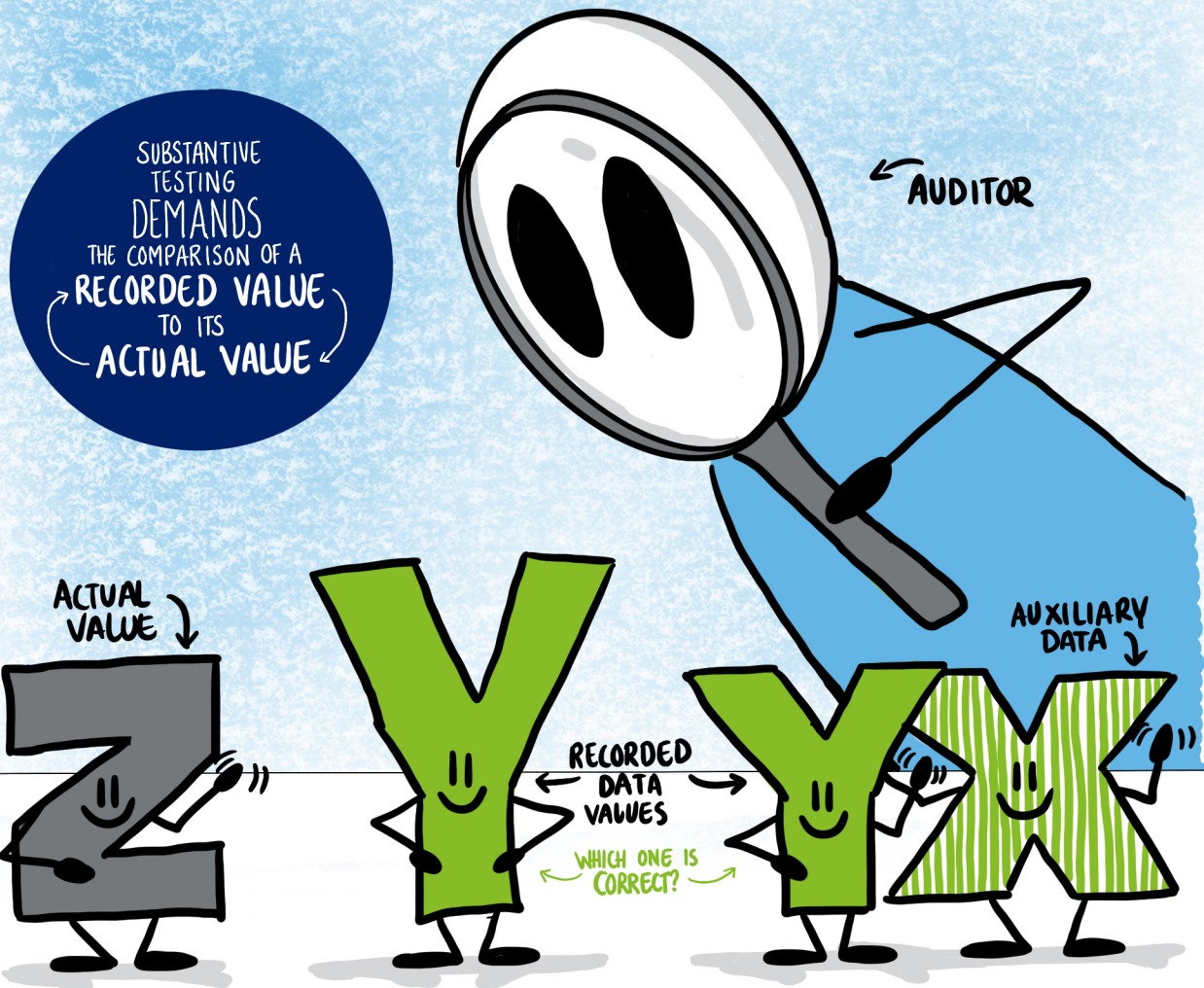


A Classification of different methods used in Data Analytics for substantive testing in auditing*



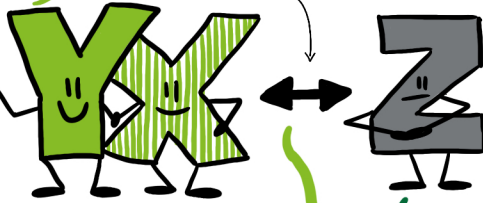
Deloitte.

