CE002 - Estatística I - Agronomia (2012) - Turma B (13/03/2012)
Prof. Walmes M. Zeviani - Departamento de Estatística - UFPR Sabatina 1
Acadêmico: grr:

1. Uma moeda justa é lançada 3 vezes e os eventos A e B são definidos conforme a seguir:

$$A = \text{“pelo menos uma face cara é observada”}$$
$$B = \text{“o número de resultados cara é um número par”}$$

a) Identifique os pontos do espaço amostral que pertencem aos eventos A, B, $A \cap B$, $A \cup B$ e A^c;
b) Encontre $P(A)$, $P(B)$, $P(A \cap B)$, $P(A \cup B)$ e $P(A^c)$ por meio da soma das probabilidades desses pontos amostrais;
c) Encontre $P(A \cup B)$ pela regra da adição de probabilidades e compare com o item b), $^*P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
d) Os eventos A e B são mutuamente exclusivos? Por que?

O espaço amostral é composto pelos seguintes resultados, nos quais o símbolo ✓ indica que pertece a um dos eventos. Considere C = cara e K = coroa.

ω	A	B	$A \cap B$	$A \cup B$	A^c
CCC	✓			✓	
CCK	✓	✓	✓	✓	
CKC	✓	✓	✓	✓	
KCC	✓	✓	✓	✓	
CKK	✓			✓	
KCK	✓			✓	
KKC	✓			✓	
KKK		✓		✓	✓

Cada elemento tem 1/8 de probabilidade de ocorrer visto que a moeda é justa, o que implica que $P(CCC) = 0.5 \cdot 0.5 \cdot 0.5 = 0.125 = 1/8$. O mesmo vale para qualquer outro resultado. Como os resultados são equiprováveis, basta somarmos o número de pontos a favor de cada evento e dividir pelo total de resultados possíveis. Dessa forma $P(A) = 7/8$, $P(B) = 4/8$, $P(A \cap B) = 3/8$, $P(A \cup B) = 1$ e $P(A^c) = 1 - P(A) = 1/8$.

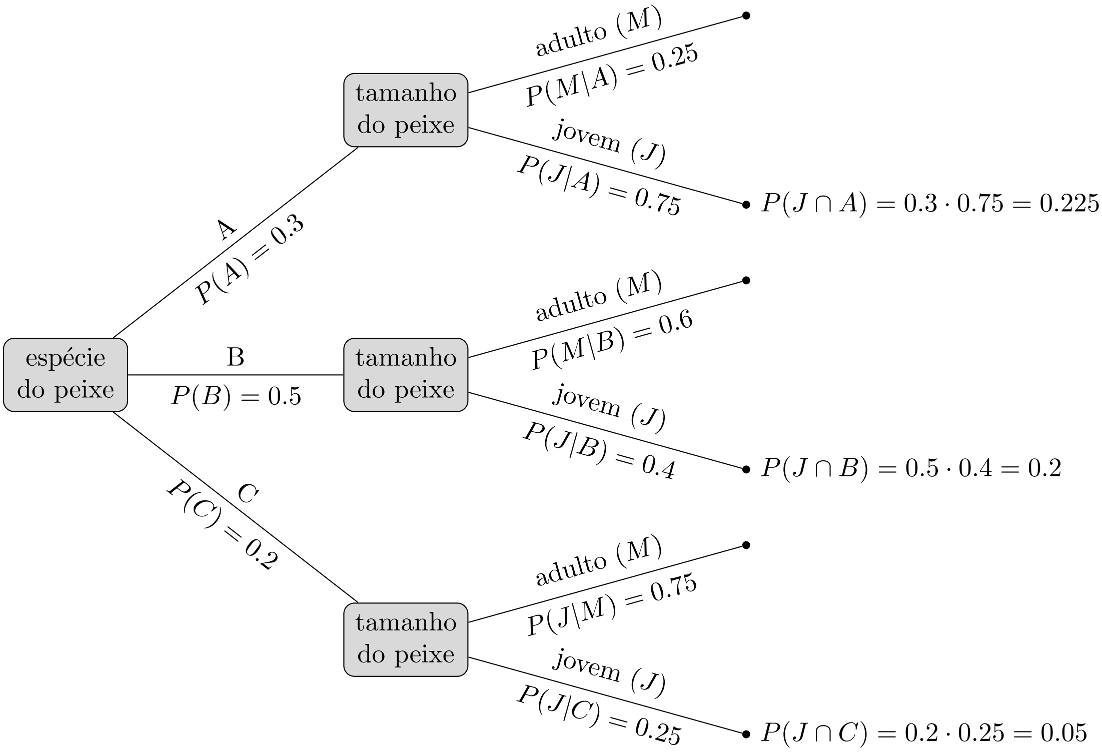
Usando a regra da adição de probabilidades, temos que $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 7/8 + 4/8 - 3/8 = 1$, como deveria de ser.

