Skewness coefficients for rounded data

Marina Iturrate–Bobes, Ignacio Montes and Raúl Pérez–Fernández

University of Oviedo

Universidad de Oviedo

12th International Conference on Soft Methods in Probability and Statistics (SMPS)
September 15th–18th 2026

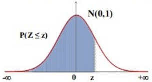
Index

Index

Index

The notion of symmetry

FUNCIÓN DE DISTRIBUCIÓN NORMAL N(0,1)



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4,0	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998

- The study of the property of symmetry can considerably simplify the calculations.
- It is a necessary assumption for some non-parametric tests (*e.g.* Wilcoxon signed-rank test).

$$\begin{aligned} H_0 &: \quad \mathrm{Me} = m \\ H_1 &: \quad \mathrm{Me} \neq m \end{aligned}$$

- And so on...

Symmetry in a random variable

Definition

A random variable $X \in \mathcal{L}(\Omega, \mathcal{A}, P)$ is called **symmetric about** $x_0 \in \mathbb{R}$ if

$$P(X \leq x_0 - x) = P(X \geq x_0 + x)\,, \text{ for any } x \in \mathbb{R}\,.$$

A skewness coefficient is a function $\gamma : \; \mathcal{L}(\Omega, \mathcal{A}, P) \longrightarrow \mathbb{R}$ that quantifies the degree of asymmetry.

An axiomatization of the notion of symmetry (Oja, 1981)

Let $X, Y \in \mathcal{L}(\Omega, \mathcal{A}, P)$ be two **absolutely continuous** random variables:

1. X is symmetric $\Rightarrow \gamma(X) = 0$.
2. $\gamma(cX + d) = \gamma(X)$ where $c, d \in \mathbb{R}$ with $c > 0$.
3. $\gamma(-X) = -\gamma(X)$.
4. $X \precsim Y$, *i.e.*, $F_Y^{-1}(F_X(x))$ is convex for all $x \in \mathbb{R} \Rightarrow \gamma(X) \leq \gamma(Y)$.

Skewness coefficients

Some skewness coefficients in the literature

Let $X \in \mathcal{L}(\Omega, \mathcal{A}, P)$:

- $S_P(X) = \dfrac{\mu - Mo(X)}{\sigma} \longrightarrow$ Pearson (1895).
- $\gamma_1(X) = \dfrac{E\left[(X-\mu)^3\right]}{\sigma^3} \longrightarrow$ Charlier and Edgeworth (1905).
- $\gamma'_{Me(X)}(X) = 3 \cdot \dfrac{\mu - Me(X)}{\sigma} \longrightarrow$ Yule (1912).

In this work...

Let $X \in \mathcal{L}(\Omega, \mathcal{A}, P)$ with $C_p(X) < C_{1-p}(X)$:

$$\gamma_p(X) = \frac{C_{1-p}(X) - 2\, C_{0.5}(X) + C_p(X)}{C_{1-p}(X) - C_p(X)} \text{ for certain } p \in]0, 0.5[\longrightarrow \text{ Hinkley (1975).}$$

Index

Rounded data: Basic concepts

Rounding digit

$d \in \mathbb{Z}$ such that

- $d > 0$: rounding to the d-th decimal digit.
- $d \leq 0$: rounding to the (1-d)-th digit to the left of the decimal point.

Rounding grid

$\mathbb{Z}_d = \{x \in \mathbb{R} \mid 10^d \cdot x \in \mathbb{Z}\}$ is the set of values that a rounded variable may take.

Step of dicretization

$h = 10^{-d} > 0$ is called the **step of discretization.**

Rounding function

$\varphi_d : \mathbb{R} \to \mathbb{Z}_d$ such that $\varphi_d(x) = \max \left\{\arg\min_{z \in \mathbb{Z}_d} |x - z|\right\}$.

Rounded data: Basic concepts

Rounded variable

Given $X \in \mathcal{L}(\Omega, \mathcal{A}, P)$, its associated **rounded variable** is $\tilde{X} = \varphi_d(X)$ for certain $d \in \mathbb{Z}$.

