Statistical process control

Rebekah Worsfold
TSB Bank

Adobe

Is this real – or is it noise?

Data in CJA → **Dashboard built** → **Business reporting**

BUT…

Lot of time spent investigating why a change has occurred – when often it's just **normal variation**

Adobe

You need statistical process control (SPC)

What is SPC?

SPC is a process where you

1. Track data
2. Use statistics to determine what's normal and what's not

Why do we need SPC?

SPC improves business decision making and effectiveness

1. You only investigate the things that need investigating
2. Trends are spotted easier and quicker because you can block out the noise

How do you do SPC?

Method will depend on data type

1. Trend data
2. A/B testing

Adobe

Trend data – look at deviation from the mean

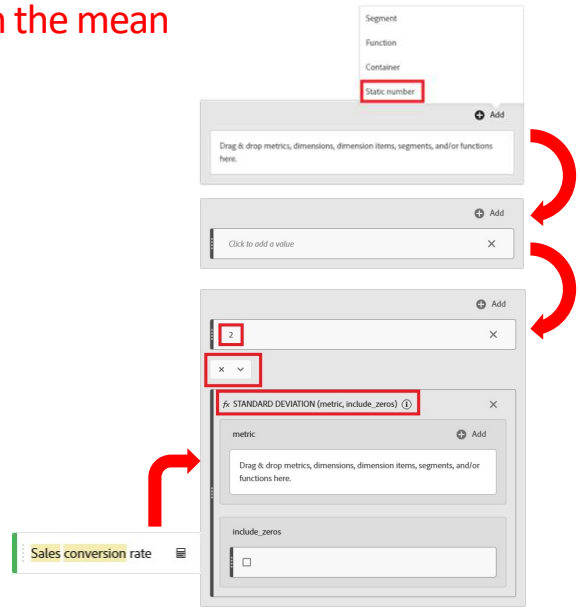
Create a metric for the standard deviation multiplied by 2

Create a metric for the mean

Create 2 metrics for the confidence intervals

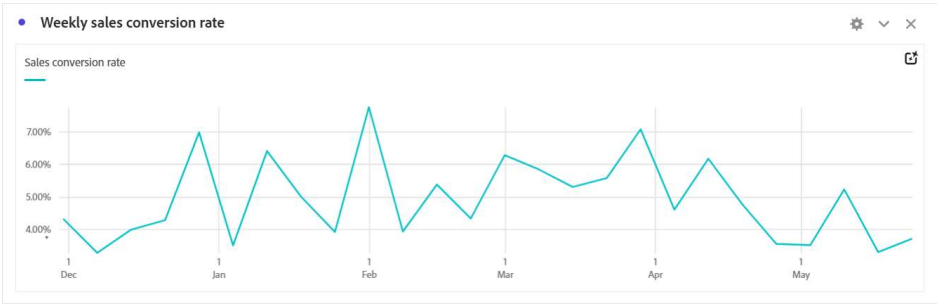
$$CI = Mean(SCM) \pm 2 * StdDev(SCM)$$

Add the mean and confidence intervals to your trend graph



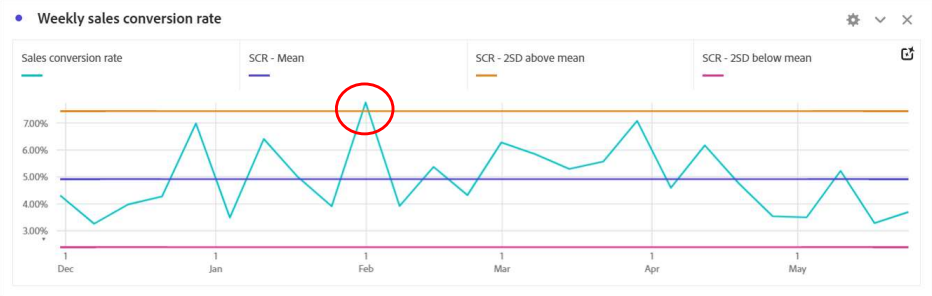
Adobe

This changes our sales conversion graph from this…



Adobe

…to this!



We combine this with a series of rules to identify anomalies:

- Points outside of confidence intervals
- 7+ points all above or all below the mean
- Patterns in the data (e.g. upward or downward trend, seasonality)

A/B testing – use a chi-squared test to assess significance

People in target audience | People in control audience

	Target Total	Target Sale	Target Sale %	Control Total	Control Sale	Control Sale %	Uplift %	Sales Uplift
Segments Page: 1 / 1 Rows: 40								
1. CS campaign	165,645	6,026	3.6%	95,099	2,195	2.3%	1.3%	2,203

BUT is this statistically significant?

