

# Capture - Recapture

## Models, Methods, and the Reality

Jens-Peder Ekros  
Anders Subotic

Bo Bergman

Linköping University, Sweden

# Outline

- Uses of inspections
- Capture-recapture methods
- Test of models
- Results
- Conclusion & Future Work

# Uses of inspections

- Primarily: *detection* of defects
- Additionally, inspection data can be used for:
  - Quality assurance, e.g. using *capture-recapture* methods
  - Quality control
  - Process improvement and organisational learning
- Additional uses of inspection data involves models of product, process and resource

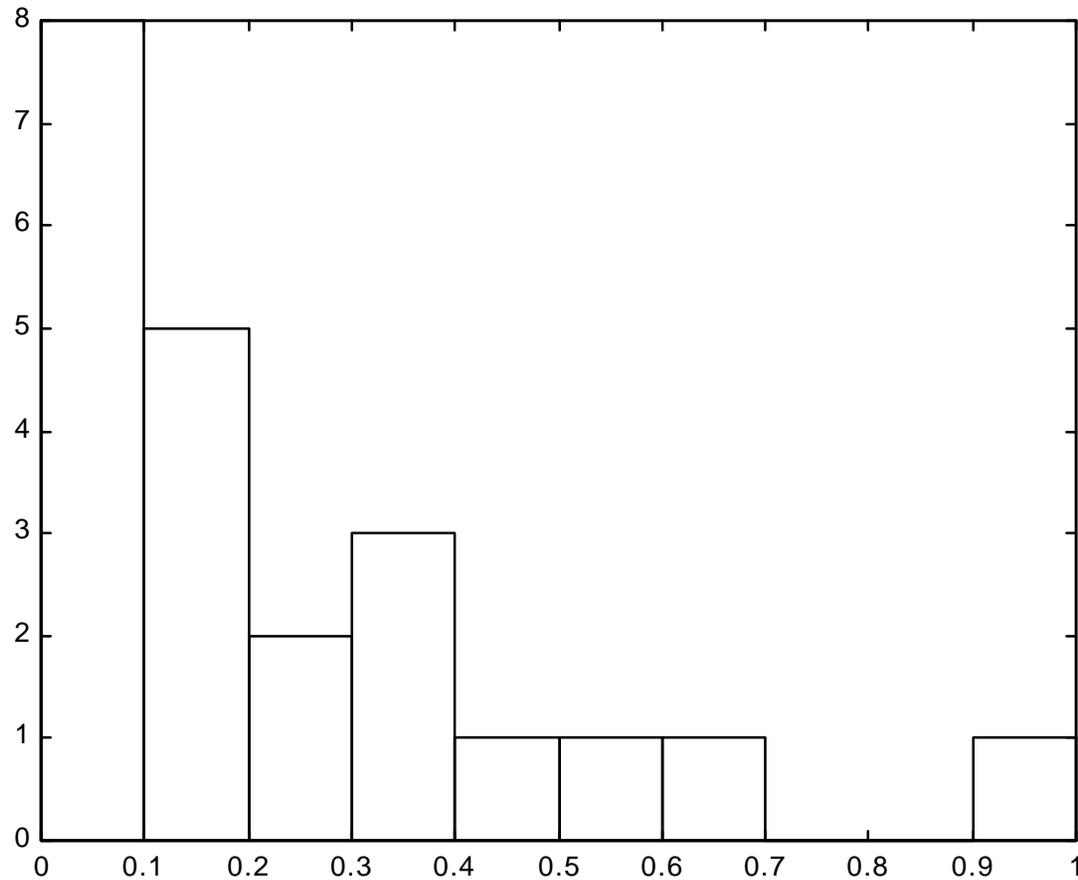
# Capture-recapture methods

- Methods for estimating size of population
- Transferred from biology to software engineering
- Estimate defect content of software artefacts
- Common methods (estimators):
  - Jack-knife
  - Maximum likelihood
  - Chao
- Different *methods assume* different *models*, i.e. characteristics of input data

