

## Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2012

### Các Phương Pháp Phân Tích Định Lượng

#### LỜI GIẢI ĐỀ NGHỊ BÀI TẬP 5

#### KIỂM ĐỊNH THỐNG KÊ

**Ngày Phát: Thứ hai 29/10/2012**

**Ngày Nộp: Thứ hai 05/11/2012**

Bản in nộp lúc **8h20 sáng**, tại Hộp nộp bài tập trong phòng Lab

Bản điện tử gửi lên <http://intranet.fetp.edu.vn:81>

---

#### **Bài 1: (25 điểm)**

Trong một hội nghị về “Bình đẳng giới” chủ tịch Liên đoàn Lao động quốc gia Hồng Lạc phát biểu rằng: “Hồng Lạc đã thực hiện bình đẳng giới trong tuyển dụng lao động, thể hiện qua mức lương khởi điểm (tính bằng HLD) của nam giới và nữ giới không còn sự khác biệt”. Để kiểm chứng điều này, Lê Dưa, phóng viên của báo Thông Tấn Thôn, đã tiến hành một cuộc điều tra dựa trên một mẫu ngẫu nhiên bao gồm 258 lao động nam và 216 lao động nữ. Dữ liệu điều tra được trình bày trong cột Salbegin của file dữ liệu **Employee Data.xls**. Với vai trò là chuyên viên thống kê, bạn hãy giúp Lê Dưa:

- Lập các giả thuyết hợp lý.
- Với mức ý nghĩa 5%, tính các thông số cần thiết và tiến hành kiểm định giả thuyết nêu trên.
- Lê Dưa và thư ký tòa soạn không biết nhiều về các khái niệm thống kê như phân phối xác suất, giá trị tới hạn, ước lượng khoảng/điểm... Nhưng họ quen thuộc với khái niệm xác suất. Bạn hãy tìm giá trị  $P_{value}$  và giải thích cho họ hiểu ý nghĩa của con số này.
- Theo bạn ở mức ý nghĩa đã chọn, phát biểu của chủ tịch Liên đoàn lao động có đáng tin cậy không?

#### **Đáp án bài 1:**

a. Thông thường, việc bình đẳng giới trong lĩnh vực lao động rất khó xác định. Sở dĩ, khó xác định vì đặc trưng sinh học của nam giới và nữ giới khác nhau nên năng suất của từng giới cũng khác nhau. Khi tuyển dụng lao động, mức lương sẽ chi trả trên cơ sở tiền lương tối thiểu và năng suất lao động. Trong trường hợp này, chúng ta nghi ngờ phát biểu “lương khởi điểm của nam giới và nữ giới không còn sự khác biệt”. Cụ thể hơn, với suy luận về năng suất lao động, có thể lương khởi điểm của nữ giới thấp hơn lương khởi điểm của nam giới. Do đó các giả thuyết hợp lý được đặt ra như ở dưới đây:

$$H_0: (\mu_{nữ} - \mu_{nam}) = 0$$

$$H_a: (\mu_{nữ} - \mu_{nam}) < 0$$

Trong đó :  $\mu_{nữ}$ : lương khởi điểm trung bình của nữ giới

$\mu_{nam}$ : lương khởi điểm trung bình của nam giới

b. Dựa vào bảng số liệu đã cho chúng ta có

- Đối với nữ giới:  $\mu_{nữ} = 13091,97$ ;  $s = 2935,59$ ;  $n_1 = 216$

- Đối với nam giới:  $\mu_{nam} = 20301,4$ ;  $s = 9111,78$ ;  $n_2 = 258$

$$\text{Giá trị kiểm định } t = \frac{(\mu_{nu} - \mu_{nam}) - 0}{\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}} = -11,98$$

