Explaining Statistical Significance

Are two results really different or could they have been obtained by chance?

Note: there is an acceptable element of chance built into statistical testing – typically 5%.

Why test for significance?

The results we observe when measuring a sample of people may not be a true reflection of the population.

Statistical significance is an objective measure that shows whether we can be 95% confident differences observed between groups being studied are "real."

When a result is statistically significant from another result it means that:

You can feel confident that the difference between two groups is real.

When a result is not statistically significant it means that:

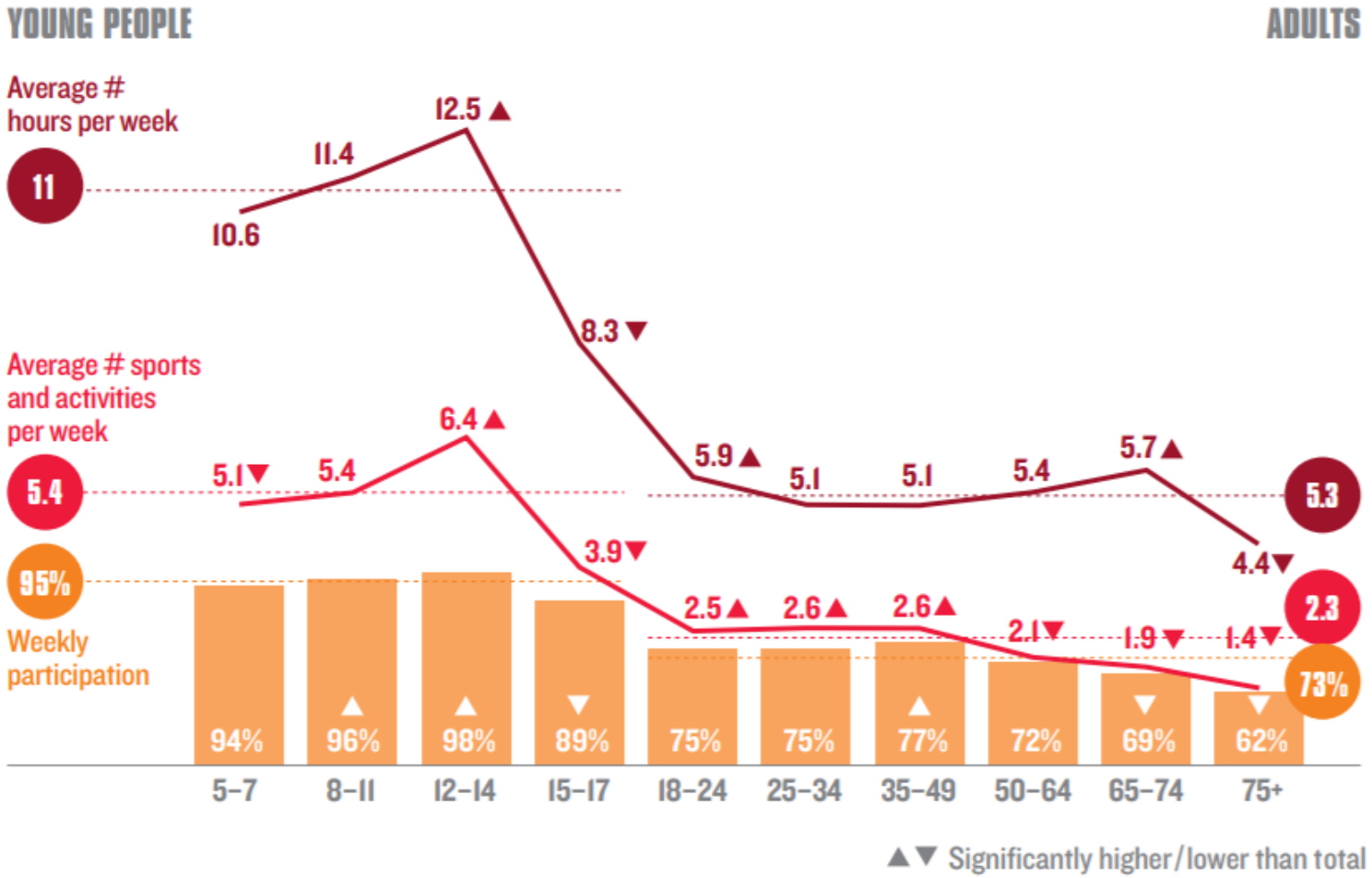
We cannot say with any confidence that there is a difference between two groups.

