

Sử dụng Minitab trong thống kê môi trường

Dương Trí Dũng

I. Giới thiệu

- Hiện nay có nhiều phần mềm (software) thống kê trên thị trường
 - Giá cao
 - Excel không đủ tính năng
 - Tính bằng công thức chậm
 - Có nhiều yêu cầu mà các software không thể hiện được
 - Cách nhập số liệu
 - Giao diện
- Minitab có thể giải quyết

II. Phương pháp

- Download phần mềm Minitab trên website
<http://www.minitab.com/en-CA/products/minitab/free-trial.aspx>
- Đăng ký, cho địa chỉ email
- Trang web sẽ gửi đường dẫn để download
- Install: điền địa chỉ email vào khi đòi hỏi
- Sử dụng thử trong 30 ngày

III. Tìm xác suất

- Tìm xác suất xuất khi $x \geq \bar{x}$ khi đã biết μ , σ và n
- Tính theo công thức $Z = \frac{x - \mu}{\sigma}$ khi n lớn và $Z = \frac{x - \mu}{\sigma_x}$ khi n nhỏ.
- Với $\sigma = \sigma_x \sqrt{n}$
- Thí dụ: Khi thu 18 mẫu CO trong không khí ở Cần Thơ, biết $\mu=10$ mg/L và $\sigma=1.019$. Tìm xác suất bắt gặp các mẫu có hàm lượng không thấp hơn 10.43 mg/L.
- Theo công thức ta được $Z= 1.79$, tra bảng Z được **$P(x>10.43)$ hay $P(z>1.79) = 0.037$**

Cách tra bảng Z

TABLE B.2 Proportions of the Normal Curve (One-Tailed)

This table gives the proportion of the normal curve that lies beyond (i.e., is more extreme than) a given normal deviate; e.g., $Z = (X_i - \mu)/\sigma$ or $Z = (\bar{X} - \mu)/\sigma/\bar{X}$. For example, the proportion of a normal distribution for which $Z \geq 1.51$ is 0.0655.

Z	0	1	2	3	4	5	6	7	8	9	Z
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641	0.0
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247	0.1
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859	0.2
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483	0.3
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121	0.4
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776	0.5
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451	0.6
0.7	0.2420	0.2389	0.2358	0.2327	0.2297	0.2266	0.2236	0.2207	0.2177	0.2148	0.7
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867	0.8
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611	0.9
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379	1.0
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170	1.1
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985	1.2
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823	1.3
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681	1.4
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559	1.5
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455	1.6
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367	1.7

III. Tìm xác suất

- Sử dụng Minitab cho thí dụ trên
- Trên Menu bar chọn Stat → Basic statistic → 1Z, Điền các giá trị vào → click vào hộp Option, chọn greater than → OK → OK được kết quả tương tự như đã tính toán
- Khi sử dụng Minitab, không cần tra bảng Z

Minitab - Untitled

File Edit Data Calc Stat Graph Editor Tools Window Help Assistant

Basic Statistics
 Regression
 ANOVA
 DOE
 Control Charts
 Quality Tools
 Reliability/Survival
 Multivariate
 Time Series
 Tables
 Nonparametrics
 EDA
 Power and Sample Size

Display Descriptive Statistics...
 Store Descriptive Statistics...
 Graphical Summary...
1Z 1-Sample Z...
 1t 1-Sample t...
 2t 2-Sample t...
 t-t Paired t...
 1P 1 Proportion...
 2P 2 Proportions...
 1P 1-Sample Poisson Rate...
 2P 2-Sample Poisson Rate...
 1 1 Variance...
 2 2 Variances...
 COR Correlation...
 COV Covariance...
 TEST Normality Test...

