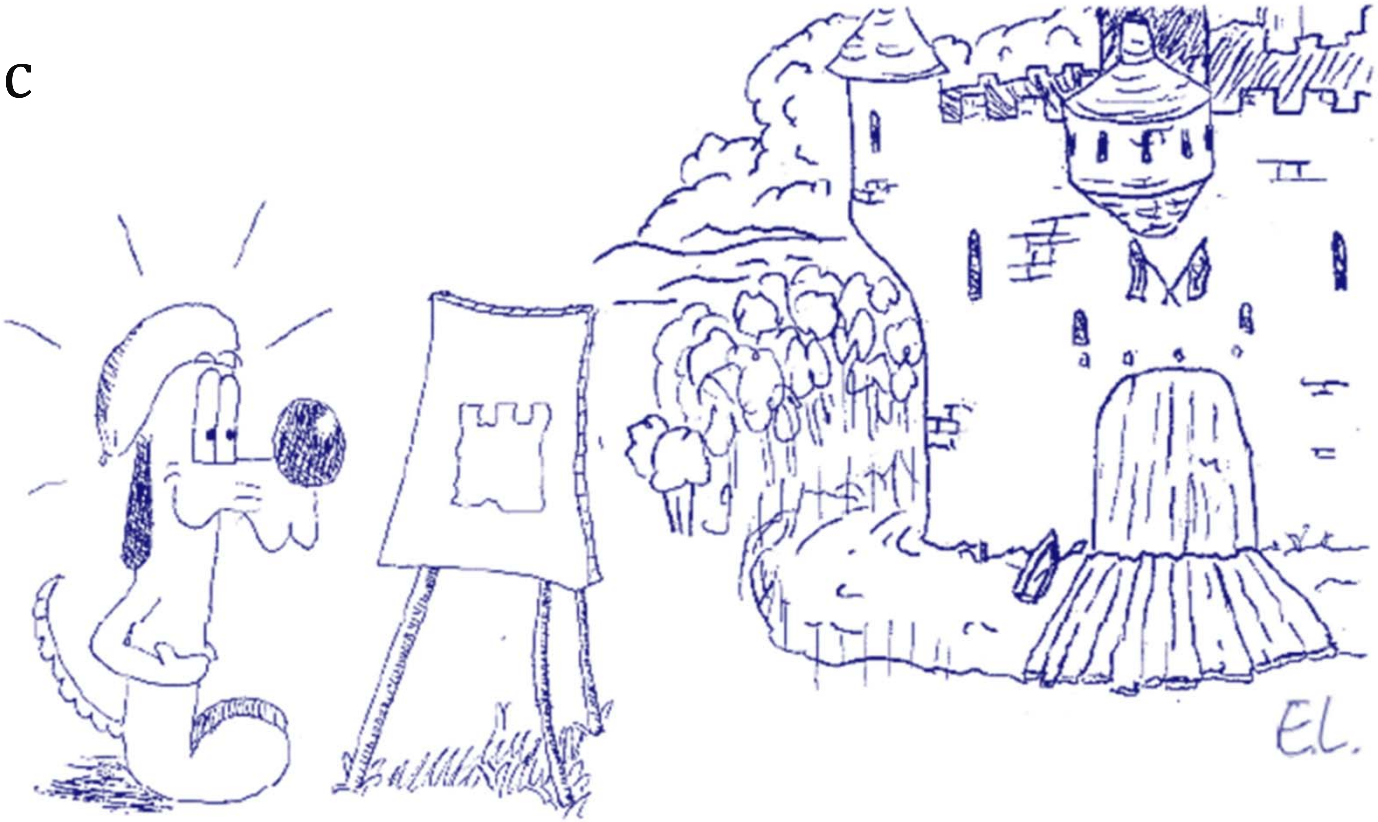
Model Fitting Bonus

JY Le Boudec

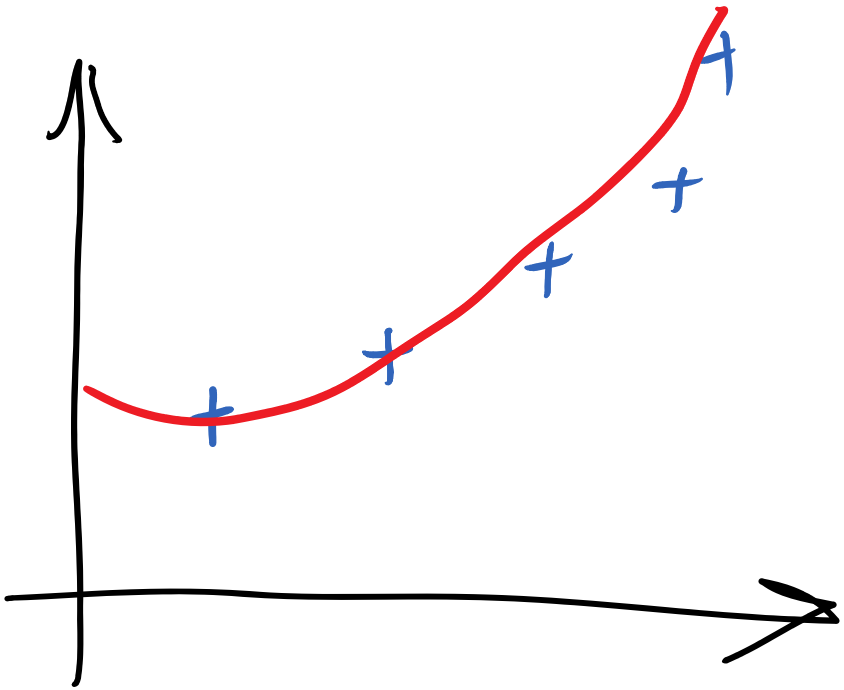
May 2015



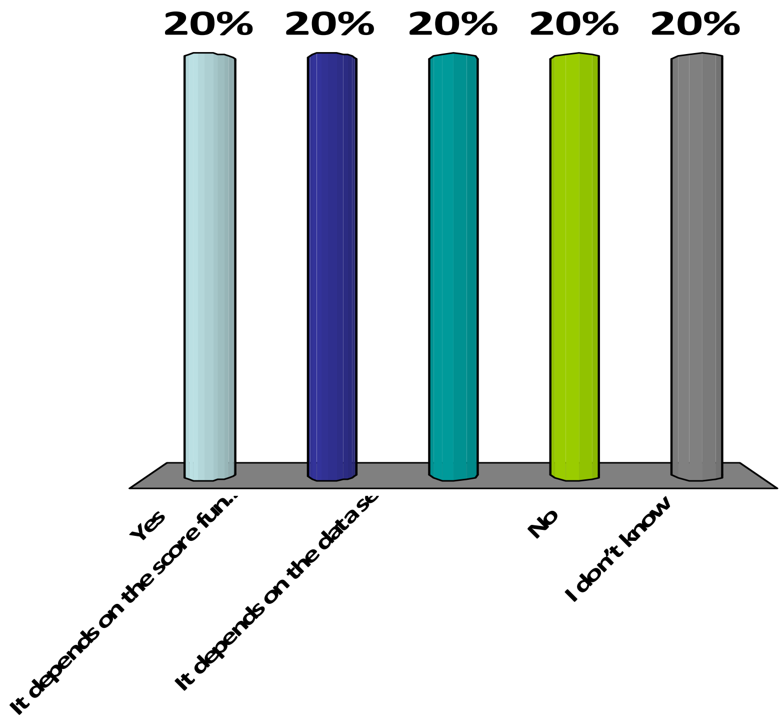
We want to fit a data set y_i to a polynomial of degree 2:

$$y_i = at_i^2 + bt_i + c.$$

Is this a linear regression model ?



