Statistics in Dyalog APL

Josh David

Statistics in Dyalog APL

- https://github.com/dyalog/pynapl

- https://github.com/kimmolinna/rsconnect

Why the need

- Heavily used in rising fields
  - Data Science
  - Machine Learning
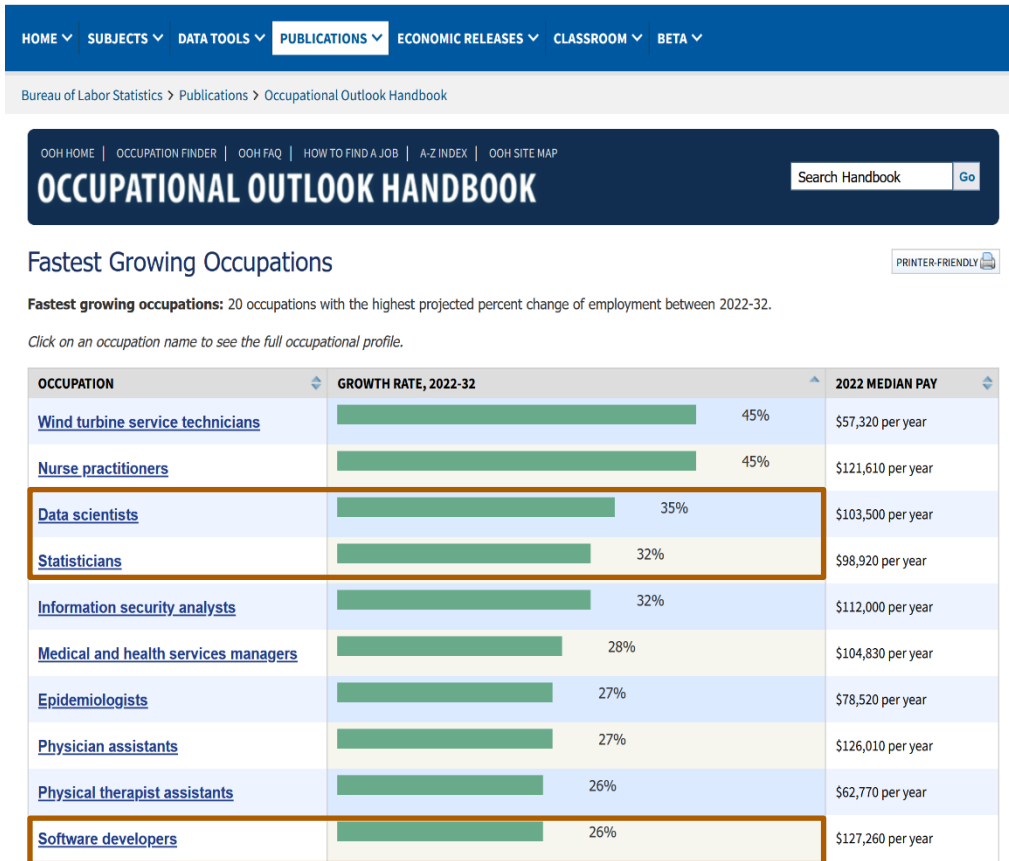- Shift in Company Cultures
  - Data driven
- Opportunity for APL

Figure 1

Where has this been?

- `+⌿` sum

Where has this been?

- `f⌿ max`

Where has this been?

- `⌈≠ max`

Where has this been?

- `⌈≠ max`

- `⌊≠ min`

- `(⌈≠-⌊≠) range`

  `y⊞x*0 mean`

Where has this been?

- ⌈╱ max

- ⌊╱ min

- (⌈╱-⌊╱) range

- (+╱÷≢) mean

Regression ⊡

Arithmetic: $y = \textcolor{red}{f}\textcolor{blue}{(x)}$
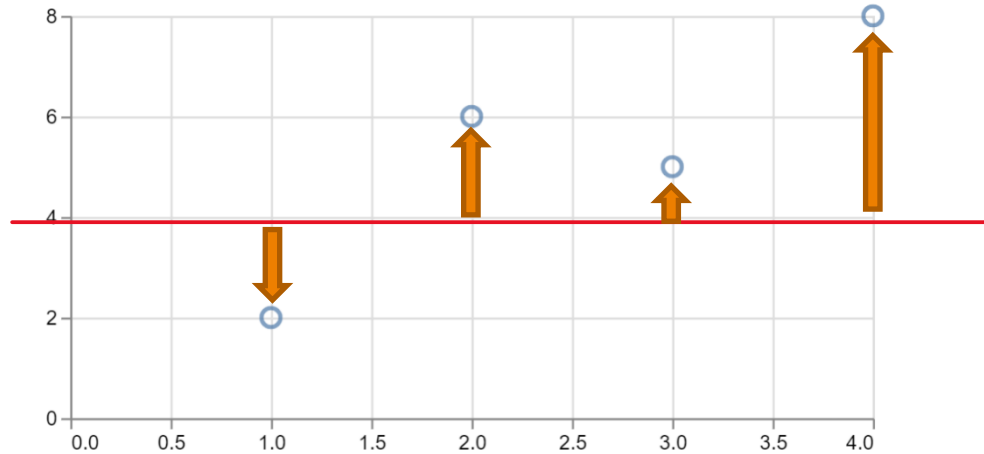
Algebra: $\textcolor{blue}{y} = \textcolor{blue}{f}\textcolor{red}{(x)}$

