.
   2. Definição complexa: população de fazendas.

Cabe à pessoa encarregada saber decidir se um determinado elemento dúbio, cabe ou não na população considerada.

Durante a amostragem podem surgir elementos que indiquem a existencia de outra população paralela à considerada. É conveniente, neste caso, colher os elementos necessários à analise desta segunda.

3.- Determinação dos dados a coligir.

1. Não omitir coisas essenciais.
2. Não colher em excesso.

4.- Metodos de medida.
Refere-se em geral a amostragem de populações humanas, relativamente a dados medicos psicologicos ou educacionais e outros.

5. Escolha da unidade de amostragem.
1. Subdivisão da população em unidades.

2. as unidades devem formar a população total.

3. ser bem determinadas.

4. Casos simples: lampadas eletricas.
5. Casos complexos: individuo, familia, quarteirão

ou bairro, numa cidade.

6.- Seleção de amostra.

1. Metodo - v. classificação

2. Dimensão - função do custo de amostragem.

7.- Organização do campo de trabalho.

1. Administração complexa - casos de pesquisas intensivas.
2. Pessoal especializado.
3. Previsão das respostas negativas - em geral de resultados imprevistos.

8.- Sumário e análise dos dados.

Diversos cuidados e processos de cálculos.

9.- Informação obtida para futuras amostragens.

Os resultados de uma amostragem fornecem material para no futuro melhorarem-se os processos.

Registrar erros e omissões.

4. Seleção das amostras.

1. Metodo.

4.1.1. Amostra randômica.

Randômico = sem objetivo, fim ou propósito.

Amostra randômica = seleção de um indivíduo de uma população é randômica quando cada membro tem a mesma probabilidade de ser escolhido.

Como obter uma amostra ao acaso ou randômica?

1. Lista telefônica.
2. Lista classificada.
3. Lançar ao ar (sobre uma plantação) um punhado de cortes.

Todos são processos viciados.
Alem disso, o <u>homem</u> tem grande tendência a viciar as amostras. É mau elemento para conduzir amostragem randômica.

Exemplos: 1. experiencia em Rothamsted.
2. experiência com leituras
3. exp. com bolas coloridas de um saco.

Analizemos cada tipo de amostra randômica.

a. Randômica pp.te dita:

Atribuir aos elementos da população números.
Marcar cartões com tais números
Tirar uma porção de cartões para constituir a amostra.

Pode-se fazer o mesmo com bolas.

Em certos casos estes processos se podem tornar extremamente complicados. →
<u>Números ao acaso</u>.

b. Sistemáticas -
Cada dez casas duma rua — problemas.
Cada tantas peças produzidas numa máquina.

c. Duplas - aplicação em contrôle de qualidade industrial.
- Fixam-se os limites de qualidade, superior e inferior.
- Retira-se a 1ª amostra. Se o resultado estiver intra ou extra limites:
- Retira-se ou não nova amostra.

d. - Sequenciais - (A. Wald) metodo novo, ainda em evolução.
- 1ª amostra randomica ppᵗᵃ dita ou sistemática, então aceitar, rejeitar ou retirar novas amostras.
- repete-se o processo.

e. - Geográficas - Amostras randômicas, intencionais, duplas ou sequenciais em diferentes localidades ou territorios, em diferentes maquinas ou fôrnos.

4.1.2 - Amostras intencionais.

Obtenção de amostra randômica típica de uma população heterogênea.

São em geral as amostras estratificadas.

Há dois processos de amostragem:

a. população discreta e finita ou infinita. - caso de beterraba, caso do cimento, caso de metais.

b. população continua e infinita - caso de materiais pulverulentos, caso idem misturados

4.1.3 - Problemas especiais que podem surgir na amostragem.

a. População infinita - saco de farinha, números entre 0 e 1, amostra de uma população normal.

b. População finita - problema de reposição.

c. População hipotética - caso dos dados.

A/ ——— x ———

5. - Amostragem de atributos (grandes amostras)

População em que há - A - presença e não A - ausência.

A = Sucesso ou acontecimento

α = falência ou não acontecimento

1. Simples amostra.

1. - cada elemento tem a mesma probabilidade p de ocorrência.
2. - probabilidade de cada acontecimento é independente das probabilidades dos anteriores.

É o caso das moedas: caso particular de amostra randômica.

Media e desvio padrão.

Sejam $\underline{n}$ tentativas

N amostras

p probabilidade de sucesso

q probabil. de fracasso

$$N(q+p)^n = N\left\{q^n + \binom{n}{1} q^{n-1} p + \dots + p^n\right\}$$

Media: $M = np$

Desvio padrão: $\sigma = \sqrt{npq}$

Tomando-se as proporções de ocorrências.

$$M = p \qquad \sigma = \sqrt{\frac{pq}{n}}$$

Sendo M o valôr esperado, da amostragem, podemos comparar o valôr obtido com aquele e decidir:

1. - Se ocorreu um evento improvável.
2. - A hipotese não é verdadeira (de que $\underline{p}$ seja a proporção na população)
3. - O processo de amostragem não é randomico.