A. Yes
B. It depends on the score function
C. It depends on the data set
D. No
E. I don’t know



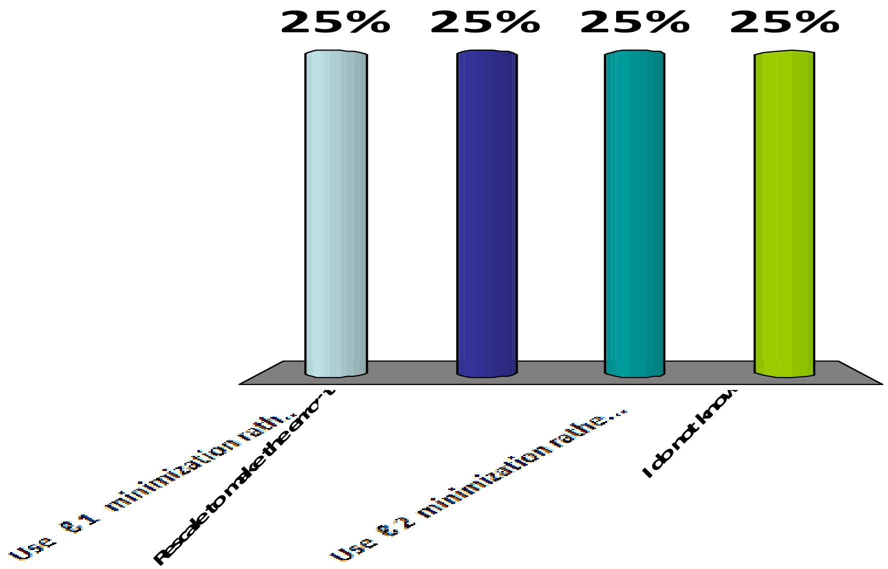
If the error terms in a fitting model are not homoscedastic, it is better to …

A. Use ℓ^1 minimization rather than ℓ^2

B. Rescale to make the error term homoscedastic

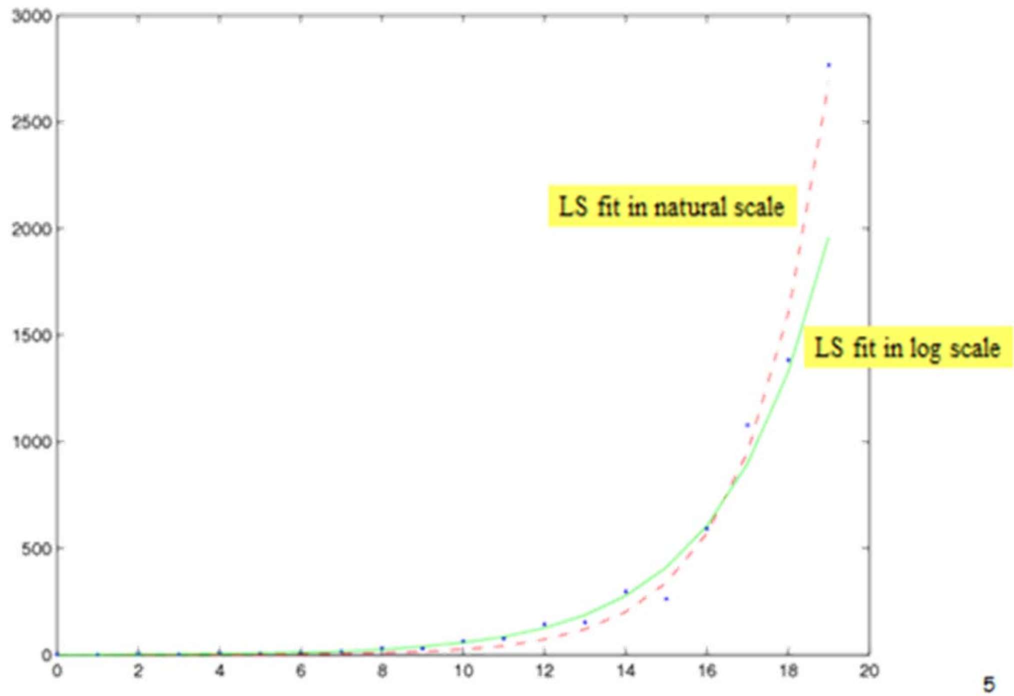
C. Use ℓ^2 minimization rather than ℓ^1

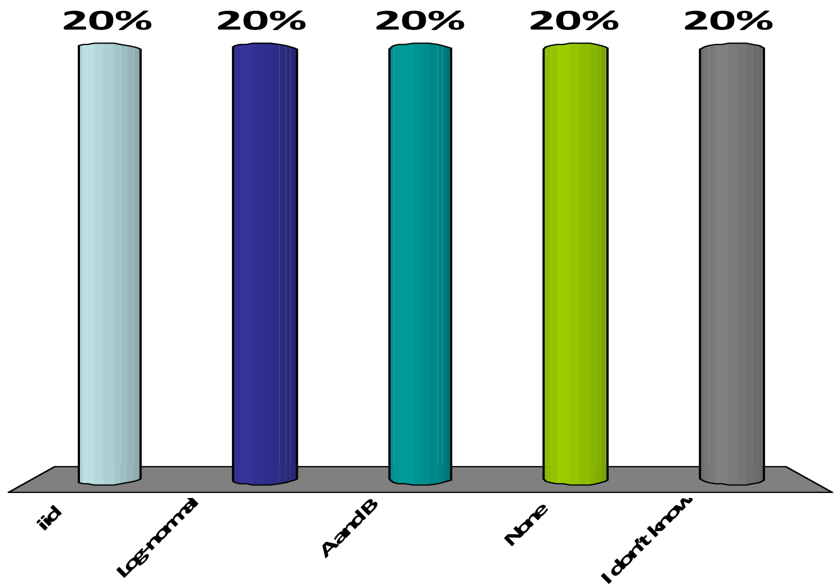
D. I do not know



The green estimation corresponds to assuming that the error terms (blue dot – green curve) are …

