



Pós-Graduação em Computação Distribuída e Ubíqua

INF612 - Aspectos Avançados em Engenharia de Software
Engenharia de Software Experimental

[Head First Statistics] Capítulos 10, 11, 12 e 13

[Experimentation in Software Engineering] Capítulos X, Y e Z

[Guide to Advanced Empirical Software Engineering]

Sandro S. Andrade
sandroandrade@ifba.edu.br

Objetivos



- Realizar um nivelamento da turma acerca dos fundamentos de estatística e probabilidade necessários à correta aplicação das técnicas de Engenharia de Software Experimental
- Apresentar a motivação para estudo e aplicação prática da Engenharia de Software Experimental
- Discutir as principais estratégias empíricas utilizadas na Engenharia de Software Experimental
- Apresentar um modelo para projeto e execução sistemáticos de experimentos em Engenharia de Software



Pós-Graduação em Computação Distribuída e Ubíqua

INF612 - Aspectos Avançados em Engenharia de Software
Engenharia de Software Experimental - Revisão de Estatística

Sandro S. Andrade
sandroandrade@ifba.edu.br

Testando Chicletes



- *RottenTeeth BubbleGum* é uma empresa de ponta na fabricação de um chiclete especial que, segundo eles, possui um sabor mais duradouro
- Eles desejam fazer um comercial na TV para atrair mais clientes, indicando quanto tempo o sabor do chiclete dura
 - Como eles podem descobrir este tempo ?

Testando Chicletes



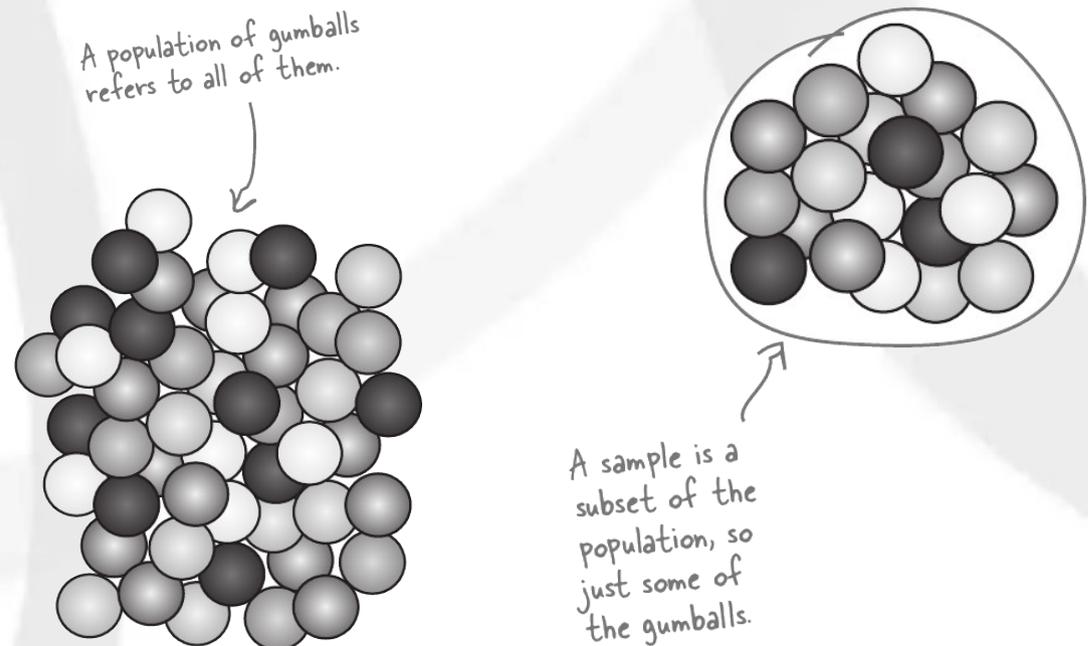
- Contratou-se um conjunto de testadores, porém dois problemas foram encontrados:
 - Os testadores estavam testando todos os chicletes da empresa
 - A empresa estava com altos gastos referentes aos planos odontológicos dos testadores



Testando Chicletes



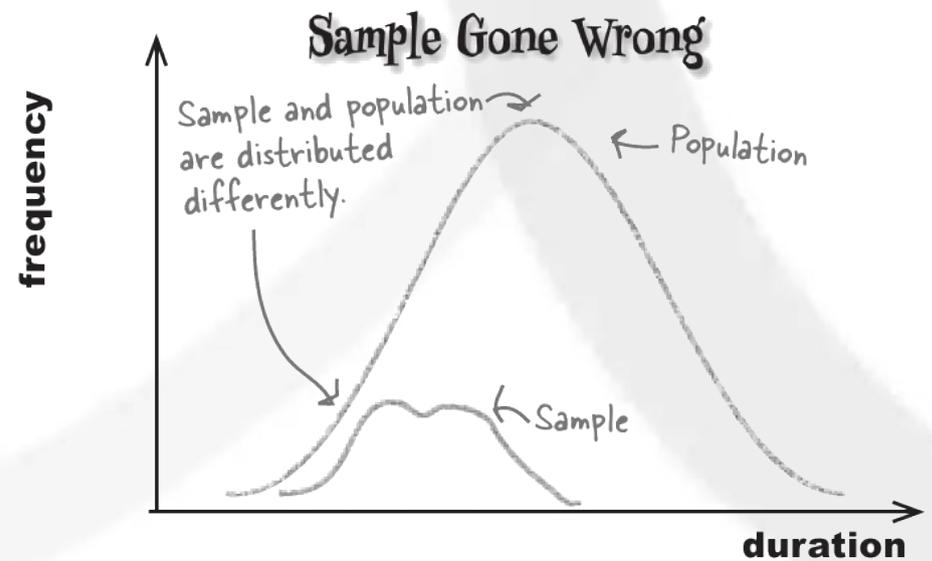
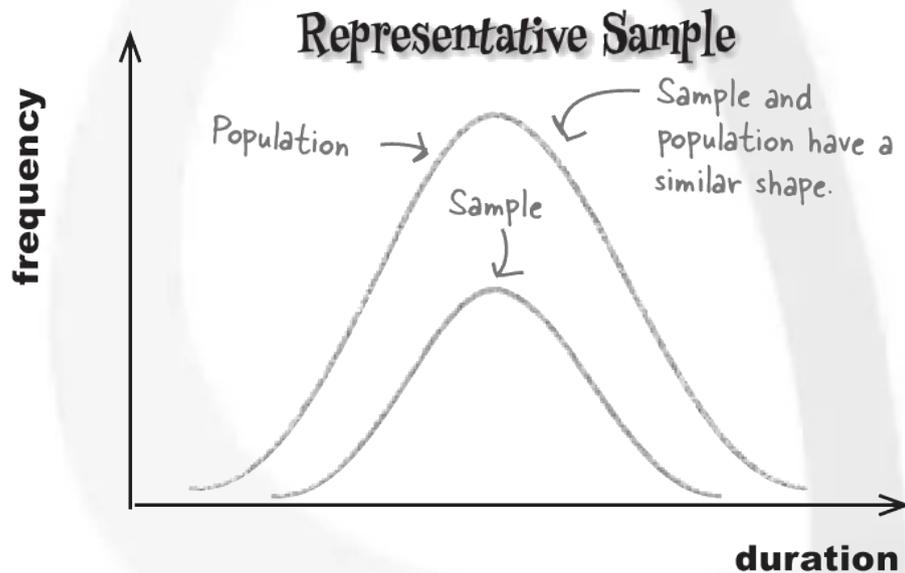
- Testar todos os chicletes custa tempo, dinheiro e dentes :)
- Pode-se testar uma amostra em vez de toda a população
- Pode-se utilizar a amostra para descobrir informações sobre a população