OBSERVED	Sale	No sale	Total
Target	6,026		165,645
Control	2,195		95,099
Total			

Create 5 more calculated metrics to complete the table

Step 1: Calculate metrics for your observed and expected values

	Target Sale	Target No Sale	Target Total	Control Sale	Control No Sale	Control Total	Sale Total	No Sale Total	Overall Total
Segments Page: 1 / 1 Rows:									
1. CS campaign	6,026	159,619	165,645	2,195	92,904	95,099	8,221	252,523	260,744

OBSERVED	Sale	No sale	Total
Target	6,026	159,619	165,645
Control	2,195	92,904	95,099
Total	8,221	252,523	260,744

EXPECTED	Sale	No sale	Total
Target			165,645
Control			95,099
Total	8,221	252,523	260,744

$$Expected\ value = \frac{Row\ total\ *Column\ total}{Overall\ total}$$

$$Expected\ Target\ Sale = \frac{Target\ total\ *Sale\ total}{Overall\ total}$$

Adobe

Step 1: Calculate metrics for your observed and expected values

	Target Sale	Target No Sale	Target Total	Control Sale	Control No Sale	Control Total	Sale Total	No Sale Total	Overall Total
Segments Page: 1 / 1 Rows:									
1. CS campaign	6,026	159,619	165,645	2,195	92,904	95,099	8,221	252,523	260,744

OBSERVED	Sale	No sale	Total
Target	6,026	159,619	165,645
Control	2,195	92,904	95,099
Total	8,221	252,523	260,744

EXPECTED	Sale	No sale	Total
Target	5,223	160,422	165,645
Control	2,998	92,101	95,099
Total	8,221	252,523	260,744

	Expected Target Sale	Expected Target No Sale	Expected Control Sale	Expected Control No Sale
Segments Page: 1 / 1 Rows: 400 1-1 of 1				
1. CS campaign	5,223	160,422	2,998	92,101

Adobe

Step 2: Calculate chi-squared critical values

$$\chi^2 = \sum \frac{(\text{Observed} - Expected)^2}{Expected}$$

- Calculate the critical values for each of the four cells
- Sum the four metrics to calculate the chi-squared critical value

OBSERVED	Sale	No sale
Target	6,026	159,619
Control	2,195	92,904

EXPECTED	Sale	No sale
Target	5,223	160,422
Control	2,998	92,101

VALUE	Sale	No sale
Target		
Control		

$$Critical\ value\ of\ Target\ Sale = \frac{(Observed\ Target\ Sale\ -\ Expected\ Target\ Sale)^2}{Expected\ Target\ Sale}$$

Adobe

Step 2: Calculate chi-squared critical values

$$\chi^2 = \sum \frac{(\text{Observed} - Expected)^2}{Expected}$$

- Calculate the critical values for each of the four cells
- Sum the four metrics to calculate the chi-squared critical value

OBSERVED	Sale	No sale
Target	6,026	159,619
Control	2,195	92,904

EXPECTED	Sale	No sale
Target	5,223	160,422
Control	2,998	92,101

VALUE	Sale	No sale
Target	124	4
Control	215	7

Difference is statistically significant if $\chi^2 > 3.841$

	Chi-squared Target Sale	Chi-squared Target No Sale	Chi-squared Control Sale	Chi-squared Control No Sale	Chi-squared	Significant
Segments Page: 1 / 1 Rows: 400 1-1 (						
1. CS campaign	124	4	215	7	350	1

Adobe

Benefits to the business

- Can identify genuine anomalies in the data
  - Avoiding kneejerk reactions
  - Not wasting time with unnecessary investigations
  - Can spot and fix genuine problems more quickly
- Can flag campaigns that are having a significant impact
  - Learn about which campaigns are particularly successful
  - Improve the quality of our marketing comms over time

Adobe

Summary

Trend data

1. Calculate the mean and +/- 2 standard deviations
2. Look for anomalies
   1. Points outside the confidence intervals
   2. 7+ points all to one side of the mean
   3. Patterns in the data

A/B testing

1. Calculate chi-squared equation

$$\chi^2 = \sum \frac{(\text{Observed} - Expected)^2}{Expected}$$

2. Look for χ^2 value > 3.841

Adobe