# Insurance, biases, discrimination & fairness

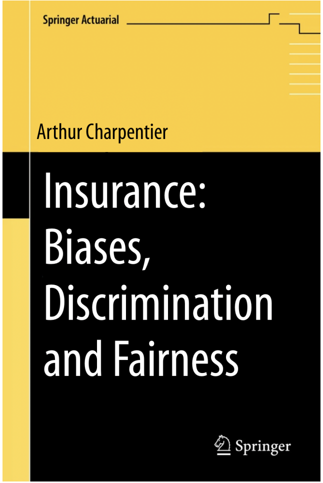
Arthur Charpentier

2024

# Reference book

Insurance, Biases, Discrimination and Fairness
ISBN : 978-3-031-49782-7

**Pitch**: Discrimination and fairness of predictive models, in insurance, in the context of data enrichment ("big data") and opaque models ("machine learning", not to say "artificial intelligence").

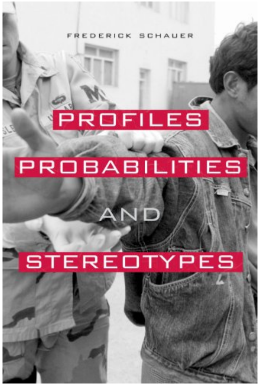


# Preliminaries

**Definition 1.1: Actuaries, Schauer (2006)**

To be an actuary is to be a specialist in generalization, and actuaries engage in a form of decision making that is sometimes called actuarial. Actuaries guide insurance companies in making decisions about large categories that have the effect of attributing to the entire category certain characteristics that are probabilistically indicated by membership in the category, but that still may not be possessed by a particular member of the category.

See Barry and Charpentier (2020) on personalization of insurance prices.




# Preliminaries

...
*– Tu la troubles, reprit cette bête cruelle,*
*Et je sais que de moi tu médis l'an passé.*
*– Comment l'aurais-je fait si je n'étais pas né ?*
*Reprit l'Agneau, je tette encor ma mère.*
*– Si ce n'est toi, c'est donc ton frère.*
*– Je n'en ai point.*
*– C'est donc quelqu'un des tiens.*
...

de La Fontaine (1668), *Le Loup et l'Agneau*.

# Preliminaries

**Definition 1.2: Discrimination, Merriam-Webster (2022)**

Discrimination is the act, practice, or an instance of separating or distinguishing categorically rather than individually.

**Definition 1.3: Prejudice, Merriam-Webster (2022)**

Prejudice is (1) preconceived judgment or opinion, or an adverse opinion or leaning formed without just grounds or before sufficient knowledge; (2) an instance of such judgment or opinion; (3) an irrational attitude of hostility directed against an individual, a group, a race, or their supposed characteristics.

# Preliminaries

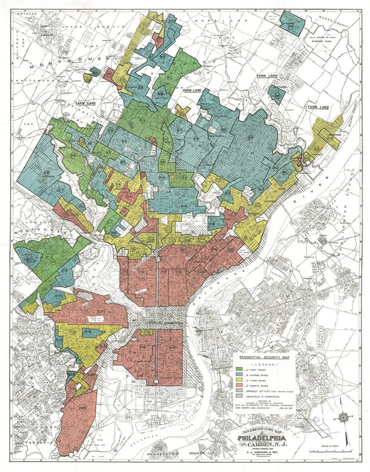
**Definition 1.4: Disparate treatment, Merriam-Webster (2022)**

Disparate treatment corresponds to the treatment of an individual (as an employee or prospective juror) that is less favorable than treatment of others for discriminatory reasons (as race, religion, national origin, sex, or disability).

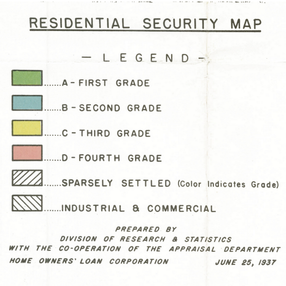
**Definition 1.5: Disparate impact, Merriam-Webster (2022)**

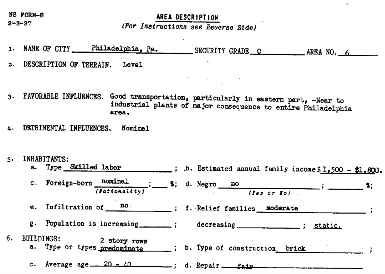
Disparate impact corresponds to an unnecessary discriminatory effect on a protected class caused by a practice or policy (as in employment or housing) that appears to be nondiscriminatory.

# Motivation (1. Redlining)




1937 HOLC (Home Owners' Loan Corporation) "residential security" map of Philadelphia



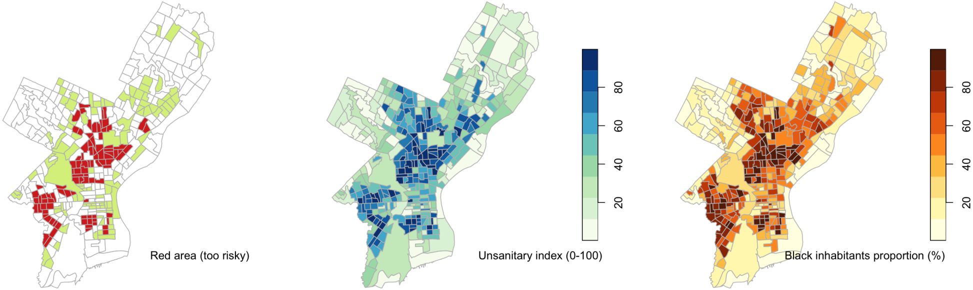



NS FORM-8
2-3-37

AREA DESCRIPTION
(For Instructions see Reverse Side)

1. NAME OF CITY Philadelphia, Pa. SECURITY GRADE C AREA NO. 6
2. DESCRIPTION OF TERRAIN. Level
3. FAVORABLE INFLUENCES. Good transportation, particularly in eastern part, -Near to industrial plants of major consequence to entire Philadelphia area.
4. DETRIMENTAL INFLUENCES. Nominal
5. INHABITANTS:
   a. Type Skilled labor ; b. Estimated annual family income $1,500 - $1,800.
   c. Foreign-born nominal (Nationality) ; %; d. Negro no (Yes or No) ; %;
   e. Infiltration of no ; f. Relief families moderate ;
   g. Population is increasing ; decreasing ; static.
6. BUILDINGS:
   a. Type or types 2 story rows predominate ; b. Type of construction brick ;
   c. Average age 20 - 40 ; d. Repair fair

# Motivation (1. Redlining)




(Fictitious maps, inspired by a Home Owners' Loan Corporation map from 1937)

- Federal Home Loan Bank Board (FHLBB) "*residential security maps*" (for real-estate investments), Crossney (2016) and Rhynhart (2020)
- Unsanitary index and proportion of Black inhabitants

# Motivation (1. Redlining)

**Definition 2.1: Redline, Merriam-Webster (2022)**

To redline is (1) to withhold home-loan funds or insurance from neighborhoods considered poor economic risks; (2) to discriminate against in housing or insurance.

