

# Analyzing a Pilot Survey on Ways Practicing Engineers Learn Statistics

Jorge Luis Romeu

Research Professor, Syracuse University

[jlromeu@syr.edu](mailto:jlromeu@syr.edu)

ASA/JSM, Seattle WA. Aug. 2006

# Outline

- Problem Statement and Assumptions
- Overview of the Survey Sample
- Main Survey Results
- Analysis of variable relationships
  - Parametric, and NonParametric
  - On Percentages, and On Ranks
  - Univariate, and Multivariate
- Caveats and interpretations
- Conclusions

# Problem Statement & Assumptions

## Undergrad Engineering Statistics is:

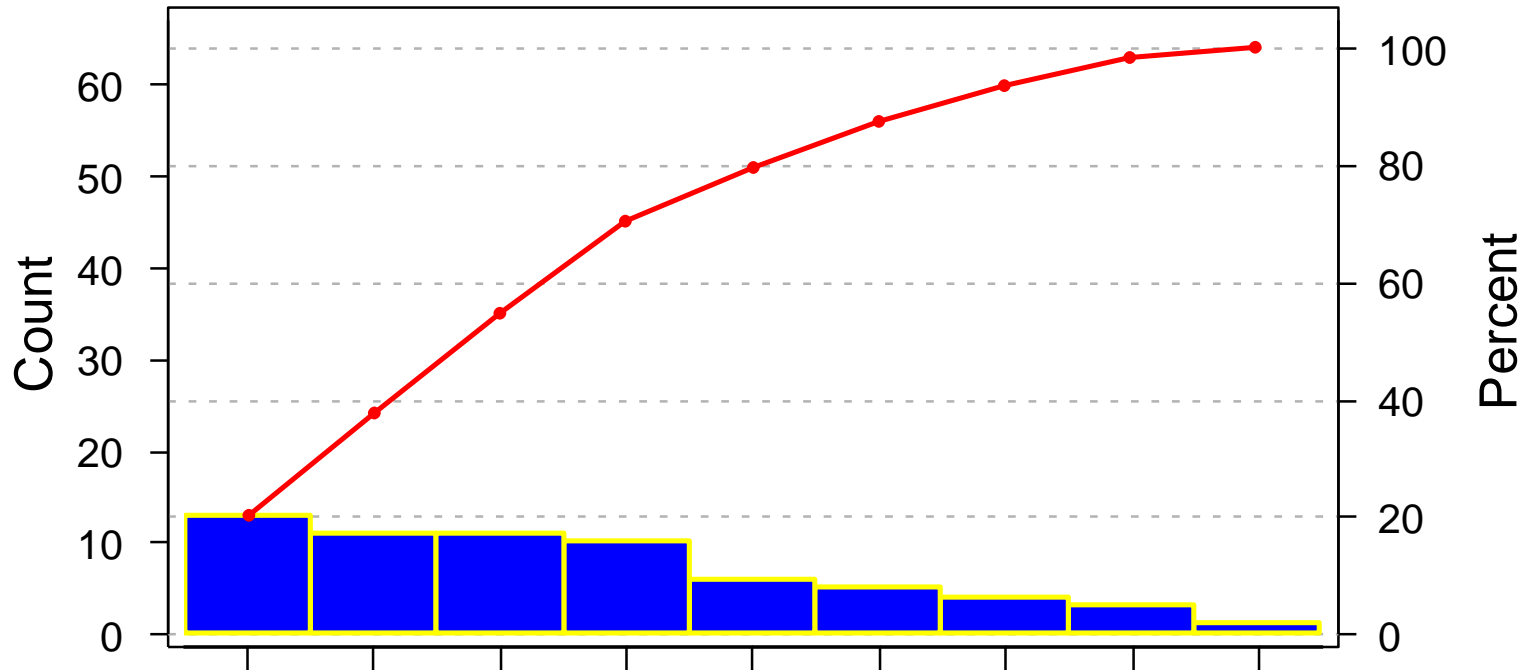
- Insufficiently and inadequately taught
  - Very few courses (mainly none or one)
  - Highly theoretical and over-stuffed
- After graduation, Engineers find they need it
- Then, they have to study Stats on their own
  - As best they can, using different means
  - As practicing professionals (after college)
- Because they need it in engineering work!