A medida que $\underline{n}$ cresce a distribuição tende para a normalidade e $\pm 3\sigma$

abrange a grande maioria de elementos. (Esta propriedade só é verdadeira quando n é grande.)

Quando p e q não são conhecidos da população, utilizam-se os valôres obtidos por amostragem.

Se p não é muito diferente de q, pode-se considerar como grande um valôr de $n = 100$; caso contrário, $n = 500$ ou um pouco menos.

2.— Limitações da hipotese de simples amostragem.

1.- Suposição de que p = constante, e que seja a mesma para todas as amostras.

2.- Suposição de que p = constante para a extração de cada elemento da amostra.

Casos em que 1 e 2 são satisfeitas:

a.- reposição dos elementos
b.- população infinita
c.- população é grande.

3.- Elementos independentes uns dos outros, como no caso dos dados.

As três condições acima limitam a aplicabilidade da simples amostragem.

3. — Aplicações da simples amostragem.

1. dado p (população) verificar se o valôr obtido difere do esperado.
   Caso das hipoteses mendelianas.

2. Duas amostras, de materiais diferentes, com proporções de A, p_1 e p_2 e número de observações n_1 e n_2.
   a) São as duas amostras oriundas de uma mesma população?

   Calculam-se $p_0 = \dfrac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$

   e, sendo $\sigma_1^2 = p_0 q_0 / n_1$

   $\sigma_2^2 = p_0 q_0 / n_2$

   $$\sigma_{12} = p_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

   Aferem-se na hipotese $\pm 3\sigma_{12}$

   b) Havendo diferença (i. é $p_1 \neq p_2$) poderia ela passar desapercebida em outra amostra?
   Então:
   $\sigma_1^2 = p_1 q_1 / n_1 \quad \sigma_2^2 = p_2 q_2 / n_2$

e $\sigma_{12} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$

Se $p_1 - p_2 \leq 3\sigma_{12}$, uma nova amostra pode levar a supôr que se trate do mesmo material.

B. - Amostragem de variáveis.

Um elemento toma qualquer valôr da variável.

Os limites da variável serão ilimitados,
limitados por considerações práticas,
limitados por considerações teóricas.

Objetos de estudo:

a) comparação das observações com os valôres esperados, afim de verificar se os desvios são devidos a amostragem ou a hipotese errada;

b) estimar, a partir da amostra, características da população originária, como a media e a variança;

c) medir a confiança que se pode depositar nas estimativas.

Distribuição das estimativas de parâmetros de uma população.

Seja $y = f(x)$ essa distribuição, então,

$$P(u_1 \leq u \leq u_2) = \frac{\int_{u_1}^{u_2} f(u)\,du}{\int_{-\infty}^{\infty} f(u)\,du}$$

$$P(u \geq u_0) = \frac{\int_{u_0}^{\infty} f(u)\,du}{\int_{-\infty}^{\infty} f(u)\,du}$$

Nem sempre se tem a forma da distribuição de estimativa de um parâmetro porque:

a) A forma da população frequentemente é desconhecida;

b) Se a forma é conhecida, podem não o ser alguns dos seus parâmetros;

c) Se a população é conhecida (na sua forma) a forma da distribuição das estimativas dos parâmetros pode ser obtida teoricamente em certas circunstancias, principalmente se a amostragem é simples. Os problemas matemáticos são muito complicados.

Hipotese para aplicação: a amostragem é simples.

Há necessidade, neste caso, de saber qual a melhor estimativa de um parametro. O processo que se adota é o de "máxima verossimilhança".

Consiste no seguinte:

Sejam $p(x; \theta)$ a distribuição de forma matemática conhecida, contendo um parâmetro desconhecido θ.

$x_1, x_2 \ldots x_n$ — uma amostra da população cuja distribuição é p.

$t = t(x_1, x_2, \ldots x_n; \theta)$ uma função dos valôres da amostra, que forneça uma estimativa de θ.

Define-se função de verossimilhança:

$$L(x_1, x_2, \ldots x_n; \theta) = p(x_1; \theta) \cdot p(x_2; \theta) \ldots$$
$$\ldots p(x_n; \theta)$$

escolhe-se como estimativa de θ, o valôr que torne L máxima, para os valôres dados x_i.

Em geral usa-se $\log L =$

$$= \sum_1^n \log p(x_i; \theta).$$

A condição de máximo se exprime

$$\frac{\partial \log L}{\partial \theta} = \sum \frac{\partial \log p(x_i; \theta)}{\partial \theta} = 0$$

A solução t_n fornece uma estimativa eficiente de θ (assintoticamente).