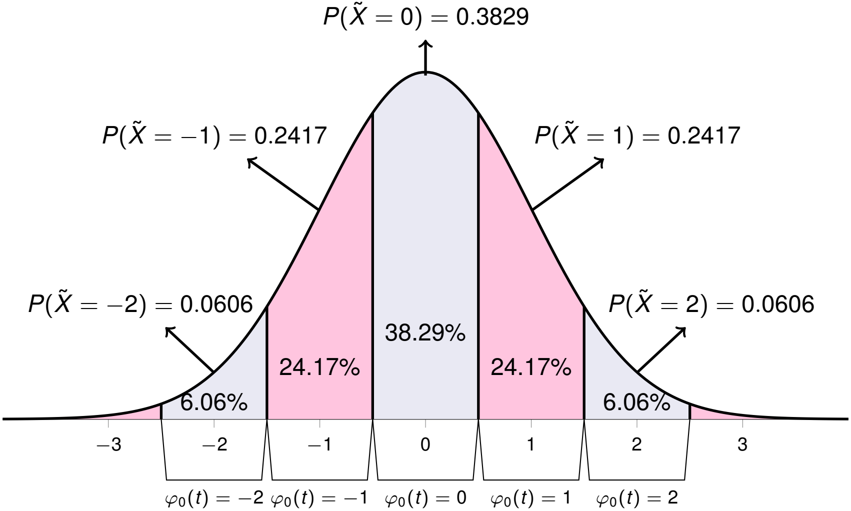
Remark: The original random variable X is typically unobservable. Instead, we have information on $\tilde{X}$.

Rounding level parameter

$r = \frac{h}{\sigma_X} \in \mathbb{R}^+$ compares the loss of information due to rounding and the scale of the data.

Large values of r indicate a significant effect of rounding

$X \equiv \mathcal{N}(0,1)$ rounded to the nearest integer ($d = 0$)



Index

Index

Theoretical framework

Random set

$\Gamma : \Omega \to \mathcal{P}(\mathbb{R})$ is a **random set** if:

- $\Gamma(\omega) \neq \emptyset$ for all $\omega \in \Omega$.
- Γ^* is strongly measurable, *i.e.*, $\Gamma^*(A) = \{\omega \in \Omega \mid \Gamma(\omega) \cap A \neq \emptyset\} \in \mathcal{A}$ for any Borel set A.

Epistemic interpretation: The random set models a real but only partially known random variable X. All we know is that $X(\omega) \in \Gamma(\omega)$ for any $\omega \in \Omega$.

Set of measurable selections

The **set of measurable selections** of the random set Γ is formed by those random variables that are compatible with Γ.

$$S(\Gamma) = \{Y \in \mathcal{L}(\Omega, \mathcal{A}, P) \mid Y(\omega) \in \Gamma(\omega)\ \forall \omega \in \Omega\}.$$

Modeling the rounding process

Given the rounded variable $\tilde{X} = \varphi_d(X)$, we define the random (interval) set:

$$\Gamma_X(\omega) = \left[\tilde{X}(\omega) - h/2, \tilde{X}(\omega) + h/2\right[\, ,$$

that models the rounding process. It is evident that both $X, \tilde{X} \in S(\Gamma_X)$.

Remark: $[\cdot, \cdot[$ denotes the left-closed and right-open interval $[\cdot, \cdot)$.

Set-valued quantiles

The set-valued quantile of order $p \in]0, 1[$ is defined as $C_p(\Gamma_X) = \{C_p(Y) \mid Y \in S(\Gamma_X)\}$.

Modeling the rounding process

Given a random sample $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1, \ldots, \tilde{\mathbf{X}}_n)^T$ from $\tilde{X}$, we denote the associated random sample of Γ_X as $\boldsymbol{\Gamma}_X = (\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_n)^T$, where:

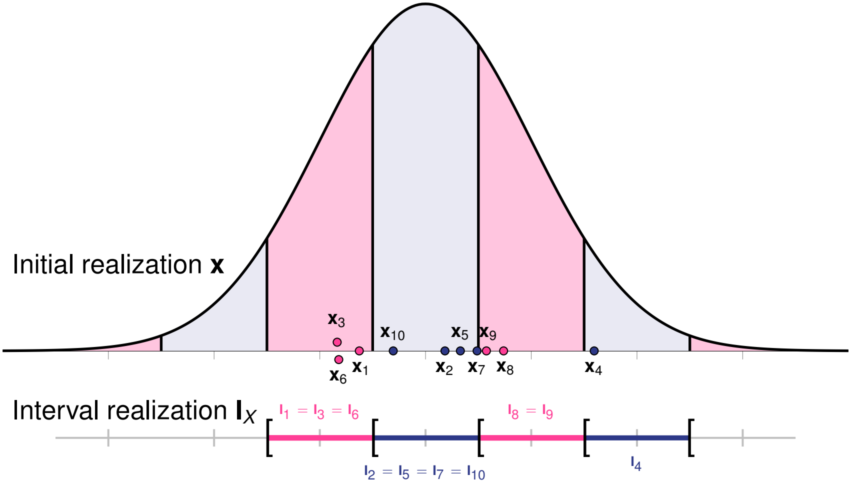
$$\boldsymbol{\Gamma}_i = \left[\tilde{\mathbf{X}}_i - {}^{h}\!/_{2}\,,\ \tilde{\mathbf{X}}_i + {}^{h}\!/_{2}\right[\quad \forall i \in \{1, \ldots, n\}\,.$$

We denote by $\mathbf{I}_X = (\mathbf{I}_1, \ldots, \mathbf{I}_n)^T$ any realization of $\boldsymbol{\Gamma}_X$.

