



DADOS REAIS NA SALA DE AULA: UMA ATIVIDADE PARA O ENSINO DE ESTATÍSTICA EM CURSOS DE ENGENHARIA

Amanda C. Martin – amanda.cmartin@hotmail.com

Andressa Fetter – andressafetter@hotmail.com

Elisa Henning – dma2eh@joinville.udesc.br

Universidade do Estado de Santa Catarina

Rua Paulo Malschitzki, s/ número - Campus Universitário Prof. Avelino Marcante – Bairro:
Zona Industrial Norte.

89219-710 – Joinville, Santa Catarina

Resumo: *Estatística é uma disciplina importante para quem lida com números. Todo graduando de engenharia deve estudar estatística – e nossa proposta é facilitar esse aprendizado. Ao utilizar softwares para o cálculo das estatísticas, temos a oportunidade de avaliar apenas os resultados, proporcionando um melhor entendimento da essência da disciplina. Neste artigo, sugerimos a utilização do software livre R, que por ser de código aberto torna é de fácil acesso a todos. Além disso, numa experiência em sala de aula, foram utilizados dados reais de uma pesquisa sobre mobilidade urbana, com dados de acidentes envolvendo ciclistas no município de Joinville – SC, entre os anos de 2008 e 2010. Os testes são facilmente aplicados através de comandos inseridos no R, e os resultados são adequados para uma boa interpretação. Assim, o aluno pode dar ênfase aos resultados, uma vez que não precisa utilizar o tempo de aula para cálculos. E, ao abranger uma situação real, o aluno é incentivado a participar ativamente deste processo de ensino-aprendizagem.*

Palavras-chave: *Primeira palavra, Segunda palavra, Terceira palavra (máximo de 5)*

1. INTRODUÇÃO

Para os engenheiros e alunos de engenharia os números representam mais do que simples valores. São dados que, após analisados corretamente, fornecem informações úteis para a tomada de decisões e resolução de problemas.

Ao verificar a variação das dimensões de uma peça, a quantidade de itens defeituosos em uma linha de produção ou até mesmo o rendimento de uma máquina em operação, surgirão resultados que deverão ser avaliados para, a partir deles, concluir se a variação das dimensões das peças é algo natural ou um problema. Assim como se o rendimento da máquina está ou não dentro dos padrões de operação e se o número de itens defeituosos corresponde a uma pequena ou grande proporção da produção.

A coleta, a apresentação, a análise e o uso dos dados para a tomada de decisões e resolução de problemas fazem parte do campo da estatística (MONTGOMERY et al., 2011). Portanto, possuir conhecimento estatístico auxilia na análise dos dados e permite avaliá-los com maior precisão.

Realização:

 **ABENGE**

Organização:



**O ENGENHEIRO
PROFESSOR E O
DESAFIO DE EDUCAR**



A estatística deve ser ensinada, então, de forma a auxiliar os estudantes de engenharia a compreenderem a significância dos números encontrados em cada processo, ou seja, identificarem as anormalidades e buscarem soluções para estas, assim como evitar conclusões precipitadas a respeito de valores que aparentemente representam problemas.

Com o intuito de ensinar aos futuros engenheiros a Estatística de uma maneira mais interessante, buscou-se levar à sala de aula exemplos atuais da aplicação da estatística. A intenção é mostrar o uso de algumas ferramentas estatísticas em situações reais e próximas aos estudantes, fugindo dos exemplos de livros e dando uma percepção melhorada da utilidade destas ferramentas, além de proporcionar, com a utilização de softwares, uma forma mais agradável e simples de análise dos dados.

2. OBJETIVOS

A estatística, por ser uma matéria importante para a maioria dos cursos de engenharia, deve ser ensinada de uma maneira atrativa e de fácil compreensão para que os graduandos se interessem pelo conteúdo e realmente o aprendam. Para isso, pensou-se em uma metodologia de ensino que leva para a sala de aula exemplos reais e de fácil acesso dos alunos, além de priorizar a compreensão dos resultados.