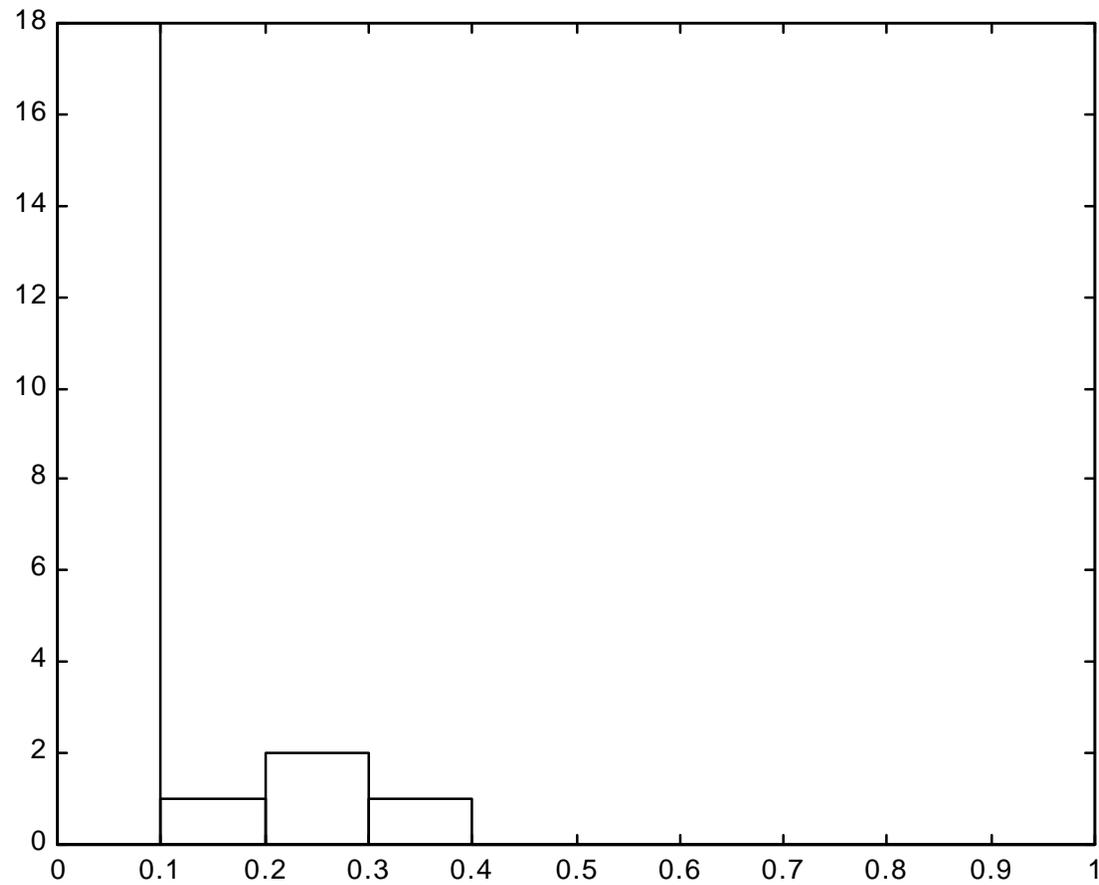
# Test of models

- Three assumptions were tested:
  - Inspectors find the same number of defects (capability)
  - Defects are equally easy to detect (detectability)
  - Inspectors find the same defects
- Assumptions tested using old and new test statistics
- Distributions of test statistics created using Monte Carlo simulation and enumeration
- 24 published data sets were used (16 from NASA)

