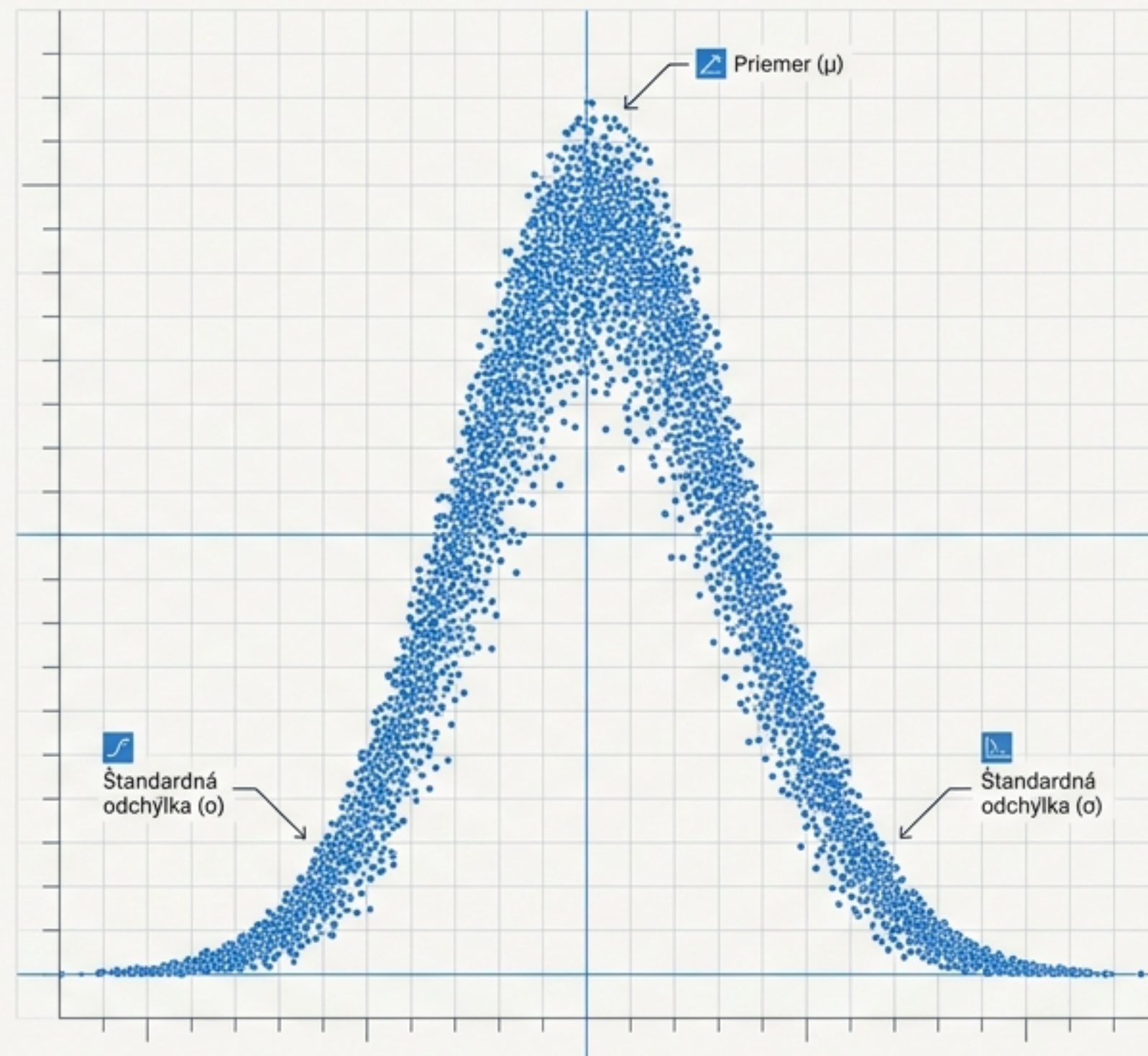


Overenie normality dát

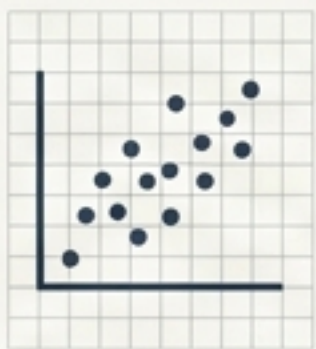
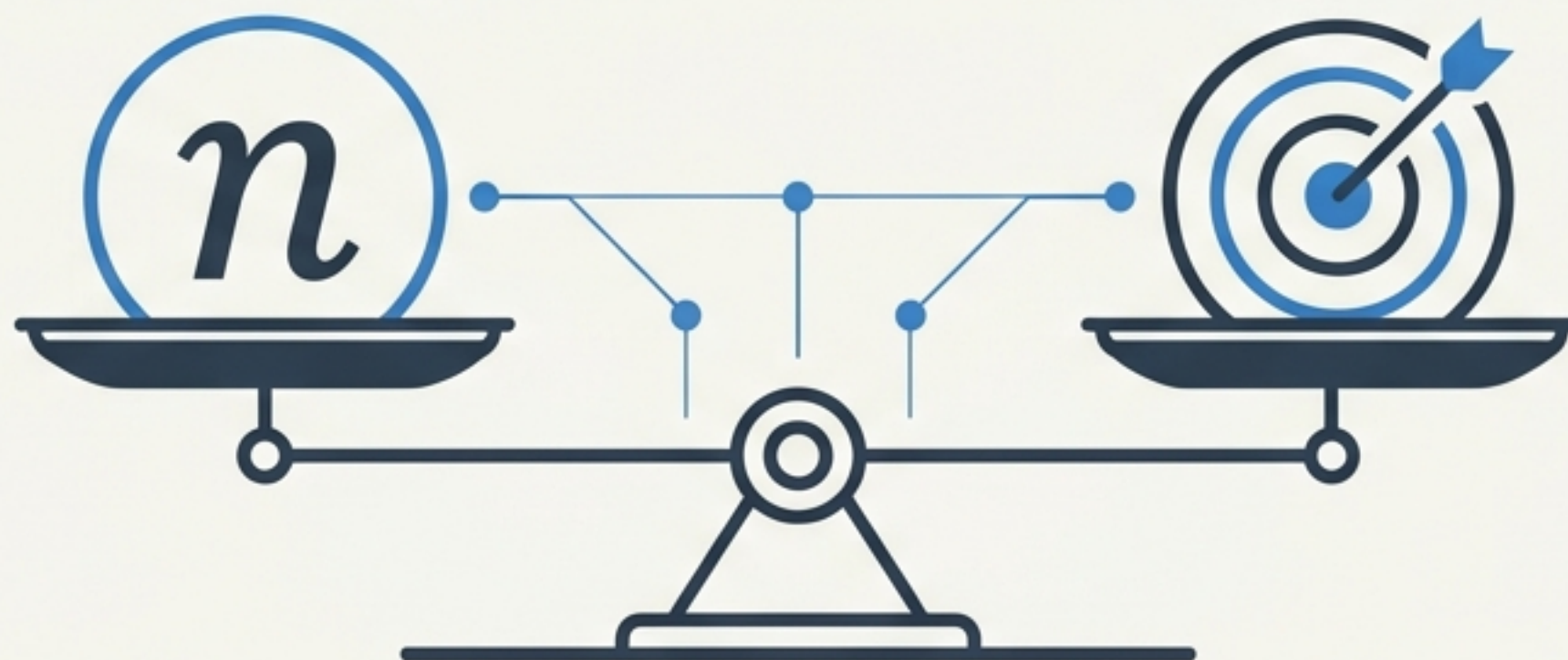
Expertný sprievodca metodikou a výberom testov v prostredí Minitab.

Diagnostický manuál pre inžinierov kvality a dátových analytikov.



NEEXISTUJE JEDEN UNIVERZÁLNY TEST NORMALITY

Voľba štatistického algoritmu nie je náhodná. Je priamo podmienená veľkosťou skúmaného súboru (n).