Com exemplos do cotidiano, os alunos conseguem perceber a aplicação dos testes estatísticos e, também, entender melhor a importância dos mesmos. Outro fator relevante é a utilização de softwares para facilitar a compreensão dos alunos. Em vez de focar em grandes cálculos, os softwares geram os valores que devem ser analisados e o foco do ensino é dado aos resultados, não às contas. Assim, os alunos aprendem de uma forma mais agradável a interpretar os resultados dos testes.

O objetivo do presente artigo é mostrar a forma encontrada para levar a estatística de uma forma simples e eficaz aos graduandos de Engenharia, com exemplos didáticos e próximos da realidade dos alunos, poupando-os dos trabalhos exaustivos de contas com o intuito de priorizar a análise e compreensão dos resultados.

3. ATIVIDADES PROPOSTAS

Foram propostas algumas atividades adicionais à atual metodologia de ensino aprendizagem aplicada na turma de “Probabilidade e Estatística” do curso de Engenharia de Produção e Sistemas da Universidade do Estado de Santa Catarina (UDESC). Os alunos desta turma seguem o conteúdo teórico de livro texto proposto pelo Plano de Ensino da disciplina.

Assim através das aulas expositivas com exemplos atuais e utilização de softwares, puderam visualizar a aplicação prática da teoria previamente explicada.

Os exemplos levados à sala de aula correspondem aos dados de uma pesquisa realizada no município de Joinville-SC, onde está localizado o campus da UDESC que possui a graduação em Engenharia de Produção e Sistemas. A pesquisa foi realizada contém informações sobre os acidentes envolvendo ciclistas no município entre os anos de 2008 e 2010 (HACKENBERG et al, 2011).

Para fazer a análise dos dados foi utilizado o software estatístico livre R (R DEVELOPMENT CORE TEAM, 2012), facilmente encontrado para *download* e com diversos manuais de utilização disponíveis na internet. Além de ser vantajoso por ser um software livre, o R possui a opção de carregar pacotes estatísticos adicionais gratuitamente, aumentando os recursos que podem ser utilizados.



Os testes estatísticos propostos foram aplicados a diferentes dados da pesquisa realizada e expostos em sala de aula através de recursos multimídia, onde os alunos puderam visualizar os dados e a forma de utilização do software, além de apreender o conteúdo teórico já visto.

4. APLICAÇÃO DOS TESTES ESTATÍSTICOS

A atividade proposta foi aplicada para complementar as atividades de ensino de dois testes estatísticos, para proporções e qui-quadrado, e para melhor entendimento do coeficiente de correlação linear de Pearson. Como os testes possuem os resultados que necessitam da análise do p -valor, o enfoque das aulas foi dado justamente a essa análise. Ou seja, os alunos foram poupados do trabalho exaustivo de longas contas e puderam dedicar-se à análise do p -valor, compreendendo o seu significado e, principalmente, compreendendo os objetivos dos testes e as suas aplicações. Além disso, foi reforçado o conceito de intervalo de confiança, tanto no teste para proporção, como no teste para o coeficiente de correlação.

4.1 Teste para comparação de proporções

Este teste compara duas proporções populacionais, a partir de dados amostrais. A análise foi feita a partir do p -valor, pode-se afirmar, com determinado nível de significância, se as proporções são iguais, diferentes, ou se uma é maior ou menor que a outra. Para este teste foi avaliado o número de acidentes envolvendo ciclistas por sexo. A tabela com os dados encontra-se a seguir.

Tabela 1 - Sexo dos ciclistas

Sexo	2009		2010		Total Geral	
	Quant.	%	Quant.	%	Quant.	%
Feminino	236	33,4	166	27,8	402	30,8
Masculino	471	66,6	417	69,8	888	68,1
Sem declaração	0	0,0	14	2,3	14	1,1
Total Geral	707	100,0	597	100,0	1304	100,0

O teste foi utilizado para verificar se houve variação na proporção de acidentes para cada sexo separadamente. Ou seja, para avaliar se entre os anos de 2009 e 2010 realmente houve redução do número de acidentes envolvendo mulheres, como aparenta a tabela, assim como se houve redução do número de acidentes envolvendo homens, como também aparenta a tabela. O tema foi levado para a turma, na forma de uma pergunta: quem sofre mais acidentes de bicicleta? Homens ou mulheres? Após uma breve discussão foi feito o teste.