- Tra bảng với  $t = -11,98$ , bậc tự do  $= n_1 + n_2 - 2 = 472$  ta có  $P_{\text{value}} = 0.0000$

c. Giá trị  $P_{\text{value}} = 0.0000$  là xác suất sai lầm lớn nhất khi ta tiến hành bác bỏ  $H_0$ . Nói cách khác: dựa trên tập dữ liệu đã cho, khi bác bỏ giả thuyết  $H_0$  thì xác suất sai lầm lớn nhất là  $P_{\text{value}}$ . Đại lượng này là chỉ số để đo lường độ phù hợp của dữ liệu với giả thuyết phát biểu. Cụ thể: Khi chúng ta dựa vào một tập dữ liệu để làm cơ sở cho một phát biểu, rồi dùng chính tập dữ liệu đó để kiểm định lại (bác bỏ) phát biểu nói trên thì  $P_{\text{value}} = 100\%$ .

d. Từ các tính toán ở trên ta thấy rằng khi bác bỏ  $H_0$  thì xác suất gặp sai lầm là 0%. Điều này đồng nghĩa với việc chúng ta công nhận rằng, lương khởi điểm trung bình của nữ giới thấp hơn lương khởi điểm trung bình của nam giới. Do đó phát biểu của chủ tịch Liên đoàn lao động quốc gia Hồng Lạc không đáng tin cậy. (lưu ý: nếu bạn đặt giả thuyết  $H_a: (\mu_{nữ} - \mu_{nam}) \neq 0$  thì phải tính toán kiểm định hai đuôi và kết luận cuối cùng không thay đổi)

### **Bài 2: (20 điểm)**

Một cửa hàng thực phẩm nhận thấy thời gian vừa qua trung bình một khách hàng mua 250 nghìn đồng thực phẩm/ngày. Để đánh giá tác động của suy thoái kinh tế tới sức mua của khách hàng, ông chủ cửa hàng quan sát ngẫu nhiên 15 khách hàng thấy trung bình một khách hàng mua 240 nghìn đồng/ngày và phương sai mẫu là  $s^2 = (20 \text{ nghìn đồng})^2$ . Với mức ý nghĩa là 5% , thử xem có phải sức mua của khách hàng hiện nay thực sự giảm sút

### **Đáp án bài tập 2:**

Gọi  $\mu$  là sức mua trung bình của khách hàng trong một ngày. Các giả thuyết hợp lý:

$$H_0: \mu = \mu_0 = 250$$

$$H_a: \mu < 250$$

$$n = 15, \bar{X} = 240, \alpha = 5\%, df = n - 1 = 14$$

Tra bảng ta có giá trị tới hạn  $t^* = 2,1448$

$$\text{Giá trị kiểm định } t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} = \frac{240 - 250}{20 / \sqrt{15}} = -1,9364$$

So sánh  $|t| < |t^*|$  nên không bác bỏ  $H_0$

Kết luận: Với mức ý nghĩa là 5%, chưa đủ cơ sở để khẳng định sức mua của khách hàng hiện nay giảm sút.

**Bài 3: (20 điểm)**

Theo một báo cáo nghiên cứu thị trường, hiện nay 80% dân cư ở thành phố Hồ Chí Minh (t.p) mua thực phẩm trong siêu thị thay vì mua ở ngoài chợ. Hỏi ngẫu nhiên 36 bà nội trợ ở T.p thì có 25 người trả lời là mua thực phẩm ở siêu thị. Với mức ý nghĩa 5%, kiểm định xem báo cáo trên có đáng tin cậy không?

**Đáp án bài 3:**

Các giả thuyết hợp lý:

$$H_0: p = p_0 = 0,8$$

$$H_a: p \neq 0,8$$

$p$  là tỷ lệ hộ dân thực sự mua thực phẩm trong siêu thị.

$p_0 = 0,8$  là tỷ lệ hộ dân mua thực phẩm trong siêu thị theo nguồn tin.

$$n = 36, \bar{p} = f = \frac{25}{36} = 0,69, \alpha = 5\%,$$

Đây là bài toán kiểm định hai đuôi, tra bảng ta có giá trị tới hạn  $Z_{\alpha/2} = 1,96$ .

$$\text{Giá trị kiểm định } Z = \frac{\bar{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{0,69 - 0,8}{\sqrt{0,8 * 0,2 / 36}} = -1,65$$

So sánh  $|Z| < |Z_{\alpha/2}|$  nên không bác bỏ  $H_0$

Kết luận: Với mức ý nghĩa là 5%, chưa đủ cơ sở để khẳng định báo cáo trên không đáng tin cậy.