# Survey Methods of Self-Learning

- (1) reading books, journals, manuals or other hard copy,*
- (2) reading Web/Internet tutorials,*
- (3) following on-line courses or learning software,*
- (4) attending conferences and chapter meeting talks,*
- (5) preparation for professional certifications,*
- (6) taking short, professional training courses,*
- (7) receiving mentoring from more experienced colleagues*
- (8) other sources: e.g. hands-on (practical) working experiences, and taking Six Sigma training*

# Sample Description

- We received 64 responses:
- 61 from the US
- 3 from abroad.
- 8% were females
- 56% had graduate degrees
- 60% had 16+ years of experience
- 90% were from industry.



Defect

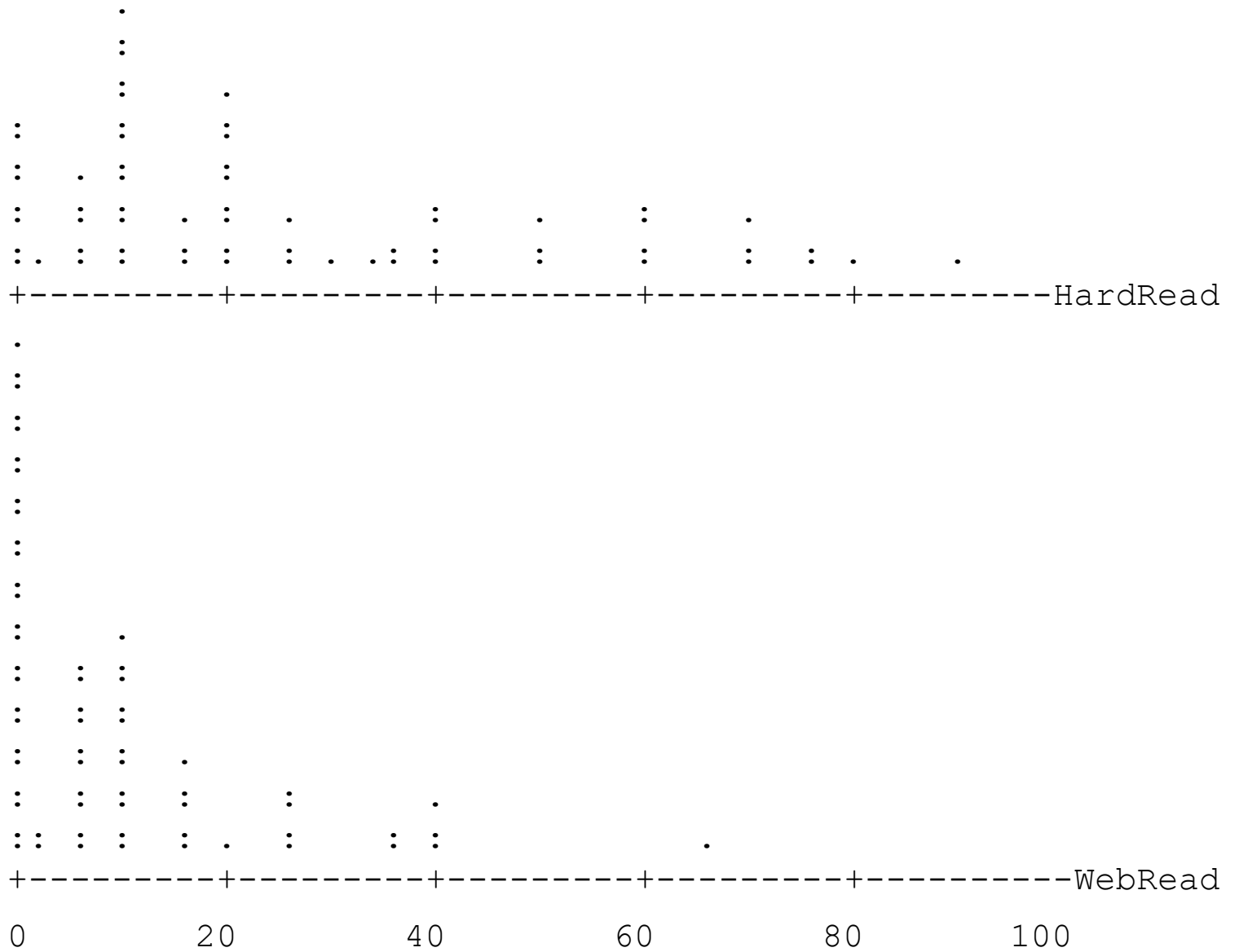
Count  
Percent  
Cum %

Defect	Count	Percent	Cum %
Q&R	13	20.3	20.3
ME	11	17.2	37.5
Chem	11	17.2	54.7
IE	10	15.6	70.3
Unkn	6	9.4	79.7
EE	5	7.8	87.5
Materials	4	6.2	93.7
Civil	3	4.7	98.4
Others	1	1.6	100.0