Testando Chicletes



- O importante é escolher amostras que são representativas da população



Testando Chicletes



Mighty Gumball's pugilistic CEO

I don't care how long flavor lasts in the sample. What I do care about is flavor duration in the population. That way, I can say how much longer our gumballs last than the competing brand.

Here's how long
flavor lasts for
in minutes

61.9 62.6 63.3 64.8 65.1
66.4 67.1 67.2 68.7 69.9



- Ok, a *RottenTeeth BubbleGum* selecionou uma amostra representativa, mas deseja-se informações da população e não da amostra

Testando Chicletes



- Como os dados abaixo podem ser utilizados para estimar a média e a variância da duração do sabor em toda a população ?
- Quão confiável você acha que a sua estimativa será ?

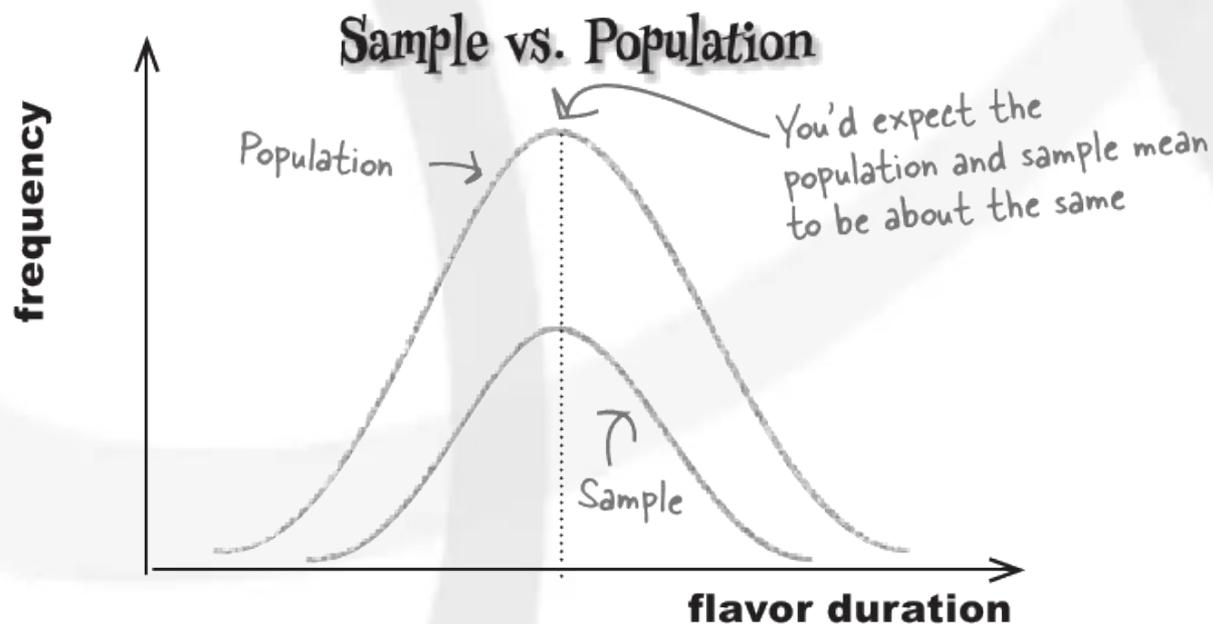
*Here's how long
flavor lasts for
in minutes*

61.9 62.6 63.3 64.8 65.1
66.4 67.1 67.2 68.7 69.9

Testando Chicletes



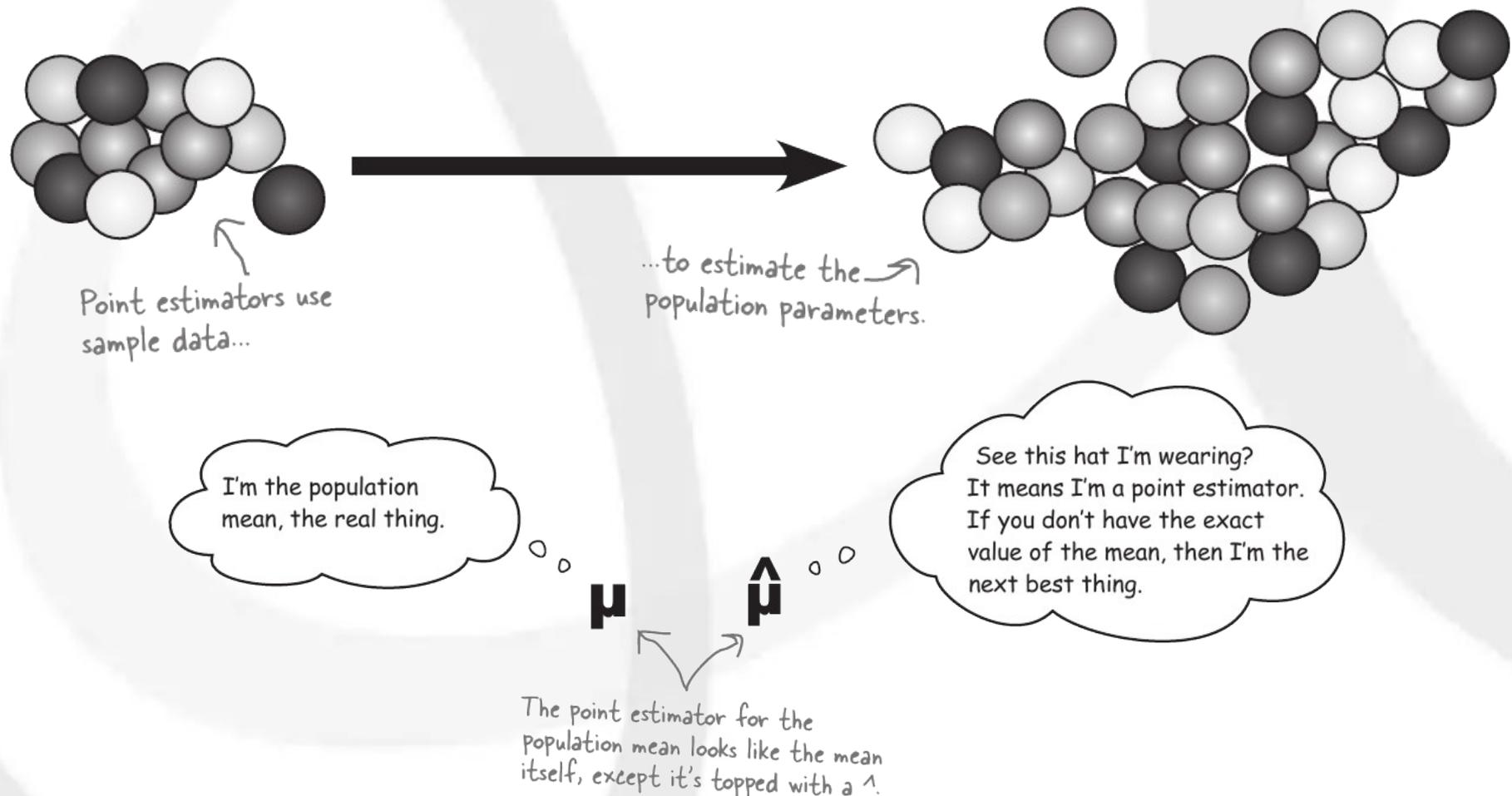
- Assume-se que a média da amostra e a média da população tenham “mais ou menos” o mesmo valor
 - A média da amostra é o melhor estimador pontual da média da população disponível



