

MAS291 PROJECT REPORT

Group members:

- | | |
|--------------------------|----------|
| 1. Nguyễn Trọng Nam | HE180258 |
| 2. Trần Ngọc Huyền Trang | HE189002 |
| 3. Đỗ Tùng Dương | HE181160 |
| 4. Trần Duy Hưng | HE181688 |
| 5. Nguyễn Đức Anh | HE182179 |
| 6. Thái Duy Phong | HE180210 |
| 7. Đinh Trần Duy Anh | HE181306 |

I. Đặt vấn đề

Chiều cao trung bình của người Việt Nam trẻ tuổi đã có những chuyển biến tích cực trong thời gian qua. Thống kê chiều cao trung bình người Việt Nam tăng hơn 3cm so với kết quả 10 năm trước, giúp chúng ta không còn thuộc nhóm “lùn nhất Đông Nam Á”. Chúng em hướng đến nghiên cứu chiều cao của nhóm đối tượng là thanh niên Việt Nam và cụ thể nhóm sinh viên đại học FPT Hà Nội.

❖ Câu hỏi:

- Chiều cao của sinh viên của sinh viên FPT và sinh viên trên toàn nước khác nhau như thế nào?
- Có những yếu tố chủ quan và khách quan nào ảnh hưởng tới chiều cao của sinh viên?
- Đưa ra giải pháp và gợi ý phù hợp để cải thiện chiều cao cho sinh viên.

II. Thu thập số liệu & phân tích

Nhóm em đã thu thập dữ liệu số của 30 bạn sinh viên trong lớp SE1843 từ ngày 19/02/2024.

2.1. Thu thập số liệu

BẢNG THỐNG KÊ CHIỀU CAO

STT	MSSV	HỌ VÀ TÊN	Chiều cao
1	HE189002	Trần Ngọc Huyền Trang	156
2	HE180210	Thái Duy Phong	159
3	HE180144	Đào Duy Hiếu	165
4	HE170112	Nguyễn Bá Bách	165
5	HS181156	Nguyễn Khánh Linh	165
6	HE172104	Nguyễn Minh Đức	165
7	HE182498	Nguyễn Trâm Anh	165
8	HE180143	Trịnh Ngọc Quang	166
9	HE163991	Nguyễn Phú Quảng	167
10	HE180475	Lê Văn Việt	168
11	HE180232	Hà Quốc Minh Đức	169

12	HE181666	Phạm Hoàng Nam	169
13	HE180173	Vũ Đặng Quang Vinh	169
14	HE181106	Đỗ Tùng Dương	169
15	HE180142	Dương Xuân Hoàng Minh	170
16	HE180171	Vũ Cường Thịnh	170
17	HE181306	Đinh Trần Duy Anh	171
18	HE171706	Nguyễn Khánh Duy	171
19	HE186653	Đới Lê Dũng	172
20	HE181547	Hoàng Tiến Dũng	172
21	HE180258	Nguyễn Tường Huy	172
22	HE181688	Trần Duy Hưng	172
23	HE180309	Ngô Quốc Huy	173
24	HE181045	Nguyễn Hoàng Phúc	173
25	HE182307	Tạ Tuấn Dũng	173
26	HE180253	Lưu Tuấn Phi	174
27	HE182179	Nguyễn Đức Anh	174
28	HE180182	Nguyễn Đức Hùng	174
29	HE180256	Nguyễn Trọng Nam	174
30	HE180240	Nguyễn Việt Hoàng	180

BẢNG THỐNG KÊ THỜI GIAN TẬP THỂ DỤC

STT	MSSV	HỌ VÀ TÊN	THỜI GIAN TẬP THỂ DỤC (PHÚT)
1	HE189002	Trần Ngọc Huyền Trang	0
2	HE180210	Thái Duy Phong	0
3	HE180144	Đào Duy Hiếu	0
5	HS181156	Nguyễn Khánh Linh	0
6	HE172104	Nguyễn Minh Đức	0
7	HE182498	Nguyễn Trâm Anh	0
10	HE180475	Lê Văn Việt	0
14	HE181106	Đỗ Tùng Dương	0
15	HE180142	Dương Xuân Hoàng Minh	0
16	HE180171	Vũ Cường Thịnh	0
18	HE171706	Nguyễn Khánh Duy	0
21	HE180258	Nguyễn Tường Huy	0
23	HE180309	Ngô Quốc Huy	0
24	HE181045	Nguyễn Hoàng Phúc	0
25	HE182307	Tạ Tuấn Dũng	0
26	HE180253	Lưu Tuấn Phi	0
30	HE180240	Nguyễn Việt Hoàng	0
9	HE163991	Nguyễn Phú Quảng	10
11	HE180232	Hà Quốc Minh Đức	10
8	HE180143	Trịnh Ngọc Quang	15
29	HE180256	Nguyễn Trọng Nam	20
4	HE170112	Nguyễn Bá Bách	30
19	HE186653	Đới Lê Dũng	30

