



Get
every one
in the picture

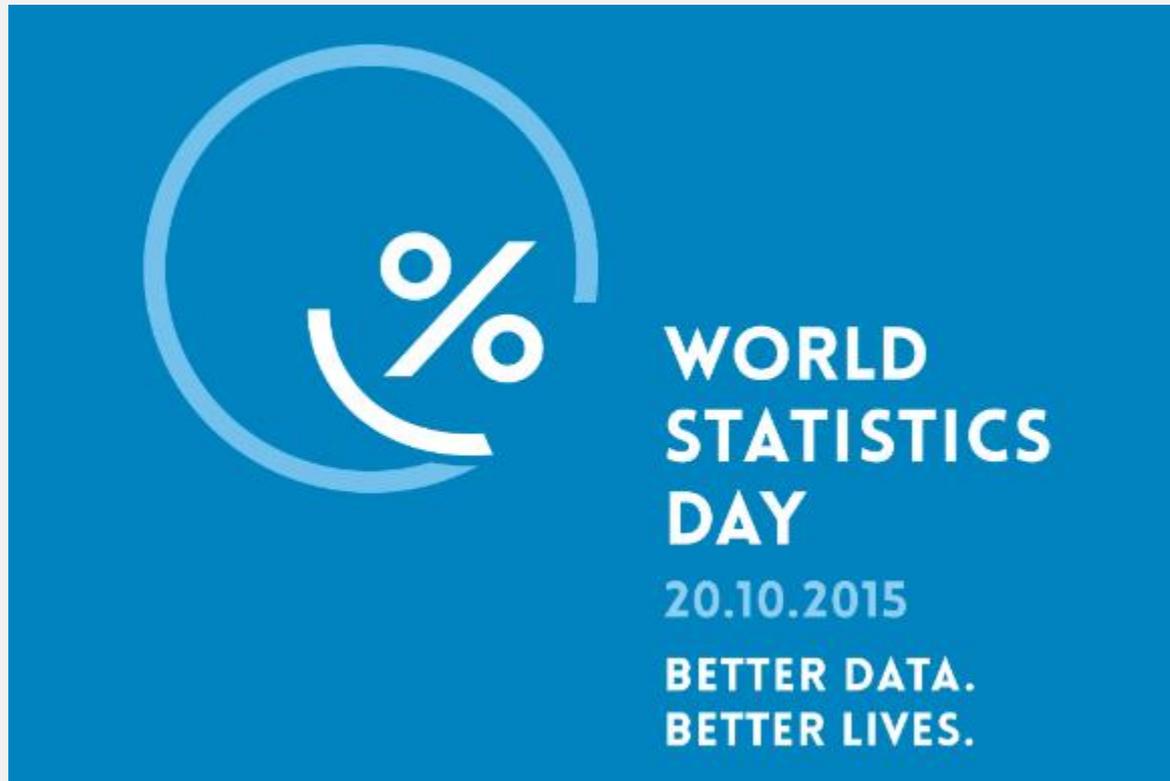
Data quality and comparison - fertility

Data analysis and Report writing
workshop for Civil registration and
vital statistics data.

Why do we need to check the quality of our results?

- ◆ Vital statistics are used to monitor and measure the health situation in countries
- ◆ Policy makers will use our data to:
 - ◆ Become aware of current situation of the country
 - ◆ Measure the effectiveness of past policies
 - ◆ Target interventions and
 - ◆ Determine future priorities

Why do we need to check the quality of our results?



It's important that our data be as accurate and reliable as possible

Why do we need to check the quality of our results?

- ◆ Data quality is something that we should consider at every step.
 - ◆ Make sure unit record data is as good as possible
 - ◆ Do tabulations/analysis and make judgements
- ◆ Aim to identify the most important sources of error
- ◆ Provide quantitative measures where possible or qualitative descriptions otherwise
- ◆ Should be informative to readers

Core data quality concepts

- ◆ Consistency
- ◆ Coverage (or scope)
- ◆ Representativeness
- ◆ Completeness
- ◆ Validity
- ◆ Reliability
- ◆ Bias



Consistency

Also known as “comparability” is:

- ◆ Description of data over time
- ◆ Does it follow a similar pattern from year to year (from month to month)?
- ◆ Are there significant gaps or peaks in our data?