Testando Chicletes



- Estimadores Pontuais x Parâmetros da População



Testando Chicletes



- Estimadores Pontuais x Parâmetros da População

\bar{x} is the mean of the sample. $\rightarrow \bar{x} = \frac{\sum x}{n}$ \leftarrow Add together the numbers in the sample, and divide by how many there are.

We estimate the mean of the population... $\rightarrow \hat{\mu} = \bar{x}$ \leftarrow ...using the mean of the sample.

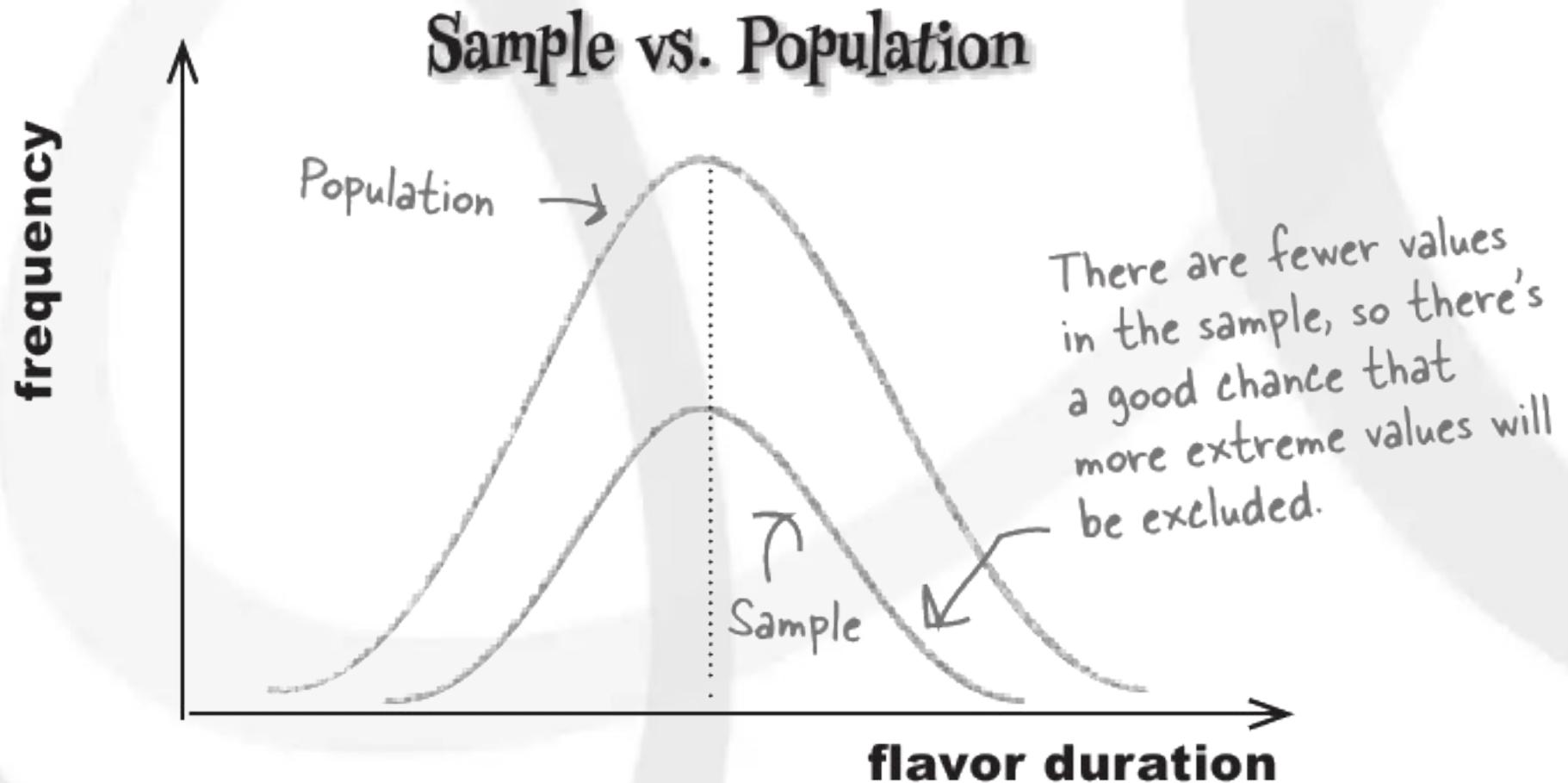
We can estimate the population mean by calculating the mean of the sample.

$$\begin{aligned} \hat{\mu} &= \bar{x} = \frac{61.9 + 62.6 + 63.3 + 64.8 + 65.1 + 66.4 + 67.1 + 67.2 + 68.7 + 69.9}{10} \\ &= 657/10 \\ &= 65.7 \end{aligned}$$

Testando Chicletes



- E quanto à variância da duração do sabor ?



Testando Chicletes



- A variância da amostra não é um bom estimador

Population variance $\rightarrow \sigma^2 = \frac{\Sigma(x - \mu)^2}{n}$ \leftarrow Population mean

n \leftarrow Size of the population

Point estimator for the population variance, based on your sample. $\rightarrow \hat{\sigma}^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1}$ \leftarrow Sample mean

$n - 1$ \leftarrow $n - 1$, not n , where n is the size of the sample. This time it's an estimate..

