SỬ DỤNG PHƯƠNG PHÁP BOOTSTRAP TRONG VIỆC PHÂN TÍCH CÁC YẾU TỐ ẢNH HƯỞNG ĐẾN KẾT QUẢ HỌC TẬP CỦA SINH VIÊN SƯ PHẠM KỸ THUẬT TP.HCM

USING BOOTSTRAP METHOD IN ANALYZING THE FACTORS AFFECTING STUDY RESULT OF HCMCUTE STUDENT

Nguyễn Hồng Nhung
Trường Đại học Sư phạm Kỹ thuật TP. Hồ Chí Minh

Ngày tòa soạn nhận bài 16/06/2015, ngày phản biện đánh giá 20/11/2015, ngày chấp nhận đăng 03/12/2015

TÓM TẮT

Bootstrap là một phương pháp Thống kê hiện đại đã và đang được sử dụng phổ biến vì phương pháp này không đòi hỏi phân phối xác suất của các tham số thống kê cần nghiên cứu như phương pháp thống kê truyền thống. Bài báo tập trung trình bày nội dung phương pháp bootstrap và sử dụng phương pháp này phân tích các yếu tố ảnh hưởng đến kết quả học tập của sinh viên Trường Đại học Sư Phạm Kỹ thuật Tp. HCM. Từ đó đưa ra các kết luận sau quá trình phân tích số liệu.

Từ khóa: *Phương pháp bootstrap; khoảng ước lượng bootstrap-t; kiểm định bootstrap.*

ABSTRACT

Bootstrap is a modern statistical method and has been used widely because it doesn't require emperical distribution function as classcical methods do. This article will focus on demonstration of bootstrap method and the use of this approach to analyze the factors which affected on the study results of students of HCMC University of Technology and Education. Then conclusion shall be stated basing on data analysis process.

Key words: *bootstrap method; the bootstrap-t interval; hypothesis testing with the bootstrap.*

1. GIỚI THIỆU

Trong các bài toán Thống kê truyền thống, ta cần có thông tin ban đầu về phân phối F của biến ngẫu nhiên X đặc trưng cho tính chất cần nghiên cứu. Tuy nhiên trong thực tế không phải lúc nào chúng ta cũng có được những thông tin này. Trong thống kê hiện đại, bootstrap là phương pháp giải quyết được các bất định của bài toán thống kê khi cỡ mẫu nhỏ mà không cần giả thuyết về phân phối xác suất của tham số thống kê cần nghiên cứu. Vì vậy, phương pháp bootstrap phù hợp cho việc phân tích số liệu trong nhiều ngành khoa học trong đó có khoa học giáo dục.

Bài báo này trình bày các kết quả thu được từ quá trình sử dụng phương pháp bootstrap phân tích các yếu tố ảnh hưởng đến kết quả học tập của sinh viên Trường Đại học Sư phạm Kỹ thuật Tp.HCM. Số liệu được thu thập ngẫu nhiên từ 888 sinh viên Trường Đại học Sư pHạm Kỹ thuật Tp.HCM dưới hình thức phiếu khảo sát hoặc trả lời bảng hỏi trên Link: https://goo.gl/FkFJPh trong khoảng thời gian từ tháng 5 đến tháng 6 năm 2015.

2. PHƯƠNG PHÁP BOOTSTRAP

2.1 Nội dung phương pháp

Phương pháp bootstrap là phương pháp coi mẫu gốc ban đầu đóng vai trò tổng thể mà từ đó nó được rút ra. Từ mẫu ban đầu lấy lại các mẫu ngẫu nhiên cùng cỡ với mẫu gốc bằng phương pháp lấy mẫu có hoàn lại. Mẫu lấy lại gọi là mẫu bootstrap. Thống kê của tham số cần quan tâm ứng với mỗi mẫu bootstrap được gọi là thống kê bootstrap. Phân phối của thống kê bootstrap được gọi là phân phối bootstrap.