Example

In the Active NZ Main Report 2017 published in 2018, triangles are used to mark where results for a group are significantly different from the total.

For example, results found that weekly participation among young people is 95% and that those aged 12-14 are significantly more likely to participate weekly (98%).

As this result is statistically significant (as indicated by the triangle) we can be 95% confident the difference is real and not due to chance.



Source: Sport New Zealand, 2018. Active NZ 2017 Participation Report. Wellington: Sport New Zealand. ISBN: 978-0-947502-73-7

What impacts statistical significance?

Sample size

The likelihood of obtaining statistically significant results increases as the sample size increases.

Why?

Because the margin of error decreases as the sample size increases (the margin of error provides a measure of variability between the actual survey estimate and the population).

Large sample sizes

- As the sample size increases so does the likelihood of seeing statistically significant results.
- With very large sample sizes, you're virtually certain to see statistically significant results for even very small differences between groups.
- Therefore, a significant difference between groups with a large sample size may be meaningless in terms of importance.

Small sample sizes

- Small samples sizes often do not yield statistical significance.
- Even when a difference seems big, if the sample is small, you may not be able to confidently say that the difference is due to anything other than chance.

Sample weighting

Why weight survey data?

- Ideally a selected sample is representative of the population it came from in all respects, e.g. age, ethnicity – this is usually not the case.
- A weighting adjustment is applied to balance up the sample, so it is reflective of the population.

What is the impact of weighting when testing for significance?

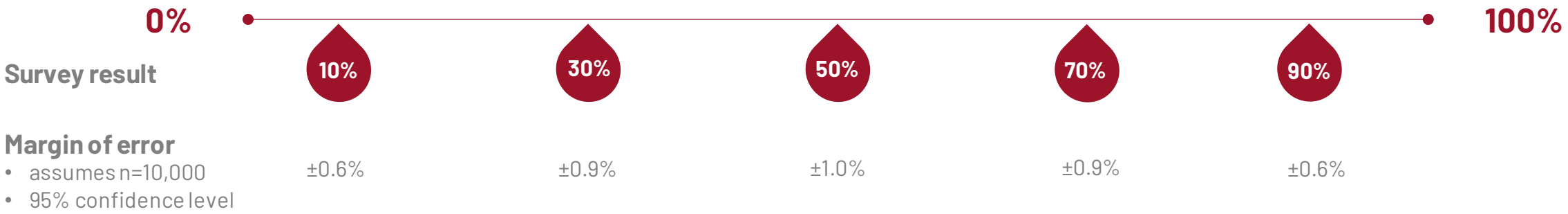
- When statistical tests are run on data which has been weighted the test uses the 'effective base'[1] size.
- The effective base takes into account adjustments made to the sample by weighting.
- It is used because it reduces the likelihood of the statistical tests producing significant results because of the adjustments made by weighting.
- The effective base is not reported in any Active NZ releases as it only applies to significance testing.

[1] The effective base =(sum of weight factors)^2 / sum of squared weight factors.

The survey result

As a result gets closer to 0% or 100%, the margin of error (a measure of variability between the actual survey estimate and the population) decreases

Because the margin of error is smaller, the difference between two results doesn't need to be as big to be significant



Note: The level of confidence is set at 95%. This means that the "true" percentage for the entire population would be within the margin of error of a surveys' reported percentage 95% of the time.

SPORT
NEW ZEALAND
IHI AOTEAROA

Knowing a difference is statistically significant does not mean the difference is important, relevant or useful

How do we determine whether a result is 'practically' significant?

That is, important, relevant or useful

- We need to determine whether the difference is meaningful within your particular context.
- There's not a one size fits all formula to guide this and it's more of a subjective measure.
- Some things to consider:
  - What is the absolute value of the percentage point difference between groups (e.g. 4% vs. 15%)?
  - What was the sample size?
  - What other patterns can you see within the overall data?
  - Contextualise the results with other sources of data or information (see the insights approach)
  - Think about the implication of each finding for the decision you're hoping to make and your organisation's objectives.

Example

An analysis of end-of-year maths test results found that girls are significantly more likely to pass their end of year school maths test than boys (88% cf 84%)

Some things you might consider:

- What is your objective/purpose for doing this analysis?
- What is the percentage point difference
  - Are the results big enough that you should change your behaviour? What about if the results were 88% cf. 64%?
- Overall results
  - How does this result compare to others (e.g. for English, science, etc); is there a trend?
- How do subgroups compare?
  - Are there differences between subgroups (e.g. those aged 10-12, low/high socioeconomic)? There might be one age group that are doing particularly well or not so well that leads to the group result being higher or lower.