O teste é realizado com os dados das colunas 2009 e 2010. A entrada de dados é feita na forma de vetor, neste caso o primeiro vetor corresponde ao número de acidentes do sexo avaliado em 2009 e em 2010, respectivamente, e o segundo contém os totais de acidentes para os dois sexos, em 2009 e 2010 respectivamente.



```
R Console
> #teste para comparar as proporções de acidentes (2009/2010)
> #envolvendo ciclistas mulheres
> prop.test(c(236,166), c(707,597))

      2-sample test for equality of proportions with continuity correction

data:  c(236, 166) out of c(707, 597)
X-squared = 4.4595, df = 1, p-value = 0.03471
alternative hypothesis: two.sided
95 percent confidence interval:
 0.004203325 0.107292390
sample estimates:
 prop 1    prop 2 
0.3338048 0.2780570

> #envolvendo ciclistas homens
> prop.test(c(471,417), c(707,597))

      2-sample test for equality of proportions with continuity
      correction

data:  c(471, 417) out of c(707, 597)
X-squared = 1.4091, df = 1, p-value = 0.2352
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.08447225 0.01987771
sample estimates:
 prop 1    prop 2 
0.6661952 0.6984925
```

Figura 1 - Teste de proporção para cada sexo

Foi possível observar que para as mulheres o p -valor é inferior a $\alpha=0.05$, ou seja, pode-se afirmar que houve redução no número de acidentes envolvendo mulheres. Em contrapartida, para os homens o p -valor é significativo, maior que α , considerando as proporções de acidentes envolvendo ciclistas do sexo masculino foram iguais nos anos 2009 e 2010.

Esse caso é um bom exemplo a ser levado à sala de aula para provar que é necessária a realização de testes estatísticos para comprovar ou negar o que se pode ver nas tabelas. Ou seja, esse exemplo mostra que é necessário realizar os testes estatísticos antes de afirmar o que é visto nas tabelas pois, apenas visualizando a tabela 1, diz-se que ambas as proporções diminuiriam, porém, após a realização do teste, ficou provado que a proporção de acidentes envolvendo homens se manteve a mesma.

Antes de encerrar, pode-se voltar à discussão do tema abrangendo outro meio de transporte, como o carro por exemplo. E, ao final, comenta-se com os alunos sobre a validade dos resultados e a importância das suposições e aleatoriedade da amostra.



4.2 Teste qui-quadrado

O teste qui-quadrado de aderência e associação permite avaliar se os dados possuem proporções diferentes ou não. Para a realização do teste qui-quadrado foram analisados os dias da semana que ocorreram os acidentes de bicicleta. Estes dados estão na tabela 2. A coluna total representa a soma dos dados de 2009 e 2010. Assim, procurou-se estabelecer se havia alguma associação entre o dia da semana e o número de acidentes. Do mesmo modo, inicia-se questionado os alunos se existe algum dia da semana em que ocorrem mais acidentes.

Tabela 2 - Dias da semana de ocorrência dos acidentes de bicicleta.

Dia da semana	Total	
	Quant.	%
Segunda-Feira	186	14,3
Terça-Feira	172	13,2
Quarta-Feira	183	14
Quinta-Feira	201	15,4
Sexta-Feira	242	18,6
Sábado	192	14,7
Domingo	128	9,8
Total Geral	1304	100

O teste realizado no software também é bastante simples. Inicialmente é criada uma variável que receberá, em forma de vetor, os valores da coluna "Quant.". Nesse caso, foi criada a variável "dias" que recebeu os valores de Segunda-Feira a domingo. Após a realização desta etapa, basta escrever `chisq.test(dias)` seguido de um "enter".

```
R Console
> #Teste qui-quadrado para os dias da semana dos acidentes
> dias=c(186, 172, 183, 201, 242, 192, 128)
> chisq.test(dias)

Chi-squared test for given probabilities

data: dias
X-squared = 37.3911, df = 6, p-value = 1.477e-06
```

Figura 2 - Teste qui-quadrado para os dias da semana dos acidentes

O p -valor é inferior a α , isso significa que a hipótese nula é rejeitada, considerando que há diferença no número de acidentes entre os dias da semana.