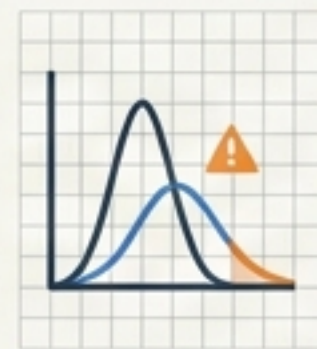


PREČO NA VEĽKOSTI ZÁLEŽÍ:

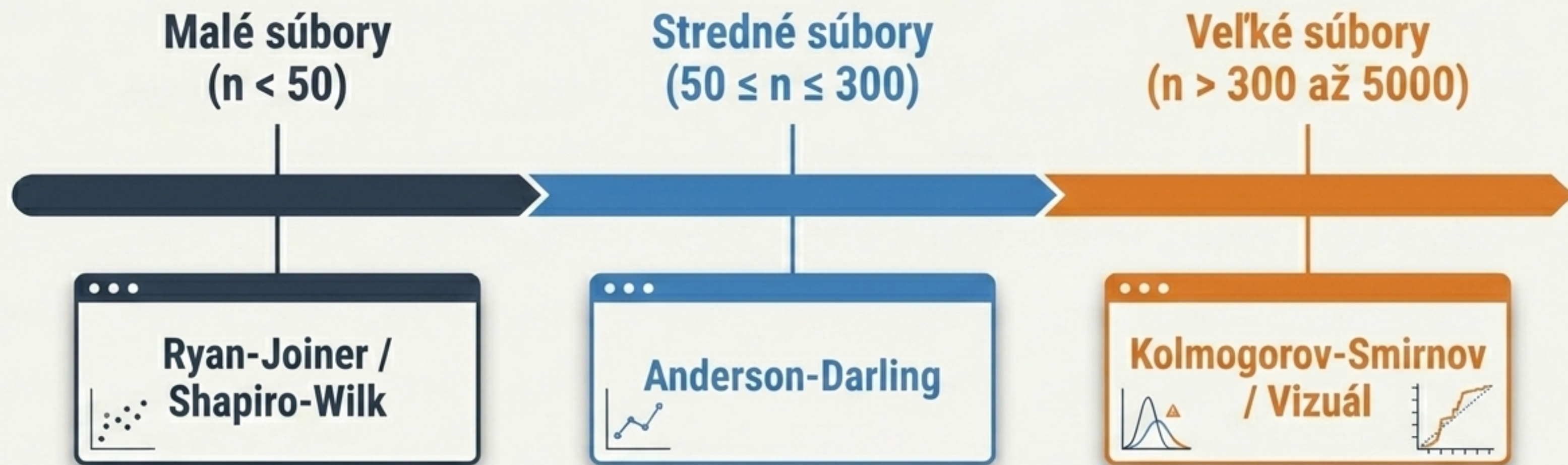
Výkon a presnosť testov normality sa mení s množstvom dát. Každý test je kalibrovaný na iné podmienky.

CIEĽ DIAGNOSTIKY:

Zvoliť test s optimálnou štatistickou silou, ktorý zabráni falošným poplachom alebo prehliadnutiu odchýlok.



SPEKTRUM VEĽKOSTI SÚBOROV



DIAGNOSTIKA: MALÉ SÚBORY ($n < 50$)



ODPORÚČANÝ TEST:

Shapiro-Wilk (v Minitabe alternatíva Ryan-Joiner).

PREČO HO POUŽIŤ?