## ICOTS Survey on Practicing Engineers Statistical Education (%).

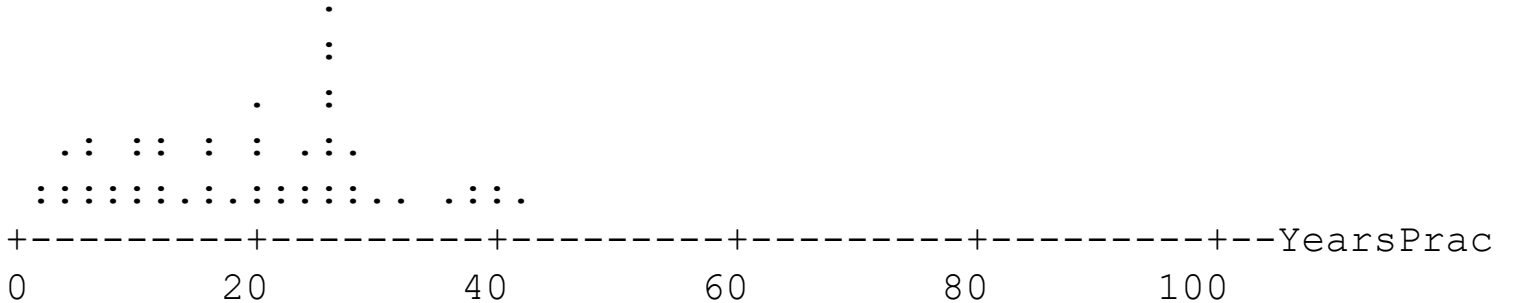
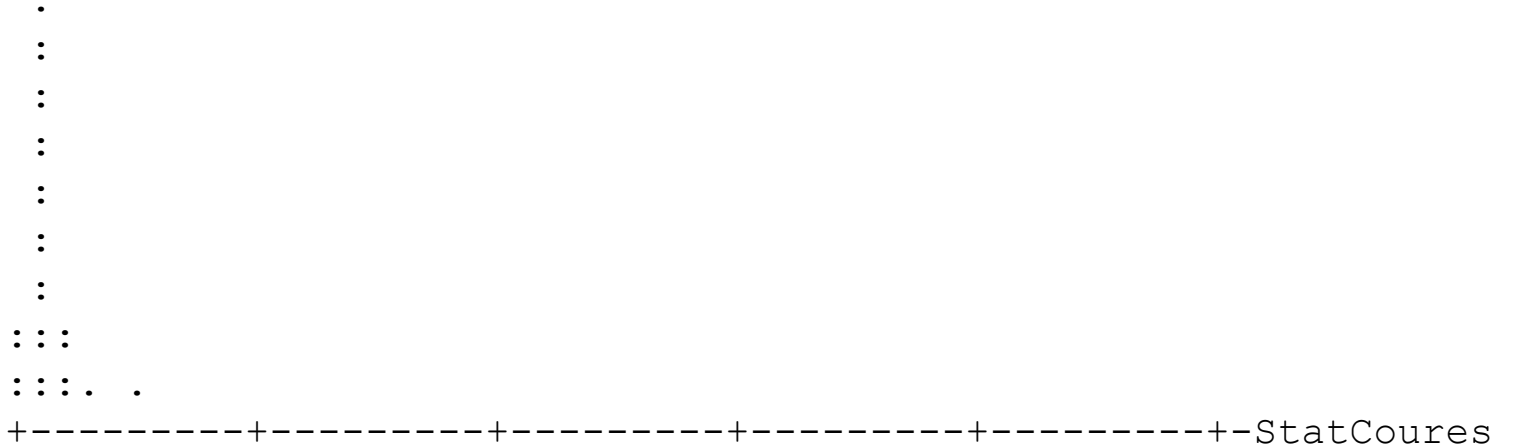
HardR	WebR	Tutor	ProfMg	Certif	ShortC	Mentor	Other	Educ	Area	Spclz	Cours	Years
15	10	5	10	10	30	20	0	BS	I	ME	2	10
35	35	0	0	0	10	20	0	BS	I	SW	2	1
70	10	10	0	0	5	5	0	PhD	I	Chm	3	7
35	15	0	0	0	25	25	0		I	Chm		
5	10	5	5	20	10	15	30		I	Chm		
5	0	0	2.5	10	2.5	0	80	BS	I	IE	1	13
25	0	0	10	15	0	10	40	MS	I	IE	2	11
10	0	0	0	75	15	0	0	BS	I	EE	0	25