A variança de t_n é:

$$V(t_n) = \frac{1}{n\,M\left[\left(\frac{\partial \log p(x;\theta)}{\partial\theta}\right)^2\right]}.$$

sendo

$$M\left\{\left[\frac{\partial \log p(x;\theta)}{\partial\theta}\right]^2\right\} =$$

$$= \int_{-\infty}^{\infty}\left(\frac{\partial \log p(x;\theta)}{\partial\theta}\right)^2 p(x;\theta)\,dx$$

$$= \int_{-\infty}^{\infty}\frac{1}{p(x;\theta)}\left(\frac{\partial p(x;\theta)}{\partial\theta}\right)^2 dx$$

$$e \quad \sum_{i=-\infty}^{\infty}\frac{1}{p(x;\theta)}\left(\frac{\partial p(x;\theta)}{\partial\theta}\right)^2.$$

Exemplos: distribuição normal.

$$p(x; m) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

$$L = (\sigma\sqrt{2\pi})^{-n} e^{-\sum \frac{(x_i-m)^2}{2\sigma^2}}$$

$$\log L = -n \log(\sigma\sqrt{2\pi}) - \sum \frac{(x_i-m)^2}{2\sigma^2}$$

$$\frac{\partial \log L}{\partial m} = \frac{\partial}{\partial m}\left[\frac{1}{2\sigma^2}\sum (x_i-m)^2\right] =$$

$$= -\frac{2}{2\sigma^2}\sum(x_i-m)$$

fazendo $\frac{\partial \log L}{\partial m} = 0 \rightarrow$

$$\sum(x_i - m) = 0 \rightarrow \boxed{m = \frac{\sum x_i}{n}}$$

———— " ————

Distribuição binomial:

$$\begin{cases} p(x=0) = 1-\theta \\ p(x=1) = \theta \end{cases}$$

sendo: em amostra de n

$$\begin{cases} x = 1 \text{ — } a \text{ vezes} \\ x = 0 \text{ — } n-a \text{ vezes} \end{cases}$$

$$L = \theta^a (1-\theta)^{n-a}$$

$$\log L = a \log \theta + (n-a) \log(1-\theta)$$

$$\frac{\partial \log L}{\partial \theta} = \frac{a}{\theta} - \frac{n-a}{1-\theta} = 0 \quad \text{ou}$$

$$\frac{a}{\theta} = \frac{n-a}{1-\theta} \rightarrow \boxed{\theta = \frac{a}{n}}$$

O método de máxima verossimilhança é utilizado em pequenas amostras.

Teste u →

$$\frac{\bar{x} - m}{\sigma/\sqrt{n}} = u$$

se $u > 1{,}96$ rejeita-se a hipotese de que $\bar{x}$ seja uma estimativa da media m.

TESTES de SIGNIFICANCIA

I. χ^2 — Concordância entre teoria e observação

II. — Casos —

a. No caso de distribuição

b. No caso de variação — dis-
tribuição de frequência, entre
intervalos e cada classe.

c. A frequência é chamada frequên-
cia da classe

Convenções

n = nº de classes
f_i = frequência da classe i observada
f_i' = frequência da classe i, hipotética
[illegible] frequência
esperada)
$f_i - f_i'$ = [illegible]
[illegible] esperadas

Os maiores χ^2 [illegible]
[illegible] teoria e fato

Notações

[illegible]

[illegible]

TESTES de SIGNIFICÂNCIA.

1. χ^2 — Concordância entre teoria e observação.

1.1. — Classes —

a. No caso de atributos.
b. No caso de variaveis — distribuição de frequência, cada intervalo é uma classe.
c. A frequência é chamada frequência de classe.

Convenções:

a.— n = nº de classes
$\bar{f}_i$ = frequência da classe i, observada
f_i = frequência da classe i, suposta uma certa hipotese H. (frequência esperada).
$\xi_i = \bar{f}_i - f_i$ = desvio em relação às frequências esperadas.

os números ξ fornecem uma idéia da discrepância entre teoria e fato.

Restrições ? (Constraints)

Suponhamos um quadro de contingencia

	A	α	
B	$\bar{f}_1$	$\bar{f}_2$	$\bar{f}_1 + \bar{f}_2$
β	$\bar{f}_3$	$\bar{f}_4$	$\bar{f}_3 + \bar{f}_4$
	$\bar{f}_1 + \bar{f}_3$	$\bar{f}_2 + \bar{f}_4$	N

Há três equações independentes:

De
$\xi_1 = \bar{f}_1 - f_1$ $\xi_2 = \bar{f}_2 - f_2$ $\xi_3 = \bar{f}_3 - f_3$

$\xi_4 = \bar{f}_4 - f_4$ obtemos:
$$\begin{cases} \xi_1 + \xi_2 = 0 \\ \xi_1 + \xi_3 = 0 \\ \xi_2 + \xi_4 = 0 \end{cases}$$

a quarta: $\xi_3 - \xi_4 = 0$ é obtida por combinação linear das precedentes.