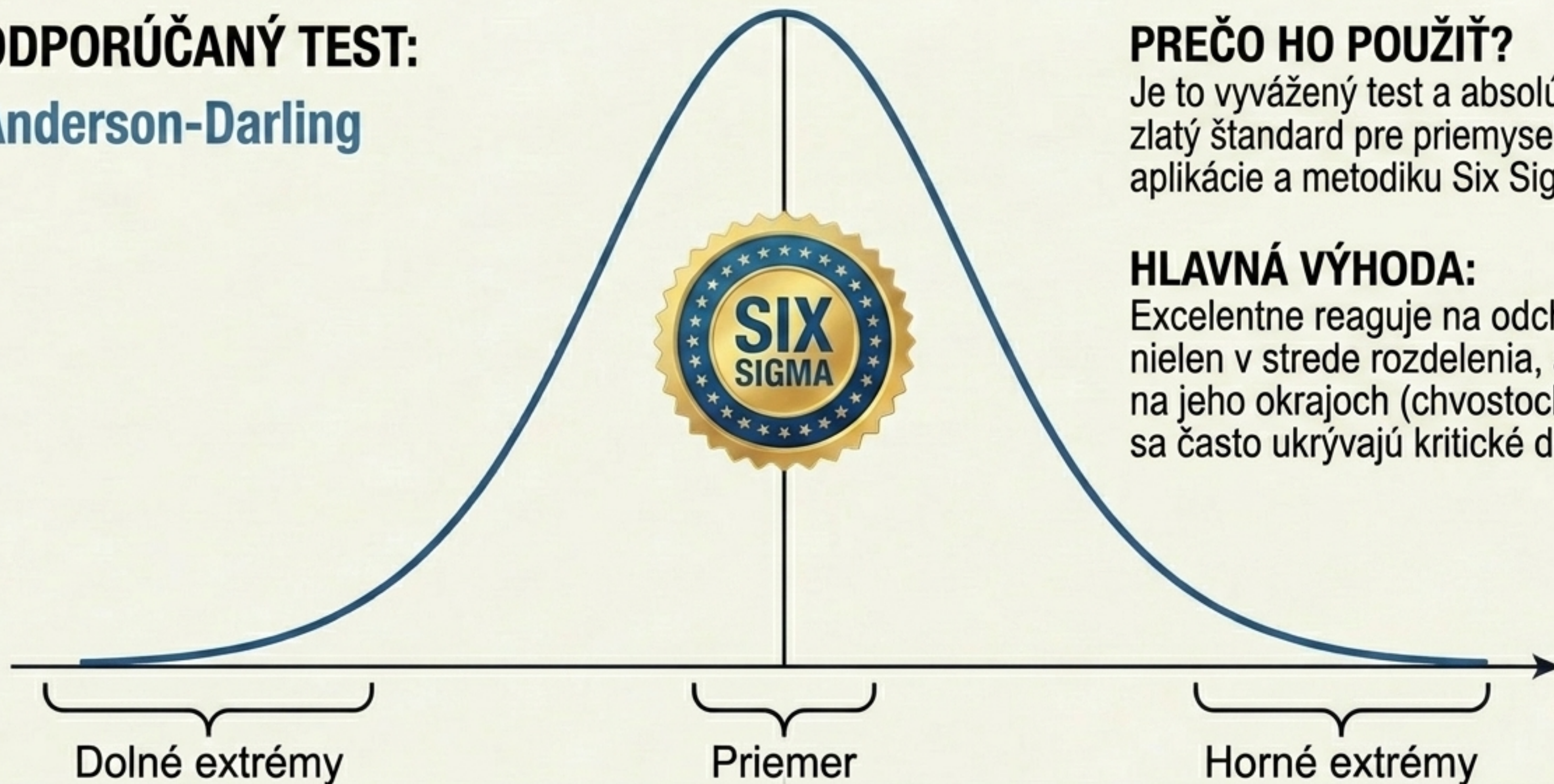
Tieto testy sú matematicky navrhnuté tak, aby mali najvyššiu možnú citlivosť aj pri veľmi malom počte dátových bodov.

! UPOZORNENIE NA RIZIKO (NÍZKA SILA TESTU):

Ak test nezamietne normalitu, nemusí to automaticky znamenať, že dáta sú normálne. Znamená to iba: Máme príliš málo dát na dôkaz opaku.

DIAGNOSTIKA: STREDNÉ SÚBORY ($50 \leq n \leq 300$)