**Bài 4: (35 điểm)**

Theo văn bản Số: 1346 /BC-UBXH12 gửi cho Quốc hội với nội dung “*Báo cáo kết quả giám sát tình hình thực hiện bình đẳng giới và việc triển khai thi hành luật bình đẳng giới*” phó chủ nhiệm Ủy ban Các vấn đề Xã hội đã phát biểu những ý sau đây:

“Vị thế việc làm của lao động nữ cũng có sự thay đổi tích cực. Trong 10 năm từ 1997 đến 2007, nhóm lao động làm công ăn lương tăng rất mạnh trong cơ cấu phân bố lao động, từ 18,6% (1997) lên tới 30% (2007), trong đó lao động nam chiếm 59,8% và lao động nữ chiếm 40,2% (2007).”<sup>1</sup>

“Ở bậc đại học và cao đẳng, tỷ lệ sinh viên, học sinh là nữ đang có xu hướng tăng dần đều qua các năm ..., trong đó năm học 2006-2007 tỷ lệ học sinh, sinh viên nữ còn cao hơn tỷ lệ nam học sinh, sinh viên (nữ 53,32%, nam 46,68)”<sup>2</sup>

“Nhìn chung, việc quan tâm chăm sóc sức khỏe luôn tăng đều ở cả nam, nữ và không có sự khác biệt giới trong việc khám bệnh, chữa bệnh nội trú, ngoại trú giữa nam và nữ (xem bảng 12 của

<sup>1</sup> Báo cáo kết quả giám sát tình hình thực hiện bình đẳng giới và việc triển khai thi hành luật bình đẳng giới, Tiêu mục 2, mục II,

<sup>2</sup> Báo cáo nói trên, Tiêu mục 3, mục II,

Phụ lục).”<sup>3</sup>

Khai thác bộ dữ liệu điều tra mức sống dân cư VHLSS 2008, các ban hãy kiểm định về tính chính xác của các con số có liên quan tới nữ giới (của báo cáo ở năm 2007) cũng như phát biểu về việc quan tâm chăm sóc sức khỏe. (Mức ý nghĩa 5%)

**Đáp án bài 4:**

Để có dữ liệu kiểm định các con số nêu trên, chúng ta phải khai thác dữ liệu VHLSS 2008 ở các mục: 123a - số liệu về giáo dục, 3b – số liệu về sức khỏe, 4a – số liệu về lao động. Cụ thể như sau;

**Giao dục**

. use "L:\DATA\VHLSS2008\muc123a.dta", clear

Tạo biến học cao đẳng, đại học (CDDH)

. gen CDDH = m2ac8 if m2ac8<=9& m2ac8>=7

(37145 missing values generated)

. tab CDDH

CDDH	Freq.	Percent	Cum.
7	98	8.84	8.84
8	246	22.20	31.05
9	764	68.95	100.00

Total | 1,108 100.00

Tạo biến nữ học cao đẳng đại học (NuCDDH)

. gen NuCDDH = 1 if m1ac2==2&( CDDH<=9& CDDH>=7)

(37729 missing values generated)

