

Is It You or Your Model
Talking ?

A Frame Work for Model
Validation

- ❑ INTRODUCTION & PREDICTIVE USES OF MODELS
- ❑ VALIDABILITY & VALIDATION
- ❑ USES OF UNVALIDATED MODELS
- ❑ CONCLUSION & EXAMPLES

LCDR Carolyn Fricke, USN

1Lt. Muzaffer Gezer, TUA

1Lt. Eray Unguder, TUA

LTJG Levent Bingol, TUN

Defense modelers,

at least generally agree,

“an investigation of the agreement of the
model with reality,”

but disagree about

“how to do it.”

C.J. Thomas

“... an investigation of agreement of the model with reality.”

--- we may not be able to define it, but we must strive to do it anyway...

Some say that...

“... all models can be validated and that one procedure and standard of validation should be devised and used across the board.”

We believe that...

“**Some** models can be validated and used to predict, while **others** can not be validated and may only be put to nonpredictive uses.”

The idea of model validation...

Model quality

- contribution to the analysis

Model's intended use

Next...

- “prediction” & “the model says X”
- Four prerequisites of validation
- What is an unvalidated model good for?

What is “prediction” ?

“I predict that Navy will beat Army this year.”

- A statement about an observable or potentially observable quantity or event is produced.
(but it carries no warranty as to its accuracy!)

What is “prediction” ?

“the model says an eclipse of the moon will occur on Thursday at 4:23 PM”

- A statement about an observable or potentially observable quantity or event is produced.
- The modeled situation is such that predictive accuracy can be measured.
- The predictive accuracy of the model in the situation has been measured.

I'll say it again...

“It requires that a prediction be a statement about what will happen with a warranted measure of its error.”

Why should our more restrictive definition
be preferable...

“Because we say so”

“Definitions make distinctions; it is
important to make useful **distinctions.**”

Examples for Predictive Uses of Models...

“weapon A will survive more often than weapon B”
BAD...

“an eclipse of the moon will occur on Thursday at 4:23
PM” **GOOD...**

“Navy is better than Army” **FUNNY...**

Important point...

- **Temporal Prediction:** an inference from the present to the future.
- **Logical Prediction:** forecast of the conformance of individual observations with theory or scientific law.

“Some temporal predictions are logical predictions, but many logical predictions are not...”

Our notion of predictive uses of models...

Temporal Predictions

Because: “Logical predictions may be part of building the science that is used, but the models themselves make temporal predictions.”

Terminology...

“Validation”

for models in predictive uses

“Evaluation”

for models in non-predictive uses

Reiteration...

“A model that cannot be validated in this sense is not always useless; it simply may not be used to make sentences like “the model says X.”

Validatability

- to take the form “the model says X ”, the model must be validated, but before that can be done, it must be possible to do so.
- If we cannot validate the model, it is not necessarily useless, but we cannot say “the model says X ”

Prerequisites

- Prerequisites apply not to models, but to the situation being modeled.
- A situation may not satisfy the prerequisites at a given time, but any progress can change that.

Prerequisites

1. (P1): It must be possible to *observe* and *measure* the situation
2. (P2): The situation being modeled must exhibit a *constancy of structure in time* (deterministic?)
3. (P3): The situation being modeled must exhibit a *constancy across variations in conditions not specified in the model*
4. (P4): It must be possible to collect ample data with which to make predictive tests of the model

The Constancy Prerequisites

- P2 is necessary if you want to validate a model for the same conditions
- P3 is necessary if you want to validate a model for a wider range of conditions
- Example: model of a Vietnam-era air-to-air missile

Accruing Validity

- Validatable models can accrue validity.
- Non validatable models cannot.
- For models that can be validated, validity is not binary; accrues along a continuum between “not valid” and “valid”

Variants and Substitutes For Predictive Tests

- Transitivity: Use the outputs from validated model B to validate a simpler model A.
- B must be valid – there is no free lunch!
- A can only be as accurate as B

Variants and Substitutes For Predictive Tests

- Other substitutes: (do not make grade)
 1. Validating a model by validating its submodels
 2. Validating a model against data used to construct it
 3. Checking the model's math and logic

Seven Uses of Unvalidated Models

If an analysis is to have the form the model says X , the notions of quality given in section 2 or 3 must apply. The model must be validated.