Coverage (or scope)

- ◆ Describes the area or population that the data set covers/includes
- ◆ Noting any group not being able to record in the current system

Representativeness

- ◆ Related to coverage
- ◆ Means how similar the covered population reflects the broader population of interest



Completeness

- ◆ Is used to assess what proportion of the events within our area of coverage that we intended to capture, did we actually manage to collect data for.
- ◆ The term can also be used to refer to the completeness of key fields within the data set.

Validity

- ◆ The plausibility of both our raw data (in terms of number of events) and of calculated measurements.

Reliability

- ◆ Is the system able to produce results of similar quality over time.

Bias

- ◆ A systematic effect on a statistic or measurement, rather than stochastic/random one.
- ◆ Generally related to some aspect of the data collection which results in being more likely to see particular answers.

ABS Data Quality Framework





**Anyway, we
should check our
data plausibility**

Checking for plausibility

- ◆ Do these results make sense? Are they believable?
- ◆ It's important to compare results to data from:
 - ◆ Previous years' vital statistics
 - ◆ Census estimates
 - ◆ Household surveys (DHS & MICS)
 - ◆ Sample registration systems
 - ◆ Academic estimates

Checking for plausibility

- ◆ Make tabulations or graphs of the total number of events by year/month
- ◆ Do the same for sub-regions to identify any reporting problems in the data
- ◆ Discuss comparability of results especially in change of methodology, concepts or definitions
- ◆ Describe effects/possible effects of benchmarking or revisions on comparability over time
- ◆ For small populations, in order to avoid poor consistency, aggregate data over several years before calculations

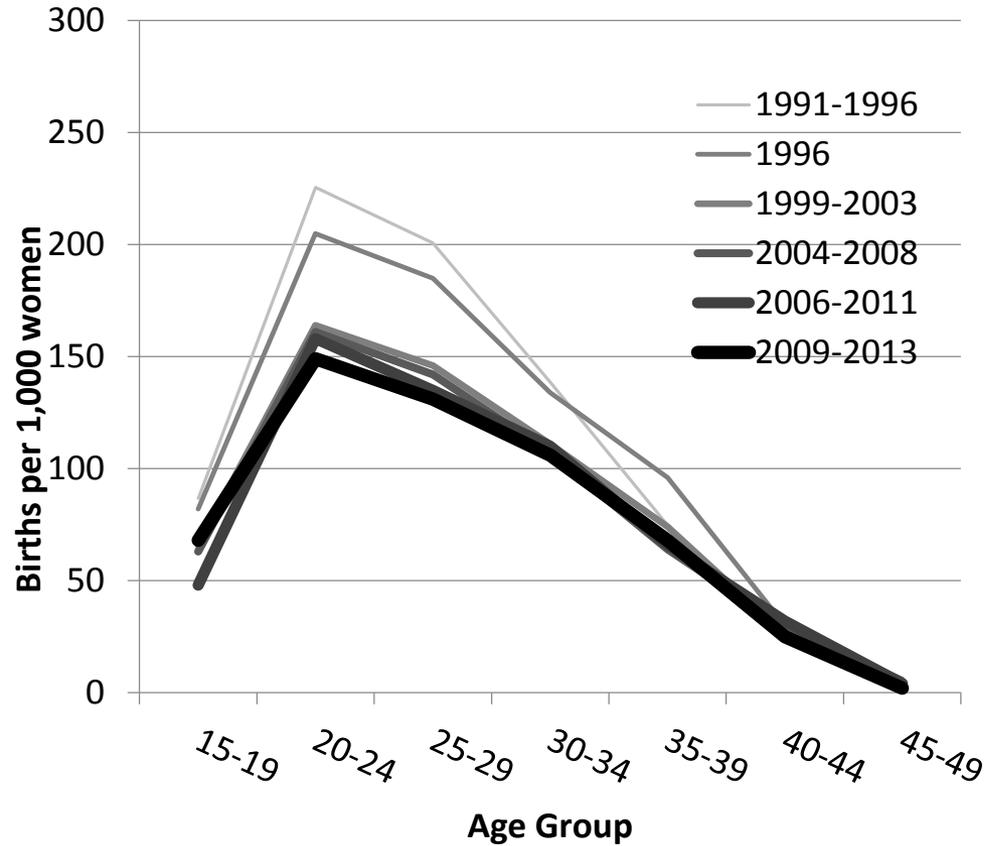
Plausibility of birth statistics

- ◆ Compare raw counts of birth and counts by Region, mother's age, sex of the baby to data from previous years and recent census data
- ◆ Can you explain any discrepancies?
 - ◆ For example, a natural disaster may destroy a hospital which causes decreased birth numbers in a certain area and increases in neighboring areas.
- ◆ Compare other statistics such as weight, length of gestation, place of birth, attendant at birth etc. to previous years' data