2.2 Khoảng ước lượng Bootstrap-t

Mục đích của thống kê học là đưa ra thông tin về các tham số của tổng thể nghiên cứu. Trong thực tế ta không biết chính xác các tham số này mà chỉ ước tính dựa vào một hay nhiều mẫu rút ra từ tổng thể. Chọn m mẫu thực nghiệm $(X_{k1},X_{k2},...X_{kn})$ có m giá trị $\hat{\theta}_k$ $=(X_{k1},X_{k2},...X_{kn})$, (k=1, 2, 3, …, m). Độ lệch tiêu chuẩn của m giá trị $\hat{\theta}_k$ gọi là sai số tiêu chuẩn của $\hat{\theta}$, ký hiệu

$$se(\hat{\theta}) = \sqrt{\frac{1}{m-1}\sum_{k=1}^{m}\left(\hat{\theta}_k - \frac{1}{m}\sum_{k=1}^{m}\hat{\theta}_k\right)^2}$$

Khi không thể lấy nhiều mẫu từ tổng thể để xác định sai số tiêu chuẩn của , ta dùng phương pháp bootstrap ước lượng sai số tiêu chuẩn của từ một mẫu gốc ban đầu. Các bước thực hiện như sau:

Bước 1: Lấy theo phương pháp có hoàn lại từ mẫu gốc ban đầu được b mẫu bootstrap độc lập cùng cỡ với mẫu gốc là $X_k^* = (X_{k1}^*, X_{k2}^*,...X_{kn}^*)$, k=1, 2, …, b.

Bước 2: Với mỗi mẫu bootstrap có được ở bước 1, ta tính giá trị thống kê

$\hat{\theta}_k^* = T(x_k^*) = T(x_{k1}^*, x_{k2}^*, \dots, x_{kn}^*)$, k=1,2,..., b.

Bước 3: Tính độ lệch tiêu chuẩn $se_b^*(\hat{\theta}^*)$ của b giá trị tính được ở bước 2. Độ lệch tiêu chuẩn $se_b^*(\hat{\theta}^*)$ là ước lượng bootstrap của sai số tiêu chuẩn $se(\hat{\theta})$.

Với mỗi mẫu bootstrap $x_1^*, x_2^*, \dots, x_b^*$ ta tính được một giá trị

$$Z_k^* = \frac{\hat{\theta}_k^* - \hat{\theta}}{se_{b_k}^*(\hat{\theta}^*)}; \text{ với } k = 1, 2, \dots, b.$$

Trong đó $se_{b_k}^*(\hat{\mu}^*)$ là sai số tiêu chuẩn cho giá trị trung bình của mẫu bootstrap thứ k; k=1, 2, …, b.

Ký hiệu giá trị phân vị mức của các giá trị bootstrap Z_k^* là $\hat{t}(\alpha)$ thỏa mãn.

$$\frac{\#\{Z_k^* \le \hat{t}(\alpha)\}}{b} = \alpha.$$

A là số phần tử của tập hợp.

Khi đó ta có khoảng ước lượng bootstrap-t với độ tin cậy 1-2a của tham số q có dạng:

$\left(\hat{\theta} - \hat{t}(1-\alpha)se_b^*(\hat{\theta}^*);\ \hat{\theta} - \hat{t}(\alpha)se_b^*(\hat{\theta}^*)\right)$. [2].

Có nhiều phần mềm có thể hỗ trợ cho việc tính toán khi sử dụng phương pháp phân tích Bootstrap như Matlab, R, Excel. [3], [4], [5].