See https://evolutionofraceandinsurance.org/ for some historical perspective, Squires and Velez (1988), or more recently Squires (2003)

... but still a concern see, e.g., Li (1996) about homosexuals.

# Motivation (2. “Gender directive”, 2004/113/EC)

Treaty on European Union (26.10.2012, C326)

– Article 2 –

The Union is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities. These values are common to the Member States in a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail.

– Article 3 –

(...) It shall combat social exclusion and discrimination, and shall promote social justice and protection, equality between women and men, solidarity between generations and protection of the rights of the child.

# Motivation (2. “Gender directive”, 2004/113/EC)

Charter of Fundamental Rights of the European Union (18.12.2000 , C364)

– Article 21 (Non discrimination) –

Any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited.

– Article 23 (Equality between men and women) –

Equality between men and women must be ensured in all areas, including employment, work and pay.
The principle of equality shall not prevent the maintenance or adoption of measures providing for specific advantages in favour of the under-represented sex.

# Motivation (2. "Gender directive", 2004/113/EC)

EU Directive (2004/113/EC), 2004 version

– Article 5 (Actuarial factors) –

1. Member States shall ensure that in all new contracts concluded after 21 December 2007 at the latest, the use of sex as a factor in the calculation of premiums and benefits for the purposes of insurance and related financial services shall not result in differences in individuals' premiums and benefits.

2. Notwithstanding paragraph 1, Member States may decide before 21 December 2007 to permit proportionate differences in individuals' premiums and benefits where the use of sex is a determining factor in the assessment of risk based on relevant and accurate actuarial and statistical data. The Member States concerned shall inform the Commission and ensure that accurate data relevant to the use of sex as a determining actuarial factor are compiled, published and regularly updated.

# Motivation (2. "Gender directive", 2004/113/EC)

- There was initially (2004) an opt-out clause (Article 5(2)).
- Where gender is a determining factor in the assessment of risk based on relevant and accurate actuarial and statistical data then proportionate differences in individual premiums or benefits are allowed.
- March 2011, the European Court of Justice issued its judgement into the "Test-Achats case". The ECJ ruled Article 5(2) was invalid.
- Insurers were no longer able to use gender as a risk factor when pricing policies, "unisex pricing".

"*Machine learning won't give you anything like gender neutrality 'for free' that you didn't explicitly ask for*", Kearns and Roth (2019)

# Motivation (2. "Gender directive", 2004/113/EC)

"Ten Oever" judgement (*Gerardus Cornelis Ten Oever v Stichting Bedrijfspensioenfonds voor het Glazenwassers – en Schoonmaakbedrijf*, in April 1993), the Advocate General Van Gerven argued that "*the fact that women generally live longer than men has no significance at all for the life expectancy of a specific individual and it is not acceptable for an individual to be penalized on account of assumptions which are not certain to be true in his specific case*," as mentioned in De Baere and Goessens (2011).
Schanze (2013) used the term "*injustice by generalization*."

# Motivation (2. "Gender directive", 2004/113/EC)

The Telegraph News Sport Money Business Opinion

**Men are still charged more than women for car insurance, despite EU rule change**

Car insurers are dodging European equality laws by making gender judgements based on people's jobs, an economist has found

*By* Kate Palmer
10 April 2015 • 12:33pm

Insurers will price by occupation, and female-dominated jobs tend to attract cheaper premiums | CREDIT: Photo: Rex Features

**CAR COSTS: Insurance according to job**

| Job | Proportion of men | Approximate average premium for a Fiat 500 driver |
|---|---|---|
| Dental Nurse | Less than 1pc male | £840 |
| Solicitor | 59pc male | £848 |
| Sports and leisure assistants | 56pc male | £880 |
| Civil engineer | 92pc male | £910 |
| Social worker | 21pc male | £920 |
| Plasterer | 98pc male | £950 |

McDonald, 'Indirect Gender Discrimination' (2015); ONS occupation data (2008)

(data source: Mcdonald (2015))

# Motivation (3. Colorado)

Andrus et al. (2021), "*What we can't measure, we can't understand*"

First Regular Session | 74th General Assembly
Colorado General Assembly

September 27, 2023, the Colorado Division of Insurance exposed a new proposed regulation entitled Concerning Quantitative Testing of External Consumer Data and Information Sources, Algorithms, and Predictive Models Used for Life Insurance Underwriting for Unfairly Discriminatory Outcomes

# Motivation (3. Colorado)

– Section 4 (Definitions) –

Bayesian Improved First Name Surname Geocoding, or "BIFSG" means, for the purposes of this regulation, the statistical methodology developed by the RAND corporation for estimating race and ethnicity.

External Consumer Data and Information Source, or "ECDIS" means, for the purposes of this regulation, a data source or an information source that is used by a life insurer to supplement or supplant traditional underwriting factors. This term includes credit scores, credit history, social media habits, purchasing habits, home ownership, educational attainment, licensures, civil judgments, court records, occupation that does not have a direct relationship to mortality, morbidity or longevity risk, consumer-generated Internet of Things data, biometric data, and any insurance risk scores derived by the insurer or third-party from the above listed or similar data and/or information source.

# Motivation (3. Colorado)

– Section 5 (Estimating Race and Ethnicity) –

Insurers shall estimate the race or ethnicity of all proposed insureds that have applied for coverage on or after the insurer's initial adoption of the use of ECDIS, or algorithms and predictive models that use ECDIS, including a third party acting on behalf of the insurer that used ECDIS, or algorithms and predictive models that used ECDIS, in the underwriting decision-making process, by utilizing:
1. BIFSG and the insureds' or proposed insureds' name and geolocation information included in the applications) for life insurance shall be used to estimate the race and ethnicity of each insured or proposed insured.
2. For the purposes of BIFSG, the following racial and ethnic categories shall be used: Hispanic, Black, Asian Pacific Islander (API), and White.