# Dotplots of the Distributions:





More Dotplots:



# College Statistical Training

*I) Among all surveyed, 16% have not taken any statistics courses in college (33% among BS), 38% took only one (38%) and 26% have taken 2 courses (24%).*

*II) 1/3 of those with a BS degree only have never taken a single statistics course in college; another 1/3 of them have taken only one course. Hence, 2/3 engineers of all surveyed had either none, or very little statistical training (i.e. taken a single course).*

*III) Engineers that pursue graduate school have a larger opportunity of taking statistics. Only 7%, in our sample, have never taken a statistics course.*

# Methods Preferred

*I) “Readings” constitute the preferred means of learning: books and journals, as well as web tutorials, provide 38% of statistics knowledge. The use of web tutorials (10%) is increasing with time: older engineers prefer hard copy, whereas younger ones read web-based material.*

*II) Short courses, exam preparations for the professional certifications, and Black Belt training, are also important methods of learning statistics (33%).*

*III) mentoring received from more experienced colleagues and hands-on (learning by doing), also constitute frequent learning activities (22%).*

# Parametric Confidence Intervals

Variable	N	Mean	StDev	SE Mean	95.0 % CI	
HardRead	64	26.08	24.31	3.04	( 20.00,	32.15)
WebRead	64	9.58	13.22	1.65	( 6.28,	12.88)
OnLnTuto	64	2.016	4.682	0.585	( 0.846,	3.185)
ProfMtgs	64	6.43	9.74	1.22	( 4.00,	8.86)
Certific	64	12.47	18.44	2.31	( 7.86,	17.08)
ShortCou	64	12.15	16.05	2.01	( 8.14,	16.16)
Mentorin	64	14.41	14.58	1.82	( 10.76,	18.05)
OtherWay	64	15.47	24.19	3.02	( 9.43,	21.51)
StatCour	51	1.725	1.686	0.236	( 1.251,	2.200)
YrsPract	55	18.93	10.80	1.46	( 16.01,	21.85)

# Wilcoxon Signed Rank C.I.

	Number	Estimated	Achieved			
	N	Missing	Median	Confidence	Confidence	Interval
HardRead	64	0	22.5	95.0	( 15.0,	30.0)
WebRead	64	0	7.50	95.0	( 5.00,	10.00)
OnLnTuto	64	0	0.00	95.0	( 0.00,	2.50)
ProfMtgs	64	0	5.00	95.0	( 2.50,	6.50)
Certific	64	0	10.0	95.0	( 5.0,	15.0)
ShortCou	64	0	10.0	95.0	( 5.0,	15.0)
Mentorin	64	0	12.50	95.0	( 10.00,	15.00)
OtherWay	64	0	10.0	95.0	( 5.0,	17.5)
StatCour	51	13	1.50	95.0	( 1.00,	2.00)
YearsPra	55	9	18.50	95.0	( 15.50,	22.50)

# Variable Parametric Correlation

	HardRd	WebRd	OnLnTutor	ProfMtgs	Certif	ShortCr	Mentor
WebRead	-0.070						
OnLnTuto	-0.137	0.122					
ProfMtgs	<b>-0.199</b>	-0.090	0.158				
Certific	-0.127	<b>-0.248</b>	0.012	-0.130			
ShortCou	<b>-0.330</b>	-0.054	-0.063	0.050	-0.160		
Mentorin	<b>-0.233</b>	0.061	-0.054	-0.090	-0.171	0.106	
OtherWay	<b>-0.334</b>	-0.230	-0.096	-0.025	<b>-0.201</b>	<b>-0.220</b>	<b>-0.228</b>
StatCour	<b>0.240</b>	0.159	0.058	-0.025	-0.164	-0.032	-0.035
YearsPrac	0.111	<b>-0.389</b>	-0.166	0.010	-0.104	<b>0.195</b>	0.016

Corr (StatCourses, OtherWay) = **-0.223**  
Corr (YearsPract, OtherWay) = -0.138  
Corr (StatCour , YearsPract) = -0.098

# Some Relevant Research Questions

- Are the groups homogeneous (merge)?
  - By engineering specialization?
  - By education levels (BS/MS/PhD)?
- Are there trends on practicing time?
  - More stats courses taken, or different readings?
- Are there multivariate associations?
  - On which survey variables (p-values)?
  - By how much (size of effects)?
- Parametric as well as Non-Parametric?