. tab NuCDDH

NuCDDH	Freq.	Percent	Cum.
1	524	100.00	100.00

Total | 524 100.00

Kiểm định tỷ lệ nữ học cao đẳng, đại học = 53,32%

. prtest NuCDDH = 0.5332

One-sample test of proportion NuCDDH: Number of obs = 524

Variable	Mean	Std. Err.	[95% Conf. Interval]	
NuCDDH	1	0	1	1

p = proportion(NuCDDH) z = 21.4183

Ho: p = 0.5332

Ha: p < 0.5332

Ha: p != 0.5332

Ha: p > 0.5332

Pr(Z < z) = 1.0000

Pr(|Z| > |z|) = 0.0000

Pr(Z > z) = 0.0000

<sup>3</sup> Báo cáo nói trên, Tiêu mục 4, mục II,

**Kiểm định tỷ lệ nữ sinh viên học ở bậc Cao đẳng và đại học cho thấy tỷ lệ trong báo cáo thấp hơn tỷ lệ thực tế.**

**Kham Benh**

```
. use "L:\DATA\VHLSS2008\muc3b.dta", clear  
. sort tinh huyen xa diaban hosomatv  
. save "C:\Users\chuongmh\Desktop\muc3b_sorted.dta", replace  
file C:\Users\chuongmh\Desktop\muc3b_sorted.dta saved
```

```
. count  
19297
```

```
. use "L:\DATA\VHLSS2008\muc123a.dta", clear  
. sort tinh huyen xa diaban hosomatv  
. count  
38253
```

```
. merge m:m tinh huyen xa diaban hosomatv using  
"C:\Users\chuongmh\Desktop\muc3b_sorted.dta"
```

```
Result # of obs.  
-----  
not matched 27,190  
  from master 24,852 (_merge==1)  
  from using 2,338 (_merge==2)  
matched 16,959 (_merge==3)  
-----
```

```
. keep if _merge==3  
(27190 observations deleted)  
Tạo biến khám bệnh ngoại trú (kbnt)
```

```
. gen kbnt = m3c10a>0  
Tạo biến khám bệnh nội trú (Kbnt)
```

```
. gen Kbnt = m3c12a>0  
Tạo biến kham bệnh (kb)  
. gen kb = (kbnt>0 | Kbnt >0)
```

```
Tạo biến nữ đi khám bệnh (nukb)  
. gen nukb = (kb>0 & m1ac2==2)  
. tab nukb
```

nukb	Freq.	Percent	Cum.
0	7,287	42.97	42.97
1	9,672	57.03	100.00
-----			
Total	16,959	100.00	

```
Tạo biến nam khám bệnh (namkb)
```

```
. gen namkb = (kb>0 & m1ac2==1)  
. tab namkb
```

namkb	Freq.	Percent	Cum.
0	9,678	57.07	57.07
1	7,281	42.93	100.00
-----			
Total	16,959	100.00	

```
Kiểm định tỷ lệ nam khám bệnh = tỷ lệ nữ khám bệnh
```

```
. prtest namkb= nukb
```



```
. prtest LCAL = 0.3
One-sample test of proportion          LCAL: Number of obs = 35154
```

Variable	Mean	Std. Err.	[95% Conf. Interval]	
LCAL	.2676225	.0023612	.2629945	.2722504

```
p = proportion(LCAL)                z = -13.2471
Ho: p = 0.3
Ha: p < 0.3          Ha: p != 0.3      Ha: p > 0.3
Pr(Z < z) = 0.0000    Pr(|Z| > |z|) = 0.0000    Pr(Z > z) = 1.0000
```

*Tạo biến nữ làm công ăn lương*

```
. gen NuLCAL = ( LCAL==1 & m1ac2==2)
```

```
. tab NuLCAL
```

NuLCAL	Freq.	Percent	Cum.
0	31,482	89.55	89.55
1	3,672	10.45	100.00
Total	35,154	100.00	

*Kiểm định tỷ lệ nữ làm công ăn lương = 40,2%*

```
. prtest NuLCAL = 0.402
One-sample test of proportion          NuLCAL: Number of obs = 35154
```

Variable	Mean	Std. Err.	[95% Conf. Interval]	
NuLCAL	.1044547	.0016312	.1012575	.1076519

```
p = proportion(NuLCAL)                z = -1.1e+02
Ho: p = 0.402
Ha: p < 0.402          Ha: p != 0.402      Ha: p > 0.402
Pr(Z < z) = 0.0000    Pr(|Z| > |z|) = 0.0000    Pr(Z > z) = 1.0000
```

***Kiểm định này cho thấy tỷ lệ phát biểu trong báo cáo cao hơn tỷ lệ thực tế.***

***Tài liệu tham khảo:***

- Lương Phan Cừ (2009). Báo cáo kết quả giám sát tình hình thực hiện bình đẳng giới và việc triển khai thi hành luật bình đẳng giới: *Quốc hội khóa XII, Ủy ban Các vấn đề xã hội.*