# Motivation (3. Colorado)

– Section 6 (Application Approval Decision Testing Requirements) –

Using the BIFSG estimated race and ethnicity of proposed insureds and the following methodology, insurers shall calculate whether Hispanic, Black, and API proposed insureds are disapproved at a statistically significant different rate relative to White applicants for whom the insurer, or a third party acting on behalf of the insurer, used ECDIS, or an algorithm or predictive model that used ECDIS, in the underwriting decision-making process.
1. Logistic regression shall be used to model the binary underwriting outcome of either approved or denied.
2. The following factors may be accounted for as control variables in the regression model: policy type, face amount, age, gender, and tobacco use.
3. The estimated race or ethnicity of the proposed insureds shall be accounted for by including Hispanic, Black, and Asian Pacific Islander (API) as separate dummy variables in the regression model.

# Motivation (3. Colorado)

4. Determine if there is a statistically significant difference in approval rates for each BIFSG estimated race or ethnicity variable as indicated by a $p$-value of less than .05.

a. If there is not a statistically significant difference in approval rates, no further testing is required.

b. If there is a statistically significant difference in approval rates, the insurer shall determine whether the difference in approval rates is five (5) percentage points or greater as indicated by the marginal effects value of each BIFSG estimated race or ethnicity variable. (...)

# Motivation (3. Colorado)

– Section 7 (Premium Rate Testing Requirements) –

Using the insureds' BIFSG estimated race and ethnicity, insurers shall determine if there is a statistically significant difference in the premium rate per $1,000 of face amount for policies issued to Hispanic, Black, and API insureds relative to White insureds for whom the insurer, or a third party acting on behalf of the insurer, used ECDIS, or an algorithm or predictive model that used ECDIS, in the underwriting decision-making process.
1. Linear regression shall be used to model the continuous numerical outcome of premium rate per $1,000 of face amount.
2. The following factors may be accounted for as control variables in the regression model: policy type, face amount, age, gender, and tobacco use.
3. The estimated race or ethnicity of the proposed insureds shall be accounted for by including Hispanic, Black, and Asian Pacific Islander (API) as separate dummy variables in the regression model.

# Motivation (3. Colorado)

4. Determine if there is a statistically significant difference in the premium rate per $1,000 of face amount for each BIFSG estimated race or ethnicity variable as indicated by a p-value of less than .05.

a. If there is not a statistically significant difference in premium rate per $1,000 of face amount, no further testing is required.

b. If there is a statistically significant difference in premium rate per $1,000 of face amount, determine whether the premium rate per $1,000 of face amount is at least 5% more than the average premium rate per $1,000 for all policies.

i. If the difference in premium rate per $1,000 of face amount is less than 5%, no further testing is required.

ii. If the difference in premium rate per $1,000 of face amount is 5% or greater, further testing is required as described in Section 8.

# Motivation (3. Colorado)

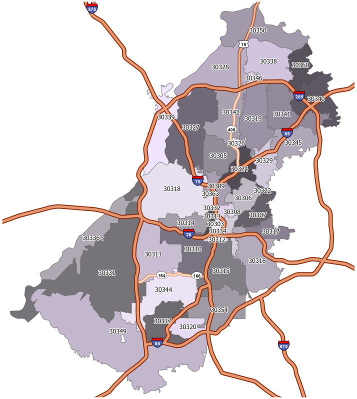
In Elliott et al. (2009), BIFSG[1], `library(eiCompare)`. , consider 12 people living near Atlanta, GA (Fulton & Gwinnett counties), and `eiCompare::wru_predict_race_wrapper`

```
         last      first   county         city zipcode whi bla his asi
1     LOCKLER  GABRIELLA   Fulton      Atlanta   30318   0   0   0   0
2      RADLEY     OLIVIA   Fulton     Fairburn   30213  14  83   1   0
3      BOORSE     KEISHA   Fulton      Atlanta   30331  97   0   3   0
4         MAZ   SAVANNAH Gwinnett     Norcross   30093   5   6  76  13
5       GAULE   NATASHIA Gwinnett   Snellville   30078  67  19  14   0
6    MCMELLEN     ISMAEL Gwinnett      Lilburn   30047  73  15   6   3
7     RIDEOUT     LUQMAN Gwinnett   Snellville   30078  77  18   2   0
8  WASHINGTON       BRYN Gwinnett     Norcross   30093   0  95   3   0
9   KULENOVIC     EVELYN Gwinnett       Buford   30518 100   0   0   0
10  HERNANDEZ   SAMANTHA Gwinnett       Duluth   30096   3   1  94   1
11       LONG     BESSIE Gwinnett       Duluth   30096  53  39   1   1
12         HE       JOSE Gwinnett Lawrenceville   30045   2   3   4  89
```

[1] Bayesian Improved First Name Surname Geocoding

# Motivation (3. Colorado)

We have 12 people,
in two counties near Atlanta
(about 10 zip-codes)

# Motivation (3. Colorado)

- Use `eiCompare::wru_predict_race_wrapper` on a revised dataset with the same name "Savannah Maz"

```
    last     first    county         city zipcode whi bla his asi
1    MAZ  SAVANNAH    Fulton      Atlanta   30318   0   0   0 100
2    MAZ  SAVANNAH    Fulton     Fairburn   30213  13  61  22   3
3    MAZ  SAVANNAH    Fulton      Atlanta   30331   3  77  19   1
4    MAZ  SAVANNAH  Gwinnett     Norcross   30093   5   6  76  13
5    MAZ  SAVANNAH  Gwinnett   Snellville   30078  13  18  69   0
6    MAZ  SAVANNAH  Gwinnett      Lilburn   30047  28  22  34  16
7    MAZ  SAVANNAH  Gwinnett   Snellville   30078  53   3  40   3
8    MAZ  SAVANNAH  Gwinnett     Norcross   30093   5   6  76  13
9    MAZ  SAVANNAH  Gwinnett       Buford   30518  79   4  14   2
10   MAZ  SAVANNAH  Gwinnett       Duluth   30096  32   8  38  22
11   MAZ  SAVANNAH  Gwinnett       Duluth   30096  55  19  22   5
12   MAZ  SAVANNAH  Gwinnett Lawrenceville  30045  15  19  62   4
```