Point estimator for the population variance $\rightarrow \hat{\sigma}^2 = s^2$

Testando Chicletes



- A variância da amostra não é um bom estimador

We can estimate the population variance by calculating s^2 .

$$\hat{\sigma}^2 = s^2$$

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

$$= \frac{(-3.8)^2 + (-3.1)^2 + (-2.4)^2 + (-0.9)^2 + (-0.6)^2 + (0.7)^2 + (1.4)^2 + (1.5)^2 + (3)^2 + (4.2)^2}{9}$$

$$= \frac{14.44 + 9.61 + 5.76 + 0.81 + 0.36 + 0.49 + 1.96 + 2.25 + 9 + 17.64}{9}$$

$$= 62.32/9$$

$$= 6.92 \text{ (to 2 decimal places)}$$

Testando Chicletes



- Outro teste:
 - Perguntou-se a uma amostra de pessoas se eles preferem os chicletes da *RottenTeeth BubbleGum* ou do concorrente
 - Como resultado, 32 pessoas de um grupo de 40 afirmaram preferir os chicletes da *RottenTeeth BubbleGum*
 - Mas qual é essa proporção em toda a população ?

Point estimator for the proportion of successes in the population $\rightarrow \hat{p} = p_s \leftarrow$ Proportion of successes in the sample

where

$$p_s = \frac{\text{number of successes}}{\text{number in sample}}$$

Testando Chicletes



- Outro teste:

We can estimate the population proportion with the sample proportion. This gives us

$$\hat{p} = p_s = 10/40 \\ = 0.25$$

The probability of choosing someone from the population who doesn't prefer pink gumballs is

$$P(\text{Preference not Pink}) = 1 - \hat{p} \\ = 1 - 0.25 \\ = 0.75$$

Testando Chicletes



- A *RottenTeeth BubbleGum* agora lança o pacote “Jumbo”, contendo 100 chicletes de cores variadas para serem mascarados em grupo
- 25% dos chicletes de toda a população são vermelhos
- Uma cliente diz que só tem interesse em comprar o pacote Jumbo se vier com, no mínimo, 40 chicletes vermelhos. Qual a probabilidade disso acontecer ?
 - Ou seja, sabendo-se que existem 100 chicletes em cada pacote qual a probabilidade de 40% dos chicletes serem vermelhos, dados que 25% da população dos chicletes são vermelhos ?

Testando Chicletes



- Este problema é diferente do anterior:
 - Antes partimos de uma amostra representativa e encontramos estimadores dos parâmetros da população
 - Agora nós conhecemos os parâmetros da população e devemos encontrar probabilidades para um pacote particular de chicletes

Testando Chicletes



- Roteiro para a solução do problema:
 - 1) Encontre todas as possíveis amostras de tamanho 100 que podem ser obtidas a partir da população
 - 2) Encontre a proporção em cada uma dessas amostras. Depois encontre a Esperança (média) e Variância destas proporções
 - 3) Utilize a distribuição das proporções em amostras para encontrar a proporção de uma amostra qualquer (aquela que será comprada pela cliente)

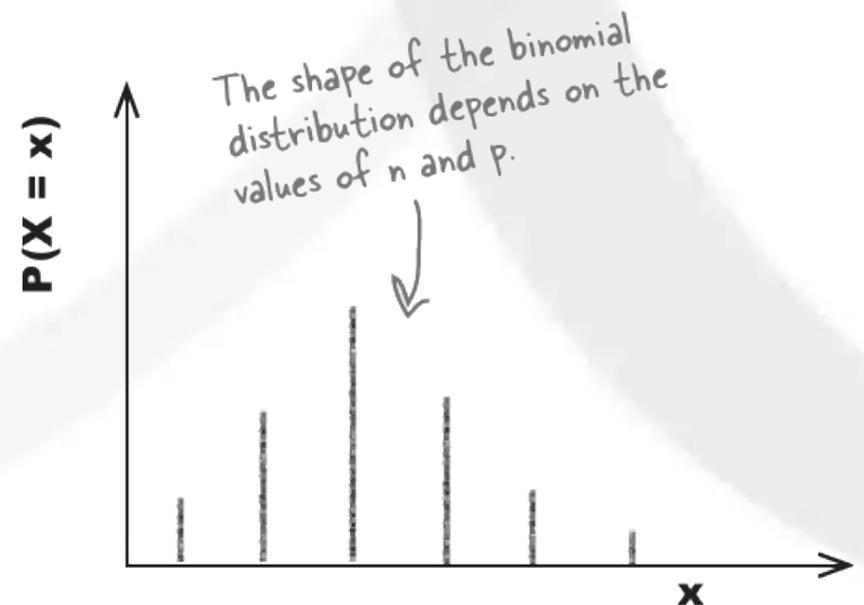
Testando Chicletes



- Seja X o número de chicletes vermelhos encontrados em um pacote específico

$$X \sim B(n, p)$$

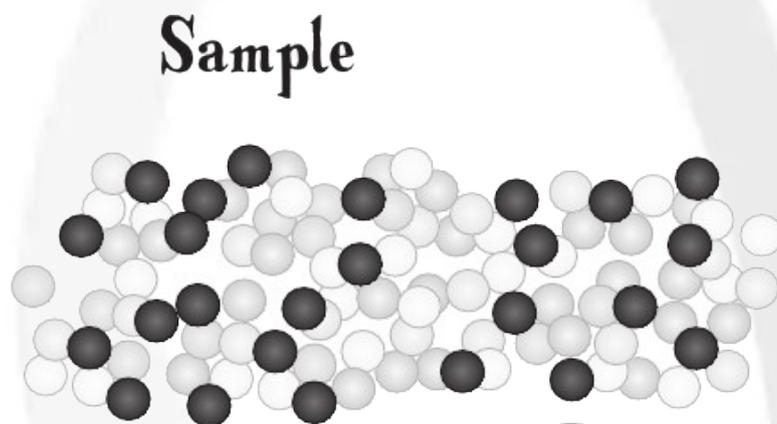
The exact shape of the binomial distribution varies according to the values of n and p . The closer to 0.5 p is, the more symmetrical the shape becomes. In general it is skewed to the right when p is below 0.5, and skewed to the left when p is greater than 0.5.



Testando Chicletes



- Qual é então a proporção de chicletes vermelhos neste pacote específico ?



$$X \sim B(n, p)$$

We don't know the exact number of red gumballs in the sample, but we know its distribution.

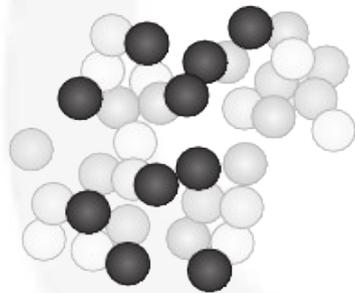
$$P_s = \frac{X}{n}$$

P_s represents the proportion of successes in the sample.