Although the model is not validated, it can be used for non predictive uses.

Seven Uses of Unvalidated Models

- Bookkeeping
- Aids to sell an idea of which the model is but an illustration
- Training Aids to induce a particular behavior
- Part of an Automatic Management System
- Aid to Communication
- Vehicle for *A Fortiori* Arguments
- Aid to Thinking and Hypothesizing

Use 1: As a Bookkeeping Device

- **Condensing Masses of Data**
 - Some models digest great volumes of inputs and produce handy numbers or pictures
 - EADSIM (Operation Desert Storm)
- **Providing a Means or Incentive to Improve Data Quality**

Use 2: As an Aid in Selling an Idea of Which the Model Is But an Illustration

- **An unvalidated model can be used to sell an idea by representing the idea and displaying the benefits.**
- **Navy Aviation System Model**
 - fine as a descriptive tool:” *Here are some things in your logistics system that you are trading off whether you know it or not.*”
 - poor as a predictive tool:” *You will save this much if you make the trade-off this other way.*”

Use 5: As an Aid to Communication

- **A model can be a systematic description of belief and knowledge about a situation. This can help:**
 - intellectual explorations
 - operating organizations
- **The Combat Analysis Group Example**

Use 3: As a Training Aid, to Induce a Particular Behavior

- **A deliberately unrealistic model can be used to induce particular behavior or to improve trainees levels**
- **National Training Center Example**
 - can not draw analytic conclusions
 - deliberately unrealistic aspect of NTC maintained because the trainers don't want Blue units to make mistakes without paying for them

Use 6: As a Vehicle for A *Fortiori* Arguments

- **Fortiori Argument**

- If condition Z were True, then policy A would be preferable to the other candidates. But the actual situation deviates from Z in ways that favor A even more. Thus, a *fortiori*, A is preferable.

- **EADSIM (Operation Desert Storm)**

- (1) Iraqi Forces assumed at their best level (Z), the model gives the attrition level as 4 percent for Blue forces.

- (2) Iraqi forces can not be better (upper limit).

- (3) In reality Iraqi forces are not in such a good condition. Our attrition level would be less than 4 percent (common sense).

Use 7: As an Aid to Thinking and Hypothesizing

- **Insight**

- The power or act of seeing into a situation: Penetration
- The unvalidated model does not give insight. It helps its user formulate questions that might be insightful or that maybe utterly ridiculous.

- **An unvalidated model can suggest but cannot reveal the truth. Can help us in two ways.**

- As a stimulus to intuition in applied research or in training
(EADSIM, JANUS)
- As a decision aid in operating organizations
(RAND)

