

## 1 ☐ Prediction

- ✓ Predicting the future allows us to prepare for it.
- ✓ Detecting causal relations and contingencies are one form of prediction.
- ✓ Another form is continuous prediction
  - Related to correlation/covariation - continuous variables.

## 2 ☐ The problem of prediction

- ✓ In making a prediction, one considers the evidence, weighs each piece according to its predictive validity, and then predicts
  - e.g., test scores, interview performance, grades used to predict job performance
  - e.g., HS GPA & ACT used to predict college grades
  - e.g., behavior in jail and personality test profile used to predict recidivism after release

## 3 ☐ Continuous prediction

- ✓ Predicting one continuous variable from another.
- ✓ Examples
  - Predicting weight from height
  - Predicting college GPA from high school GPA and ACT scores

## 4 ☐ Regression to the mean

- ✓ Extreme values tend to be followed by less extreme values
  - Very tall parents *tend* to have tall children but not children as tall as they are.
  - Very high grade on a test (by a good student) *tends* to be followed by a lower grade on the next test.
  - Years of a terrible economy are *more likely* to be followed by improvements in the economy.

## 5 ☐ People do not understand regression to the mean

- ✓ Parents expect kids to be as smart as they are.
- ✓ A business overreacts to a poorer than average sales year and turns the company upside down.
- ✓ Universities fire a coach after an especially bad season - next coach does better....

## 6 ☐ How does their behavior differ from the gambler's fallacy?

- ✓ Gambler's fallacy (improper reasoning)
  - People expect that the next string of events will balance out the previous string.
  - Extreme values predicted to be followed by opposite extreme.
- ✓ Non-regression to the mean (improper reasoning)
  - People expect extreme values to be followed by extreme values.

## 7 ☐ Relationship to probability

- ✓ You get 56% tails in first 1000 flips.
- ✓ Normative:
  - More likely than not that the next 1000 will be 50% (regression to the mean) and thus the overall expected to be 53%.
- ✓ Non-normative

- Expect next 1000 to be 44% tails to balance things out (gambler's fallacy).
- Expect next 1000 to be 56% tails (extreme values followed by more extreme values).

## 8 ☐ So how are these resolved?

- ✓ Random events
  - Gambler's fallacy
- ✓ Non-random events (or at least perceived as being non-random)
  - Non-regression to the mean
    - Analogous to an expectation of streakiness

## 9 ☐ Examples

### 10 ☐ Sports Illustrated jinx

- ✓ After appearing on Sports Illustrated cover, athletes tend to do worse.
- ✓ Also related to the "sophomore slump"
  - The tendency for the rookie-of-the-year (or other top performers) to do worse their second year.

### 11 ☐ The relative effectiveness of reward and punishment

- ✓ Psychologists know that reward is more effective than punishment
- ✓ Laypeople find this surprising - Why?
  - Rewards follow good behavior
    - Good behavior is more likely to be followed by bad behavior (regression to the mean)
  - Punishment follows bad behavior
    - Bad behavior is more likely to be followed by good behavior (regression to the mean)

### 12 ☐ Reward/punishment for arriving early/late (Schaffner, 1985)

- ✓ Computer displayed a series of student arrivals and participants chose to reward, punish, or make no comment
- ✓ Actual performance was programmed to be independent of the rewards and punishments
  - 70% of participants judged reprimand more effective than praise

### 13 ☐ The problem of hidden or missing data

#### 14 ☐ Example

- ✓ You don't know whether the people who weren't accepted could have done as well as those who were accepted
  - Here you are missing the *c* and *d* cells so you can't compute  $\Delta P$ , thus you CAN'T JUDGE the true contingency.

#### 15 ☐ This leads to errors in judgment

- ✓ You can't rely on the success rate of your accepted applicants to judge the efficacy of your selection criteria.
- ✓ Bad judgments most likely when success is increased by selection
  - Receiving the grant increases likelihood of success.
  - Getting into Yale increases the likelihood of getting a better job (institution reputation).

## 16 ☐ How to properly evaluate your selection criteria

- ✓ If other people accept your rejected individuals and those people have the same opportunities, you WILL get the  $c$  and  $d$  information you need
  - e.g., your team doesn't draft a player using your selection criteria - another team does and this player goes on to succeed.
  - e.g, Yale doesn't accept the student, but Stanford does and the student goes on to get a Nobel prize.