Seja qual fôr a hipotese H, as três equações independentes impõem restrições aos valôres ξ_1, ξ_2, ξ_3 e ξ_4. Fixado um valôr de ξ, os outros resultarão imediatamente, de acôrdo com as condições dos dados.

Analogamente, numa distribuição de frequências, obrigatoriamente a soma das freq. deve igualar N, restringindo a liberdade de calcular, mediante uma hipotese H, as freq. teoricas corresp.

Sejam então k as restrições impostas pela hipotese e n o número de celulas; denomina-se "número de gráus de liberdade" à diferença

$$f = n - k.$$

Nos casos anteriores: $f = 4 - 3 = 1$

$f = N - 1$

Havendo uma tabua de contingência $p \times q$, nós teremos:

cada coluna e cada linha impõem uma restrição, não sendo todas elas independentes, i. é:

$$k = p + q - 1$$

logo: $f = pq - (p + q - 1) =$

$= (pq - p - q + 1) =$

$= (p-1)(q-1)$ gráus de liberdade.

Define-se $\chi^2 = \sum \left(\frac{\xi_i^2}{f_i} \right) = \sum_1^n \left[\frac{(\bar{f}_i - f_i)^2}{f_i} \right]$

Se $\chi^2 = 0$ — a hipotese está de acôrdo com os dados observados.

Se um ou todos os ξ_i forem grandes, χ^2 será grande.

A medida do χ^2 fornece uma ideia da correspondência entre hipotese e observação. Deve ser lembrado que χ^2 ignora o sinal da diferença.

Conhecida a distribuição de probabilidades de χ^2 nas amostras, para diferentes gráus de liberdade, podemos determinar a probabilidade de obter um valôr de χ^2 maior ou menor que o encontrado e assim aceitar ou rejeitar uma hipotese.

A distribuição $p(\chi^2) = f(\chi^2, f)$

é dada por:

$$p(\chi^2) = \frac{e^{-\frac{\chi^2}{2}} (\chi^2)^{\frac{f-2}{2}}}{2^{\frac{f}{2}} \, \Gamma\left(\frac{f}{2}\right)}.$$

e a probabilidade de obter um valôr maior que um certo χ_0^2 é:

$$P(\chi^2 \geqslant \chi_0^2) = \frac{\int_{\chi_0}^{\infty} p(\chi^2) \, d(\chi^2)}{\int_0^{\infty} p(\chi^2) \, d(\chi^2)}$$

sendo $\int_0^{\infty} p(\chi^2) \, d(\chi^2) = 1$

Os valôres de (χ^2, f) estão tabelados, dando P.

———x———

Aplicação quando as frequências teoricas são conhecidas.

Utilização das tabelas de χ^2.

Propriedades do χ^2.

a) quando $f = 1$, a curva de distribuição é normal com desvio padrão unitário, para valôres positivos da variável.

b) quando $f > 1$ a curva tem um único ponto de máxima para $\chi^2 = f - 1$,
é tangente ao eixo do x em $\chi^2 = 0$,
e assintótica ao mesmo eixo, para $\chi^2 \to \infty$.

c) quando f é muito grande (≥ 30) a distribuição é normal com $m = \sqrt{2f-1}$, $\sigma = 1$ para os valôres de $\sqrt{2\chi^2}$

— " —

Condições de aplicabilidade do χ^2:

a) - Número N total de observações nas diversas classes, grande. ($N > 50$)

b) - Nenhuma freq. classe deve ser menor que 5. $f_i \geq 5$.
Quando isso se der $f_i < 5$, agrupam-se algumas frequências.

c) - As restrições devem ser lineares.

Aplicações quando as freqüências esperadas não são conhecidas "a priori", mas estimadas dos dados.

Observações p/ este caso:

1.- Considerar que o número de gráus de liberdade deve ser reduzido de uma unidade por cada constante calculada (parâmetros)

2.- as constantes (parâmetros) devem ser "estimativas eficientes" dos parâmetros da população (ou estimativas ótimas - J. Kingston). Tais estimativas são obtidas pelo metodo de "máxima verossimilhança".

O teste do χ^2 é aplicado para medir o ajustamento entre uma teoria e os resultados concretos de uma experiência.

- Teste de Student - t -

1. Preliminares.

1.1 - Estimativas da média aritmética.

$$\bar{x} = \frac{1}{n}\sum x$$

é a melhor no sentido de produzir a "máxima verossimilhança".

1.2 - Estimativa da variança, sendo desconhecida a média.

$$s^2 = \frac{1}{n-1}\sum(x-\bar{x})^2$$ idem.