Validating a real model

Everyone's favorite

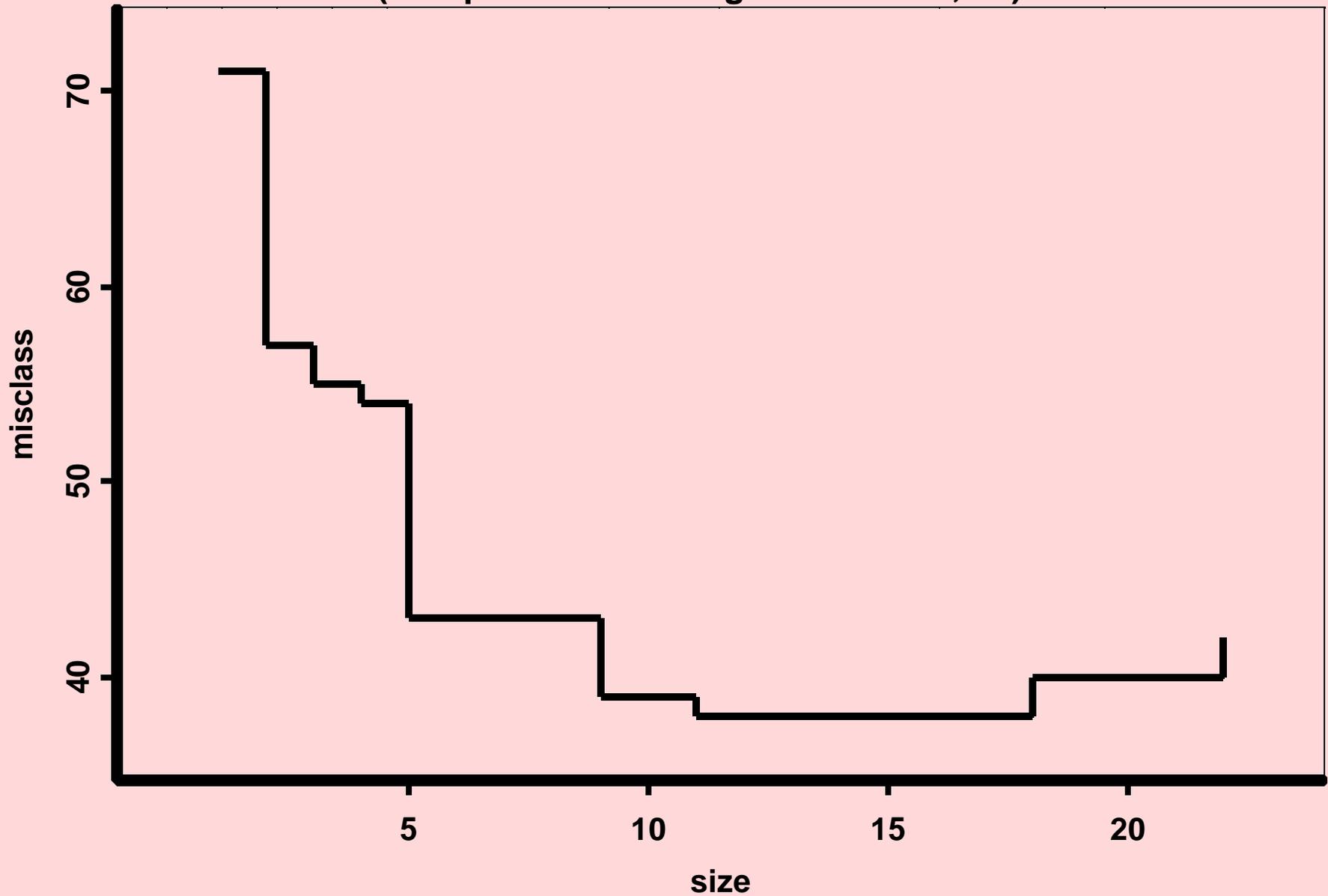
Pima tribe diabetes data

Observe and Measure

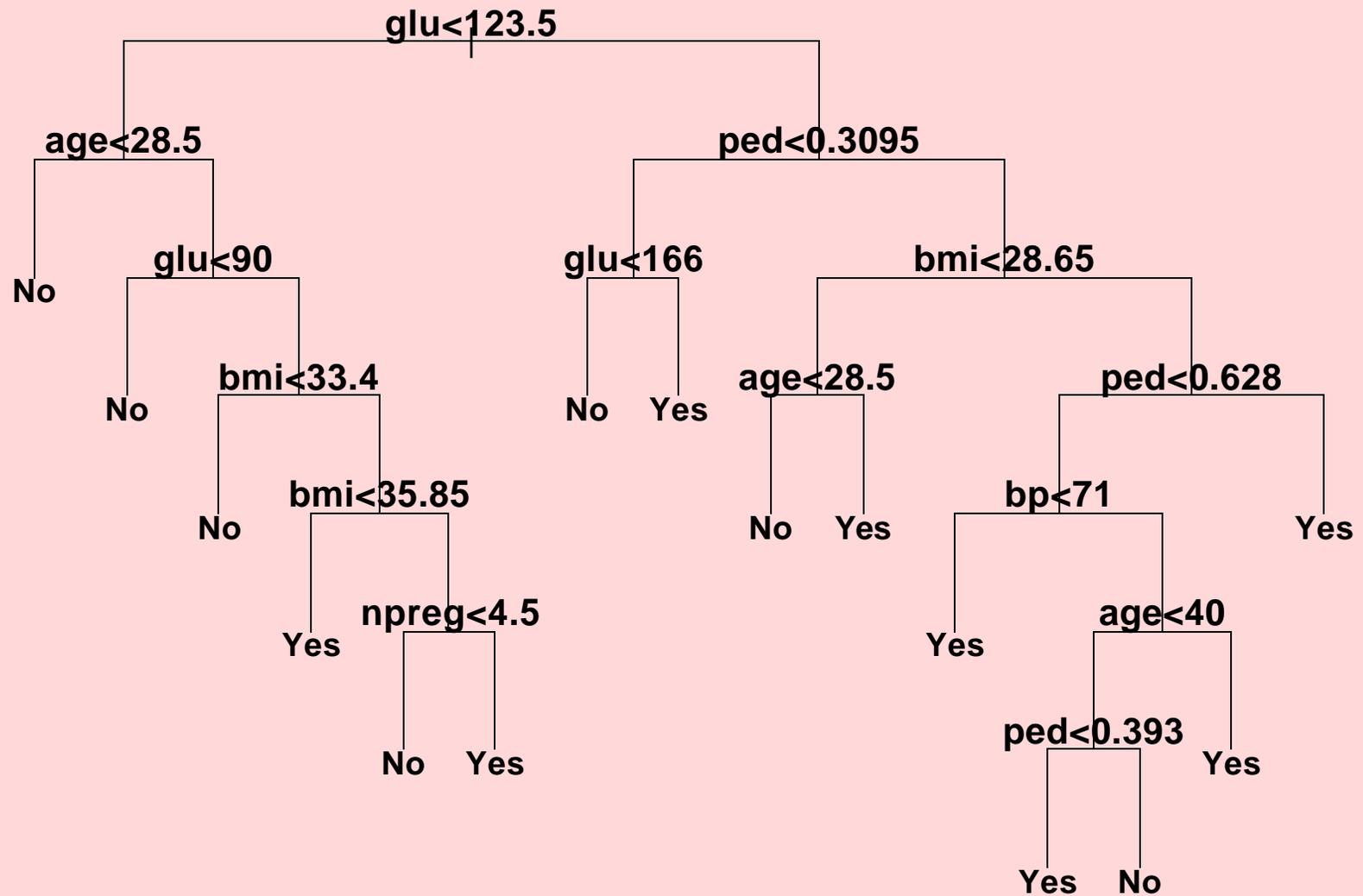
# Preg	glu	bp	skin	bmi	ped	age	type
5	86	68	28	30.2	0.364	24	No
7	195	70	33	25.1	0.163	55	Yes
5	77	82	41	35.8	0.156	35	No
0	165	76	43	47.9	0.259	26	No
0	107	60	25	26.4	0.133	23	No
5	97	76	27	35.6	0.378	52	Yes
3	83	58	31	34.3	0.336	25	No
1	193	50	16	25.9	0.655	24	No
3	142	80	15	32.4	0.2	63	No
2	128	78	37	43.3	1.224	31	Yes
0	137	40	35	43.1	2.288	33	Yes
9	154	78	30	30.9	0.164	45	No
1	189	60	23	30.1	0.398	59	Yes
12	92	62	7	27.6	0.926	44	Yes

Pima.tree Misclassification rates

(Sample run resulting in best = 11, 15)