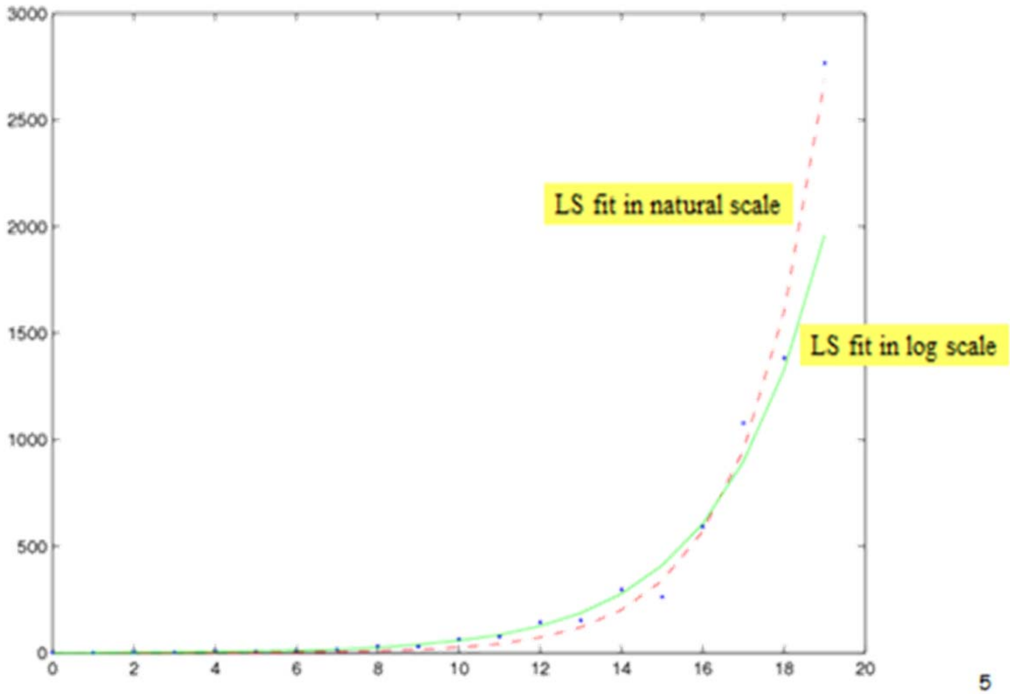
A. iid
B. Log-normal
C. A and B
D. None
E. I don't know



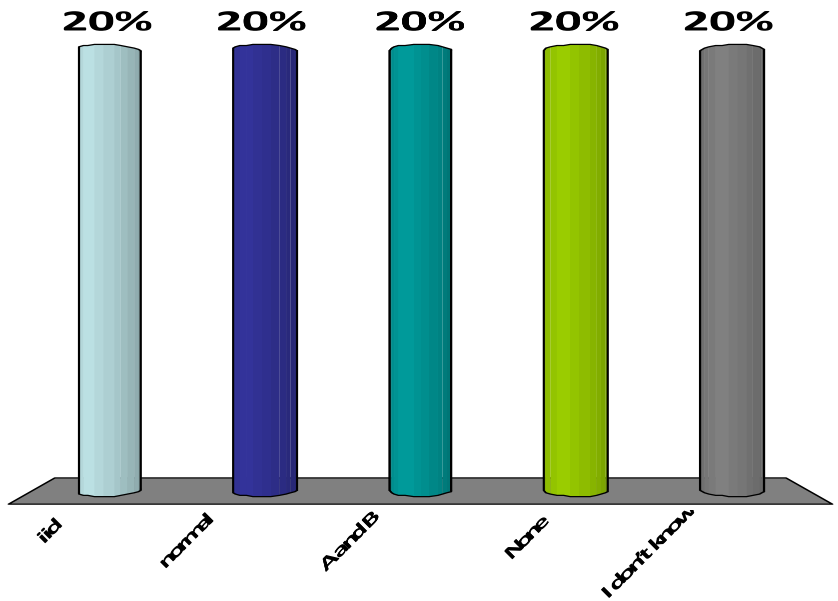


The green estimation corresponds to assuming that the *relative* error terms (blue dot – green curve) are ...

A. iid
B. normal
C. A and B
D. None
E. I don't know

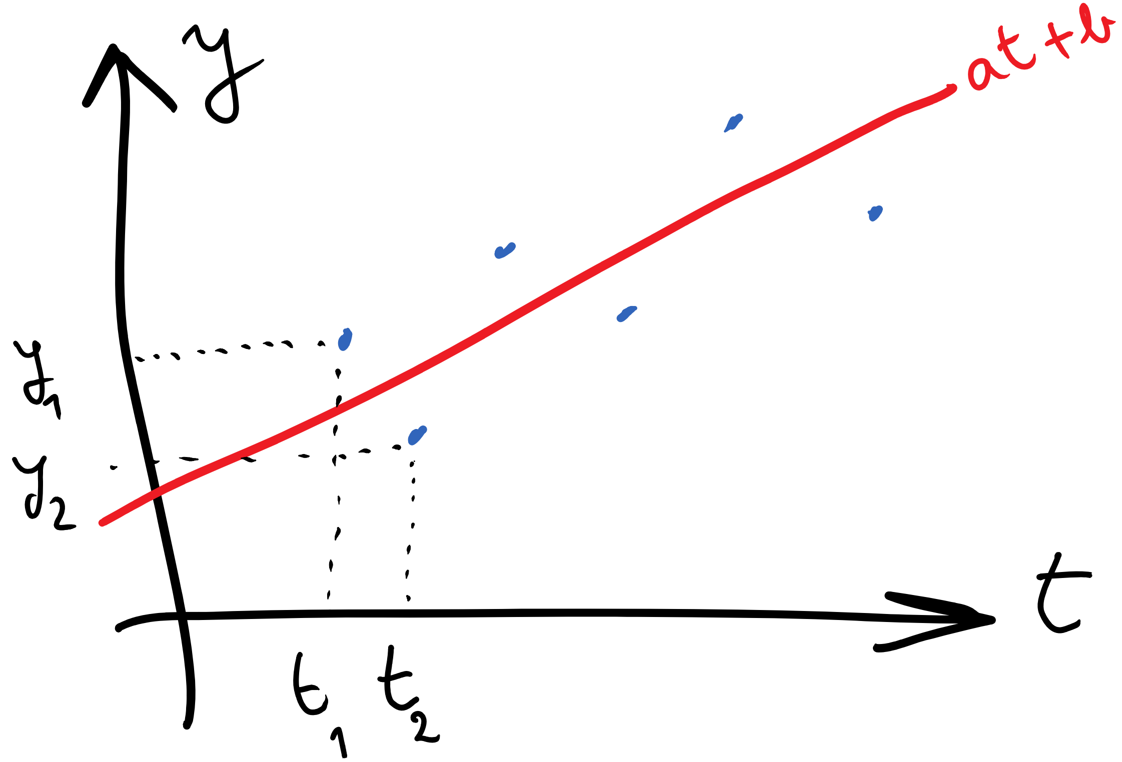


5



We fit the model $y_i = at_i + b$ using least squares.