Trong bài báo này tác giả sử dụng phần mềm Matlab để phân tích số liệu. Sau đây là thuật toán tìm khoảng ước lượng bootstrap-t với độ tin cậy 95% cho tham số tỷ lệ p.

```
>> n = length(data); % cỡ mẫu gốc.
>>b = 1000; % số lần lấy mẫu bootstrap.
>> f = mean(data) % tỷ lệ mẫu gốc
>> [bootreps, bootsam] = bootstrp(b,'mean', data);
>> sebk=zeros(size(bootreps));
>> for i = 1:b
xstark= data (bootsam(:,i));
sebk(i) = std(bootstrp(1000, 'mean', xstark));
zvals(i) = (bootreps(i) - f)/sebk(i); end
>> szval = sort(zvals);
>> gamma=0.95; % độ tin cậy
>> k=b*(1-gamma)/2;
>> tlo=szval(round(k));
>> thi=szval(round(b-k));
>> seb=std(bootreps);
>> blo = f - thi*seb % cận dưới của khoảng ước lượng
>> bhi=f - tlo*seb % cận trên của khoảng ước lượng.
```

2.3 Kiểm định Bootstrap

Một kiểm định thống kê dựa trên kiểm định T để đo sự khác biệt giữa số liệu và giả thuyết H_0. Giả sử giả thuyết H_0 là đúng. Ký hiệu giá trị quan sát được của kiểm định T là t. Ta có P giá trị được xác định theo công thức:

$$p = P(T>=t|H_0) \quad (1)$$

Tương ứng với mức ý nghĩa cho trước ta có giá trị t_a là giá trị xác định bởi phương trình $P(T>=t_\alpha|H_0)=a$. Khi đó ta sẽ bác bỏ giả thuyết H_0 với mức ý nghĩa nếu $t>=t_a$.

Giả sử giả thuyết H_0 là đúng ta xấp xỉ phân phối F bởi $\hat{F}_0$, ta tính được

$$p = P(T \ge t|\hat{F}_0) \quad (2)$$

Nếu ta không tính được chính xác (2) thì sử dụng phương pháp bootstrap ta xác định được giá trị p_{boot} xấp xỉ cho p giá trị.

Lấy theo phương pháp có hoàn lại từ mẫu gốc ban đầu được b mẫu bootstrap độc lập cùng cỡ với mẫu gốc , k=1, 2, …, b. Với mỗi mẫu bootstrap ta tính được một giá trị quan sát được của kiểm định T là . Khi đó ta có p giá trị được xấp xỉ bởi:

$$p_{boot} = \frac{1+\#\{t_k^* \geq t\}}{b+1} \quad (3)$$

[1]

Sau đây là thuật toán tìm giá trị p_{boot} cho bài toán kiểm định giả thuyết so sánh hai trung bình của hai tổng thể, H: .

```
>> xbar1=mean(data1);
>> xbar2=mean(data2);
>> n1=length(data1);% cỡ mẫu 1
>> n2=length(data2);% cỡ mẫu 2
>>zbar=(1/(n1+n2))*(sum(data1)+
sum(data2));
>> nsam1=data1-xbar1+zbar;
>> nsam2=data2-xbar2+zbar;
>> b=1000; % số lượng mẫu bootstrap
>> inds1= unidrnd(n1,n1,b);
>> xboot1 = nsam1(inds1);
>> inds2= unidrnd(n2,n2,b);
>> xboot2 = nsam2(inds2);
>> Tstar=zeros(b,1);
>>  for i=1:b
      xstar1=mean(xboot1(:,i));
      xstar2=mean(xboot2(:,i));
      sigmastar1=std(xboot1(:,i));
      sigmastar2=std(xboot2(:,i));
      Tstar(i)= abs( (xstar1 - xstar2)/
 sqrt(sigmastar1^2/n1+sigmastar2^2/
 n2));
 end
>>tval=abs((xbar1-xbar2)/
sqrt(std(data1)^2/n1+std(data2)^2/n2));
>>d=0;
 >>  for i=1:b
      if  Tstar(i)>= tval,d=d+1;end
 end
 >> Pval=(d+1)/(b+1) % P giá trị
```

3. KẾT QUẢ PHÂN TÍCH

Sử dụng các thuật toán trên phân tích số liệu thu được, ta có kết quả trình bày trong các bảng từ bảng 1 đến bảng 4 như sau.