* OF COURSE, DATA ANALYTICS CAN ALSO BE APPLIED FOR RISK ASSESSMENT OR TO INVESTIGATE OR TEST CONTROLS, BUT THE **FOCUS** IN THIS GUIDE IS ON SUBSTANTIVE TESTING.

WHERE IS THE ACTUAL VALUE?

LET US ASSUME THERE IS A SET OF DATA TO BE AUDITED Y, AND PERHAPS AUXILIARY DATA X, AND BOTH ARE AVAILABLE IN A DIGITAL FORMAT

SUBSTANTIVE TESTING MEANS THAT WE COMPARE Y TO THE ACTUAL VALUES Z



PERFORM A 100% TEST

SAMPLING = AWKWARD WHEN 100% MATCH IS POSSIBLE!

USE (AUDIT) SOFTWARE TO COMPARE Y ↔ Z!

"COMPLY OR EXPLAIN" WOULD SUIT NICELY IN ISA 500!

0101101001
1011000100
10110
0011
0011
110101010
0001000101

—YES!—

AVAILABLE IN A DIGITAL FORMAT?

NO!

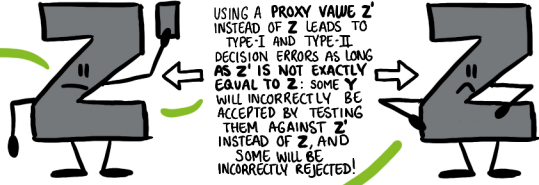
SEARCH FOR Z OUTSIDE THE COMPUTER

AUDITOR



CONSTRUCT A PROXY Z'

MANY APPLICATIONS OF DATA ANALYTICS FALL WITHIN THIS CATEGORY: THE ACTUAL VALUES Z ARE NOT, OR NOT EFFICIENTLY USEFUL AND THE AUDITOR USES A PROXY VALUE Z'



USING A PROXY VALUE Z' INSTEAD OF Z LEADS TO TYPE-I AND TYPE-II DECISION ERRORS AS LONG AS Z' IS NOT EXACTLY EQUAL TO Z. SOME Y WILL INCORRECTLY BE ACCEPTED BY TESTING THEM AGAINST Z' INSTEAD OF Z, AND SOME WILL BE INCORRECTLY REJECTED!

PERFORM A STATISTICAL SAMPLE

ONLY IF IT IS EFFICIENT!

SAMPLE SIZE → DETERMINED BY:

ALLOWED RISK

✓ vs ✗ # ✗
ERROR RATE SAMPLE ACTUAL ERROR RATE

REQUIRED PRECISION

WHEN: CLOSE TO REQUIRED PRECISION → INEFFICIENT!

PROPAGATING PROXY VALUES Z'

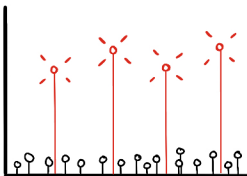
METHODS:

1 BASED ON Y ITSELF

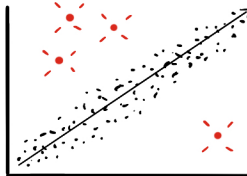
MOST SIMPLE WAY:
LOOKING AT

EXTREMES
IN Y

*OUTLIERS



*ANOMALIES



* ALL RESULTS OUTSIDE OF TWICE THE STANDARD DEVIATION FROM THE MEAN

* LARGEST AND SMALLEST RESULTS

DO NOT USE THESE APPROACHES WITHOUT A TRUE RATIONALE!

EXTREMES EXIST!

IN A POPULATION OF A "NORMALLY" DISTRIBUTED VARIABLE IT WOULD BE STRANGE IF THERE WOULD BE NO RESULTS OUTSIDE TWICE THE STANDARD DEVIATION FROM THE MEAN!