The obtained line is such that the average distance from the points to the line is 0.

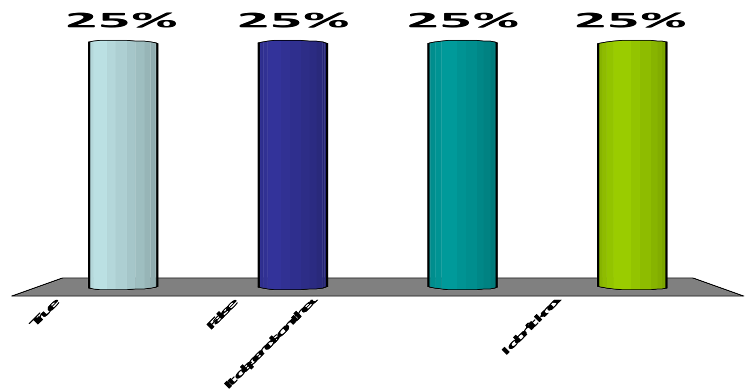


A. True

B. False

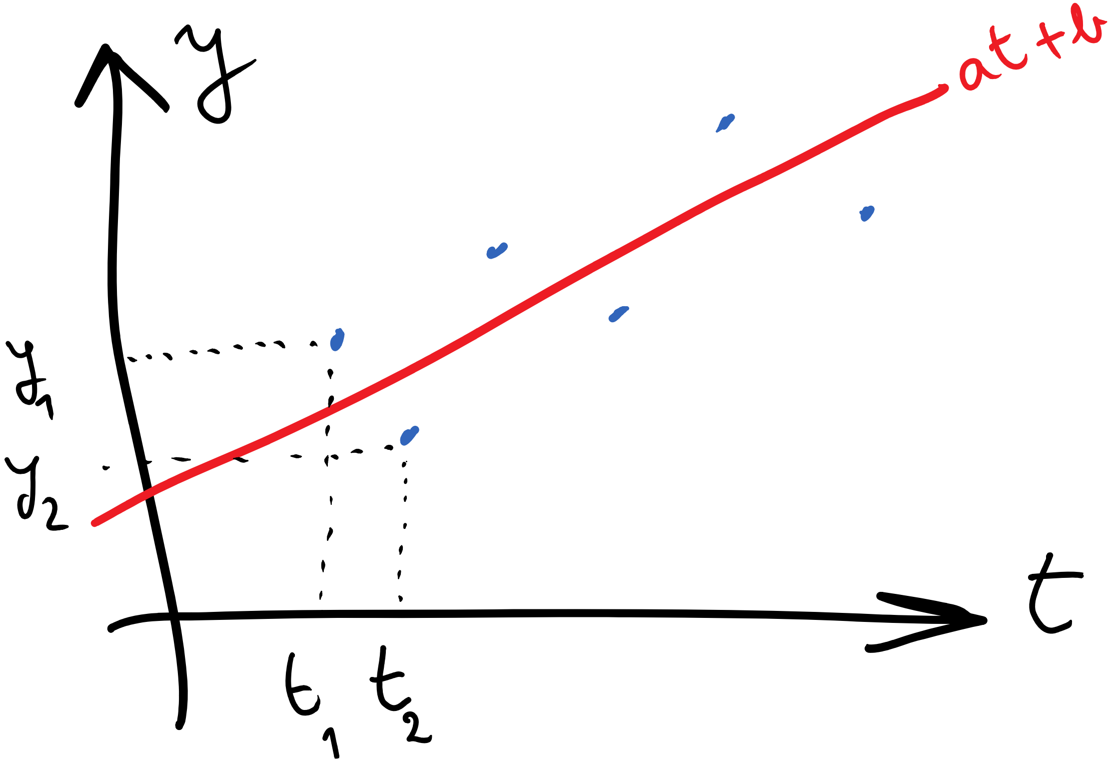
C. It depends on the data

D. I don't know

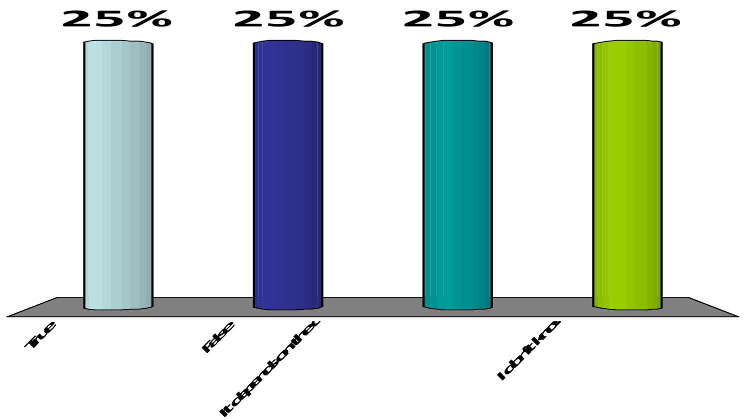


We fit the model $y_i = at_i + b$ using ℓ^1norm minimization.