## 17 ☐ Expert vs. Statistical prediction

- ✓ People fail to recognize certain statistical issues
  - They fail to regress to the mean
  - They fail to recognize the impact of hidden or missing data
- ✓ Options
  - Use “experts” to predict
  - Use statistical techniques

## 18 ☐ Definition:

### Statistical model for prediction

- ✓ Use historical evidence and actual outcomes to determine the best weighting of the evidence in the prediction of future outcomes.
  - Needs an adequate body of historical data
- ✓ Statistical technique: Regression analysis

## 19 ☐ Meehl (1954) meta-analysis

- ✓ Examined 20 studies comparing expert predictions to predictions from a statistical model.
- ✓ Both experts and the models had access to the same historical evidence in all 20 studies.
- ✓ In EVERY study, the statistical model was better than or equal to experts

## 20 ☐ Sawyer (1966) meta-analysis

- ✓ Reviewed 45 studies comparing clinical (expert) and statistical prediction
- ✓ In all 45, experts were never better than and usually worse than statistical models

## 21 ☐ Sawyer (1966), cont.

- ✓ Unlike the Meehl study, the experts were NOT limited to having the same evidence
  - in 2 studies, the expert had access to MORE information but did WORSE than the model
  - In one of these, the performance of 37,500 sailors in WWII naval schools was predicted better by test scores alone (model) than from the ratings of judges who had access to the scores and conducted interviews

## 22 ☐ Grove et al. (2000) meta-analysis

- ✓ Included 136 studies comparing expert and statistical prediction.
- ✓ Study domains included: college academic performance, success in military

training, diagnosis of gastrointestinal disorders, criminal recidivism, prediction of IQ, surgical outcomes, marital satisfaction, juvenile delinquency, ....

### 23 ☐ Grove et al. cont.

- ✓ After proper analysis, they found:
  - 8 studies with an advantage for “experts”
  - 65 studies where they tied
  - 63 studies where the statistical (aka “mechanical”) model did better.
- ✓ In 7/8 studies where experts did better, they received more data than the stat models.
- ✓ However, when that extra data included interviews, the experts did much worse!

### 24 ☐ Libby (1976) study

- ✓ Had 43 bank loan officers predict which 30 of 60 firms would go belly up within the next 3 years
  - Officers requested and were given a range of data on financial ratios
- ✓ Results
  - Bank officer judgments were 75% correct
  - Model was 82% correct!
    - A model using only the ONE best predictor was 80% correct

### 25 ☐ Stock prediction

- ✓ “October is one of the singularly most dangerous months to speculate in stocks.
- ✓ Others are November, December, January, February, March, April, May, June, July, August and September.”
  - Mark Twain

### 26 ☐ Stock prediction

- ✓ Experts vs. Averages
  - A mere 17% of managed funds bested the S&P 500 over the past 10 years (Fortune magazine).
- ✓ Experts vs. Averages
  - Over the past 10 years, the average active manager in the US and the UK has underperformed the index by 1.7% per year.
  - Underperformance increases to 3.2% per year when fees are taken into account.

### 27 ☐ Stupid statistical models

- ✓ Dawes (1988): you don’t even have to have the optimal regression weights for a statistical model to make better predictions.
- ✓ Examined three domains:
  - MMPI predicting emergence of psychosis.
  - Academic test scores and personality profiles predicting 1<sup>st</sup> year grad school GPA in psych program at U of Illinois.
  - Undergrad GPA and GRE scores predicting faculty ratings of grad school performance.

### 28 ☐ Results of Dawes study

- ✓ The improper linear models accounted for 150% more variance (i.e., about 3 times as much) on the average as did the experts' predictions.

## 29 ☐ Why do we rely so heavily on expert prediction?

- ✓ Overconfidence abhors uncertainty
  - People don't want to accept the level of uncertainty that's present in a statistical prediction
  - Expert prediction tries to take into consideration ALL of the details in an attempt to avoid uncertainty
    - This can create problems
      - Probability matching is suboptimal.
      - Rats do better than people because they maximize.

## 30 ☐ Objections to their conclusions

- ✓ Used the wrong experts...
  - Every time this has been tested and skeptics got to choose the experts, they still didn't outperform statistical models
- ✓ Experts' judgments are known to be effective, otherwise people wouldn't pay them the big bucks
  - This is social justification – we don't want to believe the money was ill spent.

## 31 ☐ Objections, cont.

- ✓ People consider additional evidence that can't be quantified
  - Studies have shown almost universally that the predictions of experts given MORE info is no better than that due to stat models
  - In fact, giving experts the predictions of the models and allowing them to try and improve on it invariably results in poorer predictions
  - If some piece of evidence is important, it CAN be quantified.