# Motivation (3. Colorado)

- Use `eiCompare::wru_predict_race_wrapper` on a revised dataset with the same name “Bryn Washington”

```
         last first   county         city zipcode whi bla his asi
1  WASHINGTON  BRYN   Fulton      Atlanta   30318   0   0   0 100
2  WASHINGTON  BRYN   Fulton    Fairburn    30213   0  99   0   0
3  WASHINGTON  BRYN   Fulton      Atlanta   30331   0  99   0   0
4  WASHINGTON  BRYN Gwinnett     Norcross   30093   0  95   3   0
5  WASHINGTON  BRYN Gwinnett   Snellville   30078   0  96   1   0
6  WASHINGTON  BRYN Gwinnett      Lilburn   30047   1  98   0   0
7  WASHINGTON  BRYN Gwinnett   Snellville   30078   6  87   2   0
8  WASHINGTON  BRYN Gwinnett     Norcross   30093   0  95   3   0
9  WASHINGTON  BRYN Gwinnett       Buford   30518   7  92   1   0
10 WASHINGTON  BRYN Gwinnett       Duluth   30096   2  96   1   0
11 WASHINGTON  BRYN Gwinnett       Duluth   30096   1  96   0   0
12 WASHINGTON  BRYN Gwinnett Lawrenceville  30045   0  98   1   0
```

# Motivation (3. Colorado)

- Use `eiCompare::wru_predict_race_wrapper` on a revised dataset with the same name "Samantha Hernandez"

```
        last    first   county         city zipcode whi bla his asi
1  HERNANDEZ SAMANTHA   Fulton      Atlanta   30318   0   0   0 100
2  HERNANDEZ SAMANTHA   Fulton     Fairburn   30213   2  12  85   0
3  HERNANDEZ SAMANTHA   Fulton      Atlanta   30331   0  16  81   0
4  HERNANDEZ SAMANTHA Gwinnett     Norcross   30093   0   0  99   0
5  HERNANDEZ SAMANTHA Gwinnett   Snellville   30078   1   1  97   0
6  HERNANDEZ SAMANTHA Gwinnett      Lilburn   30047   3   3  92   1
7  HERNANDEZ SAMANTHA Gwinnett   Snellville   30078   5   0  94   0
8  HERNANDEZ SAMANTHA Gwinnett     Norcross   30093   0   0  99   0
9  HERNANDEZ SAMANTHA Gwinnett       Buford   30518  17   1  81   0
10 HERNANDEZ SAMANTHA Gwinnett       Duluth   30096   3   1  94   1
11 HERNANDEZ SAMANTHA Gwinnett       Duluth   30096   8   4  86   0
12 HERNANDEZ SAMANTHA Gwinnett Lawrenceville   30045   1   2  97   0
```

# Motivation (3. Colorado)

- Use `eiCompare::wru_predict_race_wrapper` on a revised dataset with the same name "Jose He"

```
   last first   county         city zipcode whi bla his asi
1    HE  JOSE   Fulton      Atlanta   30318   0   0   0 100
2    HE  JOSE   Fulton     Fairburn   30213   2   9   2  84
3    HE  JOSE   Fulton      Atlanta   30331   1  27   3  55
4    HE  JOSE Gwinnett     Norcross   30093   0   0   2  98
5    HE  JOSE Gwinnett   Snellville   30078  13  18  30   0
6    HE  JOSE Gwinnett      Lilburn   30047   1   1   1  97
7    HE  JOSE Gwinnett   Snellville   30078   8   1   3  86
8    HE  JOSE Gwinnett     Norcross   30093   0   0   2  98
9    HE  JOSE Gwinnett       Buford   30518  19   1   2  78
10   HE  JOSE Gwinnett       Duluth   30096   1   0   0  98
11   HE  JOSE Gwinnett       Duluth   30096   6   2   1  85
12  HE     JOSE   Gwinnett    Lawrenceville    30045    2   3   4  89
```

# Motivation (4. Motor Insurance in the U.S.)

California

**Allowed (with applicable limitations):** driving experience, marital status, address/zip code

**Prohibited (or effectively prohibited):** gender, age, credit history, education, occupation, employment status, residential status, insurance history

**Notes & Clarifications:** California's insurance commissioner banned gender as of January 2019. Occupation and education are permitted for use in group plans (i.e. for alumni associations and other membership programs).

Georgia

**Allowed (with applicable limitations):** gender, age, years of driving experience, credit history, marital status, residential status, address/zip code, insurance history

**Prohibited (or effectively prohibited):** occupation, education, and employment status

**Notes & Clarifications:** none

Hawaii

**Allowed (with applicable limitations):** address/zip code, insurance history

**Prohibited (or effectively prohibited):** gender, age, years of driving experience, credit history, education, occupation, employment status, marital status, residential status

**Notes & Clarifications:** none

Illinois

**Allowed (with applicable limitations):** gender, age, years of driving experience, credit history, education, occupation, employment status, marital status, residential status, address/zip code, insurance history

**Prohibited (or effectively prohibited):** none

**Notes & Clarifications:** none

Massachusetts

**Allowed (with applicable limitations):** years of driving experience, address/zip code, insurance history

**Prohibited (or effectively prohibited):** gender, age, credit history, education, occupation, employment status, marital status, residential status

**Notes & Clarifications:** none

Michigan

**Allowed (with applicable limitations):** gender (group-rated policies), age, years of driving experience, credit history, education, occupation, employment status, marital status (group-rated policies), residential status, address/zip code, insurance history

**Prohibited (or effectively prohibited):** gender (non-group policies), marital status (non-group policies)

**Notes & Clarifications:** Gender and marital status are permitted only in rate-making for group plans (i.e. for alumni associations and other membership programs). **UPDATE:**Michigan lawmakers approved a major insurance reform bill in May 2019 that will ban insurers in the state from using gender, marital status, address/zipcode, residential status, education and occupation in rate setting. The ban will be enforced starting in July 2020. Insurers will be permitted to use "territory" as approved by the state regulators instead of zip code.

New York

**Allowed (with applicable limitations):** gender, age, years of driving experience, credit history, marital status, residential status, address/zip code, insurance history

**Prohibited (or effectively prohibited):** occupation, education, employment status

**Notes & Clarifications:** none

via The Zebra (2022)

# Motivation (5. Admission in Graduate Program, UC Berkeley)

## Sex Bias in Graduate Admissions: Data from Berkeley

Measuring bias is harder than is usually assumed, and the evidence is sometimes contrary to expectation.