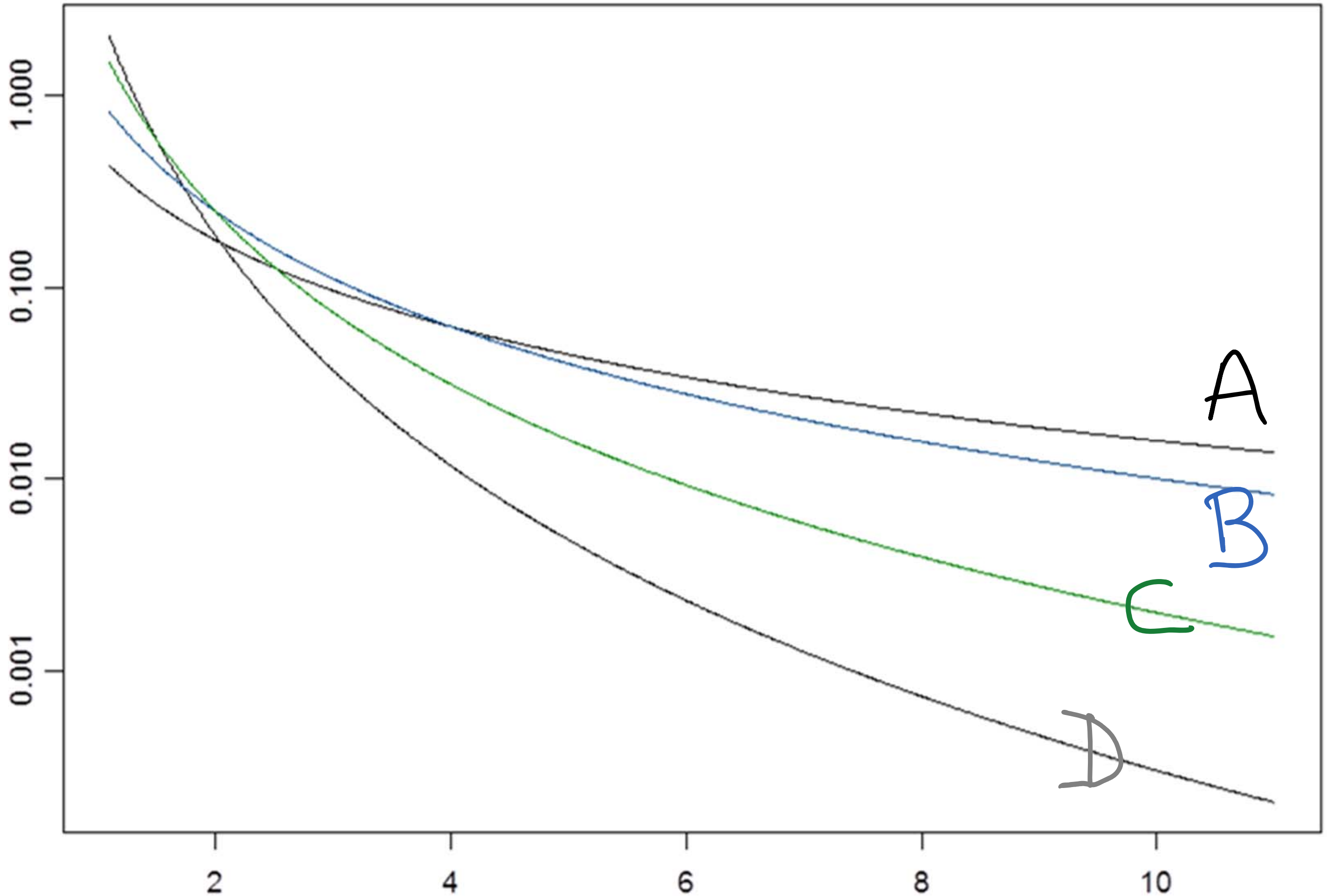
The obtained line leaves an equal number of points on each side



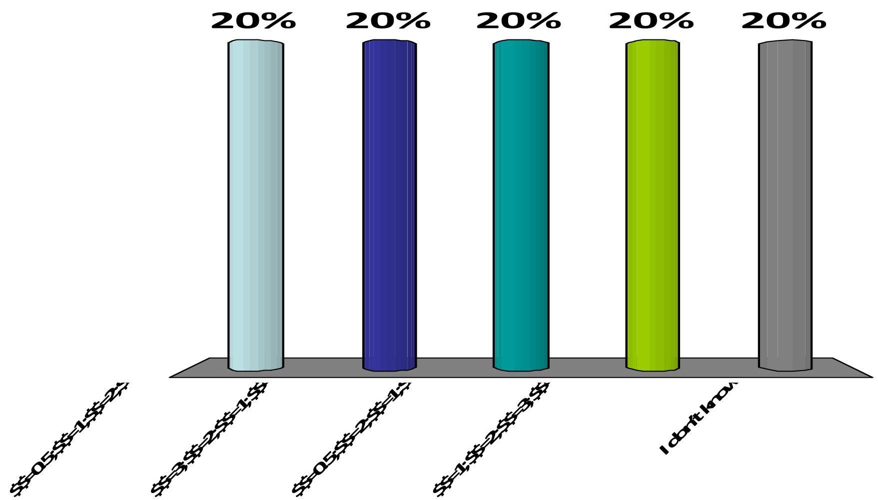
A. True
B. False
C. It depends on the data
D. I don't know



Find the parameter p for each of these standard Pareto PDFs $f(x) = \frac{p}{x^{p+1}} \mathbf{1}_{x>1}$

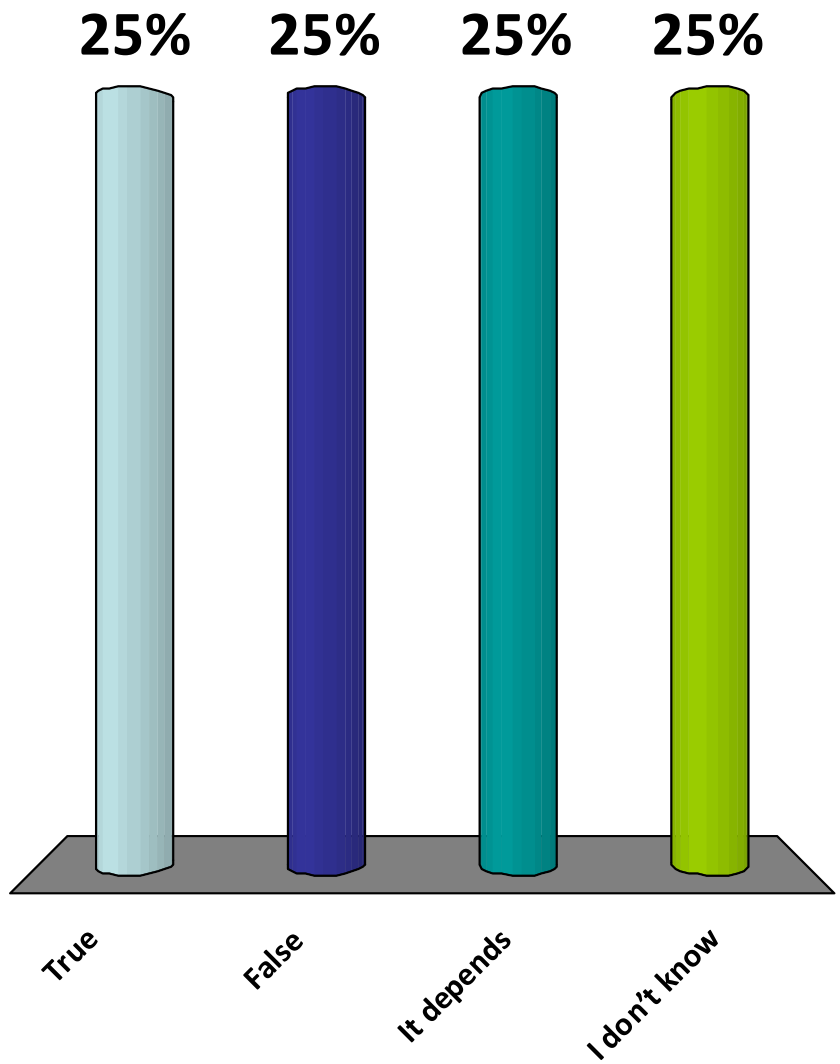


A. $A = 0.5; B = 1; C = 2; D = 3$
B. $A = 3; B = 2; C = 1; D = 0.5$
C. $A = 0.5; B = 2; C = 1; D = 3$
D. $A = 1; B = 2; C = 3; D = 0.5$
E. I don't know