## 32 ☐ Caveat: Using statistics badly....

- ✓ Must know what you're doing when you develop a statistical model.
  - Should know domain to narrow down candidate predictors.
  - People can develop poor models via data mining
    - e.g., look at lots of predictors and pick the best – may be spurious results ( $p < .05$  issue)
    - Use cross validation to test generality of model.
      - Test your model against data that has been held back.

## 33 ☐ Overconfidence abhors these results

- ✓ Meehl's and Dawes' work has been virtually ignored in the actual practice of expert judgment
  - Grad schools do use formulas now but try to go one better by examining portfolios individually
  - Med schools still interview applicants
    - Although no known predictive validity....
  - People still buy managed stock funds
  - People still rely on clinical experts in court

## 34 ☐ Resistance to changing the status quo

- ✓ Prediction is a big business

- “Experts” who make predictions are going to resist any attempts to discredit them.
- But, the best experts will adapt or become obsolete.
  - For example, an increasing proportion of investments are put in index funds (27% in 1988 to 37% in 1999).
- However, some people prefer an illusion of predictability.

### 35 ☐ The Big Business of Prediction

- ✓ Weather prediction
- ✓ Stock market prediction
- ✓ Business forecasting
- ✓ Sports handicapping
- ✓ Clinical/medical prediction
- ✓ Astrology
- ✓ Futurists
- ✓ Much more....

### 36 ☐ Why do we avoid statistical models?

- ✓ Overconfidence abhors uncertainty
- ✓ Statistical prediction seems so impersonal
  - Is it ethical to trust human lives, health, money, futures to a formula?
  - Is it ethical to use an inferior method when superior ones are available?
- ✓ It would require giving up cherished ideas about people’s abilities
  - We fail to recognize our limitations and strengths.

### 37 ☐ Solving the Prediction Problem

- ✓ Identify candidate predictors
- ✓ Measure the predictors
- ✓ Determine whether the predictors have linear or nonlinear relationships to the outcome (watch out for interactions!)
- ✓ Evaluate the predictors to determine how much weight each should get
- ✓ Consistently apply the resulting formulation

### 38 ☐ 1. Identify candidate predictors

- ✓ Driven by theoretical considerations or experience
- ✓ People drive this process - difficult to automate
  - Open to human biases
    - Limited to those predictors that readily come to mind (Availability heuristic, role of distinctiveness...)
    - Only consider predictors that confirm your expectations (Confirmation bias)

### 39 ☐ 2. Measure the predictor

- ✓ Easy for some predictors
  - Height, systolic blood pressure, winning percentage, barometric pressure
- ✓ Other predictors require quantification
  - By human judges
    - Quality of letter of rec., level of depression, probability of tumor (from ultrasound), wine quality

- By analysis
  - IQ, personality profile, HDL/LDL, RPI, expense ratio

#### 40 ☐ Measurement issues

- ✓ Lack of reliability in measures undermines the entire process
  - There's disagreement among judges regarding publication record, grant potential, patient aggressiveness, level of pain, health of an industry sector, etc.
  - Some measures may be biased (e.g., intelligence tests).

#### 41 ☐ 3. Identify the proper relationship to the outcome

- ✓ People are biased to assume simple additive models
- ✓ Simple additive models are actually good approximations for many real-world problems
  - Hence the popularity of linear regression as a statistical technique

#### 42 ☐ The Relationship between Predictors and Outcomes

- ✓ May be linear (simple additive)
  - Outcome =  $w_1 \cdot p_1 + w_2 \cdot p_2 + \text{constant}$
- ✓ May be non-linear
  - Outcome =  $w_1 \cdot p_1^2 + w_2 \cdot \log(p_2) + \text{constant}$
- ✓ May involve interactions among predictors
  - Outcome =  $w_1 \cdot p_1 + w_2 \cdot p_2 + w_3 \cdot p_1 \cdot p_2 + \text{constant}$

#### 43 ☐ People have difficulty learning some relationships (Klayman, 1988)

- ✓ Positive linear relationships learned easier than negative linear relationships.
  - Slovic (1976) gave subjects a single linear cue with +.8 or -.8 correlation with outcome.
    - Positive estimate: .76, Negative estimate: -.35.
- ✓ Nonlinear very difficult to learn
  - Deane et al. (1972) required participants to learn a three-cue relationship to outcome (max  $r = .94$ ).
    - When linear positive combo, answers correlated .72 with true outcomes
    - When a nonlinear positive combo, answers correlated .11 with true outcomes (they added a sinusoidal component).
- ✓ *Qualifier*: Many real-world nonlinear relationships are well approximated by a linear relationship (Camerer, 1981).