22	HE181688	Trần Duy Hưng	30
28	HE180182	Nguyễn Đức Hùng	30
12	HE181666	Phạm Hoàng Nam	40
20	HE181547	Hoàng Tiến Dũng	60
17	HE181306	Đinh Trần Duy Anh	90
13	HE180173	Vũ Đặng Quang Vinh	120
27	HE182179	Nguyễn Đức Anh	120

2.2. Phân tích

PHÂN TÍCH CHIỀU CAO

Bảng phân phối xác suất													
X	156	159	165	166	167	168	169	170	171	172	173	174	180
P	1	1	4	1	1	1	4	2	2	4	3	4	1
	1/30	1/30	2/15	1/30	1/30	1/30	2/15	1/15	1/15	2/15	1/10	2/15	1/30

Median (Số trung vị)	170
Mean (Kỳ vọng)	169.5517241
Variance (Phương sai)	22.52318668
Standard deviation (Độ lệch chuẩn)	4.745855952
Confidence interval with $\alpha = 0.05$ (Khoảng tin cậy đối xứng với độ tin cậy 95%)	167.8534414 ; 171.2500068 left: 168.1263797 right:170.9770685

With $n = 30$ (odd number), we have Median:

$$\begin{aligned}\tilde{x} &= avg\left(\left(\frac{n}{2}\right)^{th} + \left(\frac{n}{2} + 1\right)^{th}\right) = avg((15)^{th} + (16)^{th}) \\ &= avg(170 + 170) \\ &= 170\end{aligned}$$

Mean:

$$E(X) = \mu = \sum_i x_i f(x_i) = 169.5517241$$

Variance:

$$V(X) = \sigma^2 = \sum_i (x_i - \mu)^2 f(x_i) = 22.52318668$$

Standard deviation:

$$\sigma = \sqrt{\sigma^2} = 4.745855952$$

PHÂN TÍCH KẾT QUẢ THỜI GIAN THỂ DỤC

Bảng phân phối xác suất									
X	0	10	15	20	30	40	60	90	120
P	17	2	1	1	4	1	1	1	2
	17/30	1/15	1/30	1/30	2/15	1/30	1/30	1/30	1/15

Median (Số trung vị)	0
Mean (Kỳ vọng)	20.16666667
Variance (Phương sai)	1144.138889
Standard deviation (Độ lệch chuẩn)	33.82512216
Confidence interval with $\alpha = 0.05$ (Khoảng tin cậy đối xứng với độ tin cậy 95%)	8.062502157; 32.27083118 left:10.00781431 right:30.32551903

With $n = 30$ (odd number), we have Median:

$$\begin{aligned}\tilde{x} &= avg\left(\left(\frac{n}{2}\right)^{th} + \left(\frac{n}{2} + 1\right)^{th}\right) = avg((15)^{th} + (16)^{th}) \\ &= avg(0 + 0) \\ &= 0\end{aligned}$$

Mean:

$$E(X) = \mu = \sum_i x_i f(x_i) = 20.16666667$$

Variance:

$$V(X) = \sigma^2 = \sum_i (x_i - \mu)^2 f(x_i) = 1144.138889$$

Standard deviation:

$$\sigma = \sqrt{\sigma^2} = 33.82512216$$

III. Xây dựng khoảng tin cậy đối xứng với độ tin cậy 95% cho chiều cao trung bình của sinh viên FPT (Confidence interval with $\alpha = 0.05$)

Standard deviation:

$$\sigma = \sqrt{\sigma^2} = 4.745855952$$

Giới hạn sai số: $z_{\alpha/2} *$

Độ tin cậy: 95% $\Rightarrow 1 - \alpha = 0.95$

$$1 - \alpha / 2 = 0.975$$

$$z_{\alpha} = \Phi(0.95) \Rightarrow 1.645$$

$$z_{\alpha/2} = \Phi(0.975) = 1.96$$

Khoảng tin cậy với độ tin tưởng 95%:

$$\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \leq \mu < \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} = 8.062502157 \leq \mu < 32.27083118$$