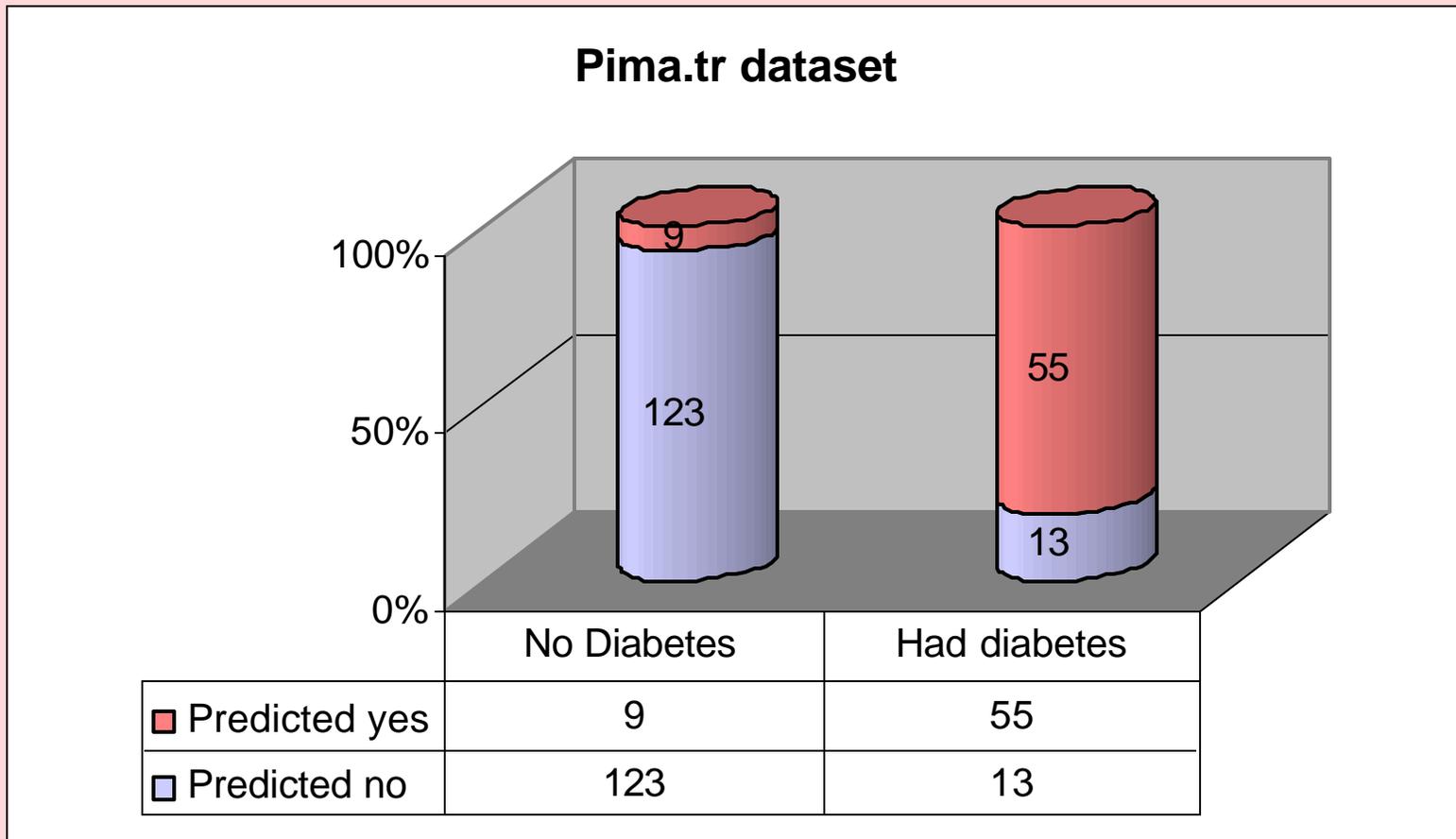
Pima.tree, size 15



Make specific predictions

**P1: We observe and
measure to be able to make
specific predictions**

Basic model

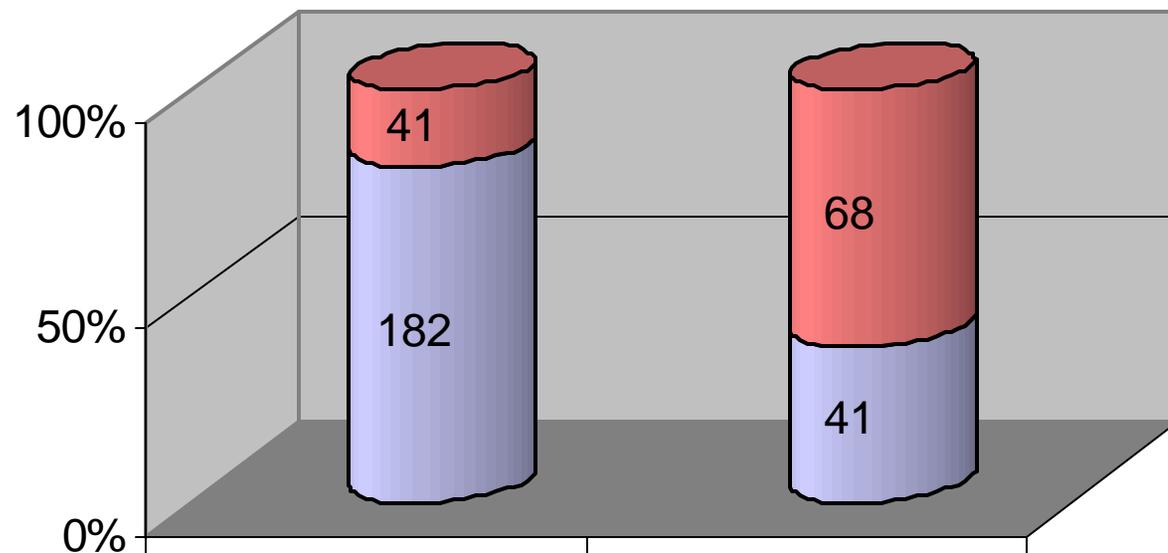


Consistency in Time

P2: We want to validate a model under the same conditions as the test