ODPORÚČANÝ TEST:
Anderson-Darling



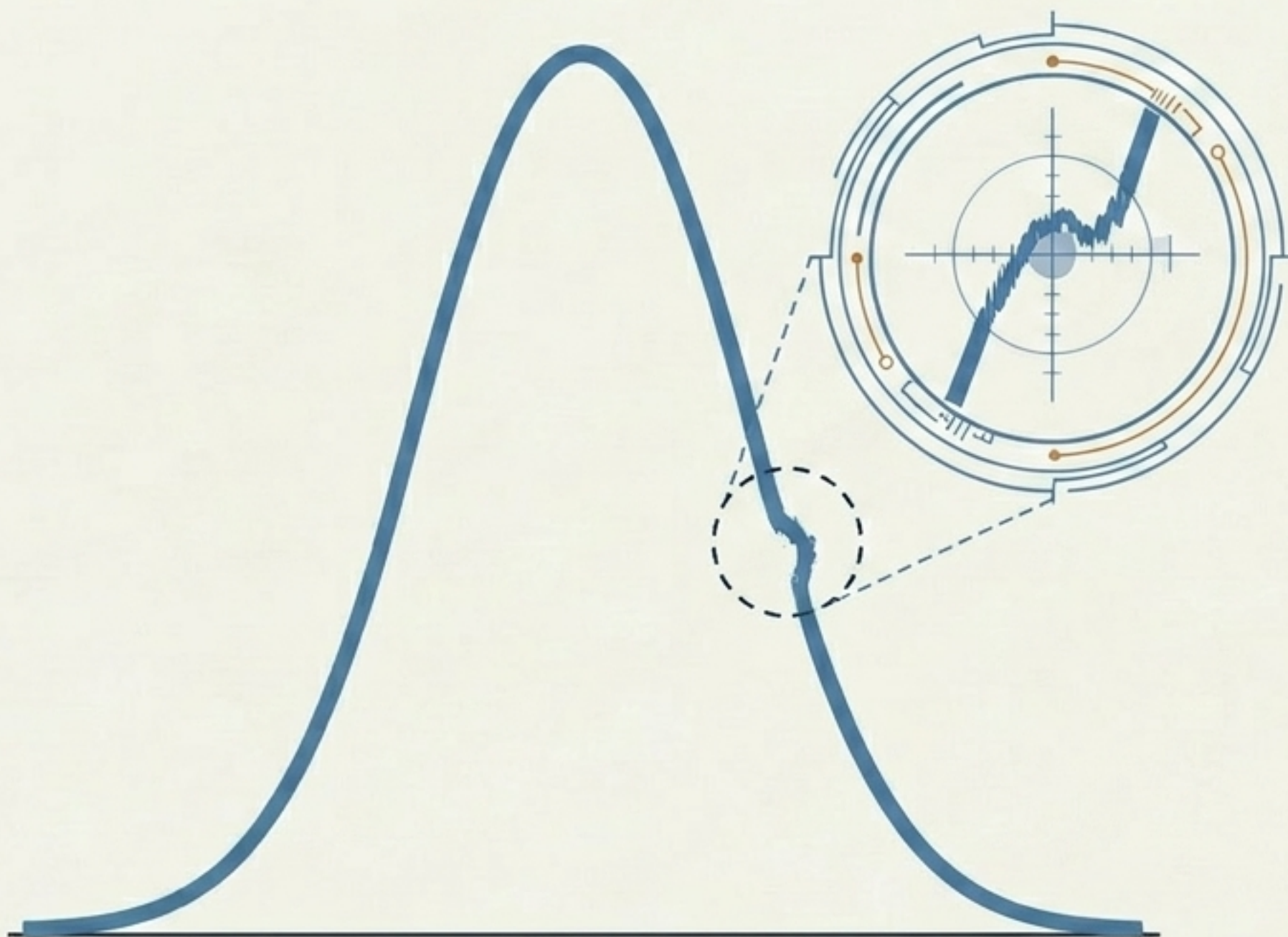
PREČO HO POUŽIŤ?

Je to vyvážený test a absolútny zlatý štandard pre priemyselné aplikácie a metodiku Six Sigma.

HLAVNÁ VÝHODA:

Excelentne reaguje na odchýlky nielen v strede rozdelenia, ale aj na jeho okrajoch (chvostoch), kde sa často ukrývajú kritické defekty.

DIAGNOSTIKA: VEĽKÉ SÚBORY ($n > 300$ až do 5000)



Odporúčaný prístup:

Kolmogorov-Smirnov alebo prechod na čisto vizuálnu analýzu.