Testando Chicletes

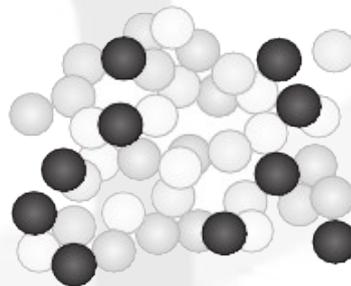


- Existe um número grande de amostras diferentes, mas o número X de chicletes vermelhos em todas elas seguem a mesma distribuição $B(n, p)$ e a proporção é sempre X/n

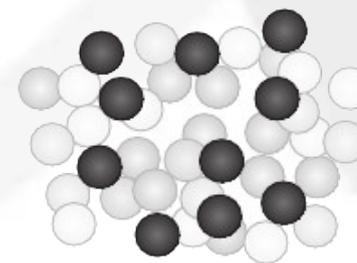


$$X \sim B(n, p)$$
$$P_s = X/n$$

Different samples



$$X \sim B(n, p)$$
$$P_s = X/n$$



$$X \sim B(n, p)$$
$$P_s = X/n$$

Each sample contains n elements, just like the previous one.

Testando Chicletes



- Pode-se formar a distribuição de todas as proporções de amostras utilizando todas as possíveis amostras
- Sabe-se que a Esperança e Variância de P_s são dados por:

$$E(P_s) = p$$

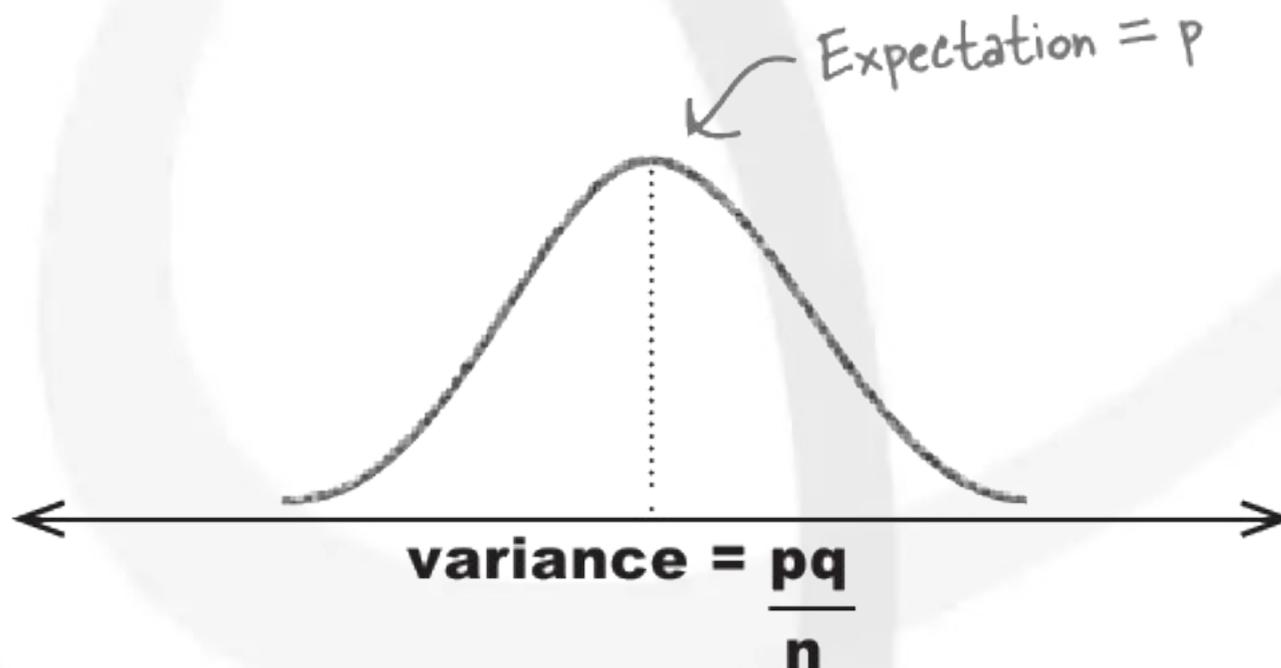
$$\text{Var}(P_s) = \frac{pq}{n}$$

- Precisa-se somente saber como estes dados estão distribuídos

Testando Chicletes



- Distribuição de P_s para valores grandes de n (> 30)
 - Ver: http://onlinestatbook.com/stat_sim/sampling_dist/index.html