Session

One-Sample Z

Test of $\mu = 10$ vs
 The assumed standa

N	Mean	SE Mea
18	10.430	0.24

Worksheet 1 ***

	C1	C2	C3	C4
+				

1-Sample Z (Test and Confidence Interval)

Samples in columns:

Summarized data

Sample size: 18

Mean: 10.43

Standard deviation: 1.019

Perform hypothesis test

Hypothesized mean: 10

Select Graphs... Options...
 Help OK Cancel

1-Sample Z - Options

Confidence level: 95.0

Alternative: greater than
 less than
 not equal
 greater than

Help Cancel

1-Sample Z - Options

Confidence level: 95.0

Alternative: greater than

Help OK

Session

One-Sample Z

Test of $\mu = 10$ vs > 10
 The assumed standard deviation = 1.019

N	Mean	SE Mean	95% Lower Bound	Z	P
18	10.430	0.240	10.035	1.79	0.037

IV. Kiểm định t

- So sánh sự khác biệt giữa một quan sát (n mẫu) với một giá trị đã biết, so sánh hai quan sát với nhau và so sánh từng cặp.
- So sánh 2 trung bình được tính theo công thức

$$T = \frac{x_1 - x_2}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

trong đó $S_p^2 = \frac{SS_1 + SS_2}{v_1 + v_2}$

và

$$SS = \sum (x - \bar{x})^2$$

IV. Kiểm định t (2 trung bình)

- Thí dụ: Kiểm tra xem tác dụng của loại phân mới trên cây trồng có tốt hơn là phân cũ hay không khi so sánh chiều cao của cây sau 1 tuần bón phân là
Phân cũ: 48.2, 54.6, 58.3, 47.8, 51.4, 52.0, 55.2, 49.1, 49.9, 52.6
Phân mới: 52.3, 57.4, 55.6, 53.2, 61.3, 58.0, 59.8, 54.8
- $X_{Obar} = 51.91$, $X_{Nbar} = 56.55$, $SS_1 = 102.23$, $SS_2 = 69.20$
- $S^2_p = (102.23 + 69.2) / (9 + 7) = 10.71$; $SE = 1.55$
- $t = -4.64 / 1.55 = -2.99$, $t_{\alpha(1), 16} = 1.746$
Vì $|t| > t_\alpha \rightarrow$ loại bỏ H_0

IV. Kiểm định t (2 trung bình) bằng minitab

- Chạy phần mềm minitab
- Nhập số liệu: copy và past từ excel hay word vào 2 cột c1 và c2. Đánh từ “phân cũ” ô dưới c1 và “phân mới” dưới ô c2
- Trên Menu bar chọn Stat → Basic statistic → 1Z, Điền các giá trị vào các ô tương ứng trong mục sample in different columns → check vào ô Assume equal variances → click vào hộp Option, chọn greater than → OK → OK được kết quả tương tự như đã tính toán
- Khi sử dụng Minitab, không cần tra bảng t và không quan tâm đến giá trị t 1 đuôi.

Minitab - Untitled

File Edit Data Calc Stat Graph Editor Tools Window Help Assistant

Basic Statistics
 Regression
 ANOVA
 DOE
 Control Charts
 Quality Tools
 Reliability/Survival
 Multivariate
 Time Series
 Tables
 Nonparametrics
 EDA
 Power and Sample Size

Display Descriptive Statistics...
 Store Descriptive Statistics...
 Graphical Summary...
 1Z 1-Sample Z...
 1t 1-Sample t...
2t 2-Sample t...
 t-t Paired t...
 1P 1 Proportion...
 2P 2 Proportions...
 1s 1-Sample Poisson Rate...
 2s 2-Sample Poisson Rate...
 σ^2 1 Variance...
 σ^2_{og} 2 Variances...
 COR Correlation...
 COV Covariance...
 TEST Normality Test...
 χ^2 Goodness-of-Fit Test for Poisson...

Worksheet 1 ***

	C1	C2	C3	C4
	phân cũ	phân mới		
1	48.2	52.3		
2	54.6	57.4		
3	58.3	55.6		

2-Sample t (Test and Confidence Interval)

Samples in one column
 Samples:
 Subscripts:

Samples in different columns
 First:
 Second:

Summarized data

	Sample size:	Mean:	Standard deviation:
First:	<input type="text"/>	<input type="text"/>	<input type="text"/>
Second:	<input type="text"/>	<input type="text"/>	<input type="text"/>

Assume equal variances

Select

Graphs... Options...