Com esse resultado e avaliando novamente a tabela, percebeu-se que os finais de semana possuíam um número menor de acidentes. Poder-se-ia pensar que os valores de Sábado e Domingo são os de maior influência sobre o p -valor, justificando a diferença entre os dias da semana. Pensando nisso, foi realizado outro teste qui-quadrado, da menos maneira que o primeiro, mas considerando apenas os dias úteis (de segunda a sexta-feira) e pode-se observar que o p -valor continuou inferior a α (p -valor=0,004387), confirmando haver diferença entre a proporção de acidentes.

As figuras 3 e 4 apresentam a forma que esse exemplo foi exposto aos alunos. A utilização de recursos multimídia possibilitou que um grande número de alunos tivesse acesso ao conteúdo, assim como a linguagem simples e metodológica facilitou a compreensão do assunto.

É possível selecionar os dados da maneira que for mais conveniente!

- Ou seja, verificamos que as proporções são realmente diferentes. Mas analisando a tabela podemos observar que nos finais de semana a quantidade de acidentes é menor.
- Será, então, que o resultado do teste qui-quadrado é devido aos números dos finais de semana?
- Podemos retirar os dados de Sábado e Domingo e trabalhar apenas entre os dias da semana.
 - A intenção ainda é verificar se a quantidade de acidente é diferente entre os dias da semana, porém, serão considerados apenas os dias úteis.

Figura 3 - Slide Introdutório do que será realizado



Dia da semana	Total	
	Quant.	%
Segunda – Feira	186	14,3
Terça – Feira	172	13,2
Quarta – Feira	183	14
Quinta – Feira	201	15,4
Sexta – Feira	242	18,6
Sábado	192	14,7
Domingo	128	9,8
Total Geral	1304	100

→ Vetor 'dias'

$H_0: p = p_0$ – a diferença **não** é significativa
 $H': p \neq p_0$ – a diferença é significativa

```
R Console
> #teste qui-quadrado para os dias úteis dos acidentes
> dias=c(186, 172, 183, 201, 242)
> chisq.test(dias)

Chi-squared test for given probabilities

data: dias
X-squared = 15.1565, df = 4, p-value = 0.004387
```

$p\text{-valor} < \alpha; \alpha = 0.05$
Rejeita H_0 em favor de H' .
Ou seja, as quantidades são diferentes

Figura 4 - Slide de apresentação dos resultados

Ou seja, apesar de aparentemente os finais de semana serem a causa das diferenças de proporção, pode-se notar, ao realizar novamente o teste, que a diferença existe ainda que sejam desconsideradas as variáveis com uma quantidade menos expressiva de acidentes.

Este exemplo foi escolhido para ser levado às aulas pois mostra a maneira que os dados podem ser tratados para fornecer as informações necessárias. É possível utilizar o mesmo conjunto de dados para obter informações diferentes. Ou seja, no primeiro caso foi possível visualizar que o número de acidentes envolvendo ciclistas é diferente entre os dias da semana. Já no segundo foram considerados apenas os dias úteis da semana.

4.3 Coeficiente de correlação de Pearson

O coeficiente de correlação permite avaliar se duas variáveis analisadas possuem alguma relação linear entre elas.

Os coeficientes de correlação podem variar entre -1 e 1, sendo que os valores próximos a -1 correspondem a uma correlação negativa (quando uma variável aumenta a outra diminui), valores próximos a 0 indicam que não há correlação e valores próximos a 1 correspondem a uma correlação positiva (quando uma variável aumenta a outra variável também aumenta).

O teste de correlação realizado no software R fornece também um p-valor. Esse p-valor fornece a significância do coeficiente de correlação encontrado. Ou seja, para p-valor >



α , a hipótese nula é aceita, considerando o coeficiente de correlação encontrado não significativo. Da mesma maneira, com um p -valor $< \alpha$, considera-se hipótese alternativa verdadeira, sendo o coeficiente de correlação significativo.

