

Báo cáo

Trà sữa và MAS291



Giáo viên hướng dẫn: Thầy Nguyễn Việt Anh (Mr)

Môn học: MAS291 - SE1609 - Team 3

Sinh viên thực hiện: Đặng Tất Thành - HE150869

Nguyễn Hữu Tiến Anh - HE150726

Nguyễn Xuân Hoàng - HE150923

Nguyễn Minh Đức - HE150901

Nguyễn Ngọc Tân - HE151074

Mục lục

- I. Lời mở đầu
- II. Thống kê
 - 1. Tìm dữ liệu
 - 2. Xây dựng biểu đồ và đánh giá chung dữ liệu và tìm khoảng tin cậy cho dữ liệu. Từ đó, thống kê, tính toán tổng lợi nhuận, sự chênh lệch giữa tổng doanh thu các ngày thường, ngày cuối tuần và ngày lễ
- III. Báo cáo
 - 1. Phân công nhiệm vụ
 - 2. Kết quả
- IV. Câu hỏi
- V. Phụ lục

I. Lời mở đầu

Trong xã hội hiện đại, con người làm việc với năng suất cao hơn, dành nhiều thời gian cho công việc. Đi cùng là thời gian cho ăn uống, nghỉ ngơi trở nên eo hẹp. Đây là lúc lên ngôi của những thức ăn nhanh, ngon, bổ , hợp vệ sinh và tiện lợi. Trà sữa là từ khóa được khá quan tâm hiện nay. Một chủ đề được giới truyền thông đề cập rất nhiều. Đây được coi là một trong những trào lưu nổi bật của giới trẻ. Số lượng các cửa hàng trà sữa ngày càng nhiều từ các hãng trà sữa nổi tiếng cho đến các cửa hàng trà sữa nhà làm. Tiện, ngon, hợp túi tiền chính là điểm thu hút của loại thức uống này. Bạn có thể chọn loại thức uống mình thích, đợi pha chế, đóng gói trả tiền ngay tại quầy, và vừa đi vừa uống. Hay bạn cũng có thể dễ dàng chọn cho mình một quán vừa nhâm nhi, vừa thư giãn ngay trong cửa hàng. Thông qua dự án này để đánh giá được thói quen tiêu dùng trà sữa của người dân hiện nay trong 1 tuần của một cửa hàng kinh doanh bán trà sữa.

II. Thống kê

1. Tìm dữ liệu

Dữ liệu được lấy trực tiếp từ Shop trà sữa TocoToco Thiên Đường Bảo Sơn. Sau đó sử dụng “máy chạy bằng trà sữa” để nhập dữ liệu vào Excel.

	B	C	D	E	F	G	H	I
1	Tổng Tiền	Giảm giá	Hoa hồng (22%)	Tiền nhận	Số cốc	Stauts	Đánh giá	
2	105000	36000	15180	53820	2	Cancel		
3	96000	48400	10472	37128	2	Cancel		
4	712000	246800	102344	362856	14	Success	5 ★	
5	92000	46800	9944	35256	2	Success		
6	112000	45200	14696	52104	2	Success		
7	111000	48800	13684	48516	2	Success	5 ★	
8	109000	47600	13508	47892	2	Cancel		
9	268000	117200	33176	117624	6	Success	3 ★	
10	343000	120000	49060	173940	7	Success		
11	104000	36800	14784	52416	2	Success		
12	275000	102000	38060	134940	5	Success		
13	102000	34400	14872	52728	2	Success		
14	325000	110800	47124	167076	6	Success	5 ★	
15	188000	75200	24816	87984	4	Success	5 ★	
16	96000	38400	12672	44928	2	Success	5 ★	
17	196000	88400	23672	83928	4	Success		
18	92000	46800	9944	35256	2	Success		

Từ đó tính ra một vài thông số cơ bản như tổng số, min, max của giá tiền, giảm giá,... của tổng số đơn.