IV. Kiểm định giả thuyết chiều cao trung bình của sinh viên Đại học FPT là 165 với mức ý nghĩa 2%

$$\sigma = \sqrt{\sigma^2} = 4.745855952$$

$$\mu = \sum_i x_i f(x_i) = 169.5517241$$

$$\bar{x} = 169.4$$

$$\alpha = 0.02$$

$$\begin{cases} H_0: \mu = 165 \\ H_1: \mu \neq 165 \end{cases}$$

$$z_0 = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = -0.17507$$

$$\alpha = 0.02$$

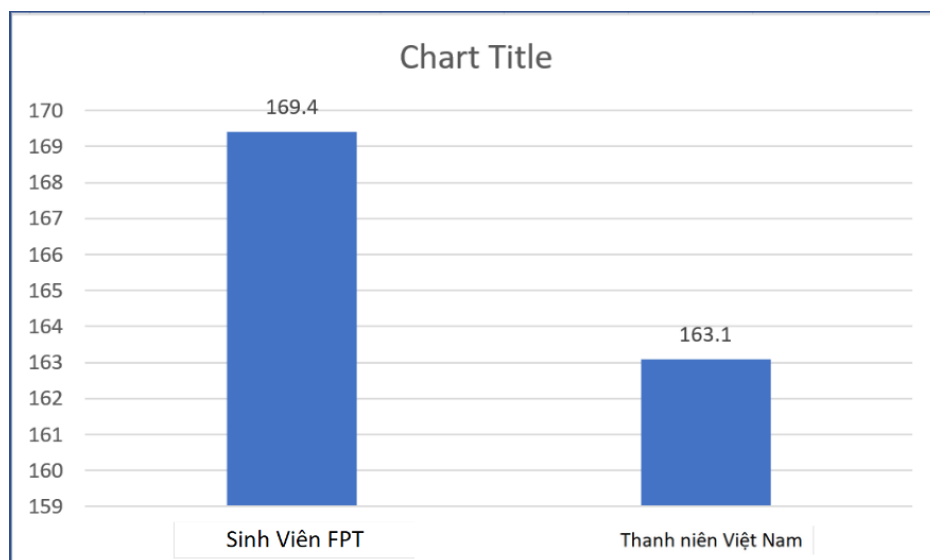
$$z_{\frac{\alpha}{2}} = \phi^{-1}(1 - \alpha) = \phi^{-1}(0.98) = 2.005$$

$$|z_0| = 0.17507$$

$$\Rightarrow |z_0| < z_{\frac{\alpha}{2}}$$

$$\Rightarrow \text{FAIL to reject } H_0$$

V. So sánh chiều cao trung bình của sinh viên FPT với chiều cao trung bình của thanh niên Việt Nam:



* Số liệu theo Báo người Lao động 10/10/2022

Qua số liệu trên, ta thấy chiều cao của sinh viên FPT cao hơn so với chiều cao trung bình của thanh niên cả nước

VI. Phân tích hệ số tương quan giữa chiều cao và thời gian tập thể dục với $\alpha = 0.05$

<p>Chiều cao :</p> $s^2 = \frac{1}{n-1} \sum n_i (x_i - \bar{x})^2 = 23.327586$ <p style="text-align: center;">$\bar{x} = 169.4$ $n = 30$</p>	<p>Thời gian tập thể dục:</p> $s^2 = \frac{1}{n-1} \sum n_i (x_i - \bar{x})^2 = 1183.5919$ <p style="text-align: center;">$\bar{x} = 20.1667$ $n = 30$</p>
---	--

$$s_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 603.4585$$

$$t_{\text{statistic}} = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = 26.366$$