1.3 - Demonstrações:

Seja uma população infinita de uma variável x com média m e variança σ^2, de que se extrai uma amostra de n elementos, distribuídos independentemente:

$$M(x) = m \quad V(x) = \sigma^2$$

$$S = x_1 + x_2 + \dots + x_n = n\bar{x}$$

$$M(S) = M(x_1) + M(x_2) + \dots + M(x) = n.m$$

$$= nM(\bar{x}) \rightarrow M(\bar{x}) = m$$

Portanto $\bar{x}$ tem como média m.

$$V(S) = V(x_1) + V(x_2) + \dots + V(x_n) = n\sigma^2$$

$$V(\bar{x}) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Análogamente para a variança:

Seja a estimativa $s^2 = \frac{1}{n-1} \sum (x - \bar{x})^2$

Ora: $x - \bar{x} = (x - m) + (\bar{x} - m)$ ou

$$(x - \bar{x})^2 = (x - m)^2 + (\bar{x} - m)^2 - 2(x - m)(\bar{x} - m)$$

$$\sum^n (x - \bar{x})^2 = \sum (x - m)^2 + \sum (\bar{x} - m)^2 - 2\sum (x - m)(\bar{x} - m)$$

ou:

$$(n-1)s^2 = \sum (x - m)^2 + \sum (\bar{x} - m)^2$$

Porém: $M(n-1)s^2 = \cancel{(n-1)}s^2$?

$$M[(x - m)^2] = V(x) = \sigma^2$$

$$M[(\bar{x} - m)^2] = V(\bar{x}) = \sigma^2/n$$

$$M[(n-1)s^2] = n\sigma^2 + n\sigma^2/n = (n-1)\sigma^2$$

$\boxed{\text{logo: } M(s^2) = \sigma^2}$

e $\underline{s}^2$ é a estimativa de σ^2.

1.4 – Objetivo do teste de significância.

a. Não se poder dispôr senão de pequena amostra.

b. Pequena confiança nos parâmetros calculados dessa amostra.

c. Procura-se, a partir do que se tem, estabelecer com um certo grau de probabilidade:

1. – se o valôr obtido poderia provir de uma população cujo parâmetro tivesse um valôr determinado;

2. – estabelecer, os limites, dentro do qual se situa o verdadeiro parâmetro.

1.5 – Hipoteses:

1. população original normal;

2. população aproximadamente normal;

3. nunca U ou J

4. Niveis de significância: 5% ou 1%.

1.6 – Teste de Student p/ pta dts.

Seja: $\bar{x} = \frac{1}{n} \sum^{n} x$

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

Define-se: $t = \frac{\bar{x} - m}{s'/\sqrt{n}}$ onde

$\underline{m}$ = media suposta conhecida

$\underline{n}$ = nº elementos da amostra.

A distribuição é dada por

$$p(t) = \frac{1}{\sqrt{\pi f}} \frac{\Gamma\left(\frac{f+1}{2}\right)}{\Gamma\left(\frac{f}{2}\right)} \left(1 + \frac{t^2}{f}\right)^{-\frac{f+1}{2}}$$

cujas propriedades são:

1. simetrica para $t = 0$

2. tem uma moda para $t = 0$

3. assintotica a $x = \pm\infty$.

4. quando $f \to \infty$ a curva tende para a normal.

5. está tabelada para valôres positivos de t.

6. Para obter a probabilidade de um desvio, isto é, de um

valôr $t \geq t_0$, calculamos a área da curva à direita de t_0:

$$P(t \geq t_0) = \int_{+t_0}^{\infty} p(t)\,dt$$

Por outro lado, a probabilidade de um valôr menor que t_0 é dada por

$$P(t \leq t_0) = \int_{-\infty}^{t_0} p(t)\,dt$$

Por razões de simetria e em virtude de ser

$$\int_{-\infty}^{\infty} p(t)\,dt = 1$$

$$P(t \geq t_0) = \int_{t_0}^{\infty} p(t)\,dt = 1 - \int_{-\infty}^{t_0} p(t)\,dt$$

Análogamente, se quizermos determinar a probabilidade de se ter

$t_1 \leq t < t_2$ →

$$P(t_1 \leq t < t_2) = \int_{t_1}^{t_2} p(t)\,dt$$

1.7 - Aplicações.

a. Verificação da prob. de uma estimativa de media provir de uma pop. com $\underline{m}$.

b. Estabelecer os limites para uma determinada media, baseando-se nas estimativas de uma amostra.

c. Comparação de duas amostras.

Sejam duas amostras:

$$x_{11}, x_{12} \dots x_{1n_1}$$
$$x_{21}, x_{22} \dots x_{2n_2}$$

$$\bar{x}_1 = \sum^{n_1} x_{1i} / n_1 \qquad \bar{x}_2 = \sum^{n_2} x_{2i} / n_2$$

$$s_1^2 = \frac{1}{n_1 - 1} \sum (x_{1i} - \bar{x}_1)^2 \qquad s_2^2 = \frac{1}{n_2 - 1} \sum (x_{2i} - \bar{x}_2)^2$$

Variança da diferença:

$$s = V(d) = \frac{1}{n_1 + n_2 - 2} \left\{ \sum (x - \bar{x}_1)^2 + \sum (x - \bar{x}_2)^2 \right.$$

se não diferem significantemente as duas ~~medias $\bar{x}_1$ e $\bar{x}_2$~~ varian-cas s_1^2 e s_2^2.