Test case

Pima.te dataset



■ Predicted yes	41	68
■ Predicted no	182	41

Consistency in variations

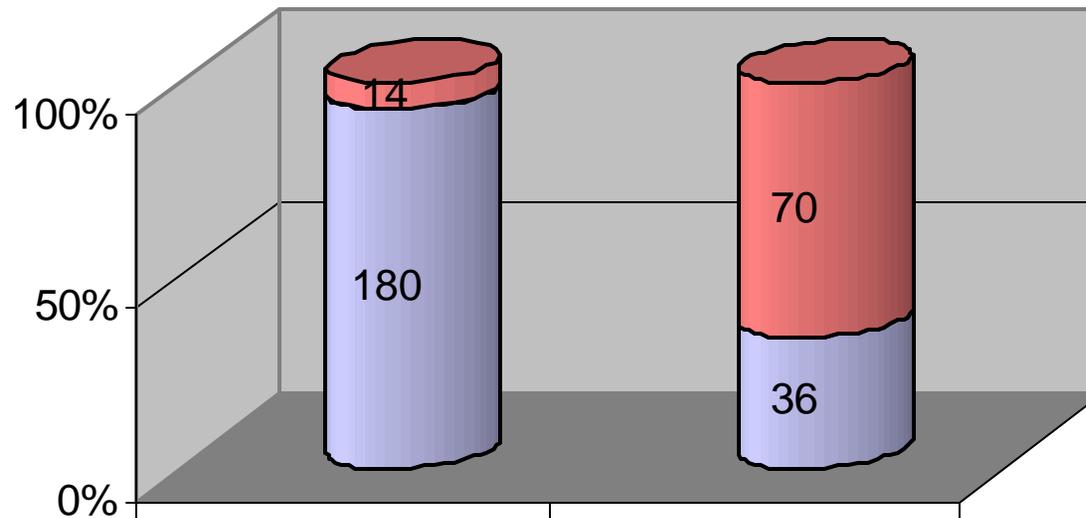
P3 We want to validate the model for a wider range of data.

What do we mean by wider range?

(a) What about varying known conditions?