P. J. Bickel, E. A. Hammel, J. W. O'Connell

Determining whether discrimination because of sex or ethnic identity is being practiced against persons seeking passage from one social status or locus to another is an important problem in our society today. It is legally important and morally important. It is also often quite difficult. This article is an exploration of some of the issues of measurement and assessment involved in one example of the general problem, by means of which we hope to shed some light on the difficulties. We will proceed in a straightforward and indeed naive way, even though we know how misleading an unsophisticated approach to the problem is. We do this because we think it quite likely that other persons interested in questions of bias might proceed in just the same way, and careful exposure of the mistakes in our discovery procedure may be instructive.

### Data and Assumptions

The particular body of data chosen for examination here consists of applications for admission to graduate study at the University of California, Berkeley, for the fall 1973 quarter. In the admissions cycle for that quarter, the Graduate Division at Berkeley received approximately 15,000 applications, some of which were later withdrawn or transferred to a different proposed entry quarter by the applicants. Of the applications finally remaining for the fall 1973 cycle 12,763 were sufficiently complete to permit a deceision to admit or to deny admission. The question we wish to pursue is whether the decision to admit or to deny was influenced by the sex of the applicant. We cannot know with any certainty the influences on the evaluators in the Graduate Admissions Office, or on the faculty reviewing committees, or on any other administrative personnel participating in the chain of actions that led to a decision on an individual application. We can, however, say that if the admissions decision and the sex of the applicant are statistically associated in the results of a series of applications, we may judge that *bias* existed, and we may then seek to find whether *discrimination* existed. By "bias" we mean here a pattern of association between a particular decision and a particular sex of applicant, of sufficient strength to make us confident that it is unlikely to be the result of chance alone. By "discrimination" we mean the exercise of decision influenced by the sex of the applicant when that is immaterial to the qualifications for entry.

The simplest approach (which we shall call approach A) is to examine the aggregate data for the campus. This approach would surely be taken by many persons interested in whether bias in admissions exists on any campus. Table 1 gives the data for all 12,763 applications to the 101 graduate departments and interdepartmental graduate majors to which application was made for fall 1973 (we shall refer to them all as departments). There were 8442 male applicants and 4321 female applicants. About 44 percent of the males and about 35 percent of the females were admitted. Just this kind of simple calculation of proportions impels us to examine the data further. We will pursue the question by using a familiar statistic, chi-square. As already noted, we are aware of the pitfalls ahead in this naive approach, but we intend to stumble into every one of them for didactic reasons.

We must first make clear two assumptions that underlie consideration of the data in this contingency table approach. Assumption 1 is that in any given discipline male and female applicants do not differ in respect of their intelligence, skill, qualifications, promise, or other attribute deemed legitimately pertinent to their acceptance as students. It is precisely this assumption that makes the study of "sex bias" meaningful, for if we did not hold it any differences in acceptance of applicants by sex could be attributed to differences in their qualifications, promise as scholars, and so on. Theoretically one could test the assumption, for example, by examining presumably unbiased estimators of academic qualification such as Graduate Record Examination scores, undergraduate grade point averages, and so on. There are, however, enormous practical difficulties in this. We therefore predicate our discussion on the validity of assumption 1.

Assumption 2 is that the sex ratios of applicants to the various fields of graduate study are not importantly associated with any other factors in admission. We shall have reason to challenge this assumption later, but it is crucial in the first step of our exploration, which is the investigation of bias in the aggregate data.

Dr. Bickel is professor of statistics, Dr. Hammel is professor of anthropology and associate dean of the Graduate Division, and Mr. O'Connell is a member of the data processing staff of the Graduate Division, at the University of California, Berkeley 94720.

### Tests of Aggregate Data

We pursue this investigation by computing the expected frequencies of male and female applicants admitted and denied, from the marginal totals of Table 1, on the assumption that men and women applicants have equal chances of admission to the university (that is, on the basis of assumptions 1 and 2). This computation, also given in Table 1, shows that 277 fewer women and 277 more men were admitted than we would have expected under the assumptions noted. That is a large number, and it is unlikely that so large a bias to the disadvantage of women would occur by chance alone. The chi-square value for this table is 110.8, and the probability of a chi-square that large (or larger) under the assumptions noted is vanishingly small.

We should on this evidence judge



that bias existed in the fall 1973 admissions. On that account, we should look for the responsible parties to see whether they give evidence of discrimination. Now, the outcome of an application for admission to graduate study is determined mainly by the faculty of the department to which the prospective student applies. Let us then examine each of the departments for indications of bias. Among the 101 departments we find 16 that either had no women applicants or denied admission to no applicants of either sex. Our computations, therefore, except where otherwise noted, will be based on the remaining 85. For a start let us identify those of the 85 with bias sufficiently large to occur by chance less than five times in a hundred. There prove to be four such departments. The deficit in the number of women admitted to these four (under the assumptions for calculating expected frequencies as given above) is 26. Looking further, we find six departments biased in the opposite direction, at the same probability levels; these account for a deficit of 64 men.

These results are confusing. After all, if the campus had a shortfall of 277 women in graduate admissions, and we look to see who is responsible, we ought to find somebody. So large a deficit ought not simply to disappear. There is even a suggestion of a surplus of women. Our method of examination must be faulty.

Table 1. Decisions on applications to Graduate Division for fall 1973, by sex of applicant—naive aggregation. Expected frequencies are calculated from the marginal totals of the observed frequencies under the assumptions (1 and 2) given in the text. $N = 12{,}763$, $\chi^2 = 110.8$, d.f. $= 1$, $P = 0$ (18).

| Applicants | Outcome | | | | Difference | |
|---|---|---|---|---|---|---|
| | Observed | | Expected | | | |
| | Admit | Deny | Admit | Deny | Admit | Deny |
| Men | 3738 | 4704 | 3460.7 | 4981.3 | 277.3 | − 277.3 |
| Women | 1494 | 2827 | 1771.3 | 2549.7 | − 277.3 | 277.3 |

### Some Underlying Dependencies

We have stumbled onto a paradox, sometimes referred to as Simpson's in this context (1) or "spurious correlation" in others (2). It is rooted in the falsity of assumption 2 above. We have assumed that if there is bias in the proportion of women applicants admitted it will be because of a link between sex of applicant and decision to admit. We have given much less attention to a prior linkage, that between sex of applicant and department to which admission is sought. The tendency of men and women to seek entry to different departments is marked. For example, in our data almost two-thirds of the applicants to English but only 2 percent of the applicants to mechanical engineering are women. If we cast the application data into a 2 × 101 contingency table, distinguishing department and sex of applicants, we find this table has a chi-square of 3091 and that the probability of obtaining a chi-square value that large or larger by chance is about zero. For the 2 × 85 table on the departments used in most of the analysis, chi-square is 3027 and the probability about zero. Thus the sex distribution of applicants is anything but random among the departments. In examining the data in the aggregate as we did in our initial approach, we pooled data from these very different, independent decision-making units. Of course, such pooling would not nullify assumption 2 if the different departments were equally difficult to enter. We will address ourselves to that question in a moment.