Plausibility of fertility statistics

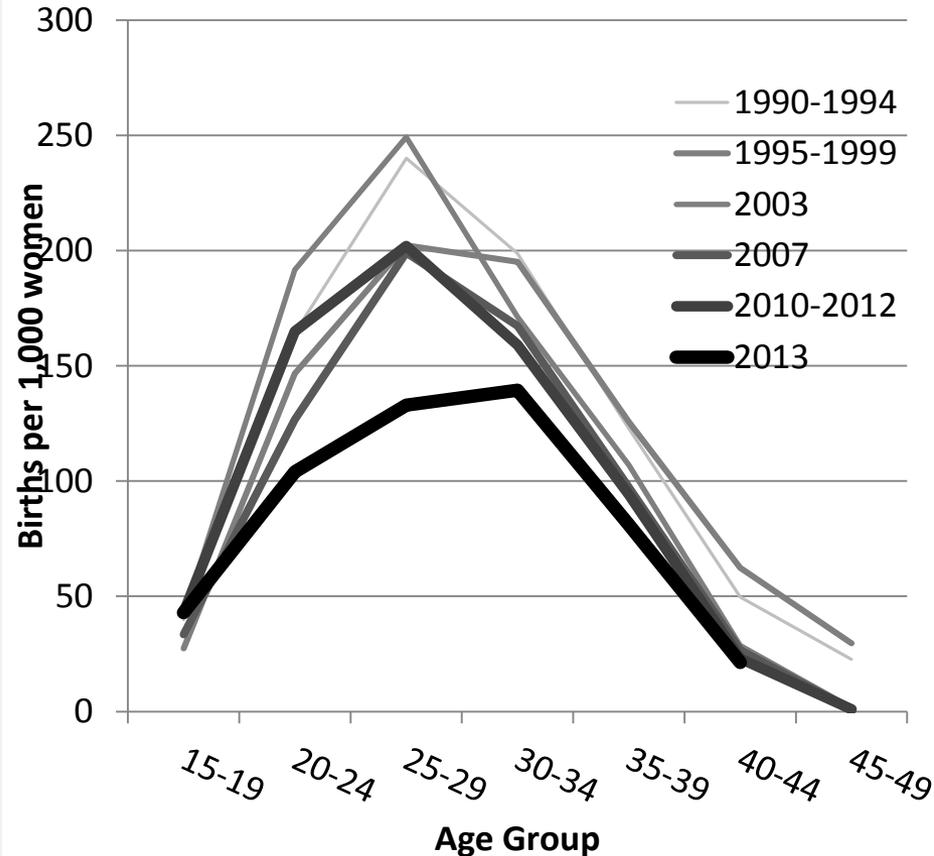
- ◆ ASFRs have a distinct shape. Does your data generally mirror this shape?
- ◆ Have the levels at fertility at each age drastically changed in a short amount of time?
- ◆ Graph your data against estimates from the census, DHS, MICS, and any other available sources. How does your data compare?

Age Specific Fertility Rates (ASFRs)



This graph has a plausible pattern with decreasing fertility over time, with the peak remaining in the 20-24 year age group

Age Specific Fertility Rates (ASFRs)