Paradox veľkých dát:

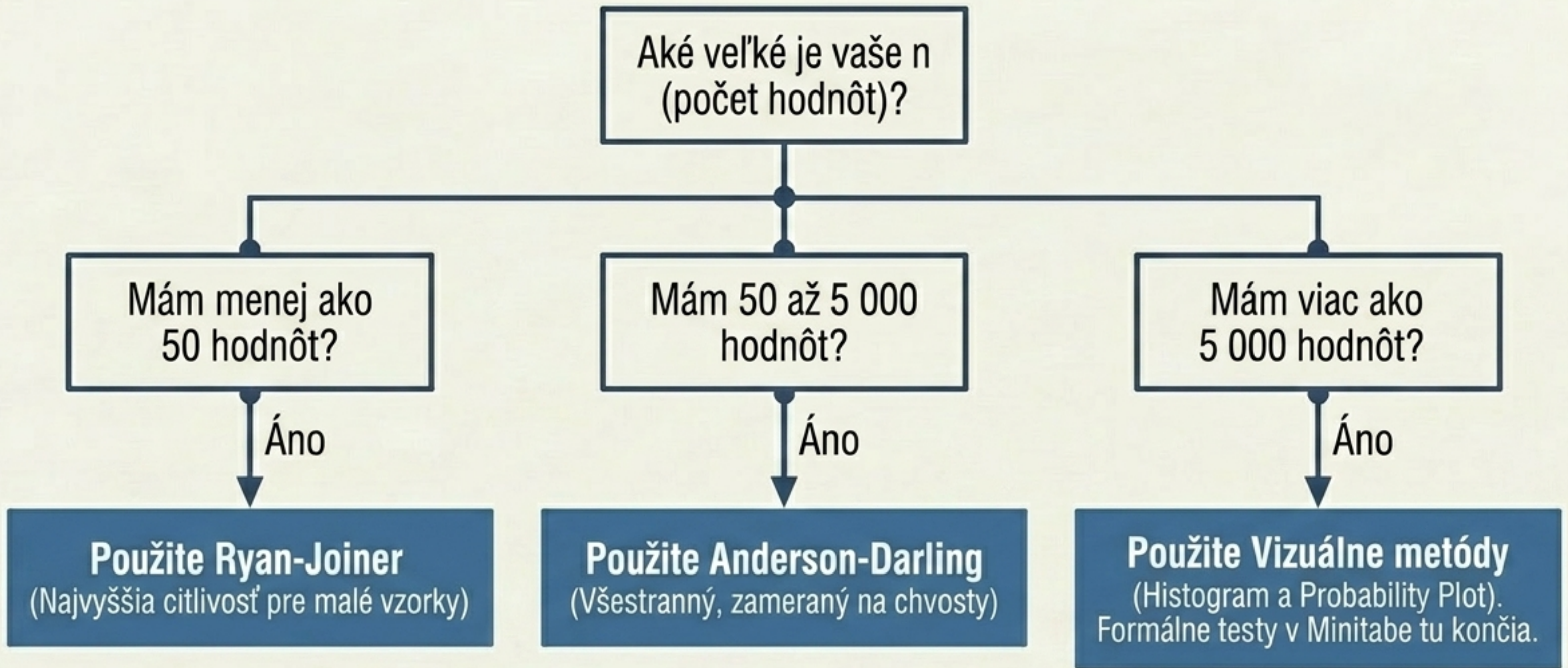
Pri veľkom n sú štatistické testy príliš prísne. Aj drobné, prakticky nevýznamné odchýlky vedú k prepadu p -hodnoty pod 0,05.

! $p < 0,05$!

Záver pre analytika:

Formálne testy tu strácajú praktický zmysel. Čísla musia ustúpiť vizuálnemu posúdeniu.

ROZHODOVACÍ STROM V PROSTREDÍ MINITAB



Porovnanie algoritmov: Čo beží na pozadí?

Detailný pohľad na tri testy z dialógového okna Minitab.