Set-valued order statistics

For any $i \in \{1, \ldots, n\}$, the i-th set-valued order statistic is $\boldsymbol{\Gamma}_{(i)} = \left[\tilde{\boldsymbol{X}}_{(i)} - {}^{h}\!/_{2},\ \tilde{\boldsymbol{X}}_{(i)} + {}^{h}\!/_{2}\right[$.

$X \equiv \mathcal{N}(0,1)$ rounded to the nearest integer ($d = 0$)



Index

Statistical inference with imprecise data

Classical approach | **Imprecise approach**

RR_0
AR_0

$\longrightarrow$

$$\mathrm{RR} = \{\mathbf{I}_X \mid (\forall \mathbf{x} \in \mathbf{I}_X)(\mathbf{x} \in \mathrm{RR}_0)\}$$
$$\mathrm{AR} = \{\mathbf{I}_X \mid (\forall \mathbf{x} \in \mathbf{I}_X)(\mathbf{x} \in \mathrm{AR}_0)\}$$
$$\mathrm{IR} = \left\{\mathbf{I}_X \mid (\exists \mathbf{x}, \mathbf{x}' \in \mathbf{I})(\mathbf{x} \in \mathrm{RR}_0 \wedge \mathbf{x}' \in \mathrm{AR}_0)\right\}$$

$$\text{p-value}(\mathbf{x}) \longrightarrow \text{p-value}(\mathbf{I}_X) = \{\text{p-value}(\mathbf{x}) \mid \mathbf{x} \in \mathbf{I}_X\}$$

$\downarrow$

$\sup \text{p-value}(\mathbf{I}_X) \leq \alpha$: **Reject H_0**
$\inf \text{p-value}(\mathbf{I}_X) > \alpha$: **No reject H_0**
$\inf \text{p-value}(\mathbf{I}_X) \leq \alpha < \sup \text{p-value}(\mathbf{I}_X)$: **Indecision**

Index

Set-valued Hinkley's family of coefficients

Let $X \in \mathcal{L}(\Omega, \mathcal{A}, P)$ such that $C_p(X) < C_{1-p}(X)$ with $p \in]0, 0.5[$. In the **classical framework**:

$$\gamma_p(X) = \frac{C_{1-p}(X) - 2\, C_{0.5}(X) + C_p(X)}{C_{1-p}(X) - C_p(X)}$$

Set-valued Hinkley's family of coefficients

In the **imprecise framework**, we define the **set-valued population coefficient** as:

$$\gamma_p(\Gamma) = \{\gamma_p(Y) \mid Y \in S(\Gamma) \wedge C_p(Y) < C_{1-p}(Y)\}\,.$$

Set-valued Hinkley's family of coefficients

Let $\mathbf{X}$ a random sample from $X \in \mathcal{L}(\Omega, \mathcal{A}, P)$ such that $\mathbf{X}_{(\lceil np \rceil)} < \mathbf{X}_{(\lceil n(1-p) \rceil)}$ with $p \in]0, 0.5[$. In the **classical framework**:

$$\hat{\gamma}_p(\mathbf{X}) = \frac{\mathbf{X}_{(\lceil n(1-p) \rceil)} - 2\,\mathbf{X}_{(\lceil 0.5n \rceil)} + \mathbf{X}_{(\lceil np \rceil)}}{\mathbf{X}_{(\lceil n(1-p) \rceil)} - \mathbf{X}_{(\lceil np \rceil)}}$$

Set-valued Hinkley's family of coefficients

In the **imprecise framework**, we define the **set-valued sample version** as:

$$\hat{\gamma}_p(\mathbf{\Gamma}_X) = \left\{ \hat{\gamma}_p(\mathbf{Y}) \mid \mathbf{Y} \in S(\mathbf{\Gamma}_X) \wedge \mathbf{Y}_{(\lceil np \rceil)} < \mathbf{Y}_{(\lceil n(1-p) \rceil)} \right\}.$$