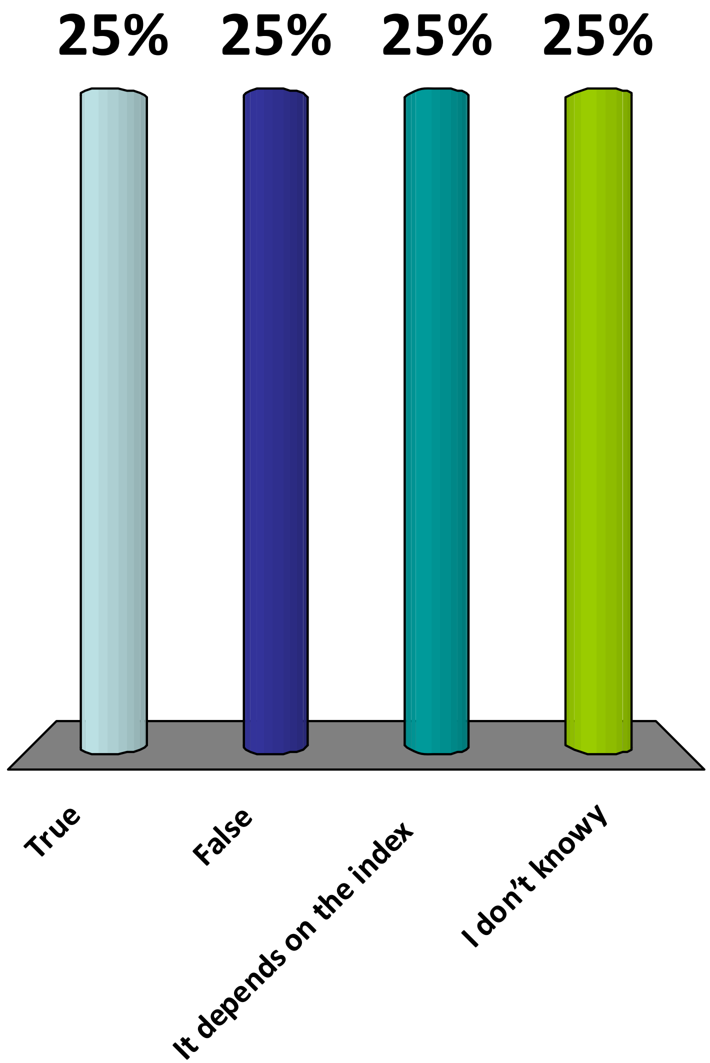
If a positive random variable has a finite mean and is heavy tailed, its variance is infinite

A. True
B. False
C. It depends
D. I don't know



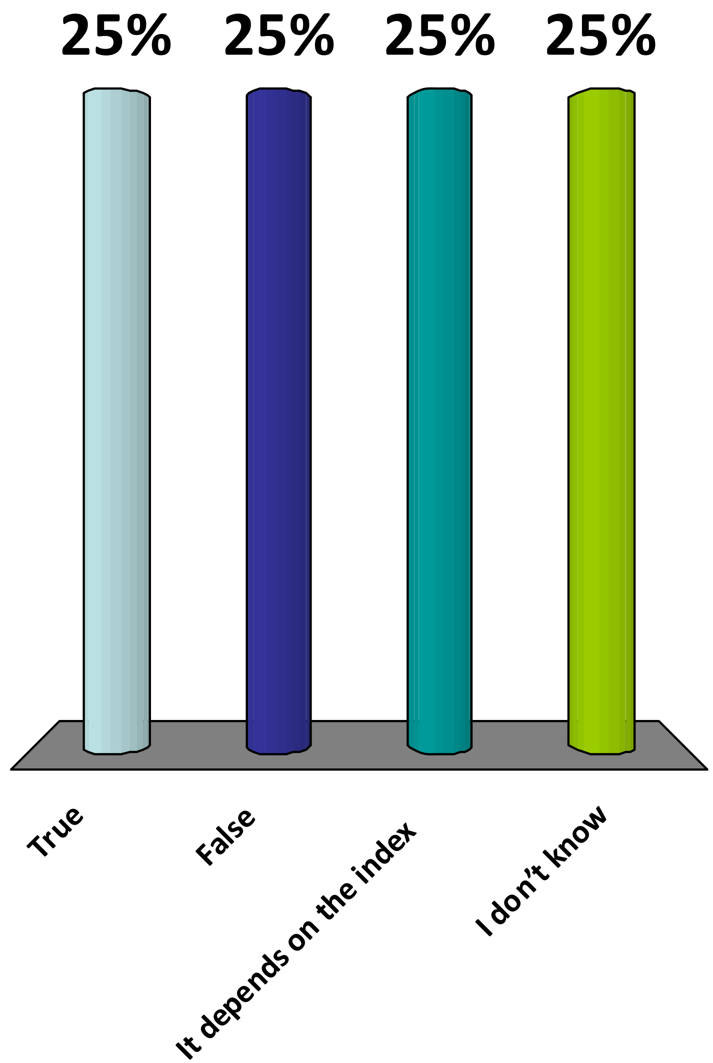
The Complementray CDF of a Pareto distribution follows a power law…

A. True
B. False
C. It depends on the index p
D. I don't knowy



A Pareto distribution is heavy tailed …

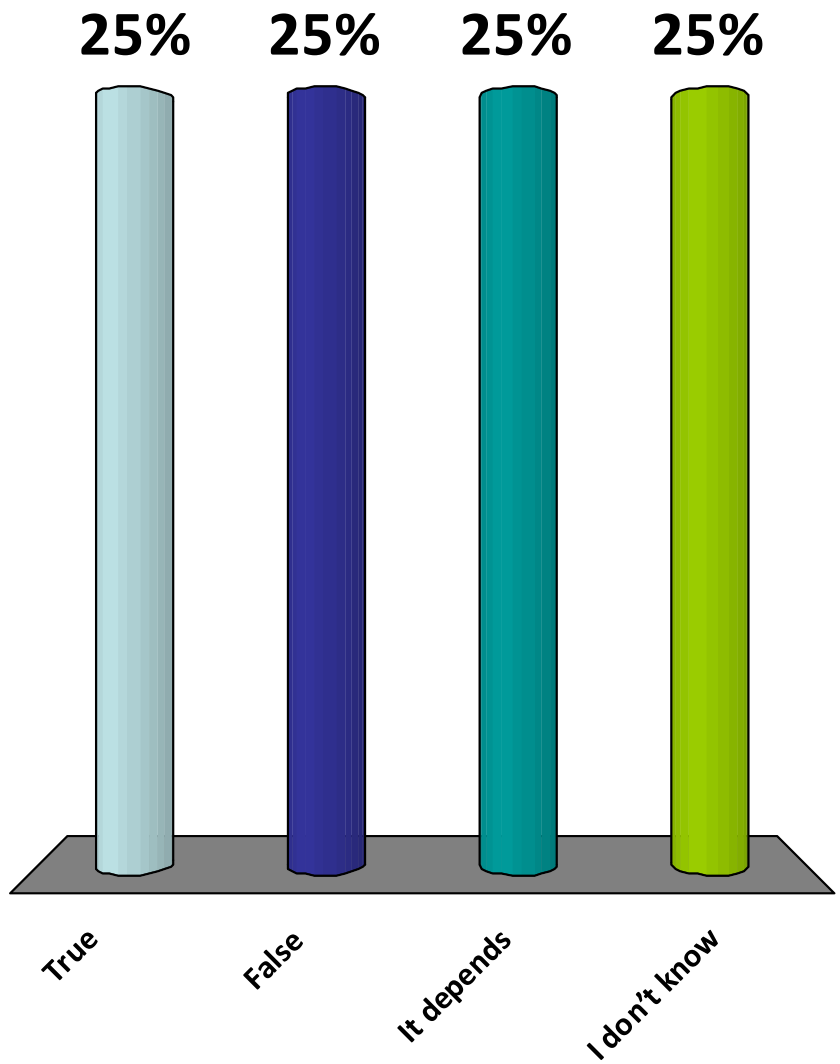
A. True
B. False
C. It depends on the index p
D. I don't know