Dois eventos são mutuamente exclusivos se a ocorrência de um implicar na não ocorrência do outro, ou seja, $A \cap B = \{\varnothing\}$. Assim, temos que A e B não são mutuamente exclusivos.

---

2. Em um tanque de criação existem três espécies de peixes (A, B e C) e animais de dois tamanhos, jovens (J) e adultos (M). Sabe-se que foram colocados 30%, 50% e 20% animais de cada espécie e que após 60 dias 25%, 60% e 75% dos peixes de cada espécie atingem o tamanho adulto e podem ser comercializados. Suponha que durante os 60 dias os animais não morreram e não se reproduziram. Considere o experimento de retirar aleatoriamente um peixe do tanque. Dado que o peixe retirado foi um peixe jovem, qual a probabilidade de ser da espécie C? Caso queira, defina os eventos e construa o diagrama de árvore de probabilidades para organizar seu raciocínio.

O diagrama de árvore de probabilidades fica da seguinte forma:



Por meio do teorema de Bayes, temos que

$$
\begin{aligned}
P(C|J) &= \frac{P(C \cap J)}{P(J)} \\
&= \frac{P(C \cap J)}{P(A \cap J) + P(B \cap J) + P(C \cap J)} \\
&= \frac{P(J|C)P(C)}{P(J|A)P(A) + P(J|B)P(B) + P(J|C)P(C)} \\
&= \frac{0.05}{0.225 + 0.2 + 0.05} \\
&= 0.1052
\end{aligned}
$$

3. Suponha que você investiu uma quantidade fixa de dinheiro em cinco empreedimentos. Assuma que 70% dos investimentos são bem sucedidos, os resultados dos investimentos são independentes uns dos outros e a distribuição de probabilidades para o número, X, de casos bem sucedidos a cada cinco investimentos é

x	0	1	2	3	4	5
$P(X = x)$	0.002	0.029	0.132	0.309	0.360	0.168

Da lista de alternativas abaixo, escolha apenas duas delas para resolver:

a) Mostre que os valores na tabela satisfazem as condições de uma distribuição de probabilidades;
b) Faça o esboço do gráfico de distribuição de probabilidades;
c) Faça o esboço do gráfico de distribuição de probabilidades acumulada;
d) Calcule o valor esperado de X e interprete o resultado;
e) Calcule $P(X = 5)$, $P(2 \leq X \leq 4)$ e $P(X > 1)$ e interprete esses resultados.

As soluções são dadas abaixo:

a) Verifica-se que os valores de $P(X = X)$ são todos não negativos e sua soma é 1, portanto, trata-se de uma distribuição de probabilidades.
b) O gráfico da distribuição de probabilidades está abaixo

c) O gráfico da distribuição de probabilidades acumulada está abaixo

d) Calcule o valor esperado de X é

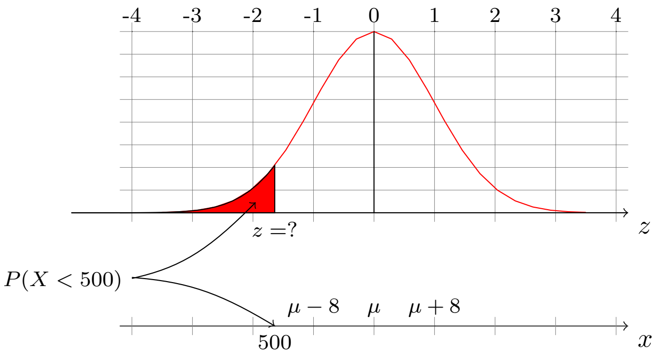
$$\mu = E(X) = \sum_{\forall x} x \cdot P(X = x) = 3.5 \quad (1)$$

que representa o número médio de investimentos bem sucedidos a cada 5 investimentos feitos e uma longa série (infinita) de aplicações.