Regression: $\textcolor{blue}{y} = \textcolor{red}{f}\textcolor{blue}{(x)}$

Regression

Regression

Regression
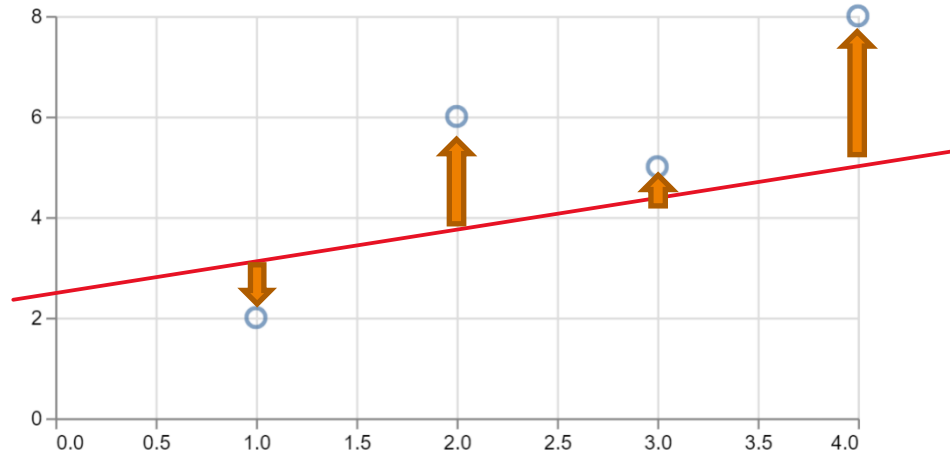
Statistics in Dyalog APL

Regression

Statistics in Dyalog APL

Regression

- `{α⍠ω*0}`   `⍝ Mean`
- `{α⌹ω∘.*0 1}`   `⍝ Linear`
- `{α⌹ω∘.*0 1 2}`   `⍝ Quadratic`
- `{α⌹ω∘.*0 1 2 3}`   `⍝ Cubic`
- `{x y←ω ⋄ y⌹x∘.*⍳1+α}`
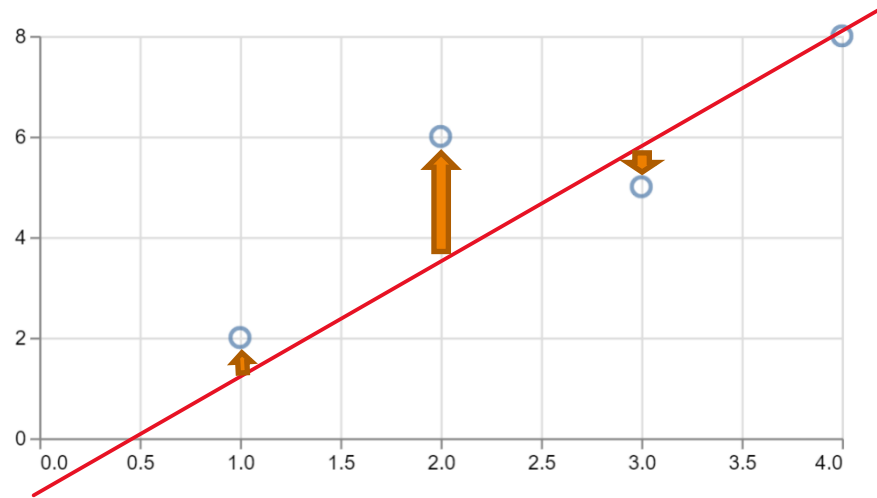
Machine Learning