#### 44 ☐ Interacting predictors

- ✓ People are biased to view predictors as independent
  - The combined effect of multiple causes is the sum of the effects of each individual cause.
  - Simple effects and additive effects are relatively easy to learn (Ellefsen & Young, in revision).
- ✓ People have a harder time spotting interactions (Ellefsen & Young, in revision; Young, Wasserman, Johnson, & Jones, 2000).
- ✓ Participants can learn some interactions from outcome feedback, but not as well as they learn additive aspects (Brehmer, 1969; Mellers, 1980; Summers et al., 1969) and sometimes not at all (Camerer, 1981).

#### 45 ☐ 4. Identify optimal weights for the predictors

- ✓ So, now you have a collection of quantified candidate predictors and the framework for a prediction model (additive linear, nonlinear, factorial...)
- ✓ Next you must determine how much weight each of these predictors should get
  - Based on historical data/experience

- Unfortunately, people are poor at this process when they rely on their judgment (e.g., dilution effect)
- People's weighting grows worse as more noise (uncertainty) is added to the data

#### 46 ☐ Quantification of weights

- ✓ Correlation, covariation...
  - Human shortcomings - seeing relationships that aren't there (illusory correlation) and missing ones that are there (invisible correlation)
- ✓ Contingency
  - Human shortcomings - overemphasis of some information and underemphasis of other information creates biased contingency estimates

#### 47 ☐ 5. Consistently apply the resulting formulation

- ✓ A large part of the problem that human judges have is that they apply their own intuitive models inconsistently
  - We are too easily swayed by the idiosyncrasies of a case....
  - There's a tendency to identify too many 'exceptions to the rule.'
- ✓ In fact, models of a non-optimal expert outperform that expert! (Dawes & Corrigan, 1974)

#### 48 ☐ Mechanical models

- ✓ Statistical models
  - Regression analysis (e.g., multiple linear, nonlinear, logistic)
  - Categorical analysis (e.g.,  $\chi^2$ , loglinear)
- ✓ Neural network models
  - Backprop, RBF, support vector machines, ....
  - Esp. good when presence of interactions and nonlinear relationships is unknown
    - Fewer data assumptions than other methods.
- ✓ Other mathematical/computer techniques

#### 49 ☐ Is prediction sufficient?

- ✓ In addition to predicting a particular value or probability, we must decide *how to act* on that prediction.
  - Often must establish a threshold for action.
    - When to sound the tornado siren.
    - When to recommend parole, accept to graduate school, search a locker, give a drug.
- ✓ Action decisions can create problems.
  - *False alarms* vs. *Misses* must be traded off.
  - See article by Hammond at end of semester.

#### 50 ☐ What *should* we do to make good predictions?

- ✓ Must recognize our strengths and weaknesses
  - Strengths
    - Good at choosing what might be important
    - Can recognize exceptions outside scope of model (but be careful - we tend to identify too many exceptions)
      - Small adjustments hurt, large adjustments helped (Nikolopolous et al., 2005) - predictors should adjust only when *sure* (broken, not bruised, legs).

– Weaknesses

- Suboptimal at identifying predictive validity
- Difficult to recognize nonlinear relationships and interactions
- Inconsistent in application of our knowledge
- Don't like uncertainty so we overreach (leave overreaching to research, not practice)

51 ☐ What *should* we do....?

✓ Use decision aids like statistical and other computational methods to overcome our weaknesses.

- We aren't afraid to wear glasses or use microscopes & telescopes to aid our vision.
- We readily accept the use of calculators, voltmeters, computers, and bulldozers.
- We *should* be willing to use any tool that will improve our predictions - do so!

52 ☐ Dawes (2002)

✓ “It is a little more difficult to accept outside aids to our perceptual and motor abilities when there is ‘nothing wrong.’ We do so, however, when we know that we should land Boeing 727’s by instruments when flying from a dark space into a lighted area -- rather than relying on our vision, which together with the shape of the cockpit creates an illusion that we are higher above the landing strip than we in fact are. Relinquishing our intellectual abilities to the external aid of an SPR meets with more resistance. But nobody ever said that behaving ethically was easy.”