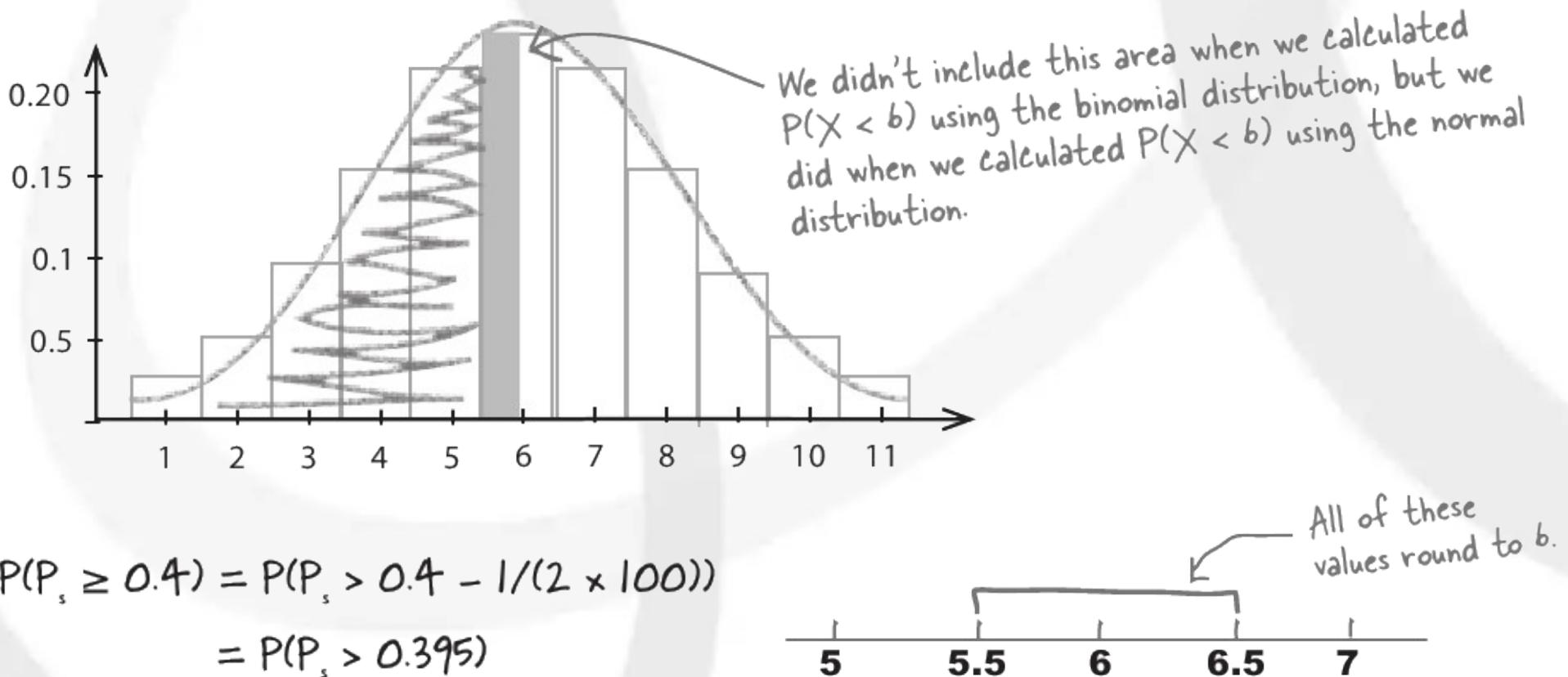


$$P_s \sim N\left(p, \frac{pq}{n}\right)$$

Testando Chicletes



- Se o valor de n não é consideravelmente grande é necessário aplicar a correção de continuidade:



Testando Chicletes



- Os livros de estatística geralmente trazem tabelas com valores cumulativos da distribuição normal com Esperança 0 e Variância 1
- É preciso portanto utilizar *z-scores* para podermos comparar os nossos dados com os dados da tabela

$$z = \frac{x - \mu}{\sigma}$$

These are the mean and standard deviation of the set of data containing the value x .

Testando Chicletes



- Calculando o *z-score*

$P_s \sim N(p, pq/n)$, where $p = 0.25$, $q = 0.75$, and $n = 100$. As pq/n is equal to $0.25 \times 0.75 / 100 = 0.001875$, this gives us

$$P_s \sim N(0.25, 0.001875)$$

As $P_s \sim N(0.25, 0.001875)$, we need to find the standard score of 0.395 so that we can look up the result in probability tables. This gives us

$$z = \frac{0.395 - 0.25}{\sqrt{0.001875}}$$

Testando Chicletes



- Calculando o *z-score*

| | | | | | | | | | | |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 2.5 | .9938 | .9940 | .9941 | .9943 | .9945 | .9946 | .9948 | .9949 | .9951 | .9952 |
| 2.6 | .9953 | .9955 | .9956 | .9957 | .9959 | .9960 | .9961 | .9962 | .9963 | .9964 |
| 2.7 | .9965 | .9966 | .9967 | .9968 | .9969 | .9970 | .9971 | .9972 | .9973 | .9974 |
| 2.8 | .9974 | .9975 | .9976 | .9977 | .9977 | .9978 | .9979 | .9979 | .9980 | .9981 |
| 2.9 | .9981 | .9982 | .9982 | .9983 | .9984 | .9984 | .9985 | .9985 | .9986 | .9986 |
| 3.0 | .9987 | .9987 | .9987 | .9988 | .9988 | .9989 | .9989 | .9989 | .9990 | .9990 |
| 3.1 | .9990 | .9991 | .9991 | .9991 | .9992 | .9992 | .9992 | .9992 | .9993 | .9993 |
| 3.2 | .9993 | .9993 | .9994 | .9994 | .9994 | .9994 | .9994 | .9995 | .9995 | .9995 |
| 3.3 | .9995 | .9995 | .9995 | .9996 | .9996 | .9996 | .9996 | .9996 | .9996 | .9997 |
| 3.4 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9998 |

$$\begin{aligned}P(Z > z) &= 1 - P(Z < 3.35) \\ &= 1 - 0.9996 \\ &= 0.0004\end{aligned}$$

Intervalos de Confiança



- A *RottenTeeth BubbleGum* está sendo processada pois uma empresa concorrente executou outros testes nos chicletes e encontraram resultados diferentes
- O que aconteceu ?
 - Os estimadores pontuais representam a melhor estimativa possível de ser utilizada, pois são obtidos a partir de uma única amostra
 - Esta amostra, mesmo sendo representativa, não deixa de ser uma amostra