For a Pareto distribution, the hazard rate $\lambda(t)$ is such that

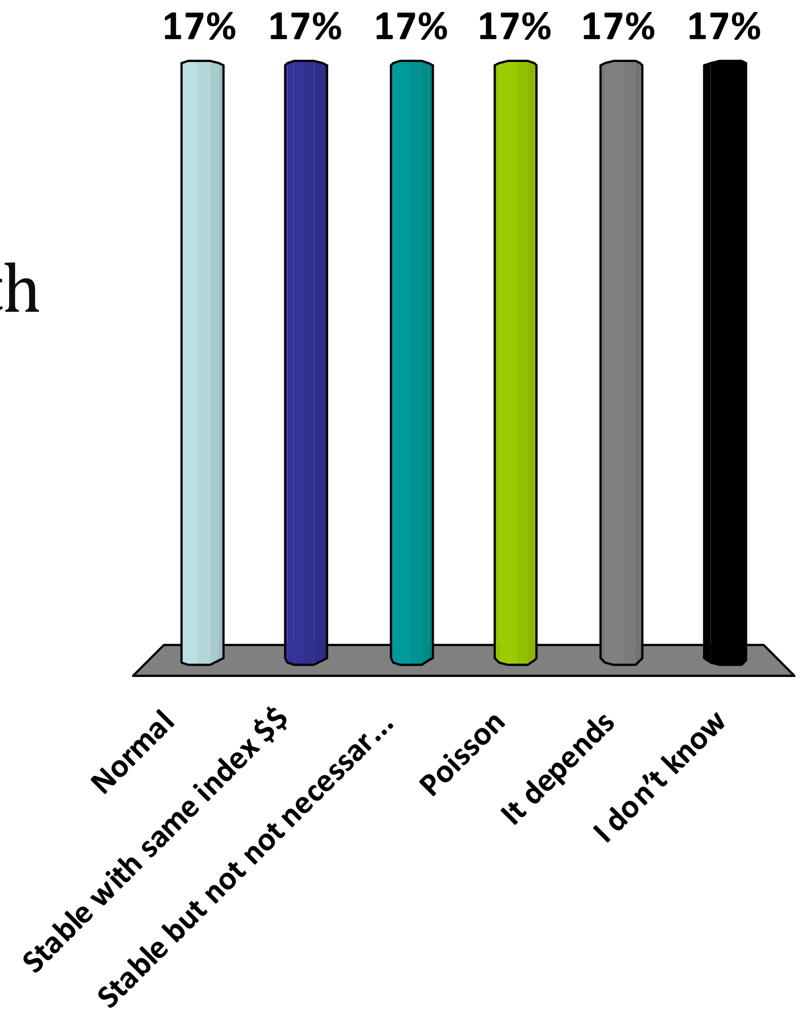
$$\lim_{t \to \infty} \lambda(t) = 0$$

A. True
B. False
C. It depends
D. I don't know



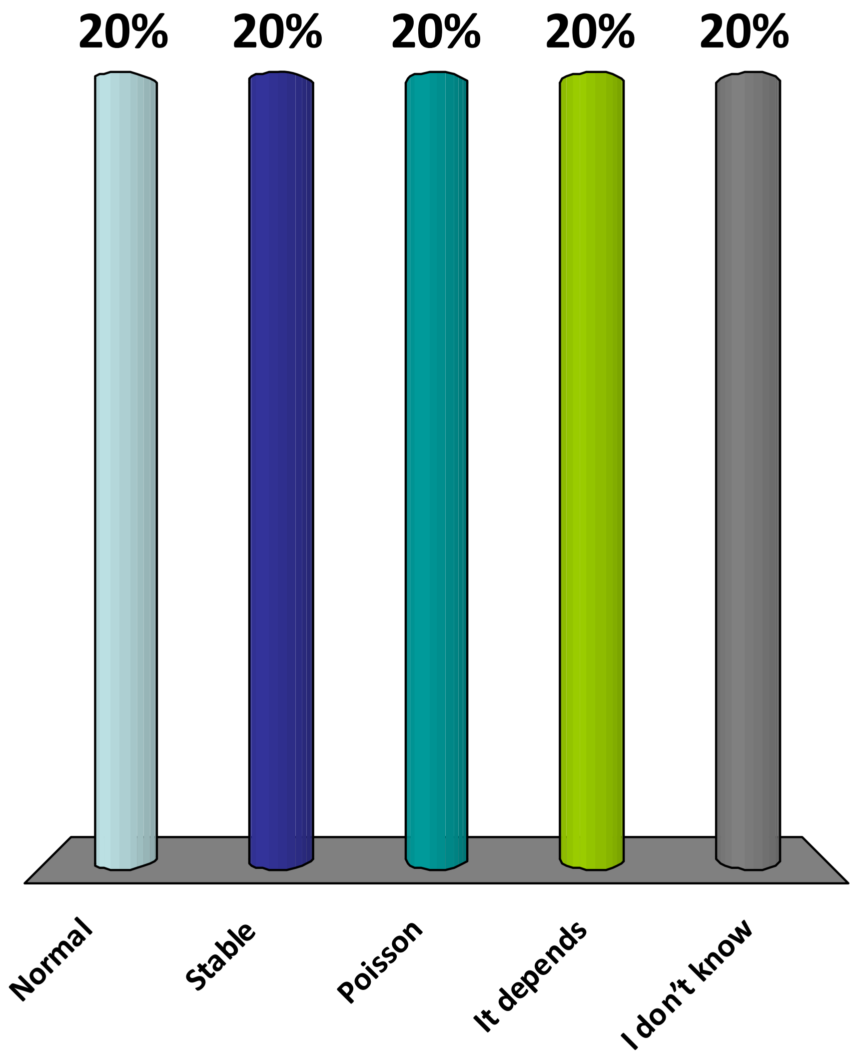
The distribution of the sum of n iid random variables with heavy tail and index $p < 2$, for large n, is approximately…

