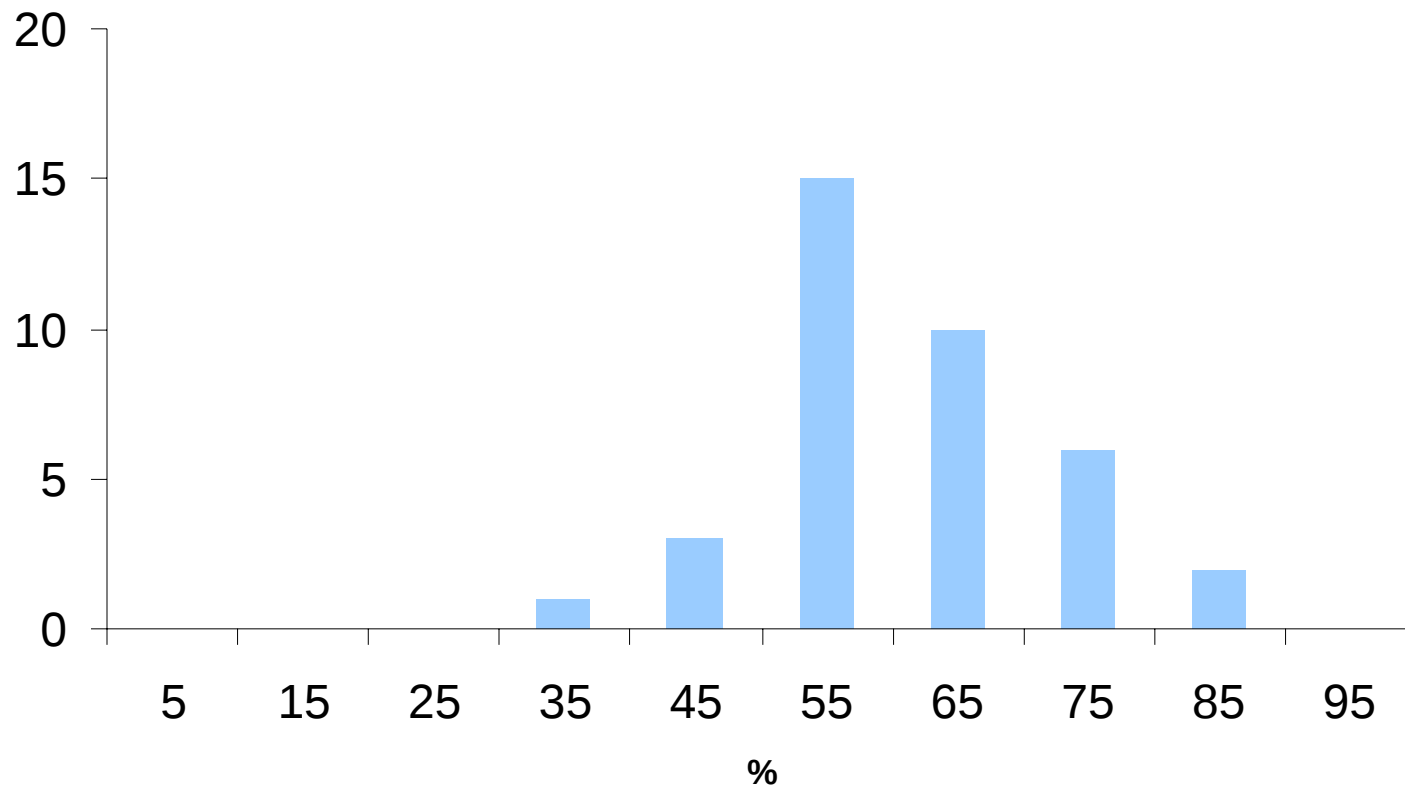


Questions?

- Project #4
 - Updated dataset
 - *If you use your own dataset please do the same*
- HW #7

Mid-term exam (mean 61%)

Mid-term grades distribution



Today

- Introduction to evaluation

Setting the stage for discovery

What does evaluation mean?

Predictions and measurements?