Let us first examine an alternative to aggregating the data across the 85 departments and then computing a statistic—namely, computing a statistic on each department first and aggregating those. Fisher gives a method for aggregating the results of such independent experiments (3). If we apply his method to the chi-square statistics of the 85 individual contingency tables, we obtain a value that has a probability of occurrence by chance alone, that is, if sex and admission are unlinked for any major, of about 29 times in 1000 (4). Another common aggregation procedure, proposed to us in this context by E. Scott, yields a result having a probability of 6 times in 10,000 (5). This is consistent with the evidence of bias in some direction purportedly shown by Table 1. However, when we examine the *direction* of bias, the picture changes. For instance, if we apply Fisher's method to the *one-sided* statistics, testing the hypothesis of no bias or of bias in favor of women, we find that we could have obtained a value as large as or larger than the one observed, by chance alone, about 85 times in 100 (6).

Our first, naive approach of examining the aggregate data, computing expected frequencies under certain assumptions, computing a statistic, and deciding therefrom that bias existed in favor of men has now been cast into doubt on at least two grounds. First, we could not find many biased decision-making units by examining them individually. Second, when we take account of the differences among departments in the proportions of men and women applying to them and avoid this problem by computing a statistic on each department separately, and aggregating those statistics, the evidence for campus-wide bias in favor of men is extremely weak; on the contrary, there is evidence of bias in favor of women.

The missing piece of the puzzle is yet another fact: not all departments are equally easy to get into. If we cast the data into a 2 × 101 table, distinguishing department and decision to admit or deny, we find that this table has a chi-square value of 2195, with an associated probability of occurrence by chance (under assumptions 1 and 2) of about zero, showing that the odds of gaining admission to different departments are widely divergent. (For the 2 × 85 table chi-square is 2121 and the probability about zero.) Now, these odds of getting into a graduate program are in fact strongly associated with the tendency of men and women to apply to different departments in different degree. *The proportion of women applicants tends to be high in departments that are hard to get into and low in those that are easy to get into.* Moreover this phenomenon is more pronounced in departments with large numbers of applicants. Figure 1 is a scattergram of proportion of applicants that are women plotted against proportion of applicants that are admitted. The association is obvious on inspection although the relationship is certainly not linear (7). If we use a weighted correlation (8) as a measure of the relationship for all 85 departments in the plot we obtain $\hat{\beta} = .56$. If we apply the same measure to the 17 departments with the largest numbers of applicants (accounting for two-



thirds of the total population of applicants) we obtain $\hat{\beta} = .65$, while the remaining 68 departments have a corresponding $\hat{\beta} = .39$. The significance of $\hat{\beta}$ under the hypothesis of no association can be calculated. All three values obtained are highly significant.

The effect may be clarified by means of an analogy. Picture a fishnet with two different mesh sizes. A school of fish, all of identical size (assumption 1), swim toward the net and seek to pass. The female fish all try to get through the small mesh, while the male fish all try to get through the large mesh. On the other side of the net all the fish are male. Assumption 2 said that the sex of the fish had no relation to the size of the mesh they tried to get through. It is false. To take another example that illustrates the danger of incautious pooling of data, consider two departments of a hypothetical university—machismatics and social warfare. To machismatics there apply 400 men and 200 women; these are admitted in exactly equal proportions, 200 men and 100 women. To social warfare there apply 150 men and 450 women; these are admitted in exactly equal proportions, 50 men and 150 women. Machismatics admitted half the applicants of each sex, social warfare admitted a third of the applicants of each sex. But about 73 percent of the men applied to machismatics and 27 percent to social warfare, while about 69 percent of the women applied to social warfare and 31 percent to machismatics. When these two departments are pooled and expected frequencies are computed in the usual way (with assumption 2), there is a deficit of about 21 women (Table 2). A discrepancy in that direction that large or larger would be expectable less than 2 percent of the time by chance; yet both departments were seen to have been absolutely fair in dealing with their applicants.

Table 2. Admissions data by sex of applicant for two hypothetical departments. For total, $\chi^2 = 5.71$, d.f. $= 1$, $P = 0.19$ (one-tailed).

| Applicants | Outcome | | | | Difference | |
|---|---|---|---|---|---|---|
| | Observed | | Expected | | | |
| | Admit | Deny | Admit | Deny | Admit | Deny |
| | *Department of machismatics* | | | | | |
| Men | 200 | 200 | 200 | 200 | 0 | 0 |
| Women | 100 | 100 | 100 | 100 | 0 | 0 |
| | *Department of social warfare* | | | | | |
| Men | 50 | 100 | 50 | 100 | 0 | 0 |
| Women | 150 | 300 | 150 | 300 | 0 | 0 |
| | *Totals* | | | | | |
| Men | 250 | 300 | 229.2 | 320.8 | 20.8 | − 20.8 |
| Women | 250 | 400 | 270.8 | 379.2 | − 20.8 | 20.8 |

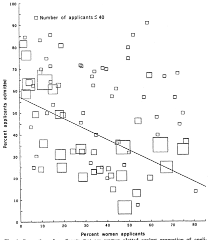



Fig. 1. Proportion of applicants that are women plotted against proportion of applicants admitted, in 85 departments. Size of box indicates relative number of applicants to the department.

The creation of bias in our original situation is, of course, much more complex, since we are aggregating many tables. It results from an interaction of the three factors, choice of department, sex, and admission status, whose broad outlines are suggested by our plot but which cannot be described in any simple way.

In any case, aggregation in a simple and straightforward way (approach A) is misleading. More sophisticated methods of aggregation that do not rely on assumption 2 are legitimate but have their difficulties. We shall have more to say on this later.