Computer_Project_MAS291

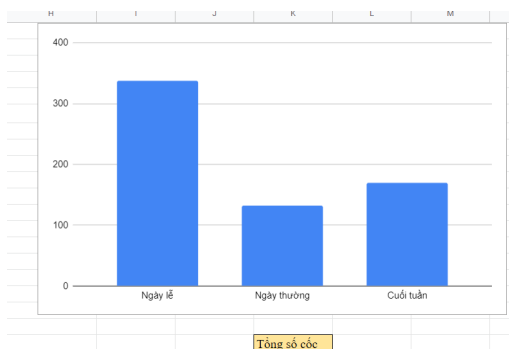
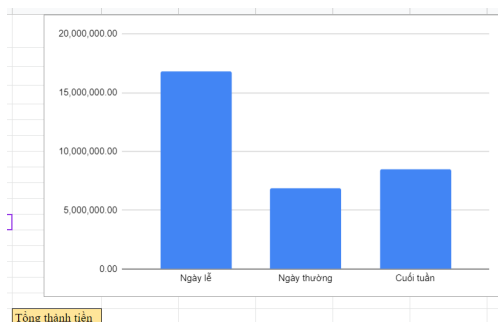
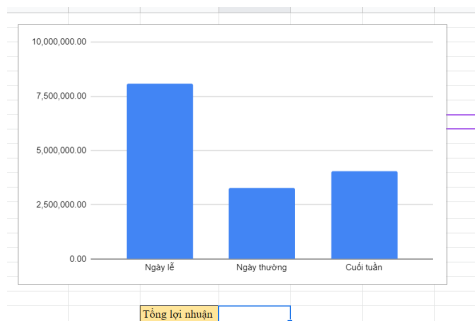
File Edit View Insert Format Data Tools Extensions Help Last edit was seconds ago

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
44	43	54000	16000	8360	29640	1	Success							
45	44	96000	38400	12672	44928	2	Success							
46	45	94000	32000	13640	48360	2	Success	5 ★						
47	46	280000	96000	40480	143520	5	Success	5 ★						
48	47	64000	19200	9856	34944	1	Success							
49	48	96000	38400	12672	44928	2	Success							
50	49	92000	36800	12144	43056	2	Success							
51	50	61000	18400	9372	33228	1	Success							
52	51	108000	37600	15488	54912	2	Success	5 ★						
53	52	384000	153600	50688	179712	8	Success	3 ★						
54	53	231000	86400	31812	112788	5	Success							
55	54	65000	19200	10076	35724	1	Success	5 ★						
56	55	215000	72800	31284	110916	4	Success							
57	56	192000	76800	25344	89556	4	Success	5 ★						
58	Tổng	8493000	3298150	1142867	4051983	170								
59	Max	712000	246800	102344	362856	14								
60	Min	48000	16000	5907	20943	1								

Ngày lễ 8/3 Ngày thường 7/3 Cuối tuần

2. Xây dựng biểu đồ và đánh giá chung dữ liệu và tìm khoảng tin cậy cho dữ liệu. Từ đó, thống kê, tính toán tổng lợi nhuận, sự chênh lệch giữa tổng doanh thu các ngày thường, ngày cuối tuần và ngày lễ.

Từ những dữ liệu thu thập được, sử dụng những công cụ tính toán của Excel để xây dựng lên những biểu đồ cũng như các bảng đánh giá chung về dữ liệu thu thập được:



Thống kê đánh giá	
Số sao	Số lượt đánh giá
1	0
2	0
3	2
4	8
5	10
Đánh giá trung bình: (5 ★) 4.40	
Số đơn có số lượng cốc lớn hơn hoặc bằng 5:	
4	
xác suất: (%)	12.50

Sau khi thống kê được tổng quát được các giá trị thu thập trước đó, tiếp đến là tính đến các giá trị kỳ vọng, phương sai và khoảng tin cậy của chúng.