# Education Levels are Homogeneous, except for Hard Read

Parametric ANOVAs Comparing Education Levels (%)

<b>Response</b>	<b>p-value</b>	<b>Lowest</b>	<b>Highest</b>
Hard Read	0.013	BS	MS
Web Read	0.352	MS	PhD
Certifications	0.210	PhD	BS
Short Courses	0.297	PhD	MS



Parametric ANOVAs Comparing Specialization (%)

<b>Response</b>	<b>p-value</b>	<b>Lowest</b>	<b>Highest</b>
Stats Courses	0.083	Material	Chemical
Certification	0.103	Mechanical	Electrical
Hard Read	0.388	Material	Civil
Web Read	0.278	Civil	Chemical
Short Courses	0.153	Civil	Mechanical
Other Ways	0.117	Q&R	Material

Non-Parametric (K-W) Comparing Specialization (%)

<b>Response</b>	<b>p-value</b>	<b>Lowest</b>	<b>Highest</b>
Hard Read	0.454	Material	Civil
Web Read	0.23	Civil	Q&R
Prof. Meetings	0.232	Civil	Chemical
Certifications	0.344	Chemical	Material
Short Courses	0.24	Civil	Material
Mentoring	0.164	Industrial	Mechanical
Stats Courses	0.048	Material	Civil

Specializations  
are  
Homogeneous,  
except in  
number of stats  
courses taken.

# Example of ANOVAs comparing Specializations

## One-Way Analysis of Variance for HardRead on Specialization:

Analysis of Variance for HardRead

Source	DF	SS	MS	F	P
Speciali	6	3505	584	1.08	0.388
Error	51	27615	541		
Total	57	31119			

Individual 95% CIs For Mean  
Based on Pooled StDev

Level	N	Mean	StDev	CI
ME	11	24.55	27.61	(-----*-----)
EE	6	35.83	21.54	(-----*-----)
Chem	11	26.36	19.38	(-----*-----)
IE	10	25.50	28.23	(-----*-----)
Material	4	7.75	6.08	(-----*-----)
Q&R	13	23.69	16.12	(-----*-----)
Civil	3	48.33	41.93	(-----*-----)

-----+-----+-----+-----  
 0 25 50

Pooled StDev = 23.27

# Equivalent Non Parametric Comparison.

Kruskal-Wallis Test on HardRead

58 cases were used

6 cases contained missing values

Speciali	N	Median	Ave Rank	Z
ME	11	15.000	26.5	-0.66
EE	6	42.500	37.2	1.19
Chem	11	20.000	31.5	0.45
IE	10	15.000	27.9	-0.32
Material	4	7.500	14.8	-1.81
Q&R	13	20.000	30.5	0.25
Civil	3	70.000	38.0	0.90
Overall	58		29.5	
H = 5.73	DF = 6	P = 0.454		
H = 5.81	DF = 6	P = 0.444	(adjusted for ties)	

## Stats Courses Taken are Significantly Different:

Kruskal-Wallis Test on StatCour

50 cases were used

14 cases contained missing values

Speciali	N	Median	Ave Rank	Z
ME	11	1.00E+00	19.5	-1.53
EE	6	1.50E+00	26.2	0.13
Chem	6	2.00E+00	35.6	1.81
IE	10	2.00E+00	31.2	1.39
Material	4	0.00E+00	7.9	-2.52
Q&R	10	1.00E+00	25.0	-0.13
Civil	3	2.00E+00	31.8	0.78
Overall	50		25.5	
H = 12.71	DF = 6	P = 0.048		
H = 13.78	DF = 6	P = 0.032	(adjusted for ties)	

## Regression Analysis for Web Readings on Years :

Regression equation:  $\text{WebRead} = 18.5 - 0.483 \text{ YearsPract}$

55 cases used 9 cases contain missing values

Predictor	Coef	StDev	T	P
Constant	18.451	3.418	5.40	0.000
YearsPra	-0.4830	0.1572	-3.07	0.003