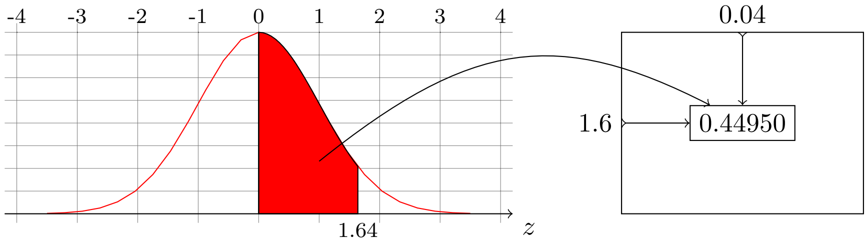
e) Temos que $P(X = 5) = 0.168$, isso representa a probabilidade de sucesso tem todos os investimetos. Temos que $P(2 \leq X \leq 4) = 0.669$ representa a probabilidade de ter de 2 à 4 investimentos bem sucedidos. Temos que $P(X > 1) = 0.969$ representa a probabilidade de ter mais de um investimento bem sucedido.

4. Uma máquina de encher sacos com café torrado e moído o faz com um peso médio μ que pode ser alterado por meio da regulagem da máquina. O peso dos sacos cheios é uma v.a. normal que apresenta variabilidade fixa e traduzida por um desvio-padrão de 8 gramas. As regras do controle de qualidade estabelecem que apenas 5% dos sacos podem ter peso inferior à 500 gramas (valor impresso na embalagem). Em quanto deve ser regulado o peso médio μ da máquina para se trabalhar dentro das regras de qualidade?

Primeira coisa é obter o valor do quantil z na tabela da distribuição normal que corresponde à uma probabilidade de 0.05. Esse valor é o $z = -1.64$, pois de $-\infty$ à -1.64 tem-se uma área de 0.05. A área do evento de nosso interesse está representado na figura abaixo



A área consultada na tabela da distribuição normal está representado na figura abaixo, uma vez que $P(Z > 1.64) = 0.50 - P(0 < Z < 1.64) = 0.50 - 0.44950 \approx 0.05$.



Então o valor de z que delimita o nosso evento de interesse é o $z = -1.64$. Partindo da conhecida equação da padronização

$$z = \frac{x - \mu}{\sigma}, \tag{2}$$

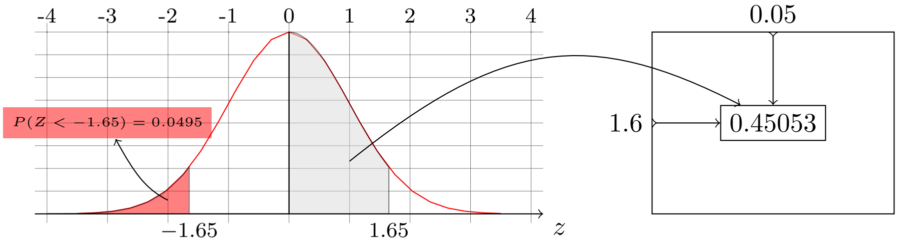
podemos descobrir o valor de μ. Dessa forma temos

$$\begin{aligned}
z &= \frac{x - \mu}{\sigma} \\
z \cdot \sigma &= x - \mu \\
\mu &= x - z \cdot \sigma \\
\mu &= 500 - (-1.64) \cdot 8 \\
\mu &= 513.12.
\end{aligned}$$

Portanto, se a máquina for regulada para encher pacotes com média de 513.12 gramas, teremos apenas 5% dos pacotes com peso inferior à 500, tal como desejávamos.