Another test case

Pima.tr2 dataset



	No Diabetes	Had diabetes
■ Predicted yes	14	70
■ Predicted no	180	36

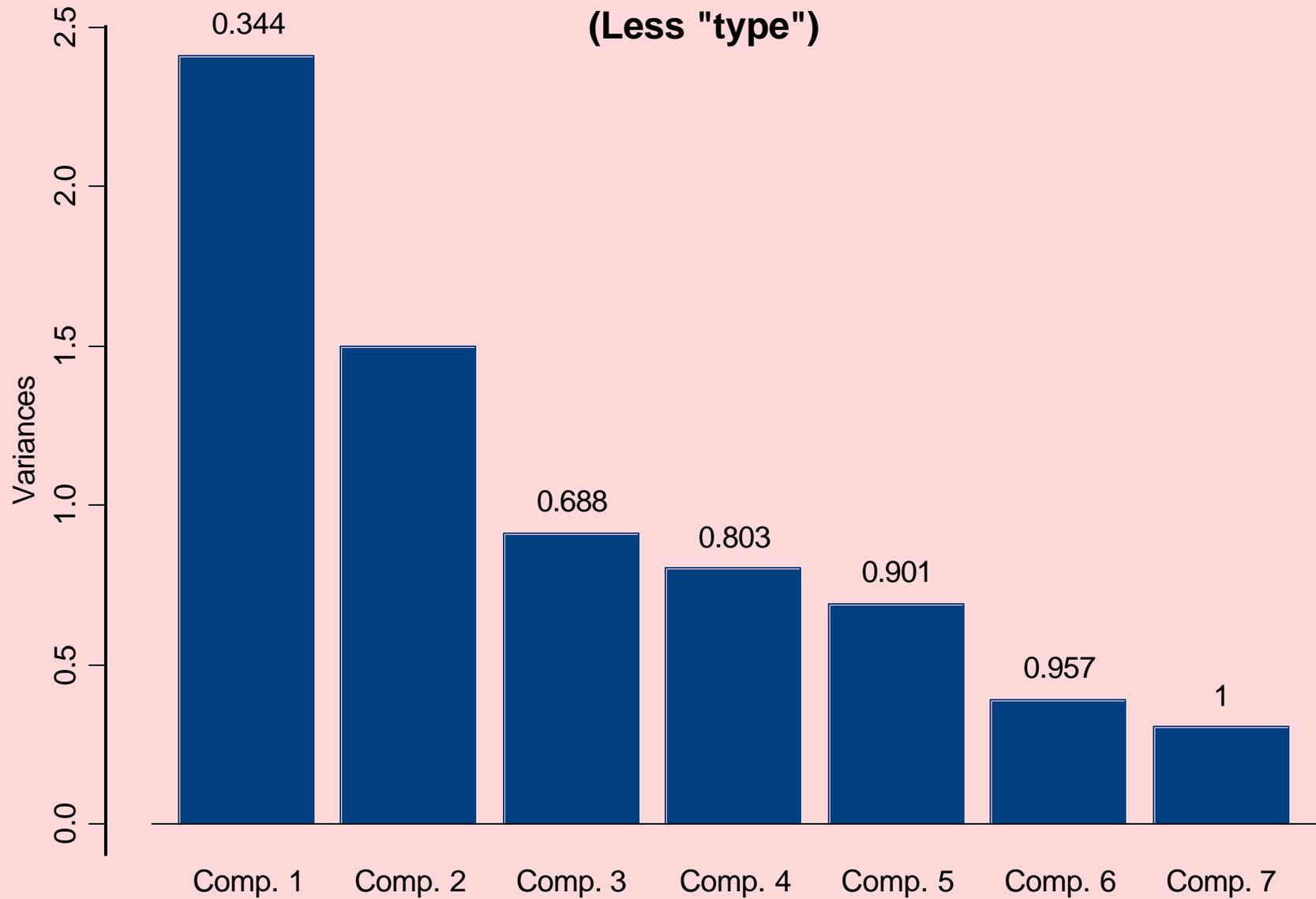
Consistency in Variations

P3 (b) How do we establish range?

What do we know?

Principal Components of Pima.tr

(Less "type")



Cor = T used: variables scaled to have unit variance

Principal Components

```
> pima.pr$loadings
```

	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6	Comp. 7
npreg	0.365	-0.463		0.501	-0.132	-0.532	-0.306
glu	0.366		0.418	-0.680	-0.445	-0.138	-0.107
bp	0.413	-0.108		-0.335	0.824	-0.108	
→ skin	0.432	0.415	-0.337	0.133	-0.154	0.381	-0.584
bmi	0.391	0.513	-0.241	0.114	-0.104	-0.414	0.575
ped		0.450	0.785	0.324	0.259		
→ age	0.471	-0.370	0.148	0.197		0.606	0.457

Consistency in Variations

What don't we know?