***Bảng 1**. Khoảng ước lượng bootstrap-t với độ tin cậy 95% cho tỷ lệ các yếu tố ảnh hưởng đến chất lượng học tập của sinh viên.*

Tính chất A	Cỡ mẫu và tỷ lệ mẫu gốc	Khoảng ước lượng bootstrap-t cho tỷ lệ xuất hiện tính chất A.
Sinh viên chọn ngành học do định hướng nghề nghiệp, đam mê yêu thích	$n = 887; f = \frac{671}{887}$	(0.7267; 0.7868)
Sinh viên cho rằng chương trình học đáp ứng yêu cầu xã hội	$n = 697; f = \frac{578}{697}$	(0.7966; 0.8598)
Sinh viên cho rằng chương trình học hấp dẫn, lý thú	$n =; f = \frac{192}{312}$	(0.5606; 0.6698)
Sinh viên sẽ giới thiệu trường cho người thân, bạn bè	$n = 888; f = \frac{645}{888}$	(0.6943; 0.7538)
Sinh viên cho rằng sĩ số lớp đông	$n = 888; f = \frac{233}{888}$	(0.2336; 0.2922)
Sinh viên cho rằng tài liệu học tập trên thư viện đầy đủ, phong phú	$n = 888; f = \frac{632}{888}$	(0.6794; 0.7410)
Sinh viên cho rằng điều kiện học tập tại các phòng thí nghiệm, thực hành tốt	$n = 888; f = \frac{779}{887}$	(0.8519; 0.8982)
Sinh viên cho rằng dụng cụ, sân tập thể dục thể thao tốt	$n = 887; f = \frac{756}{887}$	(0.8250; 0.8760)

Sinh viên cảm thấy dễ dàng, thuận lợi khi sắp xếp lịch học	$n = 887; f = \frac{649}{887}$	(0.7021; 0.7593)
Sinh viên đánh giá giáo viên dạy tốt, dễ hiểu, quan tâm đến sự tiến bộ của sinh viên	$n = 867; f = \frac{791}{867}$	(0.8925; 0.9300)
Sinh viên thường xuyên học tại các điểm tự học	$n = 888; f = \frac{98}{887}$	(0.0913; 0.1323)
Sinh viên thường xuyên lên thư viện tìm tài liệu tham khảo	$n = 886; f = \frac{54}{886}$	(0.0464; 0.0789)
Sinh viên đi học đầy đủ, ít khi vắng	$n = 861; f = \frac{824}{861}$	(0.9413; 0.9704)
Sinh viên đi học vì giáo viên điểm danh	$n = 888; f = \frac{213}{888}$	(0.2112; 0.2706)
Sinh viên thường xuyên tham gia hoạt động ngoại khóa, câu lạc bộ, thiện nguyện, tình nguyện,…	$n = 888; f = \frac{118}{887}$	(0.1118; 0.1546)
Sinh viên thường xuyên tham dự các hội thảo về học tập, nghiên cứu khoa học, nghề nghiệp,…	$n = 887; f = \frac{101}{887}$	(0.0950; 0.1359)
Sinh viên ở cùng gia đình	$n = 888; f = \frac{131}{888}$	(0.1236; 0.1721)
Sinh viên thường xuyên đi làm	$n = 888; f = \frac{105}{888}$	(0.0958; 0.1422)
Sinh viên có máy tính, điện thoại thông minh	$n = 888; f = \frac{810}{888}$	(0.8877; 0.9348)
Sinh viên thường xuyên lên mạng tìm tài liệu tham khảo	$n = 888; f = \frac{432}{888}$	(0.4508; 0.5238)
Sinh viên khi rảnh rỗi chơi game, nhắn tin, lên face	$n = 888; f = \frac{640}{888}$	(0.6985; 0.7434)
Sinh viên cảm thấy khó khăn vì thiếu thời gian học	$n = 888; f = \frac{310}{888}$	(0.3276; 0.3723)
Sinh viên cảm thấy khó khăn vì học không hiểu	$n = 888; f = \frac{533}{887}$	(0.5774; 0.6224)
Sinh viên cảm thấy không thích học	$n = 888; f = \frac{166}{888}$	(0.1652; 0.2097)