Figure 2

Open source libraries

KokoStats

TamStat

Statistics in Dyalog APL
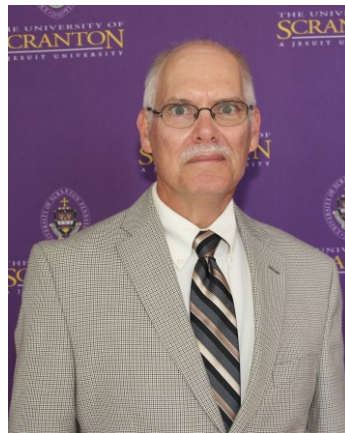
KokoStats

About

Toolkit for Data Science & Statistics

correlation · dft · linear-regression · regression · pca · fft · autocorrelation · polynomial-regression · ifft · anova · distributions · principle-component-analysis · chebyshev-polynomials · multiple-linear-regression · tukey-window · forsythe

📖 Readme
〰️ Activity
☆ 0 stars
👁 1 watching
⑂ 0 forks

Releases 1

🏷 0.0.0 Latest
2 weeks ago

https://github.com/JoshDavid/KokoStats

Demo[1] KokoStats

Correlation heatmaps & Multiple linear
regression

TamStat – Simple Linear Regression

```
⍝ Predictor Variable (X)

ADS←1 3 2 1 3

⍝ Response Variable (Y)

CARS←14 24 18 17 27

⍝ Perform linear regression

MODEL← CARS regress ADS

MODEL.B   ⍝ Intercept and Slope
10 5

  ⍝ If dealer runs 2 ads, how many sales?

  MODEL.f 2
20

  MODEL.f confInt 2   ⍝ Estimate mean sales
16.925 23.075

  MODEL.f predInt 2   ⍝ Predict dealer sales
12.469 27.531
```

```
report MODEL
```

The regression equation is:

$Y←10+(5×X1)+E$

ANOVA Table

SOURCE	SS	DF	MS	F	P
Regression	100.00	1	100.00	21.43	0.01899
Error	14.00	3	4.67		
Total	114.00	4			

$S = 2.16025$ R-Sq = 87.72% R-Sq(adj) = 83.63%

Solution

Variable	Coeff	SE	T	P
Intercept	10.00	2.37	4.22577	0.02424
B1	5.00	1.08	4.62910	0.01899

The TamStat `regress` Operator

Design	Left Argument	Operand / Function	Right Argument	Result
Simple linear regression	Response Variable*	`regress`	Predictor Variable	Intercept, Slope
Multiple Linear Regression	Response Variable*	`regress`	Vector of Predictor Variables	Intercept, Coefficients for each predictor variable.
			Matrix whose columns are predictor variables	
	Name of Response Variable* (Character string)	`regress`	Namespace containing all variables	
Simple Quadratic Regression $\quad y = b_9 + b_1(x - \bar{x}) + b_2(x - \bar{x})^2$	Response Variable	`⊥ regress`	Predictor Variable	Intercept, Coefficients for centered data and squared centered data
Multiple Quadratic Regression	Response Variable	`⊥ regress`	Vector of Predictor Variables	Intercept, Linear, Quadratic and Interaction Coefficients
Polynomial Models	Response Variable	`N ⊥ regress`	Predictor Variable	Intercept, Coefficients for all powers up to N of predictor variable
Model with Indicator Variable(s)	Response Variable*	`regress`	Vector containing Predictor Variables and at least one Character Variable	Intercept, Coefficients for each predictor variable and (k-1) coefficients for each character variable. (k = unique character values)
Variance Stabilizing Transformations	Response Variable	`fn regress [ln|sqrt|÷|arcsin]`	Predictor Variable	Intercept, Coefficients
Multiplicative Regression $\quad y = bx^a$	Response Variable	`× regress`	Predictor Variable	Constant, Powers
Indicator response variable	Boolean Variable	`≠ regress`	Predictor Variable	Intercept, Coefficients
Custom Regression	[None]	`userFn regress`	Database (Namespace)	Intercept, Coefficients

* Pseudo Left Argument – Actually an array left operand.

TamStat

EXCEL   =1-NORM.DIST(150,120,25,TRUE)

R   pnorm(150,120,25,lower.tail=FALSE)

TamStat   120 25 normal probability > 150

TamStat

- Summary Wizard

- Probability Wizard

- Distribution Wizard

- Statistical Tables

- Confidence Intervals

- Hypothesis Wizard

- Sample Size Calculator

- ANOVA

- Regression Wizard

- Chi-Square Tests

- Bayesian Analysis

Statistics in Dyalog APL

Demo[2] TamStat

Random Variable generation

Demo[3] TamStat UI

Distribution Wizard

Confidence interval

References

- Figure 1:
    - https://www.bls.gov/ooh/fastest-growing.htm

- Figure 2:
    - https://twitter.com/ingliguori/status/1644012090976354306

Statistics in Dyalog APL