Para a realização do teste de correlação foi utilizado um pacote adicional do software R, chamado RCommander (FOX, 2005) Ao abrir o software, basta clicar no ícone "Packages", em seguida em "Install package(s)" (figura 5), selecionar o local e buscar o pacote "Rcmdr".

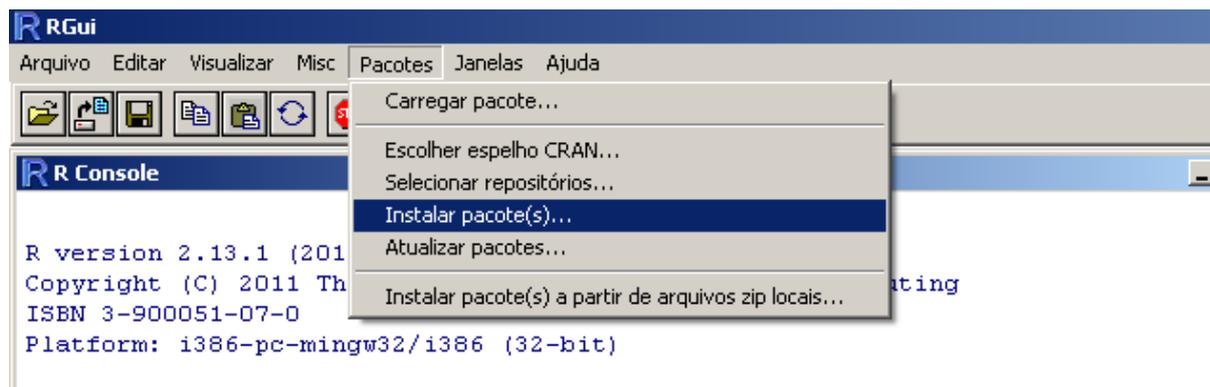


Figura 5 - Explicação para instalar o pacote adicional no software livre R

Após a instalação basta clicar novamente no ícone "Packages", "Load Packages" e, com um duplo clique, abrir o pacote "Rcmdr". Abrirá uma nova janela, onde é possível carregar os dados diretamente de um arquivo, sem necessidade de digitar os valores a serem avaliados.

Para o teste de correlação foi avaliado os dias de chuva e o número de acidentes. Os dados estão na tabela 3.



Tabela 3 - Acidentes e chuvas por mês

Ano	Mês	Acidentes	Total Mensal (mm)	Dias de chuva	Média mensal
2009	Janeiro	60	404	21	13,032
	Fevereiro	48	269,7	20	9,632
	Março	60	350,8	15	11,316
	Abril	56	81,6	13	2,720
	Maio	48	55,3	11	1,784
	Junho	45	77,1	10	2,570
	Julho	61	244,9	17	7,900
	Agosto	77	159,1	12	5,132
	Setembro	50	347,5	20	11,583
	Outubro	69	152,8	15	4,929
	Novembro	68	166,1	11	5,537
	Dezembro	65	166,1	14	5,358
2010	Janeiro	45	616,6	23	19,890
	Fevereiro	45	298	18	10,643
	Março	73	494,4	23	15,948
	Abril	73	258,8	11	8,627
	Maio	22	159,9	15	5,158
	Junho	46	141,5	10	4,717
	Julho	43	145,9	13	4,706
	Agosto	51	94,9	9	3,061
	Setembro	59	52,4	12	1,747
	Outubro	43	187,8	16	6,058
	Novembro	46	299,4	16	9,980
	Dezembro	51	248,4	20	8,013

O teste de correlação avaliou a relação entre o número de acidentes e o total mensal (mm) e obteve um coeficiente de correlação de aproximadamente 0.089, ou seja, uma correlação de 8,9%. O p-valor encontrado foi superior a α , considerando o coeficiente de correlação não significativo.

Foi avaliada também a relação entre acidentes e dias de chuva, o p-valor encontrado foi superior a α , considerando não significativo o coeficiente de correlação encontrado (-5%). O mesmo ocorreu para a relação entre acidentes e a média mensal de chuvas, o coeficiente de correlação encontrado (7,8%) não foi considerado significativo.