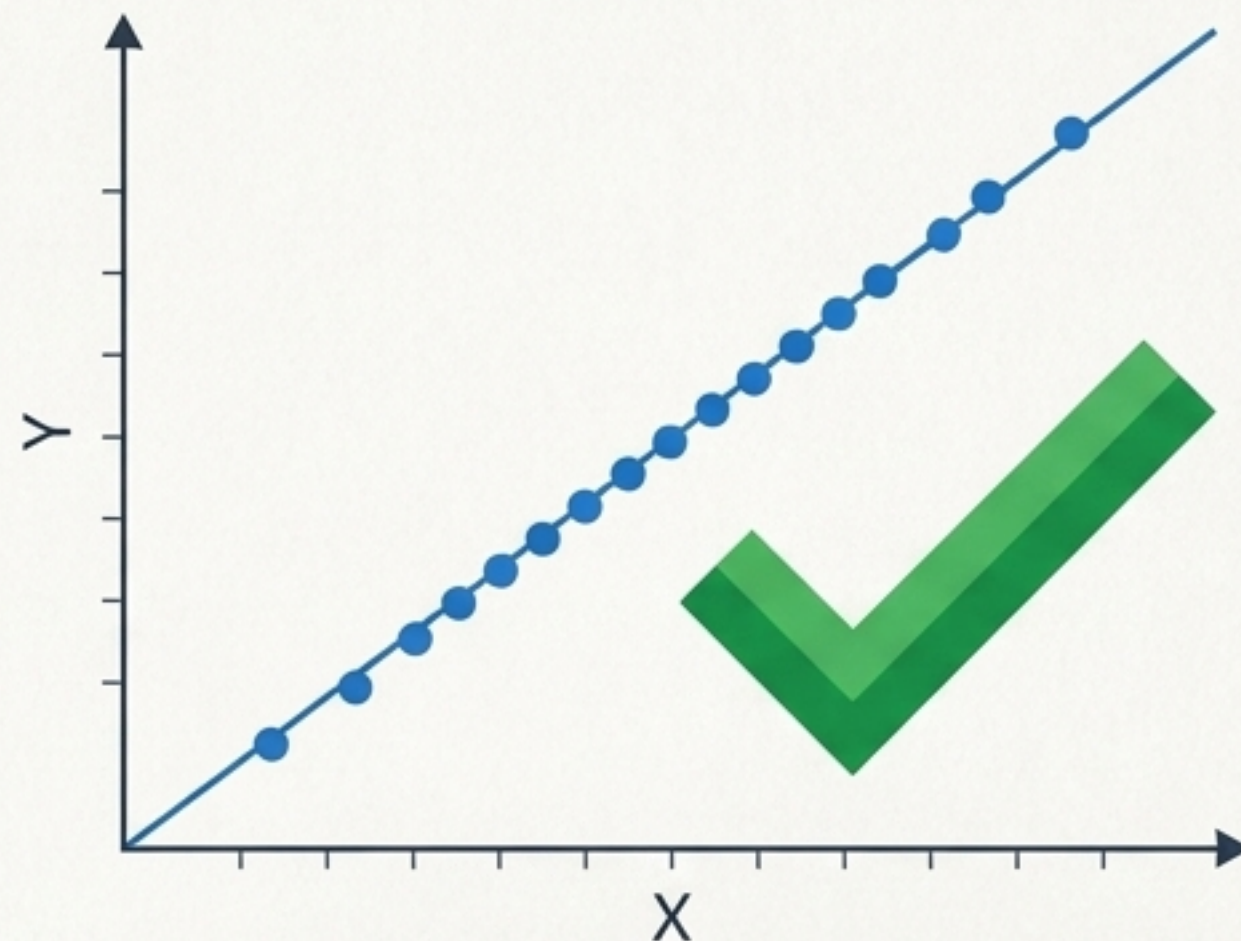
 Test	$\Sigma \int$ Matematický základ	 Silné stránky
Anderson-Darling (AD)	Kvadratická vzdialenosť	Detekcia odľahlých hodnôt a anomálií na chvostoch rozdelenia.
Ryan-Joiner (RJ)	Korelačný koeficient	Najlepšia náhrada testu Shapiro-Wilk pre malé vzorky.
Kolmogorov-Smirnov (KS)	Maximálna vzdialenosť	Konzervatívny prístup, vhodný pre veľmi veľké súbory dát.

Expertný tip #1: Pravidlo Vizual nad P-hodnotu



Limitácia formálneho testu:

Samotná p-hodnota nie je všetko. Slepá viera vo výsledok formálneho testu môže viesť k zamietnutiu funkčného procesu.