Help

OK Cancel

Session

```

phân cũ      N   Mean   StDev  SE Mean
phân mới    10  51.91   3.37   1.1

Difference = mu (phân cũ) - mu (phân mới)
Estimate for difference: -4.64
95% CI for difference: (-7.93, -1.35)
T-Test of difference = 0 (vs not =): T-Value = -2.99  P-Value = 0.009  DF = 16
Both use Pooled StDev = 3.2733
  
```

IV. ANOVA – CRD – Một nhân tố

- Nếu dùng công thức để tính ANOVA thì rất phức tạp và đặc biệt với những số liệu có nhiều chữ số thì càng phức tạp trong tính toán.
- Trong Minitab các thuật ngữ có ý nghĩa sau:
 - **Response** là số liệu ta thu thập được trong thí nghiệm
 - **Factor** là nhân tố thí nghiệm với các mức khác nhau
- Nhập số liệu theo 3 cột: lặp lại, nhân tố và kết quả
- Tiến hành phân tích ANOVA one-way

IV. ANOVA – CRD – Một nhân tố

Nghiêm thức	Lặp lại I	Lặp lại II	Lặp lại III	Lặp lại IV	Lặp lại V
Vitamin 1	4.5	5.2	6.2	3.9	4.9
Vitamin 2	5.6	4.7	4.3	4.4	6.1
Vitamin 3	6.4	6.7	6.8	6.1	6.9
Vitamin 4	5.2	5.0	6.8	3.6	5.6
Vitamin 5	4.0	4.9	4.3	4.8	4.2
Vitamin 6	7.1	6.5	6.2	6.8	6.1
Vitamin 7	6.1	4.9	4.2	3.9	6.8
Không Vit.	4.6	4.0	4.9	3.8	4.2

IV. ANOVA – CRD – Một nhân tố

Bảng phân tích phương sai

Nguồn biến động	SS	df	MS	F
Tổng cộng	45.124	39	1.157	1.905*
Nghiệm thức	25.692	7	3.670	6.044**
Sai số	19.432	32	0.607	

Minitab - Untitled

File Edit Data Calc Stat Graph Editor Tools Window Help Assistant

Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Reliability/Survival
Multivariate
Time Series
Tables
Nonparametrics
EDA
Power and Sample Size

One-Way...
One-Way (Unstacked)...
Two-Way...
Analysis of Means...
Balanced ANOVA...
General Linear Model...
Fully Nested ANOVA...
Balanced MANOVA...
General MANOVA...
Test for Equal Variances...
Interval Plot...
Main Effects Plot...
Interactions Plot...

Session

6/23/2013 6:14:20 PM

Welcome to Minitab, press F1 for help.

MTB >

Worksheet 1 ***

	C1	C2	C3
	Lặp lại	Vitamin	
1	1	1	4.5
2	1	2	5.6
3	1	3	6.4

Minitab - Untitled

File Edit Data Calc Stat Graph Editor Tools Window Help Assistant

Session

6/23/2013 6:14:20 PM

Welcome to Minitab, press F1 for help.

MTB > Oneway 'Tăng trọng' 'Vitamin';
SUBC> Tukey 5;

One-way ANOVA: Tăng trọng versus Vitamin

Source	DF	SS	MS	F	P
Vitamin	7	25.692	3.670	6.04	0.000
Error	32	19.432	0.607		
Total	39	45.124			

S = 0.7793 R-Sq = 56.94% R-Sq(adj) = 47.52%

Grouping Information Using Tukey Method

Vitamin	N	Mean	Grouping
3	5	6.5800	A
6	5	6.5400	A
4	5	5.2400	A B
7	5	5.1800	A B
2	5	5.0200	A B
1	5	4.9400	B
5	5	4.4400	B
0	5	4.3000	B

Means that do not share a letter are significantly different.

One-Way Analysis of Variance

Response: 'Tăng trọng'

Factor: Vitamin

Store residuals
 Store fits

Confidence level: 95.0

Select Comparisons... Graphs...
Help OK Cancel

One-Way Multiple Comparisons

Tukey's, family error rate:
 Fisher's, individual error rate:
 Dunnett's, family error rate:
Control group level:
 Hsu's MCB, family error rate:
 Largest is best
 Smallest is best