- Type of prediction
 - Point
 - Interval
 - Ordinal
 - Categorical
- Construct and measurement methods
 - Reliability
 - Convergent validity
 - Discriminant validity

Questioning measurements

- Are they reliable?
 - Does the experiment take into account variations between subjects?
 - *Need for testing a sample of subjects*
- Are they valid?
 - Does the experiment reflects target use?
 - *Were users typical?*
 - *Were tasks typical?*
 - *Was the setting realistic?*
 - *Was the experience biased?*
- Do they make sense?
 - Setting the stage for discovery!

Are results significant?

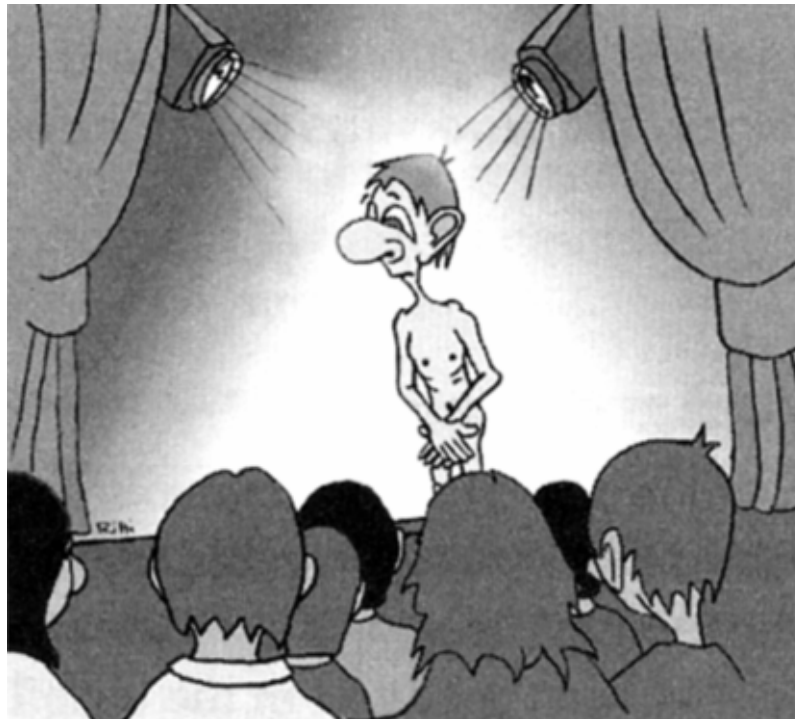
- Statistical significance
 - Comparing to the null hypothesis: “There is no effect”
 - Type I errors are the most disruptive

Researcher's Decision	Actual Situation: Null Hypothesis is	
	True	False
Accept the null hypothesis	Correct decision	Type II error
Reject the null hypothesis	Type I error	Correct decision

- Design significance?
 - 3.00s versus 3.05s?

The participant standpoint

- Testing is a distressing experience
 - Pressure to perform
 - Feeling of inadequacy
 - Looking like a fool in front of your peers, your boss,...



(from "Paper Prototyping" by Snyder)

Treating subjects with respect

- Follow human subject protocols
 - Individual test results will be kept confidential
 - Users can stop the test at any time
 - Users are aware (and understand) the monitoring technique
 - Their performance will have not implication on their life
 - Records will be made anonymous
 - *Videos*
- Use standard informed consent form
 - Especially for quantitative tests
 - Be aware of legal requirements

Conducting the experiment

- Before the experiment
 - Have them read and sign the consent form
 - Explain the goal of the experiment
 - *In a way accessible to users*
 - *Be careful about the demand characteristic*
 - *Answer questions*
- During the experiment
 - Stay neutral
 - *Never indicate displeasure with users performance*
- After the experiment
 - Debrief users
 - *Inform users about the goal of the experiment*
 - Answer any questions they have

Managing subjects

- Don't waste users time
 - Use pilot tests to debug experiments, questionnaires, etc...
 - Have everything ready before users show up
- Make users comfortable
 - Keep a relaxed atmosphere
 - Allow for breaks
 - Pace tasks correctly
 - Stop the test if it becomes too unpleasant