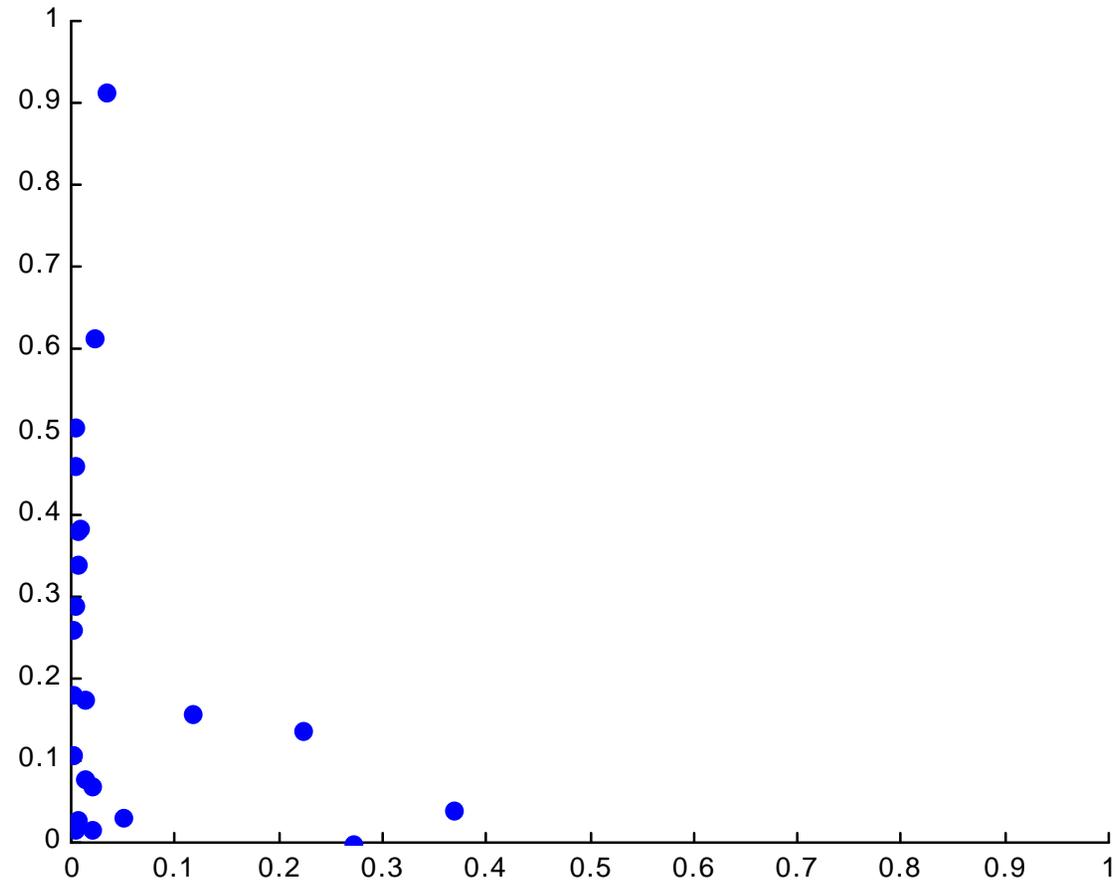
# Inspector capability varies



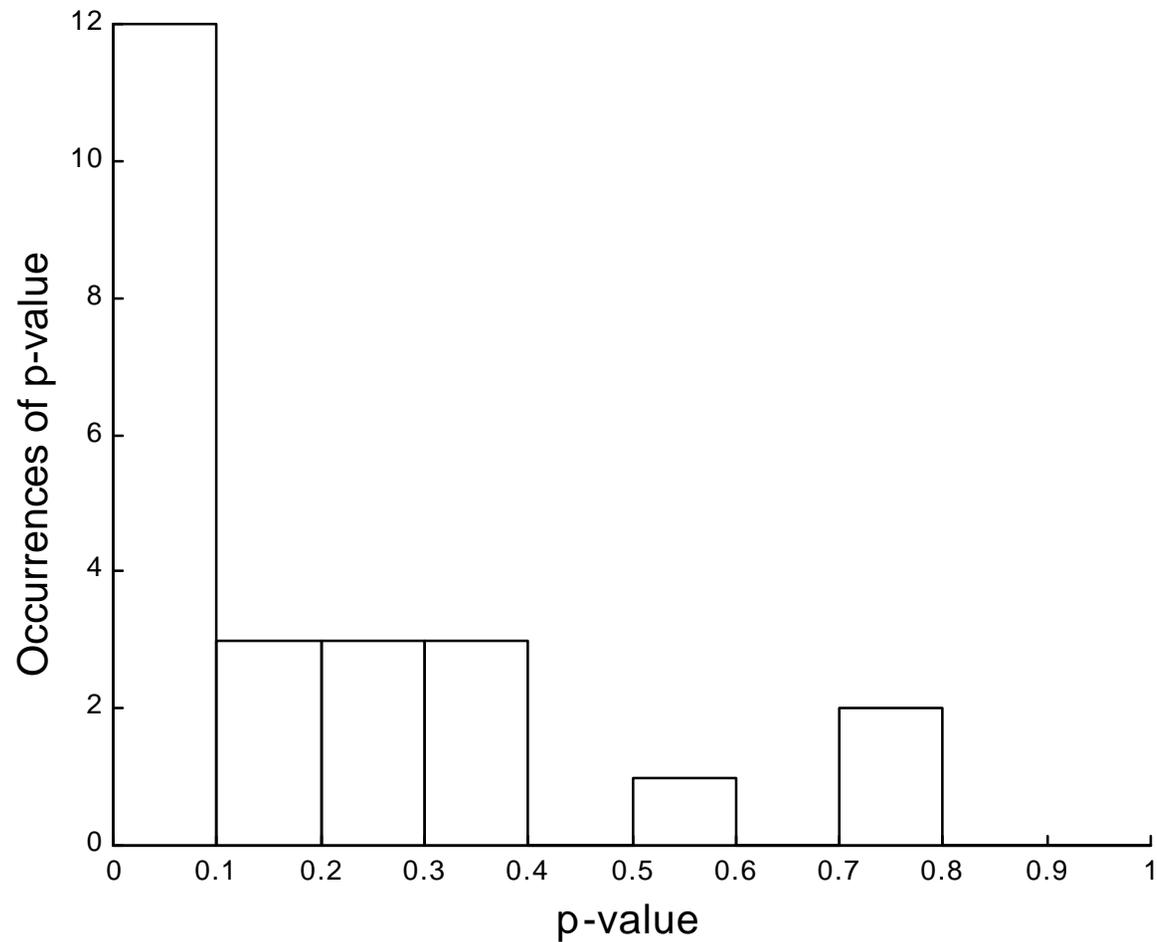
# Defect detectability varies



# Detectability varies more than capability



# Inspectors do not find the same defects



# Conclusion & Future work

- “Faults ain’t fish” - Models have to be validated
- Tests using empirical data suggest:
  - Inspectors do not find the same number of defects
  - Defects are not equally easy to detect
  - Inspectors do not find the same defects
- How can we use these results for better estimates?  
Develop new or adjust existing estimation methods
- More industrial data involved in the analysis
- Investigate the effects of broken assumptions