### Disaggregation

The most radical alternative to approach A is to consider the individual graduate departments, one by one. However, this approach (which we may call approach B) also poses difficulties. Either we must sample randomly from the different departments, or we must take account of the probability of obtaining unusual sex ratios of admittees by chance in a number of simultaneously conducted independent experiments. That is, in examining 85 separate departments at the same time for evidence of bias we are conducting 85 simultaneous experiments,

# Motivation (5. Admission in Graduate Program, UC Berkeley)

| | Total | Men | Women | Proportions |
|---|---|---|---|---|
| Total | $5233/12763 \sim 41\%$ | $3714/8442 \sim$ **44%** | $1512/4321 \sim 35\%$ | 66%-34% |
| Top 6 | $1745/4526 \sim 39\%$ | $1198/2691 \sim$ **45%** | $557/1835 \sim 30\%$ | 59%-41% |
| A | $597/933 \sim 64\%$ | $512/825 \sim 62\%$ | $89/108 \sim$ **82%** | 88%-12% |
| B | $369/585 \sim 63\%$ | $353/560 \sim 63\%$ | $17/\ 25 \sim$ **68%** | 96%- 4% |
| C | $321/918 \sim 35\%$ | $120/325 \sim$ **37%** | $202/593 \sim 34\%$ | 35%-65% |
| D | $269/792 \sim 34\%$ | $138/417 \sim 33\%$ | $131/375 \sim$ **35%** | 53%-47% |
| E | $146/584 \sim 25\%$ | $53/191 \sim$ **28%** | $94/393 \sim 24\%$ | 33%-67% |
| F | $43/714 \sim\ 6\%$ | $22/373 \sim\ 6\%$ | $24/341 \sim$ **7%** | 52%-48% |

Data from Bickel et al. (1975)

Formalize the later, $S$ is the (binary) genre, $Y$ the admission and $X$ the program (category),

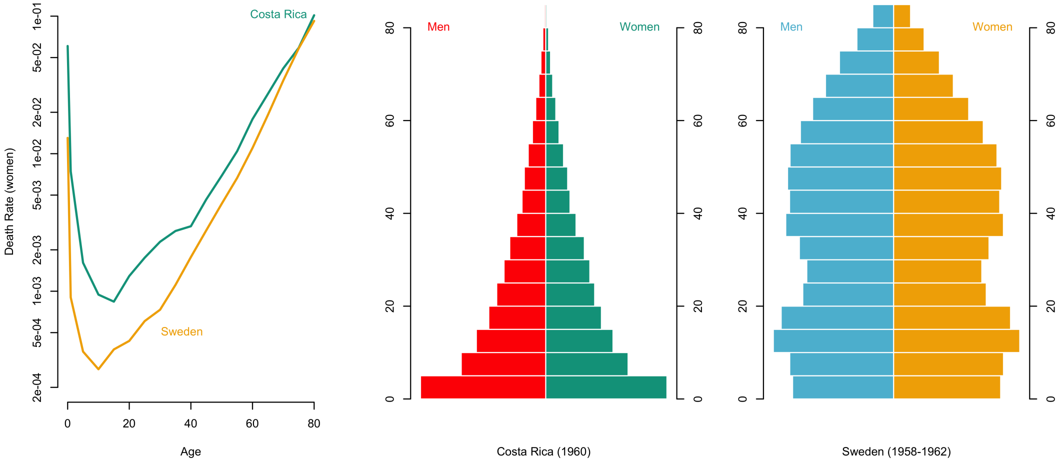
# Motivation (5. Admission in Graduate Program, UC Berkeley)

$$\mathbb{P}[\underbrace{Y=\text{yes}}_{\text{overall admission}} \mid \overbrace{S=\text{men}}^{\text{sensitive}}] \geq \mathbb{P}[\underbrace{Y=\text{yes}}_{\text{overall admission}} \mid \overbrace{S=\text{women}}^{\text{sensitive}}]$$

$$\mathbb{P}[Y=\text{yes} \mid \underbrace{X=x}_{\text{conditional on program}}, S=\text{men}] \leq \mathbb{P}[Y=\text{yes} \mid \underbrace{X=x}_{\text{conditional on program}}, S=\text{women}], \ \forall x.$$

"*the bias in the aggregated data stems not from any pattern of discrimination on the part of admissions committees, which seems quite fair on the whole, but apparently from prior screening at earlier levels of the educational system. Women are shunted by their socialization and education toward fields of graduate study that are generally more crowded, less productive of completed degrees, and less well funded, and that frequently offer poorer professional employment prospects*," Bickel et al. (1975)

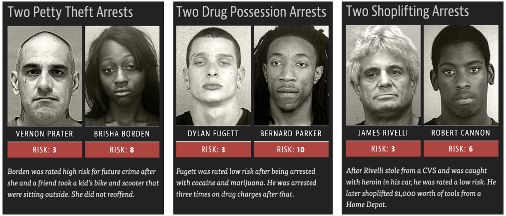
# Motivation (5'. Mortality in Costa Rica and Sweden)

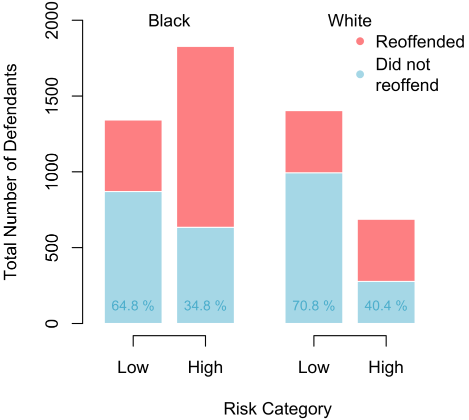



Overall mortality rate for women, 8.12‰ in Costa Rica, against 9.29‰ in Sweden.

# Motivation (6. Propublica, Actuarial Justice)

- Concept of "actuarial justice" as coined in Feeley and Simon (1994)
- **C**orrectional **O**ffender **M**anagement **P**rofiling for **A**lternative **S**anctions (**COMPAS**), Perry (2013)

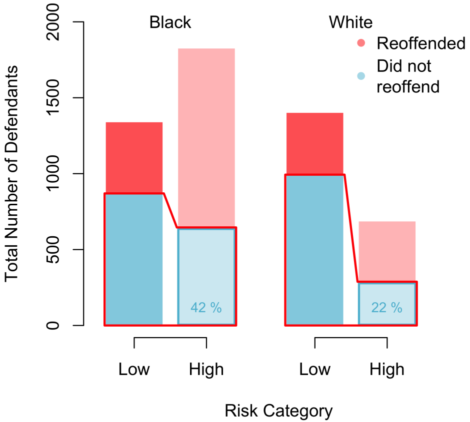



- https://github.com/propublica/compas-analysis
- Angwin et al. (2016) Machine Bias
  Dressel and Farid (2018)