Có thể dùng các công cụ của Excel cùng với áp dụng những công thức học được trên lớp để có thể tính ra được những giá trị như hình dưới đây.

t-Test: Two-Sample Assuming Equal Variances		
	<i>Số cốc</i>	<i>Đánh giá</i>
Mean	3.036036036	4.322580645
Variance	5.344144144	0.8258064516
Observations	111	31
Pooled Variance	4.375928924	
Hypothesized Mean Difference	0	
df	140	
t Stat	-3.027526016	
P(T<=t) one-tail	0.001467963656	
t Critical one-tail	1.655810511	
P(T<=t) two-tail	0.002935927313	
t Critical two-tail	1.97705372	

Test claim: $\mu_1 = \mu_2 \neq 0$

$\Rightarrow H_0: \mu_1 = \mu_2 = 0$
 $H_1: \mu_1 = \mu_2 \neq 0$

$\Delta_0 = \mu_1 - \mu_2 = 0$

Test statistic:

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

$$= \frac{(3,016 - 4,32258)}{\sqrt{4,3759 \cdot \left(\frac{1}{111} + \frac{1}{51}\right)}}$$

$$= -3,0276 \in \mathbb{R}$$

$t_{\alpha/2, df} = 1,977$

to $\notin (-t_{\alpha/2, df}, t_{\alpha/2, df})$

\Rightarrow **Reject H_0**
 \Rightarrow **Fail to reject claim**

For $\mu_1 = \mu_2$ unknown (σ_1, σ_2) $(\alpha) = 0,05$
 $n_1 = 111$ $n_2 = 51$ $\Rightarrow n = 0,05$
 $s_1^2 = 5,399$ $s_2^2 = 0,2258$ $\alpha/2 = 0,025$

(1) $(\bar{x}_1 - \bar{x}_2 \pm E)$

$$s_p^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2} = \frac{110 \cdot 5,399 + 50 \cdot 0,2258}{111 + 51 - 2}$$

$$= 4,3759$$

$E = t_{\alpha/2, df} \cdot \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$

$$= t_{0,025, 160} \cdot \sqrt{4,3759 \cdot \left(\frac{1}{111} + \frac{1}{51}\right)}$$

$$= 0,8401$$

$\Rightarrow 95\% \text{ CI for } \mu_1 - \mu_2 \text{ is } (-3,2865 + 0,8401)$
 $= (-2,1266, -0,9969)$

pool variance $s_p^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}$

degree of freedom: $df = n_1 + n_2 - 2$

$E = \text{critical value} \cdot \text{standard error}$

$$= t_{\alpha/2, df} \cdot \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

$t_{\alpha/2, df}$ **T.INV.2T(α , df)** excel

III. Báo cáo

1. Phân công nhiệm vụ

- Nguyễn Xuân Hoàng thu thập và tổng hợp dữ liệu dữ liệu
- Nguyễn Ngọc Tân xử lý, tính toán các dữ liệu thu thập được
- Nguyễn Hữu Tiến Anh tổng hợp kết quả, xây dựng Excel
- Đặng Tất Thành kiểm định giả thiết
- Nguyễn Minh Đức viết báo cáo

2. Kết quả

- Các thành viên trong nhóm cùng nhau làm bài, hoàn thành 100% dự án và đúng thời hạn được giao.

- Đánh giá: Qua những con số thống kê và dữ liệu đánh giá ở trên, ta có thể thấy được mức độ tiêu thụ trà sữa có sự chênh lệch giữa ngày thường và cuối tuần, đặc biệt có sự chênh lệch rất lớn so với ngày lễ.

IV. Câu hỏi

1. Xác suất đơn hàng bị hủy mỗi ngày ?
2. Khoảng tin cậy (Interval Confidence)
3. Kiểm tra giả thuyết
4. Hồi quy tuyến tính (Simple Linear Regression)

Answer:

1. Cách tính: tổng số đơn hủy trên tổng số đơn.

- Ngày thường: **0.0701754386**
- Ngày lễ: **0.02702702703**
- Cuối tuần: **0.05357142857**

=> **Nhận xét:** Số lượng đơn hàng tỉ lệ nghịch với tỉ lệ đơn hàng bị hủy. (Càng nhiều đơn bán ra thì tỉ lệ đơn bị hủy càng ít).