Help OK Cancel

N NHIÊN

V. ANOVA – CRD – Hai nhân tố

- Có 4 loại giống lúa mì được đưa đi trồng với chế độ bón phân khác, nhau để xác định năng suất và kỹ thuật trồng cho từng loại.
 - Thí nghiệm được tiến hành với 4 lần lặp lại, kết quả được trình bày trong bảng 1.
 - Hãy kiểm định phương sai bằng phép thử hai nhân tố để đánh giá loại giống và phương thức chăm sóc nào phù hợp nhất.
- 4 lần lặp lại ở 4 năm, mỗi năm làm thí nghiệm có đầy đủ các nghiệm thức nên đây là cách bố trí RCB. Nếu không quan tâm đến tác động của thời tiết thì không tính tác động lặp lại, chọn phân tích CRD

V. ANOVA – CRD – Hai nhân tố

Hàm lượng phân bón	Giống lúa Authur			
	Năm I	Năm II	Năm III	Năm IV
Đạm 40 lb/arce	72	74	76	70
Đạm 80 lb/arce	76	75	74	78
Đạm 120 lb/arce	72	74	73	75
Đạm 160 lb/arce	74	76	82	86
	Giống lúa Auburn			
Đạm 40 lb/arce	60	62	64	65
Đạm 80 lb/arce	61	63	69	68
Đạm 120 lb/arce	70	72	69	70
Đạm 160 lb/arce	72	7	82	86
	Giống lúa Caldwell			
Đạm 40 lb/arce	75	73	72	80
Đạm 80 lb/arce	77	78	77	82
Đạm 120 lb/arce	80	82	86	88
Đạm 160 lb/arce	84	82	84	89
	Giống lúa Compton			
Đạm 40 lb/arce	65	68	63	72
Đạm 80 lb/arce	68	72	74	76
Đạm 120 lb/arce	69	68	70	72
Đạm 160 lb/arce	73	75	74	76

V. ANOVA – CRD – Hai nhân tố

Bảng phân tích phương sai

Nguồn biến động	SS	df	MS	F
Tổng cộng	2865.9375	63	45.49107	3.736 ^{ns}
Nghiệm thức	2281.4375	15	152.0958	12.490 ^{**}
Nhân tố lúa	1282.1875	3	427.3958	35.098 ^{**}
Nhân tố phân	761.3125	3	253.7708	20.840 ^{**}
Tương tác LxP	237.9375	9	26.4375	2.171 [*]
Sai số	584.5	48	12.17708	

Minitab - Untitled

File Edit Data Calc Stat Graph Editor Tools Window Help Assistant

Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Reliability/Survival
Multivariate
Time Series
Tables
Nonparametrics
EDA
Power and Sample Size

One-Way...
One-Way (Unstacked)...
Two-Way...
Analysis of Means...
Balanced ANOVA...
General Linear Model...
Fully Nested ANOVA...
Balanced MANOVA...
General MANOVA...
Test for Equal Variances...
Interval Plot...
Main Effects Plot

Two-Way Analysis of Variance

C1 Lấp lại
C2 Phân
C3 Lúa mì
C4 Năng suất

Response: Năng suất

Row factor: Phân Display means

Column factor: Lúa mì Display means

Store residuals
 Store fits

Confidence level: 95.0

Fit additive model

Select Help OK Cancel

Session

Two-way ANOVA: M

Source	DF	SS	MS	F	P
Phân	3	761.31	253.771	20.84	0.000
Lúa mì	3	1282.19	427.396	35.10	0.000
Interaction	9	237.94	26.438	2.17	0.041
Error	48	584.50	12.177		
Total	63	2865.94			

S = 3.490 R-Sq =

MINITAB ANOVA 2.M...

C1 C2 C3 C4

Two-way ANOVA: Năng suất versus Phân, Lúa mì

Source	DF	SS	MS	F	P
Phân	3	761.31	253.771	20.84	0.000
Lúa mì	3	1282.19	427.396	35.10	0.000
Interaction	9	237.94	26.438	2.17	0.041
Error	48	584.50	12.177		
Total	63	2865.94			

S = 3.490 R-Sq = 79.61% R-Sq(adj) = 73.23%

VI. ANOVA – RCB – một nhân tố

- Cũng bài tập như ANOVA CRD một nhân tố nhưng ta chọn RCB

Bảng phân tích ANOVA

Nguồn biến động	SS	df	MS	F
Tổng cộng	45.124	39	1.157	2.15*
Nghiệm thức	25.692	7	3.67	6.81**
Lặp lại	4.349	4	1.09	2.02 ^{ns}
Sai số	19.432	28	0.54	

Minitab - Untitled

File Edit Data Calc Stat Graph Editor Tools Window Help Assistant

Basic Statistics
Regression

General Linear Model

C1 Lặp lại Responses: 'Tăng trọng'
C2 Vitamin Model:
C3 Tăng trọng

Minitab - Untitled

File Edit Data Calc Stat Graph Editor Tools Window Help Assistant

Next Brushed Row (F4)

Session

6/23/2013 7:43:24 PM

Welcome to Minitab, press F1 for help.