SPORT
NEW ZEALAND
IHI AOTEAROA

Sometimes what is and isn't statistically significant can be confusing. Here are some examples...

Example:
Why is a 1 percentage point difference from the total significant but a 9 percentage point difference is not?

Active NZ Young People 5-17 yrs (2017 results)

APPENDIX A

		Gender		
SPORT NEW ZEALAND	TOTAL	Male	Female	Indian
PLEASE NOTE: AS IS THE CASE THROUGHOUT THESE DATA TABLES, RESULTS ARE ROUNDED TO THE NEAREST WHOLE NUMBER (EG. 0.4% WOULD BE ROUNDED DOWN TO 0%). WHERE IT SAYS 0% IT DOES NOT MEAN THERE ARE ZERO YOUNG PEOPLE IN NEW ZEALAND WHO PARTICIPATE IN THIS SPORT OR ACTIVITY.				
PERCENTAGE WHO HAVE PARTICIPATED IN EACH SPORT OR ACTIVITY IN THE 7 DAYS PRIOR				
Q13. PLEASE TICK ALL THE WAYS YOU HAVE BEEN PHYSICALLY ACTIVE FOR SPORT, PE, EXERCISE OR FUN IN THE LAST 7 DAYS (NOT INCLUDING TODAY)				
Base: All respondents aged 5 to 17				
Total (Unweighted)	6004	2894	3095	118
Running, jogging or cross-country	52%	52%	51%	55%
Playing (e.g. running around, climbing trees, make-believe)	41%	42%	39%	38%
Swimming	36%	34%	37%	25%

The sample size may be very small for the subgroup - even when a difference seems big, if the sample is small, you may not be able to confidently say that the difference is due to anything other than chance

Example:
Why is one result significant but the other is not when they have the same percentage?

			Deprivation		
SPORT NEW ZEALAND IHI AOTEAROA	TOTAL	European (net)	Deprivation Low (1-3)	Deprivation Medium (4-7)	Deprivation High (8-10)
PROPORTION WHO WANT TO PARTICIPATE MORE					
Q22.WOULD YOU LIKE TO BE DOING MORE PHYSICAL ACTIVITY FOR SPORT, PE, EXERCISE OR FUN THAN WHAT YOU DO NOW?					
Base: All respondents aged 5 to 17					
Total (Unweighted)	6004	5118	2317	2194	1129
Yes - I would like to be doing more	64%	62%	62%	64%	66%
No	36%	38%	38%	36%	34%
Sum	100%	100%	100%	100%	100%

This is most likely due to effective base size of the groups (although not shown in the table above the, effective base size is smaller for the result which is not significant).

The effective base takes into account the unweighted sample size of the group and the extent to which this group has been weighted.

Example:
There is only a 1 percentage point difference from the total, why is it classed as significant?

		Gender		Age			
SPORT NEW ZEALAND IHI AOTEAROA	TOTAL	Male	Female	5 to 7 years	8 to 11 years	12 to 14 years	15 to 17 years
WEEKLY PARTICIPATION IN SPORT AND ACTIVE RECREATION (LAST 7 DAYS)							
Q12. IN THE LAST 7 DAYS (NOT INCLUDING TODAY) HAVE YOU DONE ANY PHYSICAL ACTIVITY SPECIFICALLY FOR SPORT, PHYSICAL EDUCATION (PE), EXERCISE OR FUN?							
Base: All respondents aged 5 to 17							
Total (Unweighted)	6004	2894	3095	1605	2347	1252	800
Yes	95%	95%	94%	94%	96%	98%	89%
No	5%	5%	6%	6%	4%	2%	11%
Sum	100%	100%	100%	100%	100%	100%	100%

A number of factors influence whether a result is statistically significant including:

- large sample size
- weighting of the sample
- strength of response – more likely to be significant the closer the result gets to 0% or 100%.

To summarise

Key points

- Statistical significance shows whether any differences observed between two groups being studied are "real" or whether they are simply due to chance
- Sample size, sample weighting and the strength of the response can impact whether a result is significant or not
- Statistical significance does not tell you whether the result is important, relevant or useful to you
- Practical significance is a subjective measure – you need to take into account your objectives, other data and your specific context to determine whether a result is important, relevant or useful to you.

Thank you

SPORT
NEW ZEALAND
IHI AOTEAROA