2. Khoảng tin cậy (Confidence interval):

Do đây là 1 bài toán không quá nghiêm trọng, phức tạp nên chúng ta sẽ chọn khoảng tin cậy (CI) là 95% (chín mươi lăm phần trăm). Do đây là bài toán tổng hợp nên sẽ không có xích-ma và cả 2 mẫu thử (sample) đều lớn hơn 30 nên ta sẽ dùng công thức:

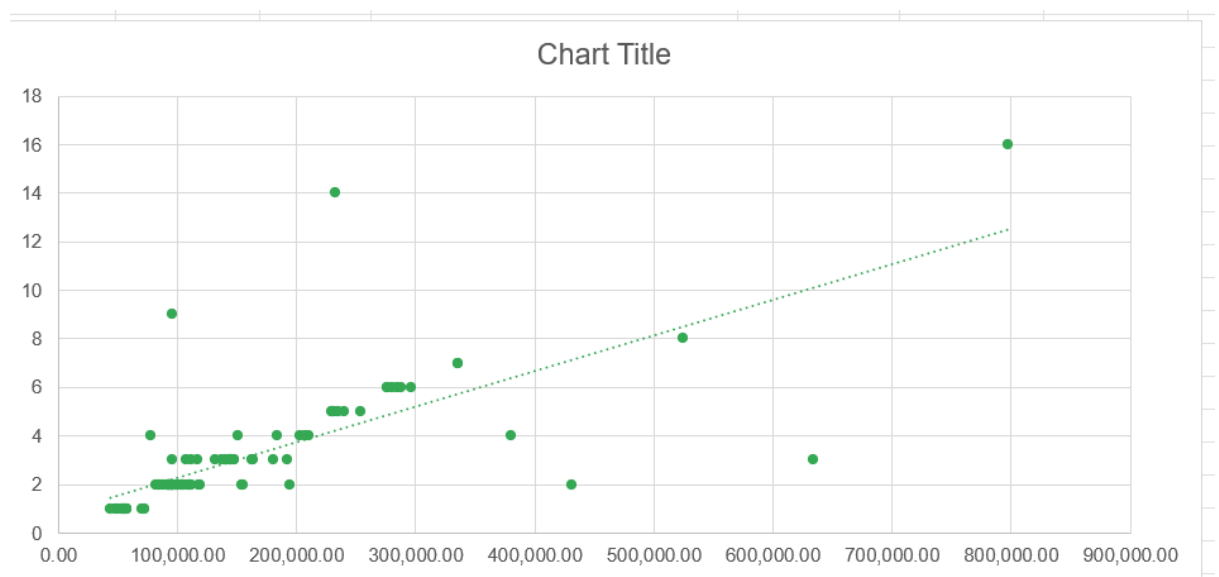
The image shows handwritten mathematical formulas on lined paper. At the top, the formula for the confidence interval is written as $(\bar{X}_1 - \bar{X}_2 \pm E)$. Below this, the pooled standard deviation S_p^2 is calculated using the formula $S_p^2 = \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2}$. The calculation shown is $S_p^2 = \frac{110 \cdot 5,344 + 30 \cdot 0,7258}{111 + 31 - 2} = 4,3759$. Finally, the margin of error E is calculated using the formula $E = t_{\alpha/2, df} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$.

Từ đó ta được khoảng tin cậy của dữ liệu là: (-2.1266 ; -0.4464)

3. kiểm tra giả thuyết (test claim): sau khi test claim với $u_1 - u_2 = 0$, ta được $H_0 : u_1 - u_2 = 0$ và $H_1 : u_1 - u_2 \neq 0$. Ta nhận được kết quả Reject H_0 hay còn được gọi là Fail to reject claim.

4. Hồi quy tuyến tính (Linear simple regression):

Ngày lễ:



V. Phụ lục

File Excel: [Computer_Project_MAS291](#)

Ảnh data: [Image](#)