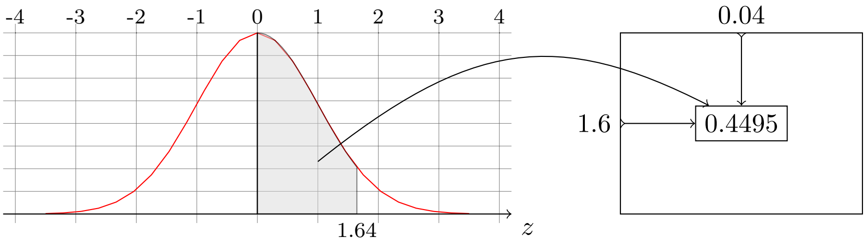
5. Fixação biológica do nitrogênio (FBN) é a principal fonte de N para a cultura da soja. Bactérias do gênero *Bradyrhizobium*, quando em contato com as raízes da soja, infectam as raízes, via pêlos radiculares, formando os nódulos. O sucesso do procedimento depende da qualidade o inóculo empregado. Uma empresa comercializa inóculo com uma população de 1×10^8 células por mL de inóculo (o mesmo que 100 células por ηL). Suponha que o número de células por ηL de inóculo seja uma v.a. com distribuição de Poisson com parâmetro $\lambda = 100$. Um teste em uma amostra comercial de inóculo contou 83 células por ηL. Então qual a probabilidade do número de células ser menor ou igual à 83? Determine os valores, simétricos à média, tal que a probabilidade de contagens nesse intervalo seja o mais próximo, mas não menor que, 90%.

Fazendo aproximação da distribuição de Poisson pela distribuição normal, temos que obter a média e o desvio-padrão que são $\mu = \lambda = 100$ e $\sigma = \sqrt{\lambda} = 10$. Agora encontramos o valor padronizado z por $z = \frac{x-\mu}{\sigma} = \frac{83.5-100}{10} = -1.65$ para então fazermos a consulta à tabela da distribuição normal padrão de onde obtemos que $P(Z < -1.65) = 0.0495$.



Portando, se o fabricante realmente comercializa inóculo com número médio de células igual à 100 (assumindo distribuição de Poisson), a probabilidade de encontrar uma amostra que apresente contagem menor ou igual a 83 é de 0.0495.

A obtenção do intervalo simétrico começa por encontrar os valores de z de delimitam uma área simétrica de 0.9 ao redor de zero. Assim, cada metade deve ter área de 0.45. Então os valores de z em questão são $z_e = -1.64$ e $z_d = 1.64$.



Com esses valores aplicamos a tranformação inversa a padronização, ou seja, $x = \sigma z + \mu$, e encontramos os valores $x_e = 83.551$ e $x_d = 116.449$. Como nossa v.a. tem suporte discreto, devemos usar valores inteiros como limites do intervalo. Assim, truncamos para o inteiro abaixo e acima dos respectivos limites, assim $x_e = 83$ e $x_d = 117$. Com isso, $P(83 \leq X \leq 117) \geq 0.9$. Portanto, espera-se que, pelo menos, 90% das amostras apresente contagem de células entre 83 e 117.

CE002 - Estatística I - Agronomia (2012) - Turma B (17/04/2012)
Prof. Walmes M. Zeviani - Departamento de Estatística - UFPR Sabatina 6
Acadêmico: grr:

6. Uma equipe de engenheiros agrônomos conduziu um experimento para selecionar híbridos de milho em um programa de melhoramento voltado para a região de Chapecó. Parte dos dados coletados em um experimento de campo com 32 híbridos estão presentes na tabela abaixo. São dados de comprimento do ciclo vegetativo (C: curto, N: normal, L: longo) e resistência à seca (S: susceptível, M: moderado, R: resistênte).

a) Classifique as variáveis quanto ao tipo.
a) Faça uma tabela de distribuição de frequências cruzadas incluindo as frequências marginais.
a) Represente por meio de um gráfico a relação entre as variáveis (use o espaço milimitrado, defina rótulos aos eixos e legendas).
a) Descreva e interprete o gráfico e conclua sobre que tipo de ciclo recomendar para uma região com riscos de seca.

h	ciclo	seca	h	ciclo	seca	h	ciclo	seca	h	ciclo	seca
1	N	S	9	C	M	17	L	M	25	C	R
2	N	M	10	N	M	18	C	S	26	C	S
3	L	R	11	N	R	19	L	R	27	L	R
4	N	M	12	L	S	20	N	M	28	N	R
5	N	M	13	L	M	21	C	S	29	C	S
6	L	R	14	N	R	22	N	S	30	L	R
7	N	R	15	C	M	23	C	M	31	N	R
8	C	S	16	L	R	24	N	S	32	C	R