MORE IMPORTANT:

IMPLAUSIBILITY \neq INCORRECTNESS
&
PLAUSIBILITY \neq CORRECTNESS

I GOT AWAY WITH IT!

WITH 1, 2, 3 & 4: DERIVING A PROXY VALUE Z' FROM THE DATA Y DOES NOT RESULT IN A Z' AS A SPECIFICATION OF A CORRECT VALUE BUT OF AN INCORRECT VALUE! WE CALL THIS:

AUSOPOOR

AUTOMATED SEARCH FOR POSSIBLE ERRORS!



AUSOPOOR



AUDITORS PERFORM RISK ASSESSMENT TO INDICATE WHAT CAN GO WRONG & INSTRUCT DATA ANALYSTS TO BUILD A ROUTINE TO FIND SYMPTOMS OF THOSE POSSIBLE ERRORS.

WHEN HAS THE AUDITOR SPECIFIED ALL ERROR TYPES AND FOUND ALL ERRORS?

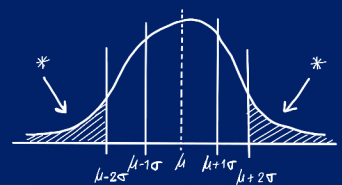


ISA 500. A2:

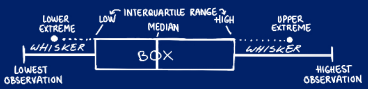
"THE ABSENCE OF INDICATIONS OF ERRORS DOES NOT LEAD TO THE CONCLUSION THAT ERRORS ARE ABSENT"

SELECT A SAMPLE?!

1 GAUSSIAN (NORMAL) DISTRIBUTION

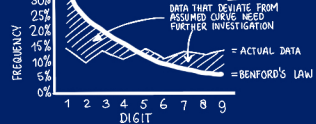


2 BOX AND WHISKER PLOTS



3 BENFORD - ANALYSIS

NO SCIENTIFIC PROOF!



4 CLUSTERING

DEVIDE DATA INTO GROUPS THAT ARE VERY SIMILAR WITHIN ANY GROUP BUT VERY DIFFERENT BETWEEN GROUPS



"WILL YOU FIND INCORRECT TRANSACTIONS THIS WAY?"
"A SMART FORWARDER WOULD FAME TRANSACTIONS THAT DO NOT STAND OUT!"

2 BASED ON THE RELATION BETWEEN Y AND X, THE AUXILIARY DATA



BASED ON HOW THE AUDITOR KNOWS OR ASSUMES THAT X INFLUENCES Y, X CAN BE USED TO BUILD Z'

CONTRARY TO AUSOPOOR-METHODS, NOW Z' IS THE BEST ESTIMATE OF THE CORRECT VALUE FOR Y!



IS X USEFUL / CORRECT?

MARGIN ANALYSIS

- SIMPLE EXAMPLE -



- OR MORE COMPLEX -
ANALYTICAL REVIEW BASED ON REGRESSION ANALYSIS OR OTHER MODEL ESTIMATION METHODS!



FOR THESE ALGORITHMS THAT YIELD Z IT IS EXTREMELY IMPORTANT THAT ("AT LEAST") THE DATA X ON WHICH THE MODEL WAS TRAINED ARE REPRESENTATIVES AS A PREDICTOR TO AUDIT Y !

THIS GUIDE IS BASED ON A COLUMN (IN DUTCH) ON [ACCOUNTANT.NL](https://www.accountant.nl/artikelen/2019/8/data-analyse...een-poging-om-de-bodem-in-het-bos-te-zien) BY PAUL VAN BATENBURG

[HTTPS://WWW.ACCOUNTANT.NL/ARTIKELN/2019/8/DATA-ANALYSE...EEN-POGING-OM-DE-BODEM-IN-HET-BOS-TE-ZIEN](https://www.accountant.nl/artikelen/2019/8/data-analyse...een-poging-om-de-bodem-in-het-bos-te-zien)

Deloitte.

FEEROY TEKENT.NL
#TEKENENDE ACCOUNTANT