Então: $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

Se $s_1^2 \neq s_2^2$ (significante)

Então $V(d) = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$

e $$t = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ onde

$\delta = m_1 - m_2$.

---

d. Significância do coeficiente de regressão.

$$y - \bar{y} = b(x - \bar{x})$$

Supondo β ser o valôr na população:

$$t = \frac{(b-\beta)\, s_1 \sqrt{n-2}}{\sqrt{(s_2^2 - b^2 s_1^2)}}$$

Distribuição de Fisher.

É a distribuição da razão de estimativas de varianças.

$v^2 = \dfrac{s_1^2}{s_2^2}$, cuja distribuição teórica é dada por:

$$p(v^2) = \frac{\Gamma\left(\frac{f_1+f_2}{2}\right)}{\Gamma\left(\frac{f_1}{2}\right)\Gamma\left(\frac{f_2}{2}\right)} f_1^{\frac{f_1}{2}} f_2^{\frac{f_2}{2}} \frac{(v^2)^{\frac{f_1}{2}-1}}{(f_2+f_1 v^2)^{\frac{f_1+f_2}{2}}}$$

A distribuição devida a Fisher era a de $z = \frac{1}{2} \ln \frac{s_1^2}{s_2^2}$

Aplicações análogas às anteriores.

Análise da variança.

Aplicabilidade:

~~a~~ amostra classificada em grupos, ou sub-amostras.

Determinar se há diferenças significantes entre os grupos.

— x —

1º caso — uma variável independente.

Sejam p famílias, cada uma com n_p elementos:

1ª família — $u_{11}, u_{21}, u_{31}, \dots u_{n_1 1}$

2ª família — $u_{12}\ u_{22}, u_{32}, \dots u_{n_2 2}$

3ª família — $u_{13}, u_{23}, u_{33} \dots u_{n_3 3}$

— — — — — — —

pª família — $u_{1p}, u_{2p}, u_{3p} \dots u_{n_p p}$

e mais $u_{..}$ a media total

$u_{.j}$ a media da família j.

i — variação dentro de cada família

j — variação entre as famílias.

i — varia de 1 a n_j

j — varia de 1 a p.

Então:

$$\sum_{ij}(u_{ij}-u_{..})^2 = \sum_{ij}(u_{ij}-u_{.j}+u_{.j}-u_{..})^2$$
$$= \sum (u_{ij}-u_{.j})^2 +$$
$$+ \sum (u_{.j}-u_{..})^2 +$$
$$+ 2\sum (u_{ij}-u_{.j})(u_{.j}-u_{..})$$

Extendendo a somatoria a $\underline{i}$ apenas:

$$2\sum (u_{ij}-u_{.j})(u_{.j}-u_{..}) =$$
$$= (u_{.j}-u_{..})\sum_i (u_{ij}-u_{.j}) = 0$$

logo:

$$\sum (u_{ij}-u_{..})^2 = \sum_{ij}(u_{ij}-u_{.j})^2 + \sum_{ij}(u_{.j}-u_{..})^2$$
$$= \sum_{ij}(u_{ij}-u_{.j})^2 + \sum_j n_j (u_{.j}-u_{..})^2$$

Observações:

1. o primeiro membro é a soma dos quadrados dos desvios de todos os elementos em relação à media geral;

2. a primeira parcela do segundo membro é a soma dos quadrados dos desvios <u>dentro de cada familia</u>.

3. a segunda parcela é a soma dos quadrados dos desvios das medias das familias relativamente à media total e exprime variação <u>entre</u> familias.

Se $n_1 = n_2 = \dots = n_p \rightarrow$

$$\sum (x_{ij} - x_{..})^2 = N . V(x) = np\, V(x)$$

$$\sum_j n_j (x_{.j} - x_{..})^2 = n \sum (x_{.j} - x_{..})^2$$

Portanto:

$$np\, V(x) = \sum (x_{ij} - x_{.j})^2 + n \sum (x_{.j} - x_{..})^2$$

ou:

$$V(x) = \frac{1}{np} \sum (x_{ij} - x_{.j})^2 + \frac{1}{p} \sum (x_{.j} - x_{..})^2$$

Resulta de uma forma geral.

1. $\dfrac{\sum (x_{ij} - x_{..})^2}{np-1}$ é uma estimativa de $V(x)$.

2. $\dfrac{1}{np-p} \sum (x_{ij} - x_{.j})^2$ idem.

3. $\dfrac{1}{p-1} \sum n_j (x_{.j} - x_{..})^2$ idem.

4. $np-1 = (np-p) + (p-1)$
os graus de liberdade se somam.