**Probably can't use for anybody but
Pima Women**

Accreditation is a political issue

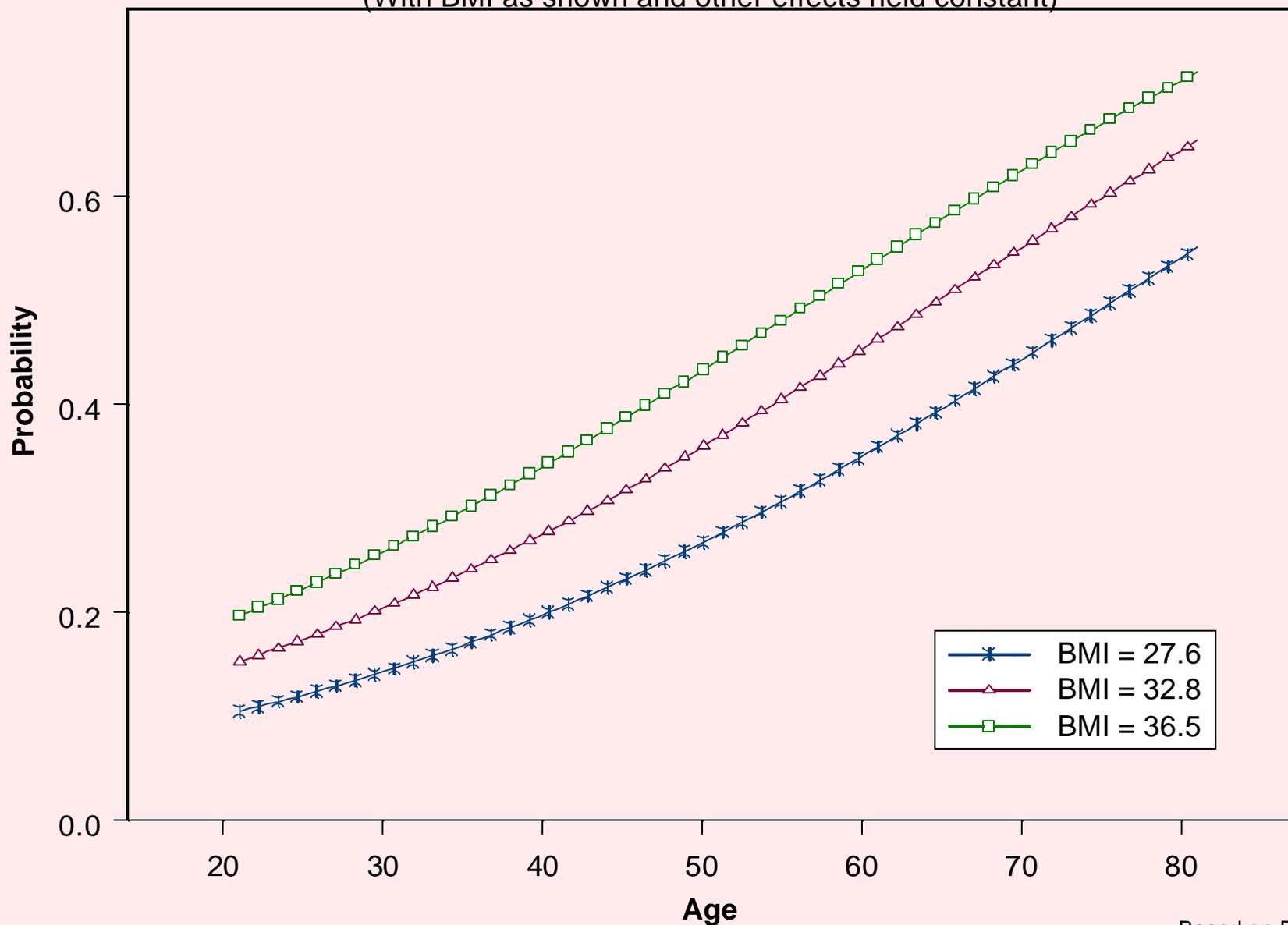
Collect Ample Data

**P4 Collect ample data to
be able to predict**

622 Subjects

Probability of Diabete vs Age

(With BMI as shown and other effects held constant)



Based on Pima.tr data

When you can't use to predict

Incentive to improve data quality

Sell an idea

Aid to communication

Stimulate thinking

The End

Questions??

3 Questions

1. Does checking the math and logic help validation?
 - A. No, that is verification not validation.
2. What are two of the 4 prerequisites of validation?
 - A. Observable & Measurable, Constancy in time (recreate under same conditions), Constancy in variations (wider range of conditions), Collect ample data.
3. What are 3 of the 7 uses of unvalidated models?
 - A. Bookkeeping, to sell an idea, training aid, automatic management system, communication, fortiori arguments, thinking aid.