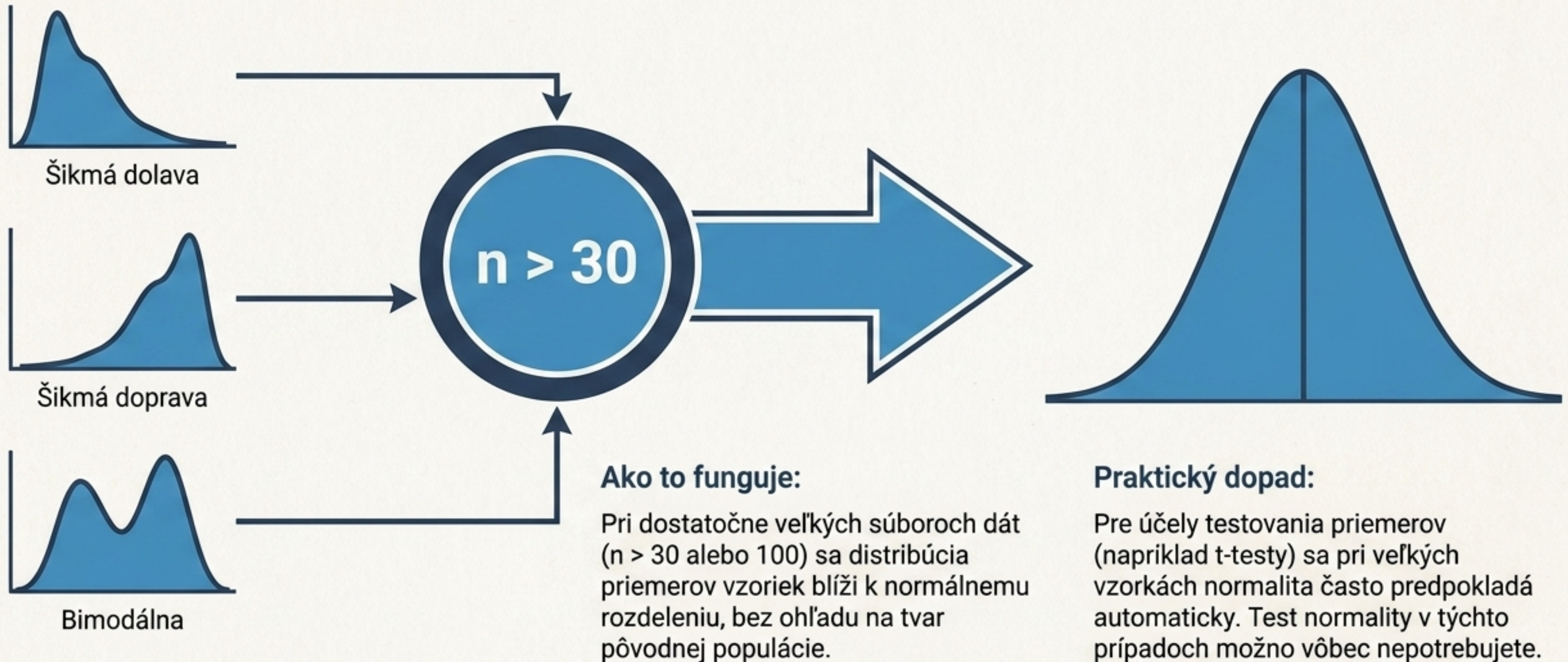


Riešenie a konečný arbiter:

Graf pravdepodobnosti (Probability Plot) je dôležitejší. Ak dátové body vizuálne ležia na priamke, proces je v poriadku a stabilný, aj keď je p-hodnota tesne pod hranicou 0,05.

Expertný tip #2: Štatistický Lifehack (CLT)

Centrálna limitná veta (Central Limit Theorem)



Diagnostický ťahák: Zhrnutie pravidiel

1 $\lceil n^x \rceil$

Zvoľte test podľa veľkosti n.

Prispôbajte nástroj objemu dát.
Pre $n < 50$ voľte Ryan-Joiner.

2 ★

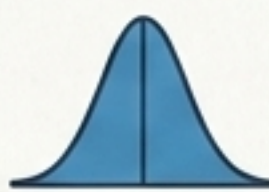
Zlatý štandard.

Pre stredné vzorky a priemyselné aplikácie
je Anderson-Darling najvšestrannejšia voľba.

3 

Vizuál vždy dopĺňa test.

Probability Plot je konečným arbitrom.
Body na priamke prebývajú p-hodnotu.

4 

Pamätajte na CLT.

Pri testovaní priemerov nad $n > 30$ sa
normalita generuje sama. Neanalyzujte
zbytočne.