Ambas variáveis são qualitativas ordinais. A ordem dos níveis é C<N<L e S<M<R. A tabela cruzada de distribuição de frequências absolutas e relativas está abaixo. Como faz mais sentido pensar que a resistência à seca depende do ciclo, obtivemos a tabela de distribuição relativa dos níveis de seca condicionais aos níveis de ciclo. Com base nessa tabela obtemos o gráfico de mosaico onde observamos que maior resistência à seca está associado à maiores ciclos. Dessa maneira, para uma região de risco de seca deve-se cultivar híbridos de ciclo longo.

Frequências absolutas

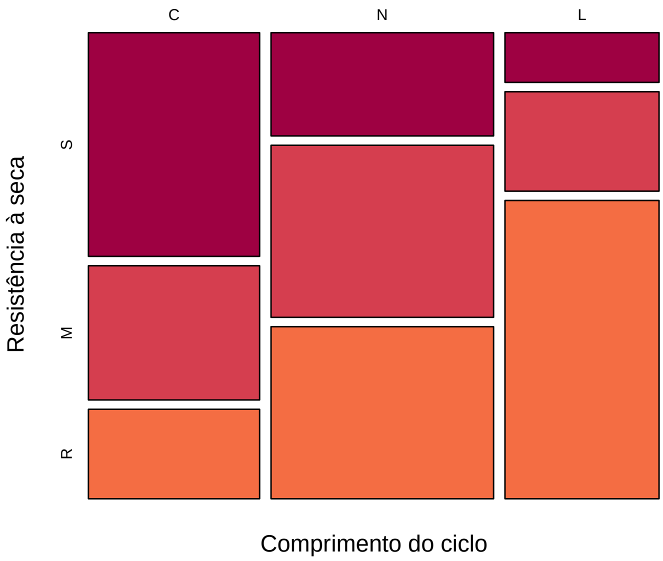
	S	M	R	tot
C	5	3	2	10
N	3	5	5	13
L	1	2	6	9
tot	9	10	13	32

Frequências relativas

	S	M	R	tot
C	0.156	0.094	0.062	0.312
N	0.094	0.156	0.156	0.406
L	0.031	0.062	0.188	0.281
tot	0.281	0.312	0.406	1.000

Frequências condicionais

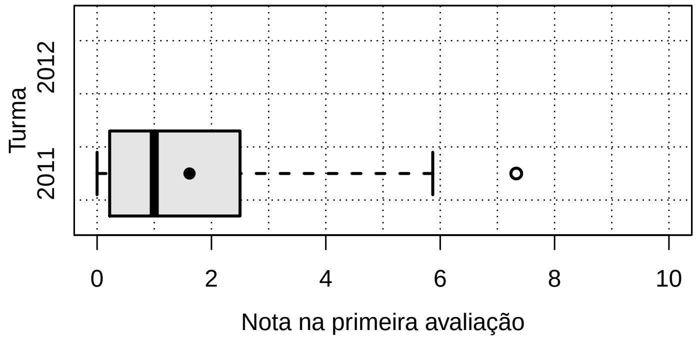
	S	M	R	tot
C	0.5000	0.3000	0.2000	1
N	0.2308	0.3846	0.3846	1
L	0.1111	0.2222	0.6667	1



7. As notas na primeira avaliação de Estatística de duas turmas de agronomia estão representadas abaixo. O desempenho da turma de 2011 está representado em um gráfico de caixas e bigodes (*boxplot*) e da turma de 2012 está em um diagrama de ramos-e-folhas. No *boxplot* o ponto preenchido é a média amostral. Obtenha o *boxplot* para turma de 2012, tal como temos para turma de 2011, a partir do diagrama de ramos-e-folhas e compare os desempenhos das turmas.