Resumo:

Variacões	Gráus de liberdade	Soma de desvios	Estimativa da variança da população (suposta homogeneidade).
Entre familias	$p-1$	$\sum_j n_j (u_{.j} - u_{..})^2$	$\frac{1}{p-1} \sum n_j (u_{.j} - u_{..})^2 \quad v_1$
Dentro das fam.	$np-p$	$\sum_{ij} (u_{ij} - u_{.j})^2$	$\frac{1}{np-p} \sum (u_{ij} - u_{.j})^2 \quad v_2$
Total	$np-1$	$\sum_{ij} (u_{ij} - u_{..})^2$	$\frac{1}{np-1} \sum (u_{ij} - u_{..})^2 \quad v_3$

Interessa a verificação da homogeneidade aplicando o teste de Fisher para v_1/v_2 (numerador > denominador)

Assim por diante.

ALGUMAS APLICAÇÕES.

Ponto 20.

Planejamento experimental. —

Cap. II de "The Design of Experiments."
R. A. Fisher.

Contrôle de qualidade industrial.

Característica de qualidade
- Propried. física ou química
  - Temperatura
  - Pressão
  - Vida (lamp.)
- Dimensão ou dimensões
- Outras

Qualidade é consequência da evolução da organização economica.

1ª fase: $\boxed{P.}\boxed{C.}$ P = produção
C = consumo

um mesmo produtor e consumidor

2ª fase $\boxed{P.}$ $\boxed{C.}$

um produz, outro consome

3ª fase M = comerciante

$\boxed{P}$ $\boxed{M}$ $\boxed{C}$

um produz, outro ~~co~~ movimenta levando ao consumidor.

4ª fase.

	P				
p_1	p_2	i	v	M	C

Especialização interna do produtor, que:

1.- projeta (p_1)

2.- produz (p_2)

3.- inspeciona (i)

4.- vende (v)

x

O Controle de qualidade é criado por necessidades sociais:

1. Como consequência da lei de Aristóteles (divisão dos ofícios e convergência de esforços) e do conhecimento científico, particularmente físicos, químicos e biológicos:

   a. Desenvolvimento mecânico: máquinas e motores.

   b. Desenvolvi.^to instrumentos de medidas de propr. físicas e químicas

   c. Produção massiça → ciência e indústria ao alcance de todos:

      Especialização = maior produtivid

2.- Possibilidade de intercambiar manufaturas.

3.- Necessidade de manter a precisão da produção

4.- Crescimento gigantesco de companhias individuais.

——→ × ←——

Objetivos da Estatística no Controle de qualidade industrial.

1. Observar e estimar a qualidade. Manter as especificações dentro de limites razoáveis.

2. Acompanhar as falhas de um produto acabado, até a matéria prima.

3. Prover meios de manter o controle. Dar aviso das alterações. Ajudar a identificar as alterações e a localizar suas causas.

4. Estudar os melhores processos.

5. Estudar os melhores processos sob o ponto de vista de aplicabilidade economica.

São de principal interesse para o Controle Qualid.ᵉ Industrial o conhecimento de:

metodos estatisticos: media aritm., variança e desvio padrão, conhecimento das distribuições, particularmente normal, binomial e de Poisson; calculo do intervalo de variação, etc.

→ + ←

Exemplo de problema a resolver:

1. 1100 resistencias para radio inspecionados apresentaram 37 defeituosas;

1600 de outro fabricante 35.

São elas diferentes de fabricante a fabricante?

2. Velas experimentais ~~que~~ operaram em um motor de avião com um desgaste medio do eletrodo de 0.0049 pol. durante 100 horas de operação. Um grupo de controle das velas anteriormente em operação mostrou desgaste de 0.0064 pol.

São aquelas melhores que as ~~pr~~ antigas?

Geralmente:

Verificar a significancia da diferença entre	Quando σ da populaç.	
$\bar{x}$ e m	conhecida	(1)
$\bar{x}$ e m	descon.h.	(2)
σ e s	conh.	(3)
p e π	conh.	(4)
$\bar{x}$ de 2 amostras	conh.	(5)
$\bar{x}$ " " "	desc. mas comuns	(6)
$\bar{x}$ de " " emparelhadas	desconh.	(7)
s de duas amostras	desconh.	(8)
p de 2 amostras	desconh.	(9)

(1) Usa-se o teste t

$$t = \frac{\bar{x} - m}{\sigma/\sqrt{n}}$$

(2) a. Calcula-se $s' = s\sqrt{\frac{n}{n-1}}$

b. aplica-se $t = \frac{\bar{x} - m}{s'/\sqrt{n}}$

(3) - Calcula-se

$$n\left(\frac{s'^2}{\sigma^2}\right) = \chi^2$$

(4) - Sendo $\underline{n}$ suficientemente gran de, de forma que $n\pi \geq 5$, calcu lam-se os parâmetros como bino mial e determinam-se os li- mites ~~como normal~~ pelo teste $\underline{t}$:-

$$t = \frac{p - \pi}{\sqrt{\pi(1-\pi)/n}}$$

sendo p a proporção de defe tuosos.