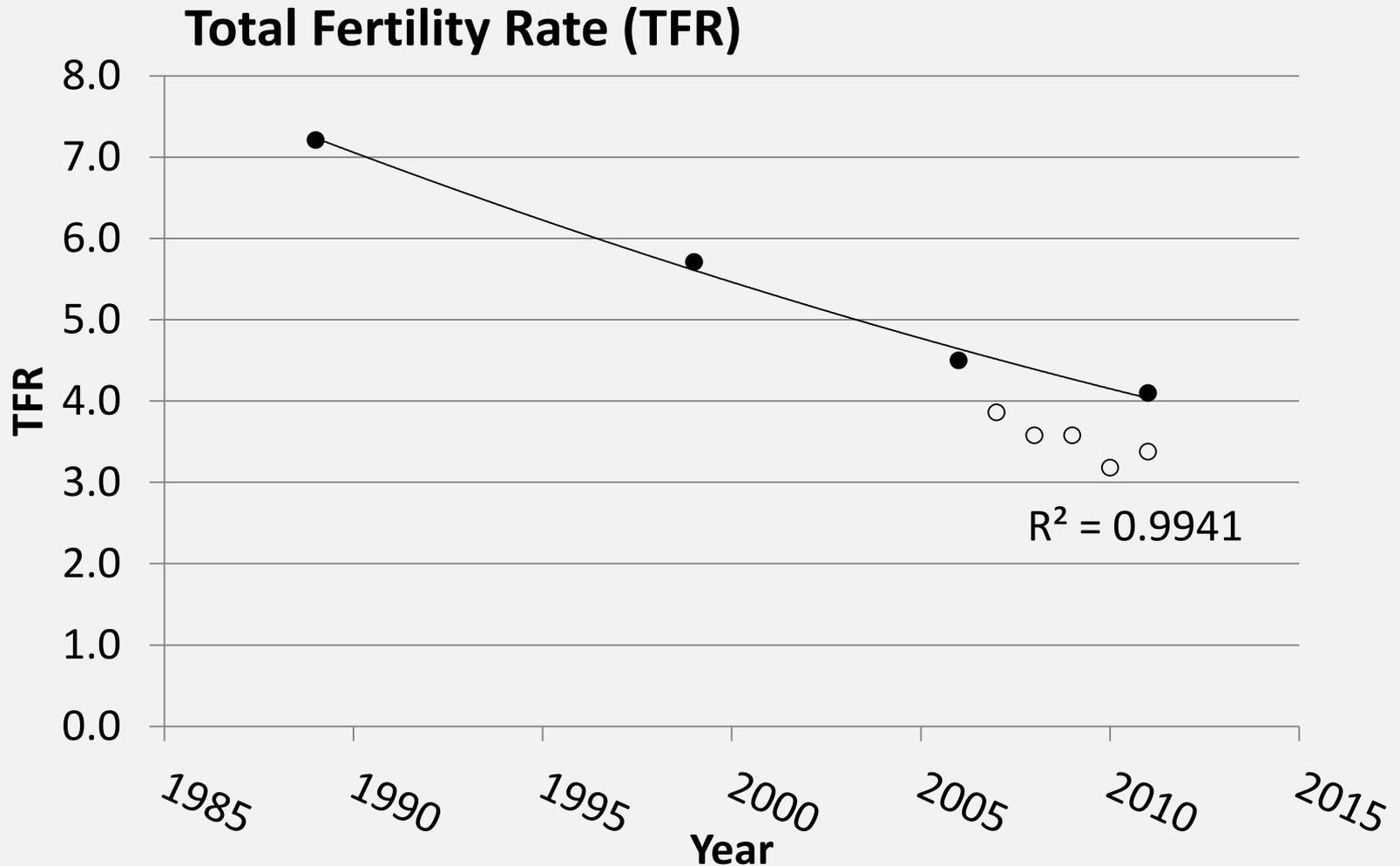


The bottom graph has a suspect curve for 2013, fertility decreased at most ages and the peak shifted from 25-29 to 30-34 years old, which is unusual.