Set-valued Hinkley's family of coefficients

(a) If $\mathbf{\Gamma}_{(\lceil np \rceil)} \neq \mathbf{\Gamma}_{(\lceil 0.5n \rceil)} \neq \mathbf{\Gamma}_{(\lceil n(1-p) \rceil)}$:

$$\hat{\gamma}_p(\mathbf{\Gamma}_X) = \left] \hat{\gamma}_p(\tilde{\boldsymbol{X}}) - \frac{2h}{\tilde{\boldsymbol{X}}_{(\lceil n(1-p) \rceil)} - \tilde{\boldsymbol{X}}_{(\lceil np \rceil)}},\ \hat{\gamma}_p(\tilde{\boldsymbol{X}}) + \frac{2h}{\tilde{\boldsymbol{X}}_{(\lceil n(1-p) \rceil)} - \tilde{\boldsymbol{X}}_{(\lceil np \rceil)}} \right[.$$

(b) If $\mathbf{\Gamma}_{(\lceil np \rceil)} = \mathbf{\Gamma}_{(\lceil 0.5n \rceil)} \neq \mathbf{\Gamma}_{(\lceil n(1-p) \rceil)}$:

$$\hat{\gamma}_p(\mathbf{\Gamma}_X) = \left] 1 - \frac{2h}{\tilde{\boldsymbol{X}}_{(\lceil n(1-p) \rceil)} - \tilde{\boldsymbol{X}}_{(\lceil np \rceil)}},\ 1 \right] .$$

(c) If $\mathbf{\Gamma}_{(\lceil np \rceil)} \neq \mathbf{\Gamma}_{(\lceil 0.5n \rceil)} = \mathbf{\Gamma}_{(\lceil n(1-p) \rceil)}$:

$$\hat{\gamma}_p(\mathbf{\Gamma}_X) = \left[-1,\ -1 + \frac{2h}{\tilde{\boldsymbol{X}}_{(\lceil n(1-p) \rceil)} - \tilde{\boldsymbol{X}}_{(\lceil np \rceil)}} \right[.$$

(d) If $\mathbf{\Gamma}_{(\lceil np \rceil)} = \mathbf{\Gamma}_{(\lceil 0.5n \rceil)} = \mathbf{\Gamma}_{(\lceil n(1-p) \rceil)}$: $\hat{\gamma}_p(\mathbf{\Gamma}_X) = [-1,\ 1] .$

Index

Skewness-based tests of normality

Setting up the hypotheses of the test

Given a random sample $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)^T$ from an absolutely continuous random variable $X \in \mathcal{L}(\Omega, \mathcal{A}, P)$, we test:

$$\begin{cases} H_0 : X \text{ is normally distributed}\,, \\ H_1 : X \text{ is not normally distributed}\,. \end{cases}$$

Skewness-based tests of normality in the classical framework

In the **classical framework**, given a random sample $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)^T$, we define the following test statistic:

$$T(\mathbf{X}) = \sqrt{n}\frac{\hat{\gamma}_p(\mathbf{X}) - 0}{\tau_p(\mathcal{N})} \xrightarrow[H_0]{\mathcal{D}} \mathcal{N}(0,1)\,.$$

Let $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ be a realization of $\mathbf{X}$. Its associated p-value is

$$\text{p-value}(\mathbf{x}) = 2\,\Phi(-|T(\mathbf{x})|) = 2\,\Phi\left(-\left|\sqrt{n}\frac{\hat{\gamma}_p(\mathbf{x}) - 0}{\tau_p(\mathcal{N})}\right|\right)\,,$$

where Φ is the cdf of $\mathcal{N}(0,1)$.

Skewness-based tests of normality with rounded data

In the **imprecise framework**, given a random sample $\boldsymbol{\Gamma}_X = (\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_n)^T$ of Γ_X, the test statistic is

$$T(\mathbf{X}) \in \left[\sqrt{n}\frac{\inf_{\mathbf{Y}\in\mathbf{S}(\boldsymbol{\Gamma}_\mathbf{X})} \hat{\gamma}_p(\mathbf{Y}) - 0}{\tau_p(\mathcal{N})}, \sqrt{n}\frac{\sup_{\mathbf{Y}\in\mathbf{S}(\boldsymbol{\Gamma}_\mathbf{X})} \hat{\gamma}_p(\mathbf{Y}) - 0}{\tau_p(\mathcal{N})} \right].$$