Phần lớn sinh viên cảm thấy hài lòng với điều kiện học tập tại nhà trường thể hiện bởi tỷ lệ sinh viên cho rằng tài liệu thư viện phong phú, đầy đủ; tỷ lệ sinh viên cảm thấy dễ dàng, thuận lợi khi sắp xếp lịch học khá cao khoảng 70%. Tỷ lệ sinh viên cho rằng điều kiện học tập tại các phòng thí nghiệm, thực hành tốt; giáo viên dạy tốt, dạy hay dễ hiểu, quan tâm đến sự tiến bộ của sinh viên là cao, trên 80% với độ tin cậy 95%.

Số liệu cho thấy, tỷ lệ sinh viên thường xuyên lên thư viện tìm tài liệu tham khảo thấp chỉ từ 4.64% đến 7.89%. Mặt khác, tỷ lệ sinh viên thường xuyên lên mạng tìm tài liệu tham khảo lại từ 45.69% tới 50.31%. Qua đó thể hiện xu thế mới trong việc tìm tài liệu tham khảo của sinh viên. Khoảng tin cậy 95% cho tỷ lệ sinh viên có máy tính và điện thoại thông minh là (88.77%; 93.48%). Đây là điều kiện thuận lợi cho sinh viên tiếp cận với nguồn tài liệu số vô cùng phong phú trên mạng. Tuy nhiên, máy tính và điện thoại thông minh cũng là công cụ giải trí của sinh viên với tỷ lệ sinh viên khi rảnh rỗi chơi game, nhắn tin, lên face thuộc khoảng (69.85%; 74.34%).

***Bảng 2**. Kiểm định giả thuyết H: $p_1=p_2$ với mức ý nghĩa 5%, kiểm định yếu tố tác động đến khó khăn trong học tập của sinh viên: thiếu thời gian học, học không hiểu, không thích học.*

p_1	p_2	Cỡ mẫu và tỷ lệ hai mẫu gốc	p_{boot}	Kết luận
Tỷ lệ sinh viên nữ cảm thấy không thích học	Tỷ lệ sinh viên nam cảm thấy không thích học	$n_1 = 178;\ f_1 = \frac{29}{178};$ $n_2 = 710;\ f_2 = \frac{137}{710}$	0.3356	$p_1=p_2$
Tỷ lệ sinh viên nữ cảm thấy khó khăn vì học không hiểu	Tỷ lệ sinh viên nam cảm thấy khó khăn vì học không hiểu	$n_1 = 178;\ f_1 = \frac{106}{178};$ $n_2 = 710;\ f_2 = \frac{427}{710}$	0.8861	$p_1=p_2$
Tỷ lệ sinh viên nữ cảm thấy khó khăn vì thiếu thời gian học	Tỷ lệ sinh viên nam cảm thấy khó khăn vì thiếu thời gian học	$n_1 = 178;\ f_1 = \frac{74}{178};$ $n_2 = 710; f_2 = \frac{236}{710}$	0.0470	$p_1>p_2$
Tỷ lệ sinh viên năm 1, 2 cảm thấy không thích học	Tỷ lệ sinh viên năm 3 trở lên cảm thấy không thích học	$n_1 = 548;\ f_1 = \frac{102}{192};$ $n_2 = 340;\ f_2 = \frac{64}{340}$	0.9361	$p_1=p_2$
Tỷ lệ sinh viên năm 1, 2 cảm thấy khó khăn vì học không hiểu	Tỷ lệ sinh viên năm 3 trở lên cảm thấy khó khăn vì học không hiểu	$n_1 = 548;\ f_1 = \frac{350}{548};$ $n_2 = 340;\ f_2 = \frac{183}{340}$	0.0070	$p_1>p_2$
Tỷ lệ sinh viên năm 1, 2 cảm thấy khó khăn vì thiếu thời gian học	Tỷ lệ sinh viên năm 3 trở lên cảm thấy khó khăn vì thiếu thời gian học	$n_1 = 548;\ f_1 = \frac{194}{548};$ $n_2 = 340; f_2 = \frac{116}{340}$	0.7123	$p_1=p_2$
Tỷ lệ sinh viên không thích học trong số các sinh viên bắt buộc phải theo học ngành mình đang học.	Tỷ lệ sinh viên không thích học trong số các sinh viên học ngành học do đam mê, định hướng nghề nghiệp.	$n_1 = 216; f_1 = \frac{59}{216};$ $n_2 = 671; f_2 = \frac{107}{671}$	$9.99x10^{-4}$	$p_1>p_2$