General Linear Model: Tăng trọng versus Lặp lại, Vitamin

Factor	Type	Levels	Values
Lặp lại	random	5	1, 2, 3, 4, 5
Vitamin	fixed	8	0, 1, 2, 3, 4, 5, 6, 7

Analysis of Variance for Tăng trọng, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Lặp lại	4	4.3490	4.3490	1.0872	2.02	0.119
Vitamin	7	25.6920	25.6920	3.6703	6.81	0.000
Error	28	15.0830	15.0830	0.5387		
Total	39	45.1240				

S = 0.733947 R-Sq = 66.57% R-Sq(adj) = 53.44%

Unusual Observations for Tăng trọng

Obs	Tăng trọng	Fit	SE Fit	Residual	St Resid
20	6.80000	5.42250	0.40200	1.37750	2.24 R

VII. ANOVA – RCB – Hai nhân tố

- Có 4 loại giống lúa mì được đưa đi trồng với chế độ bón phân khác, nhau để xác định năng suất và kỹ thuật trồng cho từng loại.
 - Thí nghiệm được tiến hành với 4 lần lặp lại ở 4 năm khác nhau, kết quả được trình bày trong bảng 1.
 - Hãy kiểm định phương sai bằng phép thử hai nhân tố để đánh giá loại giống và phương thức chăm sóc nào phù hợp nhất.
- 4 lần lặp lại ở 4 năm, mỗi năm làm thí nghiệm có đầy đủ các nghiệm thức nên đây là cách bố trí RCB.

V. ANOVA – RCB – Hai nhân tố

Bảng phân tích phương sai

Nguồn biến động	SS	df	MS	F
Tổng cộng	2865.9375	63	45.49107	3.502**
Lặp lại	257.5625	3	85.85417	6.610**
Nghiệm thức	2281.4375	15	152.0958	11.710**
Nhân tố lúa	1282.1875	3	427.3958	32.905**
Nhân tố phân	761.3125	3	253.7708	19.537**
Tương tác LxP	237.9375	9	26.4375	2.035**
Sai số	584.5	45	12.98889	

Minitab - Untitled

File Edit Data Calc Stat Graph Editor Tools Window Help Assistant

Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Reliability/Survival
Multivariate
Time Series
Tables
Nonparametrics
EDA
Power and Sample Size

One-Way...
One-Way (Unstacked)...
Two-Way...
Analysis of Means...
Balanced ANOVA...
General Linear Model...
Fully Nested ANOVA...
Balanced MANOVA...
General MANOVA...
Test for Equal Variances...
Interval Plot...
Main Effects Plot...
Interactions Plot...

Session

Welcome to Minitab

Worksheet 1 ***

	C1	C2	C3-T	C4
	Lấp lại	Phân	Lúa mì	Năng suất
1	1	40	Authur	72

General Linear Model

Responses: 'Năng suất'

Model:
'Lấp lại' Phân 'Lúa mì' Phân* 'Lúa mì'

Random factors:
'Lấp lại'

Covariates... Options... Comparisons...
Graphs... Results... Storage...
Factor Plots...
Select
Help
OK Cancel

Session

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Lấp lại	3	257.56	257.56	85.85	11.82	0.000
Phân	3	761.31	761.31	253.77	34.93	0.000
Lúa mì	3	1282.19	1282.19	427.40	58.83	0.000
Phân*Lúa mì	9	237.94	237.94	26.44	3.64	0.002
Error	45	326.94	326.94	7.27		
Total	63	2865.94				

S = 2.69542 R-Sq = 88.59% R-Sq(adj) = 84.03%

**CẢM ƠN SỰ QUAN TÂM
CỦA QUÝ VỊ**

KHOA MÔI TRƯỜNG VÀ TÀI NGUYÊN THIÊN NHIÊN