$$d_f = n - 2 = 28 \text{ with } \alpha = 0.02$$

$$\Rightarrow t_{\text{critical}} = 1.701$$

$$t_{\text{statistic}} > t_{\text{critical}}$$

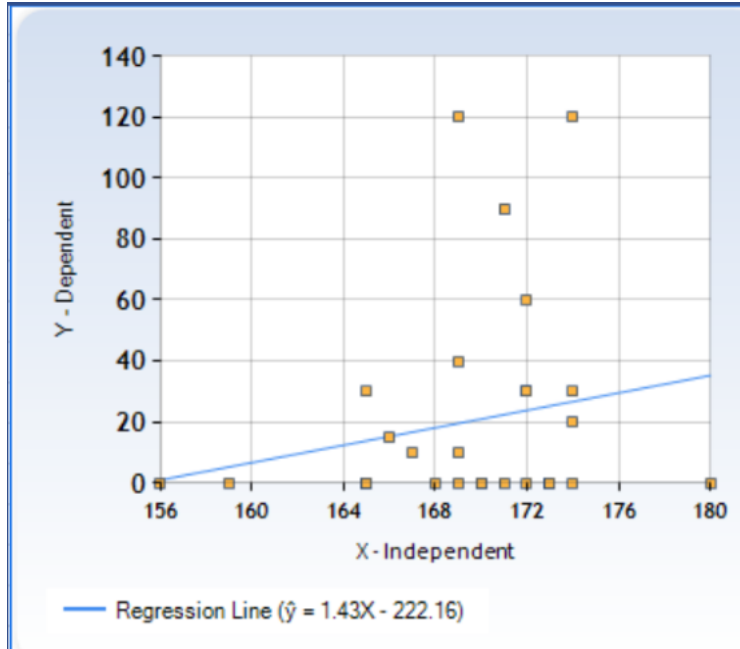
Kết luận : Chiều cao và thời gian tập thể dục của sinh viên có ảnh hưởng đến nhau

VII. Tiến hành kiểm định giả thuyết cho Regression Slope

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.168120025							
R Square	0.028264343							
Adjusted R Squar	-0.007725867							
Standard Error	4.188480104							
Observations	29							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	13.77740524	13.77741	0.785334	0.383340008			
Residual	27	473.6708706	17.54337					
Total	28	487.4482759						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	169.4415175	0.911126572	185.9692	1.68E-43	167.5720402	171.310995	167.5720402	171.3109948
0	0.020158666	0.022747544	0.886191	0.38334	-0.02651544	0.06683277	-0.02651544	0.066832771

Kết luận : Chiều cao và thời gian tập thể dục của sinh viên không ảnh hưởng đến nhau

VIII. Simple Linear Regression



$$s_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 963$$

$$s_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = 673.2$$

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} = 1,432609342$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = -2227347$$

IX. Đánh giá, kết luận thông qua các số liệu :

Chiều cao không phụ thuộc hoàn toàn vào vận động thể lực, mà còn phụ thuộc vào nhiều yếu tố như: giới tính, dinh dưỡng, môi trường (bệnh tật), tâm lý, gene, giấc ngủ.

Chúng em đã thu thập thêm thời gian đi ngủ của các bạn học sinh trong lớp SE1843:

Fullname	Giờ đi ngủ
Nguyễn Minh Đức	22
Nguyễn Đức Anh	22
Nguyễn Phú Quảng	23
Dương Xuân Hoàng Minh	23
Vũ Cường Thịnh	23
Ngô Quốc Huy	23
Hà Quốc Minh Đức	23.5
Phạm Hoàng Nam	23.5
Nguyễn Khánh Linh	24
Trịnh Ngọc Quang	24
Đỗ Tùng Dương	24
Nguyễn Khánh Duy	24
Hoàng Tiến Dũng	24
Trần Duy Hưng	24
Tạ Tuấn Dũng	24
Lưu Tuấn Phi	24
Nguyễn Đức Hùng	24
Nguyễn Trâm Anh	24.5
Đinh Trần Duy Anh	24.5
Đới Lê Dũng	24.5
Trần Ngọc Huyền Trang	1
Nguyễn Bá Bách	1
Lê Văn Việt	1
Nguyễn Tường Huy	1
Nguyễn Hoàng Phúc	1
Nguyễn Trọng Nam	1
Vũ Đặng Quang Vinh	2
Nguyễn Việt Hoàng	2
Thái Duy Phong	3
Đào Duy Hiếu	3

Bảng phân phối xác suất								
X	22	23	23.5	24	24.5	1	2	3
P	2	4	2	9	3	6	2	2
	17/30	1/15	1/30	1/30	2/15	1/30	1/30	1/30

Tuy không ảnh hưởng nghiêm trọng đến phát triển chiều cao, tuy nhiên các bạn sinh viên đang có xu hướng đi ngủ rất muộn. Việc này ảnh hưởng không nhỏ đến sức khỏe và chất lượng học tập của các bạn.

X. Bài học nhận ra khi làm dự án:

Sau khi làm việc cùng nhau trong dự án này, chúng em đã học hỏi được rất nhiều kiến thức mới cũng như củng cố kiến thức cũ.- Ôn lại các kiến thức về xác suất và thống kê như xác suất, phân phối chọn mẫu, thống kê mô tả, khoảng tin cậy, kiểm định các giả thuyết, hồi quy tuyến tính. - Tính đơn giản, tương quan, ...

- Tìm hiểu thêm về cách thu thập, làm sạch, xử lý dữ liệu và tính toán.
- Nâng cao khả năng làm việc nhóm
- Hiểu được tình hình dịch bệnh ở Việt Nam