Além da realização dos testes para cada item relacionado individualmente com o número de acidentes, foi construída uma matriz de correlação e o gráfico de dispersão. Assim, os alunos conseguiram visualizar melhor a forma que os dados estão correlacionados. A matriz de correlação (figura 5) possibilitou avaliar todas as possíveis correlações e os seus coeficientes, assim como o gráfico de dispersão (figura 6) permitiu a compreensão visual da forma que os dados estão relacionados.



	Acidentes	Dias.de.chuva	Média.mensal	Total.Mensal..mm.
Acidentes	1.00000000	-0.05236989	0.07769129	0.08894862
Dias.de.chuva	-0.05236989	1.00000000	0.84295694	0.83859358
Média.mensal	0.07769129	0.84295694	1.00000000	0.99816585
Total.Mensal..mm.	0.08894862	0.83859358	0.99816585	1.00000000

Figura 6 - Resultados da Matriz de correlação

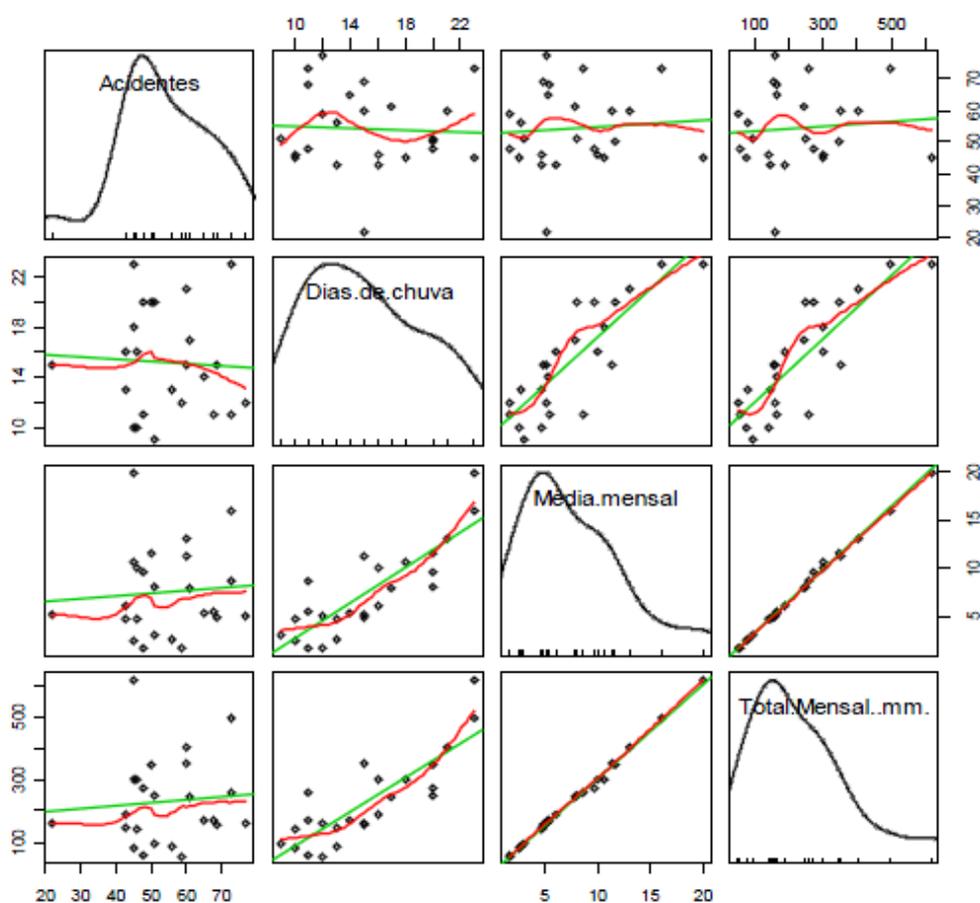


Figura 7 - Gráfico de Dispersão

Os testes de correlação para a chuva foram interessantes para mostrar aos alunos a aplicação e utilização dos testes estatísticos. Para esse caso, foi possível mostrar claramente a necessidade de conhecimento das ferramentas estatísticas para desmistificar alguns pré-conceitos criados, como o aumento do número de acidentes em dias de chuva. Encerra-se a aula com uma questão: em dias de chuva há menos pessoas andando de bicicleta?