S = 12.47      R-Sq = 15.1%      R-Sq(adj) = 13.5%

## Regression Analysis for Hard Copy Readings on Years

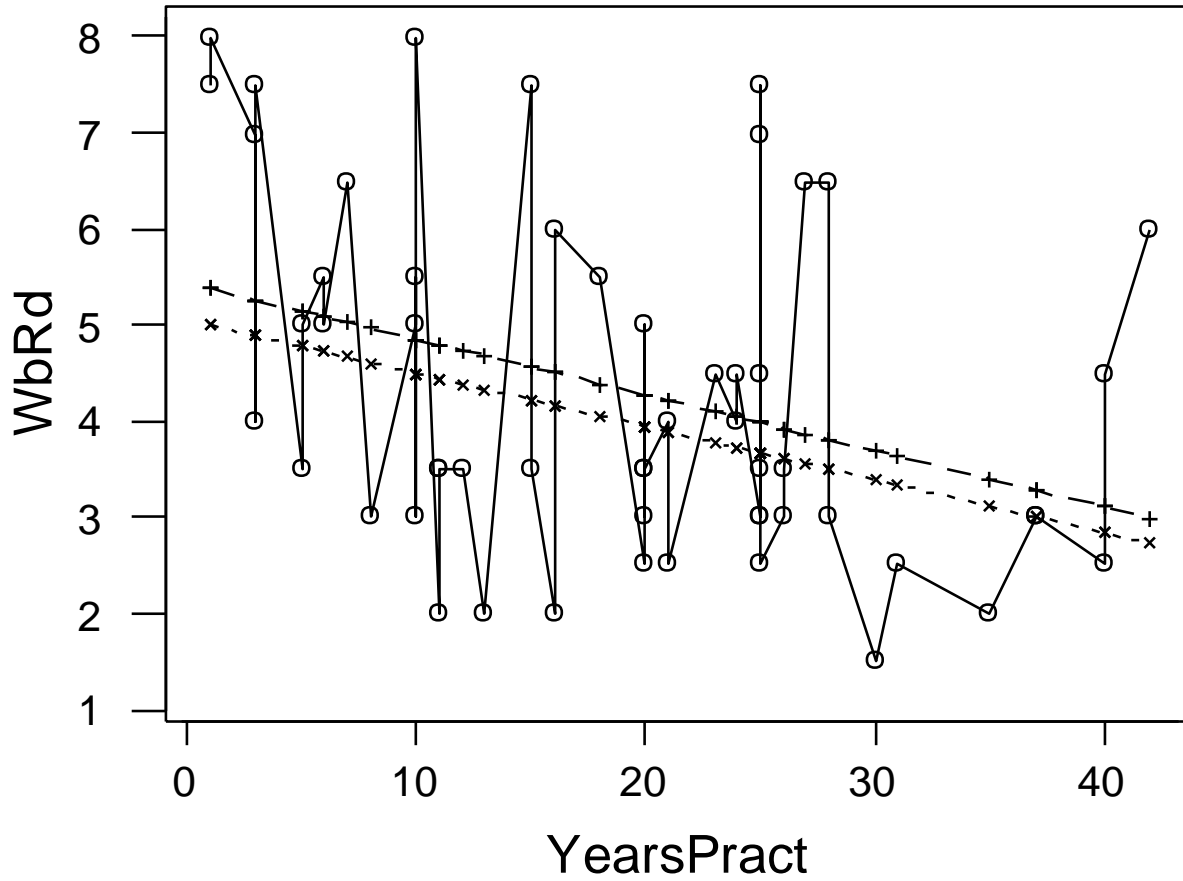
Regression equation:  $\text{HardRead} = 22.5 + 0.255 \text{ YrsPrct}$

55 cases used 9 cases contain missing values

Predictor	Coef	StDev	T	P
Constant	22.518	6.790	3.32	0.002
YearsPra	0.2551	0.3123	0.82	0.418

S = 24.78      R-Sq = 1.2%      R-Sq(adj) = 0.0%

# Parametric vs. Non Parametric Regression Analysis of Ranks:



## Sen's Non Parametric Regression:

### Intercepts:

Parametric = 5.4428

NonParametric = 5.05556

### Slopes:

Parametric = -0.05846

NonParametric = -0.0555556

## **Sign Non-Parametric CI for the Regression Slope:**

	N	N*	Median	Confidence	Confidence Interval
<b>Slopes</b>	<b>1431</b>	<b>585</b>	<b>-0.05556</b>	<b>0.9496</b>	<b>(-0.06818,-0.04054)</b>