Intervalos de Confiança

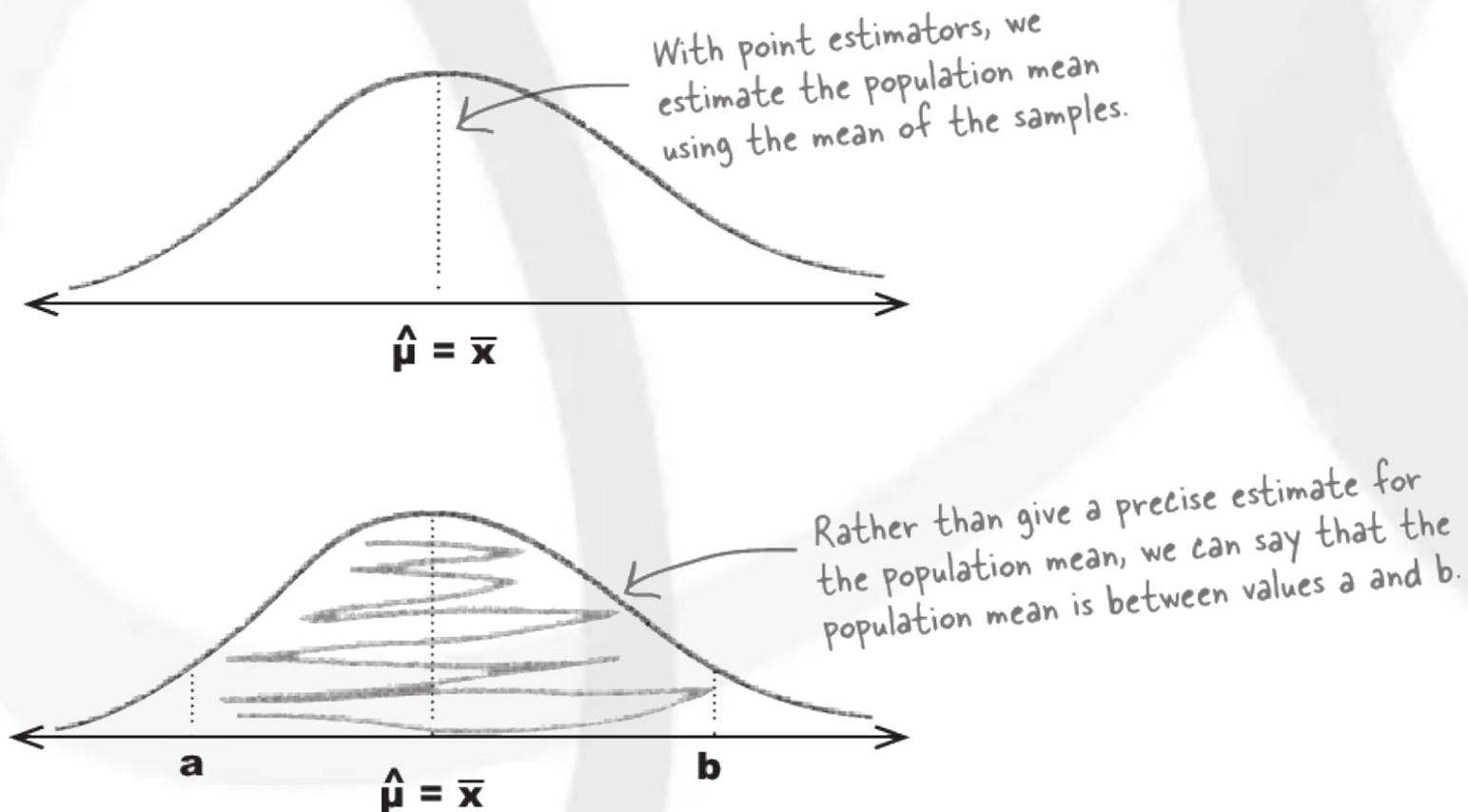


- Uma melhor solução é não afirmar valores precisos para a população e sim intervalos que acomodam margens de erro. Ex: o chiclete dura entre 55 e 65 minutos
- Mas como chegamos neste intervalo ?
 - Ele depende basicamente de quão confiantes queremos estar acerca da afirmação

Intervalos de Confiança



- Intervalos de Confiança



Intervalos de Confiança



- Roteiro para a solução do problema:
 - 1) Escolha a estatística da população para a qual você deseja construir o intervalo de confiança (neste caso a média da duração do sabor)
 - 2) Encontre a distribuição desta estatística nas amostras
 - 3) Escolha o nível de confiança a ser utilizado
 - 4) Encontre os limites de confiança

Intervalos de Confiança



- Encontrando a distribuição das médias das amostras



Assume that $X \sim N(\mu, \sigma^2)$ and that the number in the sample is large. What distribution does \bar{X} follow? Use $E(\bar{X})$ and $\text{Var}(\bar{X})$ above to help you.

If X follows a normal distribution, then \bar{X} does too. Substituting in the point estimator for σ^2 , we get:

$$\bar{X} \sim N(\mu, s^2/n)$$

or:

$$\bar{X} \sim N(\mu, 0.25)$$

Intervalos de Confiança

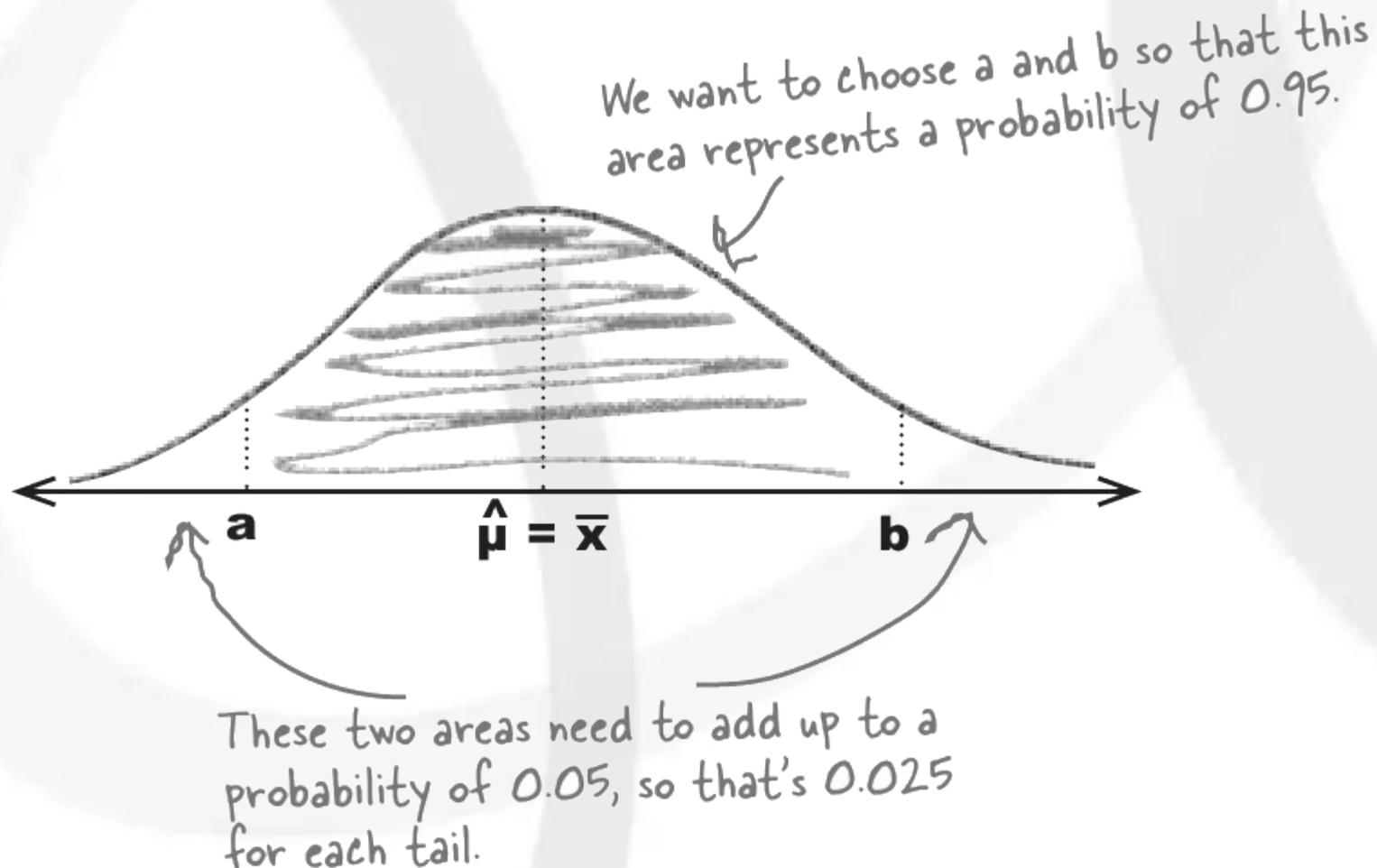


- A escolha do nível de confiança depende do risco que deseja-se correr
- 95% é um nível de confiança geralmente utilizado
- Quanto maior o nível de confiança maior se torna o intervalo de confiança obtido, o que pode produzir resultados menos significativos
 - O que significaria dizer que o sabor do chiclete dura entre 0 minutos e 3 dias ?