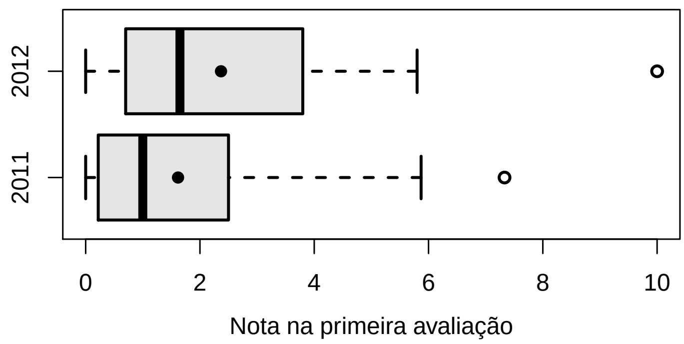
```
  The decimal point is at the |

   0 | 0000235777
   1 | 0012567
   2 | 0255
   3 | 022
   4 | 4578
   5 | 448
   6 |
   7 |
   8 |
   9 |
  10 | 0
```



No gráfico de ramos e folhas a amostra já esta ordenada e dessa forma podemos calular os extremos e quartis necessários para fazer o gráfico de caixas. A média também deve ser obtida bem como a amplitude interquartílica e os limites para discriminação de pontos. Essas estatísticas estão na tabela abaixo. Com elas faz-se o gráfico apresentado a seguir.

min	q_1	m_d	q_2	max	$\bar{x}$	aiq	$q_1 - 1.5 \cdot \text{aiq}$	$q_3 + 1.5 \cdot \text{aiq}$
0.00	0.70	1.65	3.80	10.00	2.37	3.10	-3.95	8.45



De modo geral entende-se que 2012 foi melhor que 2011 pois as duas medidas de posição foram superiores. A média foi aproximadamente 0.75 de unidade superior. Observa-se assimetria a direita com a discrinação de duas notas superiores a 7, sendo superior a da turma 2012. A amplitude bigodes foi praticamente a mesma, porém a amplitude da caixa para 2012 foi maior o que descreve uma distribuição de notas mais uniforme ou menos assimétrica que a 2011. Em 2011 temos que 25% das notas estiveram abaixo de 0.25 sendo que para 2012 esse limite é 0.75. Em 2011 25% dos alunos tiveram notas superiores à 2.5 em para 2012 a mesma proporção esteve acima da nota 4.

CE002 - Estatística I - Agronomia (2012) - Turma B (15/05/2012)
Prof. Walmes M. Zeviani - Departamento de Estatística - UFPR Sabatina 8
Acadêmico: grr:

1. Considerando as propriedades de média e variância e o teorema do limite central. Se repetidas amostras de $n = 10$ elementos são feitas de uma variável aleatória com $\mu = 100$ e $\sigma = 5$, a distribuição amostral da média ($\bar{X}$) será aproximadamente

( ) $N(\mu_{\bar{X}} = 100, \sigma_{\bar{X}} = 5/10)$; ( ) $N(\mu_{\bar{X}} = 100, \sigma_{\bar{X}} = \sqrt{5/10})$; (x) $N(\mu_{\bar{X}} = 100, \sigma_{\bar{X}} = 5/\sqrt{10})$;
( ) $N(\mu_{\bar{X}} = 100, \sigma_{\bar{X}} = 5^2/10)$; ( ) $N(\mu_{\bar{X}} = 100, \sigma_{\bar{X}} = 5)$; ( ) $N(\mu_{\bar{X}} = 100/10, \sigma_{\bar{X}} = 2.5)$;