## Non significant effect of practicing times:

### Regression Analysis for Stats Courses Taken

Regression equation:  $\text{StatCours} = 2.02 - 0.0158 \text{ YearsPrat}$

51 cases used 13 cases contain missing values

Predictor	Coef	StDev	T	P
Constant	2.0195	0.4866	4.15	0.000
YearsPra	-0.01578	0.02280	-0.69	0.492

S = 1.695      R-Sq = 1.0%      R-Sq(adj) = 0.0%

### Regression Analysis for Short Courses Taken

Regression equation:  $\text{ShortCours} = 6.45 + 0.305 \text{ YearsPrct}$

55 cases used 9 cases contain missing values

Predictor	Coef	StDev	T	P
Constant	6.445	4.578	1.41	0.165
YearsPra	0.3055	0.2105	1.45	0.153

S = 16.71      R-Sq = 3.8%      R-Sq(adj) = 2.0%



Factors that appear to substitute for practical time & experience:

**Multivariate Regression Analysis (Percents):**

The regression equation is:

$$\text{YearsPract} = 26.7 - 0.407 \text{ WebRead} - 0.159 \text{ Certificats} - 0.141 \text{ OtherWays}$$

55 cases used 9 cases contain missing values

Predictor	Coef	StDev	T	P
Constant	26.676	2.327	11.46	0.000
WebRead	-0.4070	0.1023	-3.98	0.000
Certific	-0.15919	0.07110	-2.24	0.030
OtherWay	-0.14147	0.06306	-2.24	0.029

S = 9.524

R-Sq = 26.5%

R-Sq(adj) = 22.2%

## ICOTS Survey on Practicing Engineers Statistical Education (Ranks)

HardR	WebR	Tutor	ProfMg	Certif	ShortC	Mentor	Other	Educ	Area	Spclz	Cours	Years
1.5	5	3	5	5	8	7	1.5	BS	I	ME	2	10
7.5	7.5	2.5	2.5	2.5	5	6	2.5	BS	I	SW	2	1
8	6.5	6.5	2	2	4.5	4.5	2	PhD	I	Chm	3	7
8	5	2.5	2.5	2.5	6.5	6.5	2.5		I	Chm		
2	4.5	2	2	7	4.5	6	8		I	Chm		
6	2	2	4.5	7	4.5	2	8	BS	I	IE	1	13
7	2	2	4.5	6	2	4.5	8	MS	I	IE	2	11
6	3	3	3	8	7	3	3	BS	I	EE	0	25
7.5	7.5	3	3	3	3	6	3	PhD	I	Chm		
4	4	1.5	4	7	8	6	1.5	BS	I	Matr	0	21
5.5	5.5	2.5	2.5	2.5	2.5	7	8	PhD	I	Matr	0	10

Note: the highest the rank, the most preferred method.

**Sign confidence interval for median of rank data (1):**

	N	N*	Median	Confidence	Confidence Interval	Pos
HrdRd	64	0	6.000	0.9392	( 5.500, 7.000)	25
				0.9500	( 5.500, 7.000)	NLI
				0.9664	( 5.500, 7.000)	24
WbRd	64	0	4.000	0.9392	( 3.500, 4.500)	25
				0.9500	( 3.500, 4.641)	NLI
				0.9664	( 3.500, 5.000)	24
OnLnTut	64	0	2.500	0.9392	( 2.500, 3.000)	25
				0.9500	( 2.500, 3.000)	NLI
				0.9664	( 2.500, 3.000)	24
PrfMtgs	64	0	3.500	0.9392	( 3.000, 4.500)	25
				0.9500	( 3.000, 4.500)	NLI
				0.9664	( 3.000, 4.500)	24

Ranks are 1 through 8, the highest being the most preferred.