(5) - Aplica-se o $\underline{t}$:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{n_1 + n_2}{n_1 n_2}}}$$

(6) - Calcula-se

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$ e em

seguida $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s}$

(7) - Dados emparelhados.

Calculam-se as diferenças d_i e em seguida $\bar{d}$ e s_d

para finalmente obter

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

(8) Usa-se o teste F ou z (de Fisher:

$$\left(\frac{s_1}{s_2}\right)^2 = F$$

(9) Calcula-se a proporção π ponderada:

$$p' = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

testa-se mediante $\underline{t}$:

$$t = \frac{p_1^* - p_2}{\sqrt{p'(1-p')\frac{n_1+n_2}{n_1 n_2}}}$$

Cartões de controle.

Procura-se, atravez do controle:

1. - preencher as especificações;

2. - manter a uniformidade.

Afim de constatar se o processo está sob controle:

1. Se um processo desconhecido está sob controle. Estabilidade.

2. Se um processo que estava sob controle permanece sob controle.

Há um metodo contínuo de manter sob controle um processo. Foi desenvolvido por W. A. Shewhart, constitue o metodo dos cartões de controle.

Usam-se cartões

a. De medias de amostras.

b. De amplitudes de medidas de amostras. Cartões R

c. De porcentagem de defeituosos na amostra. p

d. De números de defeitos na amostra. (C)

As melhores são as cartões de media e amplitude.

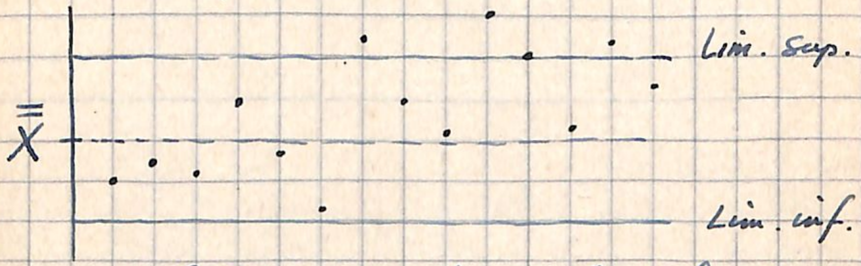
O de medias: indica quando uma alteração ocorreu tendente a afetar todas as peças.

O de amplitudes: indica quando uma perda ou melhoria de uniformidade ocorreu.

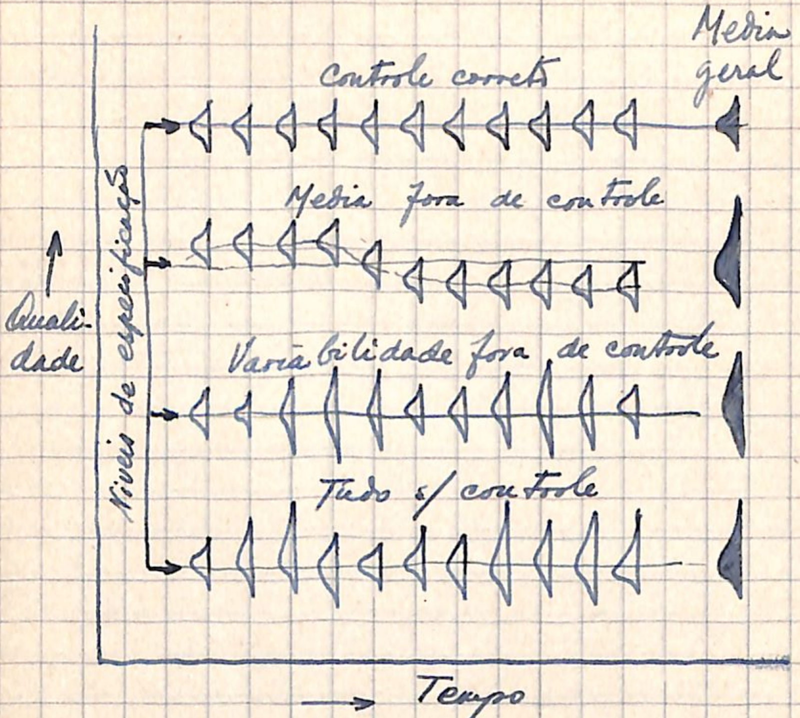
O de proporção de defeituosos está para a tendência da proporção de defeituosos como o de medias para a tendência de medias.

O de c (numero de defeitos) controla-os ao invez de sua proporção.

<u>Tipos de cartões de controle.</u>



Cartas de medias. ~~Cont~~ Qualidade fóra de controle. Limite superior excedido quatro vezes.



Esquêma de L. H. C. Tippet.