2. Classifique em verdadeiro (V) e falso (F) as sentenças abaixo.

(F) O teorema do limite central estabelece que a distribuição amostral da média ($\bar{X}$) é Normal sob qualquer condição. (é normal à medida que o tamanha da amostra aumenta)
(F) A estimação pelo método da máxima verossimilhança consiste em encontrar uma estimativa que torne máximo o poder da amostra. (máxima a probabilidade de obter a particular amostra)
(V) O melhor estimador é aquele que apresenta não tendenciosidade e variância mínima em relação aos concorrentes. ()
(F) Um estimador tendencioso tem sempre uma variância menor que um estimador não tendencioso. (essas características não estão associadas)
(F) O parâmetro de uma distribuição de probabilidades muda a cada amostra realizada. (parâmetro é uma quantidade fixa)
(V) Erro-padrão é uma medida de dispersão associada à estimativa de um parâmetro. ()
(V) A média amostral é um estimador não tendencioso para a média de uma distribuição Normal. ()
(F) Por distribuição amostral entende-se a distribuição de uma estatística em uma série infinita de amostras dependentes. (amostras independentes)
(F) Estimador pontual é uma função dos valores da amostra que fornece uma média para o parâmetro. (fornece uma estimativa para o parâmetro)
(F) Por distribuição amostral entende-se a distribuição de um parâmetro em uma série infinita de amostras independentes. (distribuição de uma estatística)
(V) Para um parâmetro é possível ter mais de um estimador possível. ()
(V) Um estimador não tendencioso apresenta valores mais próximos do parâmetro à medida que o tamanho da amostra aumenta. ()
(F) Estimador pontual é uma função dos valores da amostra que fornece uma medida de incerteza para o parâmetro. (a estimação pontual é apenas um valor, logo, não fornece medida de incerteza)
(F) Um parâmetro é tendencioso toda vez que é diferente da esperança do seu estimador. (um estimador é tendencioso toda vez que sua esperança é diferente do parâmetro)
(V) A estimação pelo método dos momentos consiste em igualar os momentos populacionais aos respectivos momentos amostrais. ()
(V) Parâmetro é uma quantidade fixa associada à uma distribuição de probabilidades. ()

1. Uma pequena usina hidrelétrica foi instalada nas proximidades de uma região produtora de feijão. Os agricultores dessa região estão insatisfeitos pois afirmam que a usina teve impactos negativos na produtividade que historicamente era de 800 kg ha^{-1}. Para avaliar a veracidade da acusação, uma equipe e engenheiros agrônomos realizou um levantamento amostral da produtividade de feijão onde foram observados os seguintes valores (kg ha^{-1}):

767.8	750.2	736.1
764.1	756.0	746.1
716.8	692.5	731.4

Qual a produtividade média estimada? Qual a conclusão ao nível de significância de 5%? Aplique o teste de hipótese considerando $H_0 : \mu = 800$ e $H_1 : \mu < 800$, sendo $\alpha = 0.05$ temos $t_{\alpha,8} = 1.86$.

O problema é de teste de hipótese para a média com variância desconhecida. Neste caso usa-se o teste t de Student. Assume-se que os dados têm distribuição normal. A estimativa da média da amostra é $\bar{x} = 740.11$ e o desvio-padrão é $s = 24.07$. A estatística do teste é

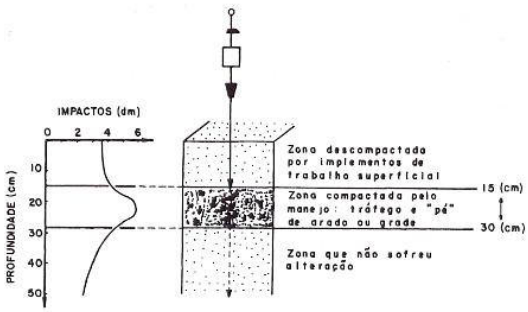
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{740.11 - 800}{24.07/\sqrt{9}} = -7.46.$$

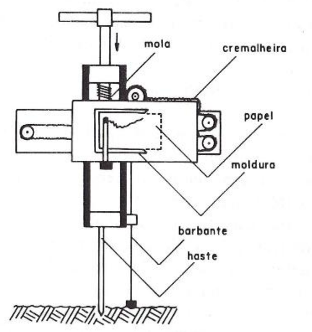
Como o valor calculado está à esquerda do valor tabelado para o teste unilateral, -1.86, portando na região de rejeição da hipótese nula, rejeitamos a hipótese H_0. No entando, é valido ressaltar que essa diferença de produtividade não pode ser atribuída apenas à influência da hidrelétrica, pois, de fato existe influência de outros fatores, mesmo que sejam mínimas, como o clima do presente ano agrícola, a época de plantio, a presença de forma mais intensa de uma praga ou doença, dentre outras possibilidades. Um acompanhamento da produtividade ao longo das safras deve ser conduzido para melhor julgar a hipótese.