Như vậy, sinh viên nữ cảm thấy thiếu thời gian học hơn sinh viên nam; sinh viên năm 1, 2 cảm thấy khó khăn vì học không hiểu nhiều hơn so với sinh viên năm 3 trở lên. Lý do chọn ngành học là yếu tố ảnh hưởng đến hứng thú học tập của sinh viên.

***Bảng 3.** Kiểm định giả thuyết H: với mức ý nghĩa 5%, kiểm định các yếu tố ảnh hưởng đến điểm học tập trung bình của sinh viên.*

μ_1	μ_2	Cỡ mẫu và trung bình, sai số chuẩn hai mẫu gốc	p_{boot}	Kết luận
Điểm trung bình của sinh viên nữ	Điểm trung bình của sinh viên nam	$n_1 = 178;\ \bar{x}_1 = 7.0370;$ $se(x_1) = 0.7498;$ $n_2 = 710;\ \bar{x}_2 = 6.5106;$ $se(x_2) = 0.8325$	$9.99x10^{-4}$	$\mu_1 > \mu_2$
Điểm trung bình của sinh viên khoa chất lượng cao	Điểm trung bình của sinh viên không thuộc khoa chất lượng cao	$n_1 = 205;\ \bar{x}_1 = 6.6217;$ $se(x_1) = 0.8779$ $n_2 = 683;\ \bar{x}_2 = 6.6145;$ $se(x_2) = 0.8329$	0.9301	$\mu_1 = \mu_2$

Điểm trung bình của sinh viên đi làm thường xuyên	Điểm trung bình của sinh viên không đi làm thường xuyên	$n_1 = 105;\ \bar{x}_1 = 6.5823;$ $se(x_1) = 0.7887;$ $n_2 = 783;\ \bar{x}_2 = 6.6207;$ $se(x_2) = 0.8504$	0.6424	$\mu_1 = \mu_2$
Điểm trung bình của sinh viên có phương pháp học chủ yếu là tự học	Điểm trung bình của sinh viên không có phương pháp học chủ yếu là tự học	$n_1 = 525;\ \bar{x}_1 = 6.7846;$ $se(x_1) = 0.8313;$ $n_2 = 363;\ \bar{x}_2 = 6.3726;$ $se(x_2) = 0.8001$	$9.99x10^{-4}$	$\mu_1 > \mu_2$
Điểm trung bình của sinh viên có phương pháp học chủ yếu là tham khảo tài liệu	Điểm trung bình của sinh viên không có phương pháp học chủ yếu là tham khảo tài liệu	$n_1 = 637;\ \bar{x}_1 = 6.6708;$ $se(x_1) = 0.8183;$ $n_2 = 251;\ \bar{x}_2 = 6.4773;$ $se(x_2) = 0.8893$	0.0070	$\mu_1 > \mu_2$
Điểm trung bình của sinh viên có phương pháp học chủ yếu là nghe giảng	Điểm trung bình của sinh viên không có phương pháp học chủ yếu là nghe giảng	$n_1 = 616;\ \bar{x}_1 = 6.6720;$ $se(x_1) = 0.8481$ $n_2 = 272;\ \bar{x}_2 = 6.4896;$ $se(x_2) = 0.8187$	0.0030	$\mu_1 > \mu_2$

Ta thấy điểm trung bình của sinh viên nữ cao hơn sinh viên nam. Các sinh viên có phương pháp học tập chủ yếu là tự học, tham khảo tài liệu và nghe giảng có điểm trung bình học tập cao hơn các sinh viên khác. Sinh viên thuộc khoa chất lượng cao hay không thì không ảnh hưởng đến điểm trung bình học tập của họ. Yếu tố đi làm thường xuyên không tác động đến kết quả học tập của sinh viên.