**Sign confidence interval for median of rank data (2):**

	N	N*	Median	Confidence	Confidence Interval	Pos
Certif	64	0	4.000	0.9392	( 3.000, 5.500)	25
				0.9500	( 3.000, 5.500)	NLI
				0.9664	( 3.000, 5.500)	24
ShrtCrs	64	0	4.750	0.9392	( 3.500, 5.500)	25
				0.9500	( 3.500, 5.641)	NLI
				0.9664	( 3.500, 6.000)	24
Mentor	64	0	6.000	0.9392	( 4.500, 6.000)	25
				0.9500	( 4.500, 6.000)	NLI
				0.9664	( 4.500, 6.000)	24
OthrWys	64	0	3.500	0.9392	( 3.000, 4.500)	25
				0.9500	( 2.859, 4.641)	NLI
				0.9664	( 2.500, 5.000)	24

Ranks are 1 through 8, the highest being the most preferred.

# Preferred means of post-college education depends on Level.

Non-Parametric (Mood) Comparing Education (Ranks)

<b>Response</b>	<b>p-value</b>	<b>Lowest</b>	<b>Highest</b>
Stats Courses	0.028	PhD	MS
Hard Read	0.080	BS	PhD
Web Read	0.052	MS	PhD
Certifications	0.043	PhD	BS
Short Courses	0.114	PhD	MS

Parametric ANOVAs Comparing Specialization (Ranks)

<b>Response</b>	<b>p-value</b>	<b>Lowest</b>	<b>Highest</b>
Hard Read	0.268	Material	Electrical
Web Read	0.569	Material	Chemical
OnLine Tutors	0.236	Material	Industrial
Prof. Meetings	0.033	Chemical	Industrial
Certifications	0.349	Mechanical	Material
Short Courses	0.378	Industrial	Mechanical
Mentoring	0.121	Industrial	Material
Other Ways	0.223	Q&R	Material

Non-Parametric (K-W) Comparing Specialization (Ranks)

<b>Response</b>	<b>p-value</b>	<b>Lowest</b>	<b>Highest</b>
Hard Read	0.268	Material	Civil
Web Read	0.588	Civil	Chemical
OnLine Tutor	0.138	Material	Industrial
Prof. Meetings	0.019	Chemical	Q&R
Certifications	0.475	Mechanical	Material
Short Courses	0.477	Civil	Mechanical
Mentoring	0.166	Industrial	Mechanical

Again, type of specialization does not have a significant effect on the preferred means of learning.

# Non-Parametric Example of Specialization Comparisons:

## Kruskal-Wallis Test of HrdRd on Specialization

58 cases were used

6 cases contained missing values

Speciali	N	Median	Ave Rank	Z
ME	11	5.500	23.9	-1.22
EE	6	7.750	40.2	1.63
Chem	11	6.000	31.2	0.37
IE	10	6.500	30.3	0.16
Material	4	4.750	14.0	-1.90
Q&R	13	6.500	30.5	0.24
Civil	3	8.000	36.2	0.70
Overall	58		29.5	

H = 7.61 DF = 6 P = 0.268

H = 7.79 DF = 6 P = 0.254 (adjusted for ties)

# A different Non Parametric method of comparisons:

## Mood median test for HardRead on Specialization:

Chi-Square = 6.24    DF = 6    P = 0.397

Speciali	N<=	N>	Median	Q3-Q1	Individual 95.0% CIs
Chem	6	5	20.0	30.0	(---+-----)
Civil	1	2	70.0	75.0	(-----)
EE	2	4	42.5	42.5	(-----+-----)
IE	6	4	15.0	25.0	(---+-----)
Material	4	0	7.5	11.8	(---+---)
ME	8	3	15.0	50.0	(-----+-----)
Q&R	8	5	20.0	26.5	(---+-----)

+-----+-----+-----  
0                      25                      50

Overall median = 20.0



Factors obtained with ranks coincide with those obtained with the percentages. Web and Certifications make-up for “years”.

**Multivariate Regression Analysis on Ranks:**

The regression equation is:

$$\text{YearsPract} = 35.3 - 2.61 \text{ WbRd} - 1.17 \text{ Certif}$$

55 cases used 9 cases contain missing values

Predictor	Coef	StDev	T	P
Constant	35.340	5.293	6.68	0.000
WbRd	-2.6068	0.8018	-3.25	0.002
Certif	-1.1658	0.6263	-1.86	0.068

S = 9.983

R-Sq = 17.7%

R-Sq(adj) = 14.5%

# Conclusions

- Statistics not taught sufficiently in school
  - Survey shows 2/3 get none or a single course
  - Study it later, on their own, at their own pace.
- Pilot Survey sheds light on learning tools
- Web Read more; but Hard read, decreasing
- Short courses; certifications; mentoring
- No difference by specialization; but by level
- Learning methods used are complementary.