CE002 - Estatística I - Agronomia (2012) - Turma B (02/10/2012)
Prof. Walmes M. Zeviani - Departamento de Estatística - UFPR Sabatina 10
Acadêmico: grr:

1. Engenheiros agrônomos estão avaliando uma metodologia para medição da compactação do solo que seja fácil de aplicar e barata a nível de campo. A metodologia avaliada é a do penetrômetro, aparelho de fácil fabricação e uso, que é baseado na resistência do solo à penetração de uma haste, após recebimento de um impacto provocado pelo deslocamento vertical de um bloco de ferro colocado na parte superior da haste, por uma distância conhecida em que o número de impactos por intervalo de profundidade é usado para calcular a resistência a compactação. A metodologia padrão é a do penetrógrafo, que difere do penetrômetro por fornecer diretamente o gráfico do índice de cone (medida da resistência à penetração) sem a necessidade de tabulação de dados (de número de impactos). Após a aplicação de ambos, tem-se uma medida de resistência a compactação do perfil do solo. Em 10 pontos distribuídos numa área agrícola foi avaliado a compactação pelos dois métodos e deseja-se saber se eles apresentam o mesmo valor médio de compactação do solo. Avalie essa hipótese com os dados disponíveis na tabela, de resistência à penetração (MPa), considerando o nível de significância igual à 5% ($t_{5\%;9} = 2.262$).

penetrômetro	penetrógrafo	diferença
1.396	1.449	0.052
2.139	2.139	0.000
2.542	2.565	0.022
0.827	0.934	0.106
2.215	2.411	0.196
2.253	2.342	0.089
1.713	1.762	0.049
1.727	1.736	0.009
1.718	1.734	0.016
1.555	1.897	0.342





Dada situação apresentada a hipótese nula é $H_0 : \mu_1 - \mu_2 = 0$ e a hipótese alternativa é bilateral, assim $H_1 : \mu_1 - \mu_2 \neq 0$. Para avaliar tal hipótese aplicaremos o teste t pareado, o que é equivalente a aplicar o teste t normal sobre as diferenças observadas. A estatística do teste é

$$t_0 = \frac{\bar{X}_d - \Delta_0}{\sqrt{\frac{S_d^2}{n}}} = \frac{0.088 - 0}{\sqrt{\frac{0.107^2}{10}}} = 2.613.$$

Como a estatística do teste se encontra na região de rejeição da hipótese nula, rejeitamos H_0 e concluímos que as metodologias para medir a compactação do solo apresentam valores médios diferentes. A título de curiosidade, o p-valor do teste foi 0.0281.

CE002 - Estatística I - Agronomia (2012) - Turma B (09/10/2012)
Prof. Walmes M. Zeviani - Departamento de Estatística - UFPR Sabatina 11
Acadêmico: grr:

1. Segundo descobertas de Mendel, se duas caractarísticas são controladas por um gene cada, têm segregação independente, cada um com dois alelos com dominância completa, então a proporção esperada de fenótipos resultantes esperado deve ser 9:3:3:1. Um geneticista afirma que as mesmas condições se aplicam para duas características de uma espécie vegetal recém inserida no programa de melhoramento de uma cultura agrícola. As características avaliadas foram: RAT (resistência à alumínio tóxico) e RN (resistência ao nematóide). Ao submeter 300 plantas aos testes (+: resultado positivo, −: resultado negativo), observou-se os resultados da tabela abaixo.

fenótipo	frequência observada	frequência esperada
RAT+ e RN+	155	
RAT+ e RN−	74	
RAT− e RN+	52	
RAT− e RN−	19	

Há evidências de que os resultados desse experimento estão de acordo com a distribuição de probabilidades proposta por Mendel? Avalie essa hipótese com $\alpha = 5\%$. Dado: $\chi^2_1 = 3.84$, $\chi^2_2 = 5.99$, $\chi^2_3 = 7.81$, $\chi^2_4 = 9.49$.

Dado que as proporções esperadas são 9:3:3:1, os valores esperados são calculados, para o primeiro fenótipo, por $300 \times 9/16 = 168.75$.

fenótipo	frequência observada	frequência esperada
RAT+ e RN+	155	168.75
RAT+ e RN-	74	56.25
RAT- e RN+	52	56.25
RAT- e RN-	19	18.75

Com os valores na tebela obtem-se a estatística do teste chi-quadrado

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = 7.046.$$

Como o valor calculado não é superior ao valor tabelado $\chi^2_3 = 7.81$, aceita-se a hipótese de que as probabilidades de cada genótipo são aquelas previstas por Mendel.