A. Normal
B. Stable with same index p
C. Stable but not not necessarily with same index p
D. Poisson
E. It depends
F. I don't know



The distribution of the sum of n iid random variables with finite variance, for large n, is approximately…

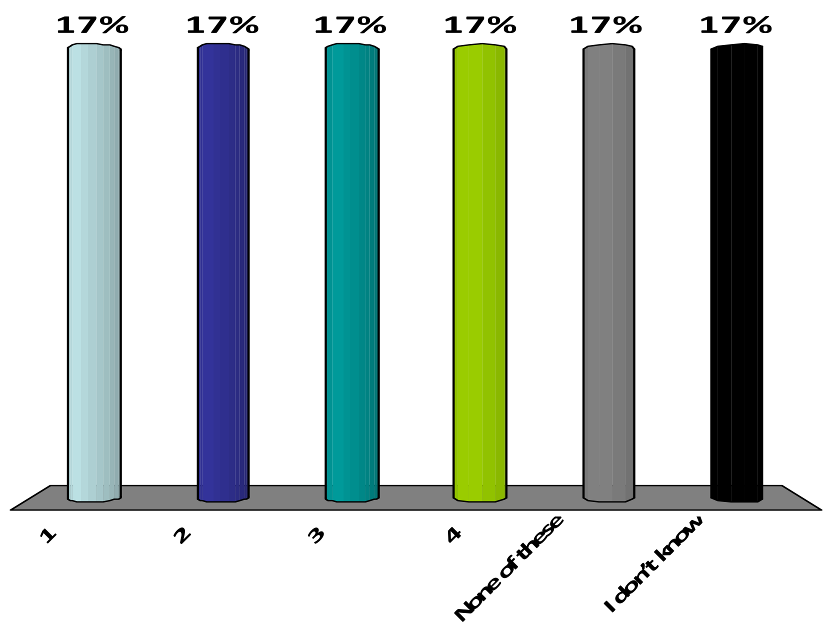
A. Normal
B. Stable
C. Poisson
D. It depends
E. I don't know



We want to estimate some quantity μ. We have $m+n$ independent measurements $X_1, \dots, X_m \sim iid\ N(\mu, \sigma^2)$ and $Y_1, \dots, Y_n \sim iid\ N(\mu, \lambda^2\sigma^2)$ λ is known. We do weighted least square estimation of μ. What do we obtain ?

1. $\hat{\mu}_1 = \frac{X_1+\dots+X_m+Y_1+\dots+Y_n}{m+n}$
2. $\hat{\mu}_2 = \frac{X_1+\dots+X_m+\lambda(Y_1+\dots+Y_n)}{m+\lambda n}$
3. $\hat{\mu}_3 = \frac{X_1+\dots+X_m+\frac{Y_1+\dots+Y_n}{\lambda}}{m+\frac{n}{\lambda}}$
4. $\hat{\mu}_4 = \frac{X_1+\dots+X_m+\frac{Y_1+\dots+Y_n}{\lambda^2}}{m+\frac{n}{\lambda^2}}$

A. 1
B. 2
C. 3
D. 4
E. None of these
F. I don't know



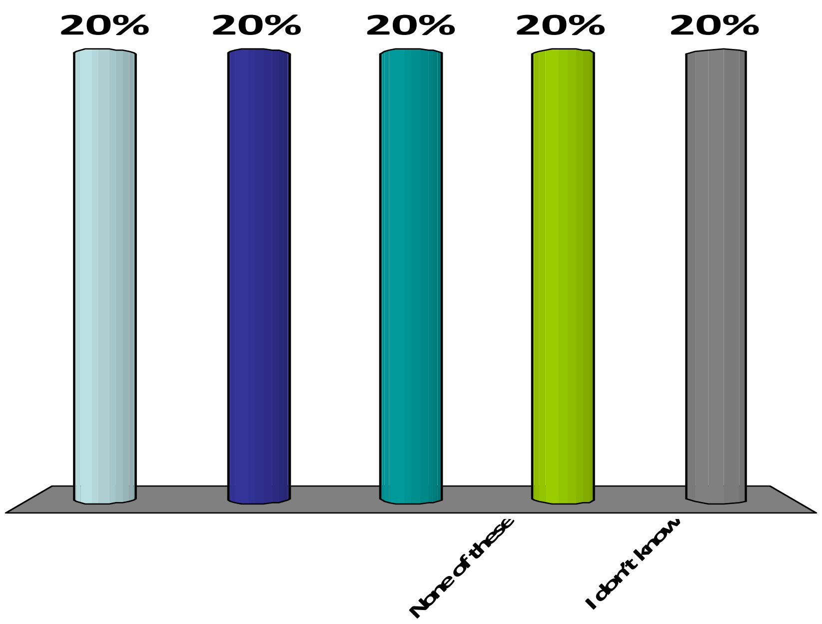
We want to estimate some quantity μ. We have $m + n$ independent measurements $X_1, \ldots, X_m \sim iid\ N(\mu, \sigma^2)$ and $Y_1, \ldots, Y_n \sim iid\ N(\mu, \lambda^2\sigma^2)$

σ and λ are unknown but we think that $\lambda >> 1$. i.e. Y is very noisy; $m \approx n$

To estimate μ, say which formula you prefer:

1. $\hat{\mu}_1 = \frac{X_1+\cdots+X_m}{m}$
2. $\hat{\mu}_2 = \frac{Y_1+\cdots+Y_n}{n}$
3. $\hat{\mu}_3 = \frac{X_1+\cdots+X_m+Y_1+\cdots+Y_n}{m+n}$

1. 1
2. 2
3. 3
4. None of these
5. I don't know



We want to estimate some quantity μ. We have $m + n$ independent measurements $X_1, \ldots, X_m \sim iid\ N(\mu, \sigma^2)$ and $Y_1, \ldots, Y_n \sim iid\ N(\mu, \lambda^2\sigma^2)$

σ and λ are unknown but we think that $\lambda >> 1$.

To estimate μ, say which formula you prefer:

1. $\hat{\mu}_1 = \frac{X_1 + \cdots + X_m}{m}$
2. $\hat{\mu}_2 = \frac{Y_1 + \cdots + Y_n}{n}$
3. $\hat{\mu}_3 = \frac{X_1 + \cdots + X_m + Y_1 + \cdots + Y_n}{m+n}$
4. $\hat{\mu}_4 = \frac{X_1 + \cdots + X_m + \frac{Y_1 + \cdots + Y_n}{\lambda^2}}{m + \frac{n}{\lambda^2}}$

A. 1
B. 2
C. 3
D. 4
E. None of these
F. I don't know