Bảng 4. *Kiểm định giả thuyết H: $p_1=p_2$ với mức ý nghĩa 5%, kiểm định các yếu tố ảnh hưởng đến việc học lại nhiều của sinh viên.*

p_1	**p_2**	**Cỡ mẫu và tỷ lệ hai mẫu gốc**	**p_{boot}**	**Kết luận**
Tỷ lệ sinh viên nữ học lại nhiều	Tỷ lệ sinh viên nam học lại nhiều	$n_1 = 178;\ f_1 = \frac{8}{178};$ $n_2 = 710;\ f_2 = \frac{92}{710}$	$9.99x10^{-4}$	$p_1<p_2$
Tỷ lệ sinh viên đi làm thường xuyên học lại nhiều	Tỷ lệ sinh viên không đi làm thường xuyên học lại nhiều	$n_1 = 105;\ f_1 = \frac{16}{105};$ $n_2 = 783;\ f_2 = \frac{84}{783}$	0.2488	$p_1=p_2$
Tỷ lệ sinh viên chất lượng cao học lại nhiều	Tỷ lệ sinh viên không thuộc khoa chất lượng cao học lại nhiều	$n_1 = 205;\ f_1 = \frac{19}{205};$ $n_2 = 683;\ f_2 = \frac{81}{683}$	0.2817	$p_1=p_2$
Tỷ lệ sinh viên lúc rảnh đi chơi với bạn học lại nhiều	Tỷ lệ sinh viên lúc rảnh không đi chơi với bạn học lại nhiều	$n_1 = 206;\ f_1 = \frac{35}{206};$ $n_2 = 682;\ f_2 = \frac{65}{682}$	0.0060	$p_1>p_2$

Có thể thấy rằng tỷ lệ sinh viên nữ phải học lại nhiều ít hơn sinh viên nam. Các yếu tố về việc sinh viên đi làm thường xuyên, sinh viên thuộc khoa chất lượng cao không tác động đến việc học lại nhiều của sinh viên. Tỷ lệ sinh viên thời gian rảnh đi chơi với bạn thì học lại nhiều cao hơn các sinh viên khác.

4. KẾT LUẬN

Tác giả đã thực hiện được công việc thu thập số liệu mẫu và sử dụng phương pháp

bootstrap viết thuật toán phân tích số liệu thu được. Các kết luận thu được từ quá trình phân tích là những thông tin tham khảo hữu ích giúp các thầy cô giáo cũng như các phòng ban trong Trường Đại học Sư phạm Kỹ thuật Tp.HCM có được những nhận định về các yếu tố ảnh hưởng đến kết quả học tập của sinh viên. Từ đó có thể đưa ra những giải pháp nhằm nâng cao hơn nữa chất lượng học tập của sinh viên trong trường.

TÀI LIỆU THAM KHẢO

[1] A.C. Davidson and D.V.Hinkley. *Bootstrap methods and their Application.* Cambridge Series in Statistical and Probabilistic Mathematics, page141, (1997).

[2] Bradley Efron. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, Inc., Publication, page 160. (1994).

[3] Michael R. Chernick, Robert A. LaBudde. *An introduction to bootstrap methods with applications to R* . Wiley, New Jersey. (2011).

[4] Rochowicz, John A. Jr. *Bootstrapping Analysis, Inferential Statistics and EXCEL*. Spreadsheets in Education (eJSiE): Vol. 4: Iss.3, Article 4. (2011).

[5] Wendy L. Martinez, Angel R. Martinez. *Computational Statistic Handbook with Matlab*. Chapman&Hall/CLC Boca Raton London New York Washington, D.C (2002).