5. CONCLUSÕES E CONSIDERAÇÕES FINAIS

A disponibilidade de dados condizentes com a realidade do aluno permitiu que se criasse uma discussão a respeito do assunto, tornando os graduandos participativos nas aulas, o que sugere um interesse pela disciplina que está sendo ministrada.



Além disso, sem as contas exaustivas, os alunos puderam manter o foco na análise dos resultados, priorizando a importância destes. A visualização do p-valor com o auxílio dos softwares poupou tempo e proporcionou mais dedicação à compreensão da utilização das ferramentas.

Por se tratar de um software livre, todos os alunos têm acesso e podem fazer uso do programa em seus computadores pessoais. Sendo assim, são passíveis de serem avaliados por meio de trabalhos realizados através desse software, pois é uma ferramenta que visa o conhecimento da disciplina e interpretação de dados, não somente conhecimento de fórmulas.

A estatística é uma disciplina essencial para os cursos de engenharia e, por isso, deve ser ensinada de forma clara, simples e objetiva. Com esse propósito, as aulas com exemplos práticos e com a utilização do software permitiram que o aluno realmente entendesse o conteúdo, afastando-o da necessidade de decorar fórmulas e proporcionando o aprendizado efetivo da disciplina.

Esta iniciativa surgiu do trabalho em conjunto de duas acadêmicas de um curso de Engenharia de Produção. Uma era bolsista de iniciação científica trabalhando na pesquisa sobre mobilidade urbana, enquanto que a outra era monitora da disciplina de Probabilidade e Estatística. Dividindo o mesmo espaço, e assim trocando algumas ideias, nasceu esta proposta, que foi prontamente aceita pela professora da disciplina. Embora este trabalho ainda esteja numa fase embrionária, será possível ampliá-lo de modo a utilizar outros resultados da pesquisa de mobilidade na sala de aula, em outros momentos da disciplina.

Ressalta-se que a temática de mobilidade urbana, e todos os problemas subjacentes a ela incentivam a participação do aluno nas discussões, sendo positiva neste processo de ensino-aprendizagem. Congestionamento, acidentes, transporte público e a (falta) de infraestrutura e política adequada para incentivar meios de transporte com maior sustentabilidade são problemas reais e que fazem parte do dia-a-dia dos acadêmicos.

6. REFERÊNCIAS

FOX, J. The R Commander: A Basic Statistics Graphical User Interface to R. *Journal of Statistical Software*. **Journal of Statistical Software**. v. 14, n 9, 2005.

HACKENBERG, A. M. et al. Transporte por Bicicleta em Cidades Catarinenses: Metodologia para levantamento da realidade e recomendações para incremento da sua participação na mobilidade urbana. Joinville, v.1, 2011.

MONTGOMERY, D. C., RUNGER, G. C., HUBELE, N. F. Estatística aplicada à engenharia. 2ed. Rio de Janeiro: LTC, 2011. 336 p.

R DEVELOPMENT CORE TEAM (2012). **R: A language and environment for statistical computing**. Disponível em: < <http://www.R-project.org> > Acesso em: 20 mai. 2012.



REAL DATA IN THE CLASSROOM: AN ACTIVITY FOR TEACHING STATISTICS IN ENGINEERING COURSES

Abstract: *Statistics is an important subject for anyone who deals with numbers. All undergraduate engineering student learn statistics - and our proposal is to facilitate this learning. By using software to calculate the statistics, we have the opportunity to evaluate only the results, providing a better understanding of the essence of statistics. In this paper, we suggest using R, a free software environment for statistical computing, which, being open source makes it easily accessible to all. Furthermore, is shown an experiment conducted in the classroom, using actual data from a survey on urban mobility, with data from accidents involving cyclists in the city of Joinville - SC, between 2008 and 2010. Statistical test are easily implemented by commands contained in R, and the results are suitable for a correct interpretation. Thus, students can add emphasis to the results, since they do not need to use class time for calculations. And by covering a real situation, the students are encouraged to participate actively in this learning process.*

Key-words: *Statistics, Learning, Software*