Let $\mathbf{I}_X = (\mathbf{I}_1, \ldots, \mathbf{I}_n)^T$ be a realization of $\boldsymbol{\Gamma}_X$. Its associated p-value is such that

$$\text{inf p-value}(\mathbf{I}_X) = 2\,\Phi\Big(- \sup_{\mathbf{y}\in\mathbf{I}_X} |T(\mathbf{y})| \Big).$$

$$\text{sup p-value}(\mathbf{I}_X) = 2\,\Phi\Big(- \inf_{\mathbf{y}\in\mathbf{I}_X} |T(\mathbf{y})| \Big).$$

Skewness-based tests of normality with rounded data

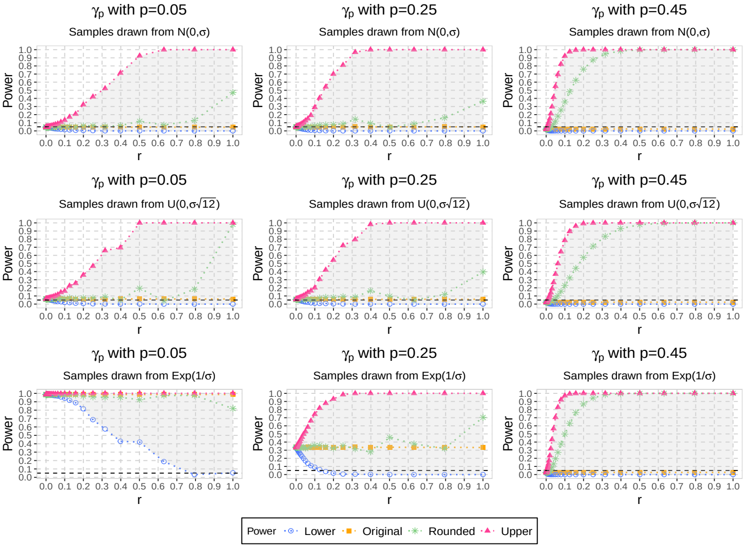
Experimental framework

- 3 scenarios:

$$\boxed{X \equiv \mathcal{N}(0,\sigma)} \qquad \boxed{X \equiv \mathcal{U}(0,\sigma\sqrt{12})} \qquad \boxed{X \equiv \exp\left(^{1}/_{\sigma}\right)}$$

- $\mathrm{Var}(X) = \sigma^2$ with $\sigma = 10^t$, for $t = \mathtt{seq}(-1, 3, \mathtt{by} = 0.3)$.
- $d = 1 \Rightarrow r = 10^{-(1+t)}$.
- $n = 50$.
- Significance level $\alpha = 0.05$.
- Powers: $\underbrace{\textit{original}, \textit{rounded}}_{\textbf{classical framework}}, \underbrace{\textit{lower}, \text{and } \textit{upper}}_{\textbf{imprecise framework}}$ where:
  - *lower* power refers to $P(\mathrm{RR})$.
  - *upper* power refers to $1 - P(\mathrm{AR})$.
- Powers are estimated by Monte Carlo simulation with 10^5 replications.

Skewness-based tests of normality with rounded data



Index

Conclusions

Conclusions of the experimental setup

- **Under** H_0: Expected behavior when the imprecision due to rounding is not too large.
- The tests have difficulty distinguishing distribution families whenever the skewness coefficient takes the same value (*e.g.*, symmetric distributions), even in the absence of imprecision.
- The indecision region remains relatively small **up to approximately** $r = 0.2$.
- **Small values of** $p \in]0, 0.5[$ capture more differences between distributions.

Future work

To be continued...

1. To perform a **more detailed theoretical analysis** of the properties of the set-valued Hinkley's coefficient for rounded data.
2. To generalize for **goodness-of-fit tests to location-scale families**.
3. To apply this framework to **tests of symmetry** with rounded data.

↓

Asymmetry in rounded data: Skewness coefficients and tests of goodness-of-fit and of symmetry
(Recently submitted)

Funding

Funding from grant PID2022-140585NB-I00 funded by MICIU/AEI/10.13039/501100011033 and “FEDER/UE”.

MINISTERIO
DE CIENCIA, INNOVACIÓN
Y UNIVERSIDADES

Cofinanciado por
la Unión Europea

AGENCIA
ESTATAL DE
INVESTIGACIÓN

Skewness coefficients for rounded data

Marina Iturrate–Bobes, Ignacio Montes and Raúl Pérez–Fernández

University of Oviedo

Universidad de Oviedo

12th International Conference on Soft Methods in Probability and Statistics (SMPS)
September 15th–18th 2026