Intervalos de Confiança



- Encontrando o intervalo de confiança



Intervalos de Confiança

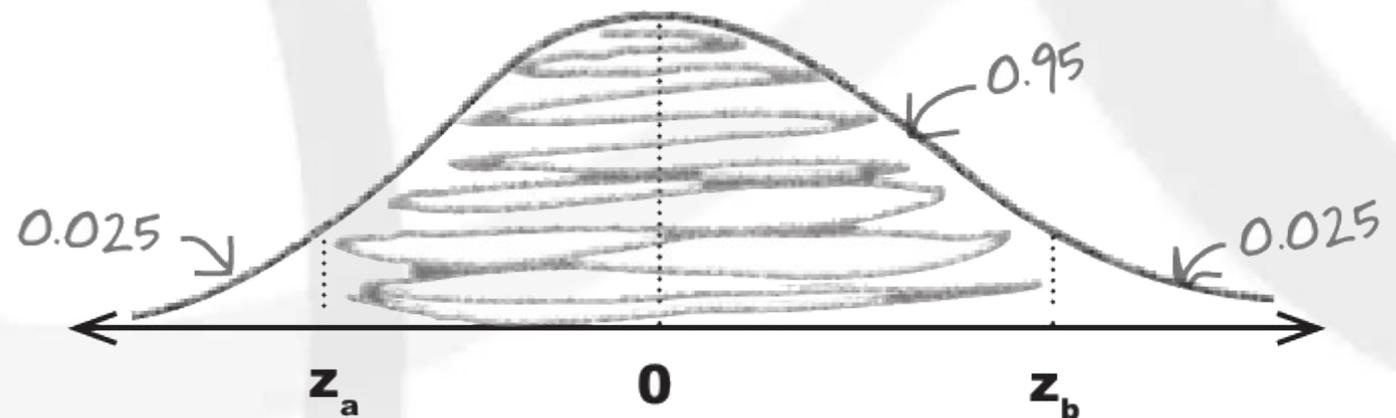


- Comece calculando o *z-score* como uma função da média

$$Z = \frac{\bar{X} - \mu}{\sqrt{0.25}}$$

where

$$Z \sim N(0, 1)$$



Intervalos de Confiança



- Use a tabela para encontrar os valores de Z_a e Z_b
 1. Use probability tables to find the value of z_a where $P(Z < z_a) = 0.025$.
If we look up the probability 0.025 in standard normal probability tables, this gives us $z_a = -1.96$
 2. Use probability tables to find the value of z_b where $P(Z > z_b) = 0.025$.
To find z_b , we need to look up a value of 0.975. This gives us $z_b = 1.96$.

Intervalos de Confiança



- Re-escreva a inequação em função da média

$$P\left(-1.96 < \frac{\bar{X} - \mu}{0.5} < 1.96\right) = 0.95$$

$$P(\bar{X} - 0.98 < \mu < \bar{X} + 0.98) = 0.95$$

The confidence limits are given by $\bar{X} - 0.98$ and $\bar{X} + 0.98$. If we substitute in the mean of the sample, we get confidence limits of $62.7 - 0.98$ and $62.7 + 0.98$. In other words, our confidence interval is $(61.72, 63.68)$.

Estratégias Empíricas



- Pesquisa Qualitativa x Pesquisa Quantitativa
- Principais estratégias empíricas:
 - Survey
 - Estudo de Caso
 - Experimento

Estratégias Empíricas



- *Survey*:
 - Geralmente utilizado *a posteriori*
 - Os principais meios de obtenção de dados quantitativos e qualitativos são questionários e entrevistas, aplicados a uma amostra representativa da população
 - São derivadas, a partir dos resultados, conclusões descritivas e explanatórias
 - Tais conclusões são então generalizadas para a população
 - Quanto ao propósito:
 - Descritivo
 - Explanatório
 - Exploratório

Estratégias Empíricas



- Estudo de Caso:
 - Realizados através do monitoramento de projetos, atividades e atribuições
 - Os dados são coletados para um propósito bem definido e, a partir deles, estatísticas são obtidas
 - Geralmente tem como objetivo rastrear um atributo específico ou estabelecer relacionamentos entre atributos diferentes
 - Possui baixo nível de controle
 - Análise estatística multivariada é frequentemente aplicada (regressão linear, PCA, etc)

Estratégias Empíricas



- Experimento:
 - Geralmente realizados “em laboratório” com um nível maior de controle
 - O objetivo é manipular uma ou mais variáveis e manter todas as outras em níveis fixos
 - O efeito da manipulação é medido e análises estatísticas são realizadas
 - Em alguns casos o experimento é, na verdade, um *quasi-experimento*
 - Exemplo: avaliar com significância estatística se um determinado método de inspeção é melhor que outro

Processo Sistemático de Experimentação



- Fases de um experimento:
 - Definição
 - Planejamento
 - Operação
 - Análise e Interpretação
 - Apresentação e Empacotamento

Processo Sistemático de Experimentação



- Definição:
 - A hipótese deve ser claramente apresentada, mesmo que informalmente
 - Elementos constituintes da definição:
 - Objeto de estudo (o que será estudado ?)
 - Propósito (qual a intenção do estudo ?)
 - Foco de Qualidade (qual efeito está sendo estudado ?)
 - Perspectiva (o estudo será realizado a partir da visão de quem ?)
 - Contexto (onde o estudo será conduzido ?)

Processo Sistemático de Experimentação



- Planejamento:
 - O contexto (ambiente, participantes, etc) do experimento é determinado com detalhes
 - A hipótese é afirmada em termos de uma hipótese nula e uma hipótese alternativa
 - Determinação das variáveis dependentes e independentes:
 - Valores que elas podem assumir
 - Escala de medição (nominal, ordinal ou intervalar) utilizada em cada variável

Processo Sistemático de Experimentação



- Planejamento:
 - Projeto do experimento:
 - Define-se o tipo de projeto (um fator – dois tratamentos, um fator – mais de dois tratamentos, dois fatores – dois tratamentos, etc)
 - Avalia-se as validades interna, externa, de construção e de conclusão
- Operação:
 - Subfases: preparação, execução e validação dos dados

Processo Sistemático de Experimentação



- Análise e Interpretação:
 - Compreensão dos dados através de estatística descritiva
 - Avaliação da necessidade de redução do conjunto de dados
 - Realização do teste de hipótese (o tipo de teste depende das escalas de medição, valores dos dados de entrada e tipo de resultado desejado)
- Principais mecanismos de teste de hipótese:
 - t-test, Mann-Whitney, F-test, Paired t-test, Wilcoxon, Sign test, ANOVA, Kruskal-Wallis, Chi-2

Processo Sistemático de Experimentação



- Apresentação e Empacotamento:
 - Documentação dos resultados (*paper, lab package* para replicação ou elemento de uma base de conhecimento)

Conclusões



- Engenharia de Software Experimental como arcabouço teórico de apoio à tomada de decisão
- Engenharia de Software Experimental como mecanismo sistemático de avaliação de projetos de desenvolvimento de sistemas
- As diversas estratégias empíricas nos permite um maior conhecimento sobre nossas atividades, com a possibilidade de levantamento de informações quantitativas e qualitativas



Pós-Graduação em Computação Distribuída e Ubíqua

INF612 - Aspectos Avançados em Engenharia de Software
Engenharia de Software Experimental

Sandro S. Andrade
sandroandrade@ifba.edu.br