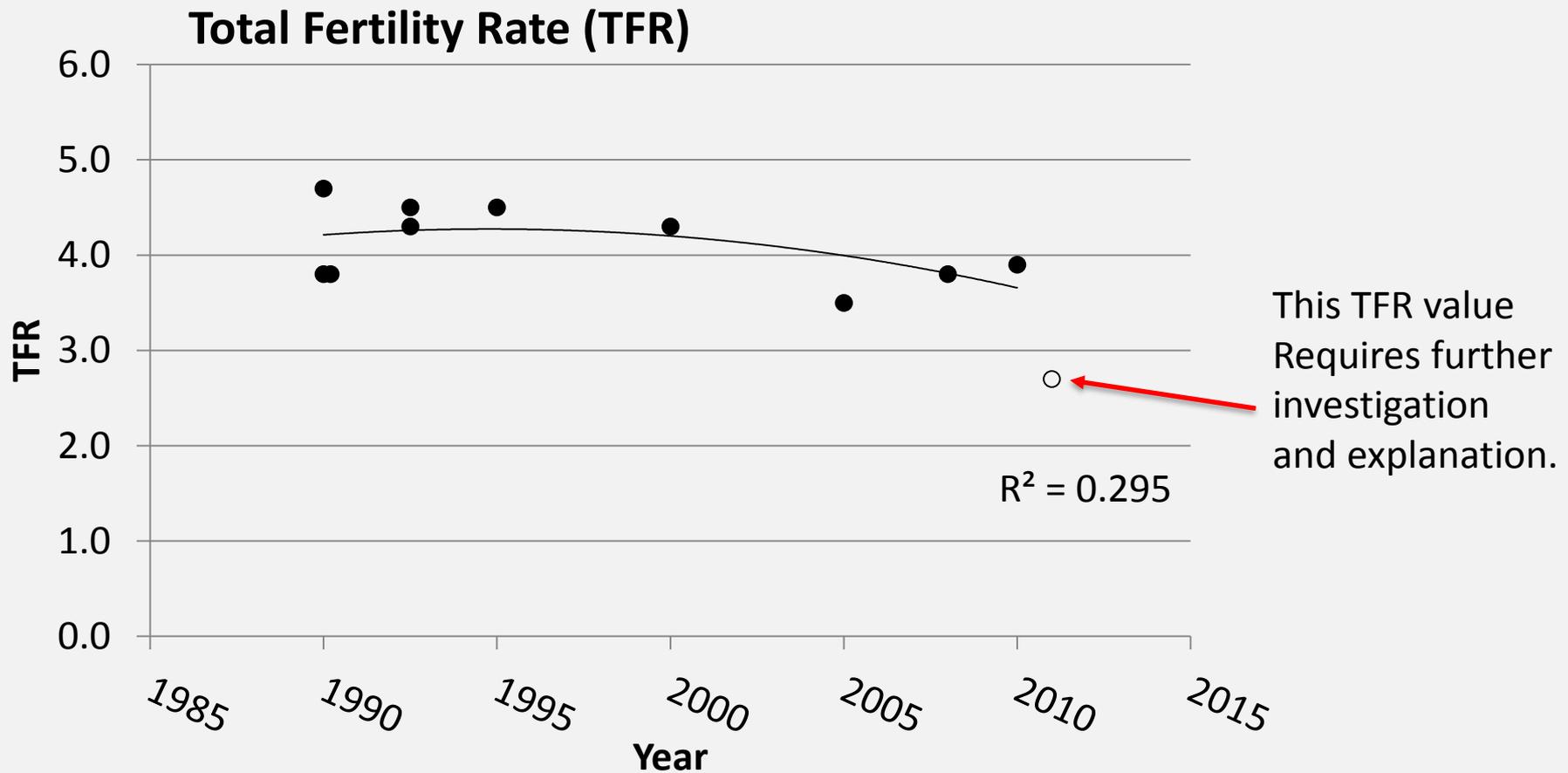
Plausibility of TFR values

- ◆ TFR does not usually change drastically in a short period of time.
- ◆ Compare to previous years' VS data, data from the census, DHS, MICS, and other sources.
- ◆ It's helpful to graph values over time

Are your TFR values plausible?



Are your TFR values plausible?



What to do with implausible values?

- ◆ Do your research - was there a policy, event, or other cause that could have affected your data?
- ◆ If so, highlight in the text the reason why your number may be different from other estimates, but why you believe it's still accurate.
- ◆ If the explanation is more likely due to incomplete data or poorer data quality, you can still publish your results, but be sure to caution the reader in the text about their low value and mention possible causes for this.
 - ◆ It's helpful if you publish aberrant results alongside other estimates for comparison so the reader can interpret them correctly.

Exercises

- ◆ Graph the test data ASFRs against other ASFR sources and years. What do you see? Does your data look plausible?
- ◆ Repeat this exercise for TFR values
- ◆ Find other sources of ASFR and TFR data in your country for the last 20 years. Plot these values with your vital statistics data. What kind of trends do you see?