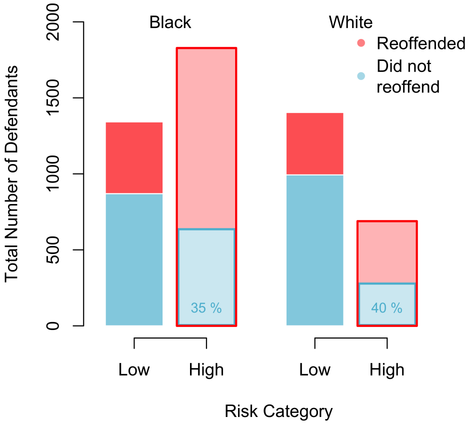
# Motivation (6. Propublica, Actuarial Justice)

- From Feller et al. (2016),
  - for White people, among those who did not re-offend, 22% were wrongly classified,
  - for Black people, among those who did not re-offend, 42% were wrongly classified,
  - problem, since $42\% \gg 22\%$



# Motivation (6. Propublica, Actuarial Justice)

- From Dieterich et al. (2016),
  - for White people, among those who where classified as high risk, 40% did not re-offend,
  - for Black people, among those who where classified as high risk, 35% did not re-offend,
  - no problem, since 40% $\approx$ 35%



# Motivation (6. Propublica, Actuarial Justice)

Formalize the later,

$$\begin{cases} S: & \text{race (binary), black \& white} \\ Y: & \text{re-offense (binary), no \& yes} \\ \widehat{Y}: & \text{classifier (risk category), low \& high} \end{cases}$$

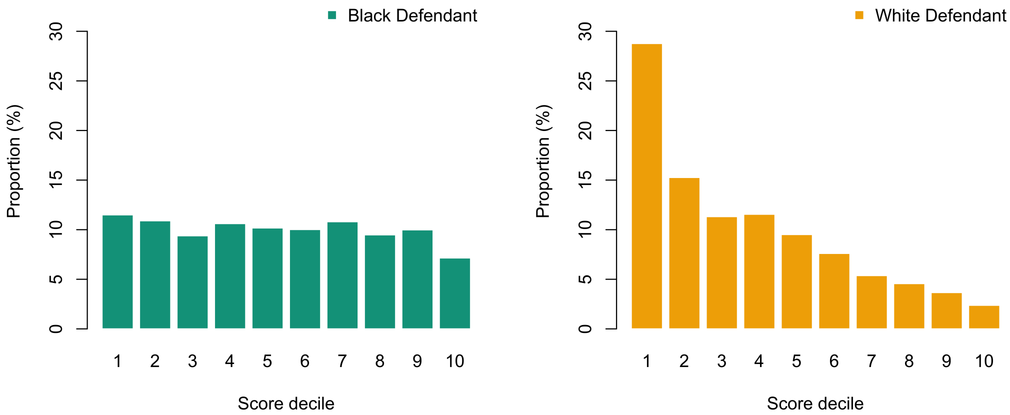
$$\mathbb{P}[\widehat{Y}=\text{high}|Y=\text{no}, \overbrace{S=\text{black}}^{\text{sensitive}}] = 42\% \overset{?}{=} \mathbb{P}[\widehat{Y}=\text{high}|Y=\text{no}, \overbrace{S=\text{white}}^{\text{sensitive}}] = 22\%,$$

false positive rate

$$\mathbb{P}[Y=\text{no}|\widehat{Y}=\text{high}, S=\text{black}] = 35\% \overset{?}{=} \mathbb{P}[Y=\text{no}|\widehat{Y}=\text{high}, S=\text{white}] = 40\%.$$

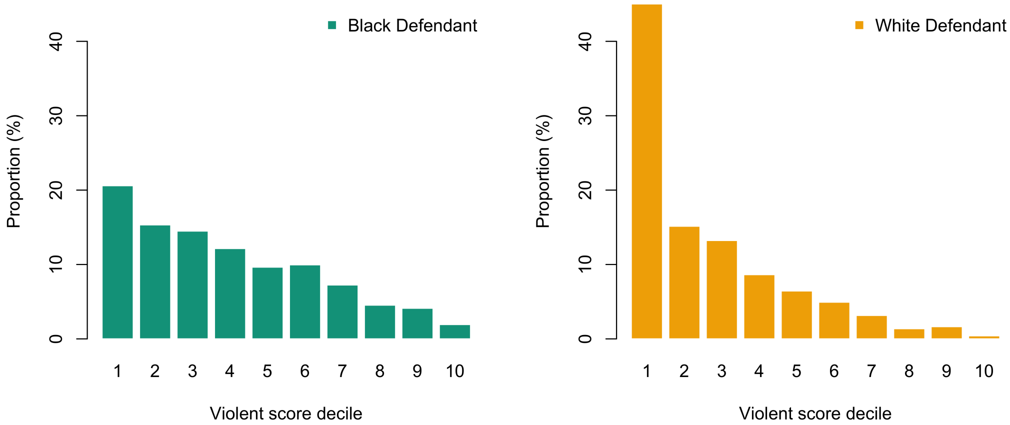
false discovery rate

# Motivation (6. Propublica, Actuarial Justice)



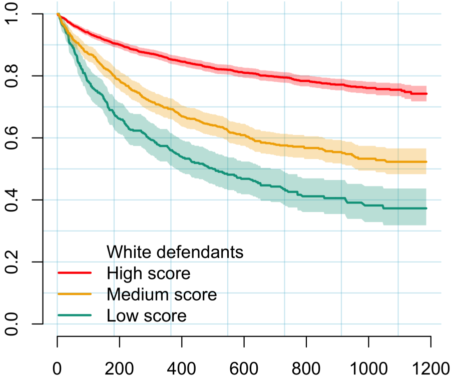

Look at score distributions, black and white defendant, Larson et al. (2016).

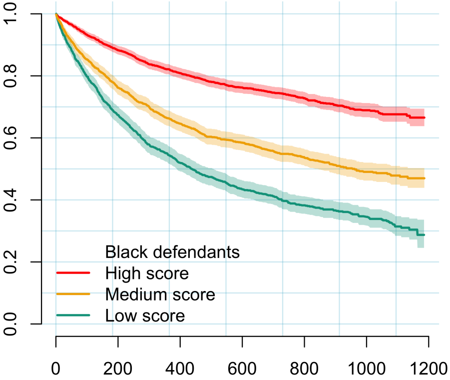
# Motivation (6. Propublica, Actuarial Justice)




Look at score distributions, black and white defendant, Larson et al. (2016).

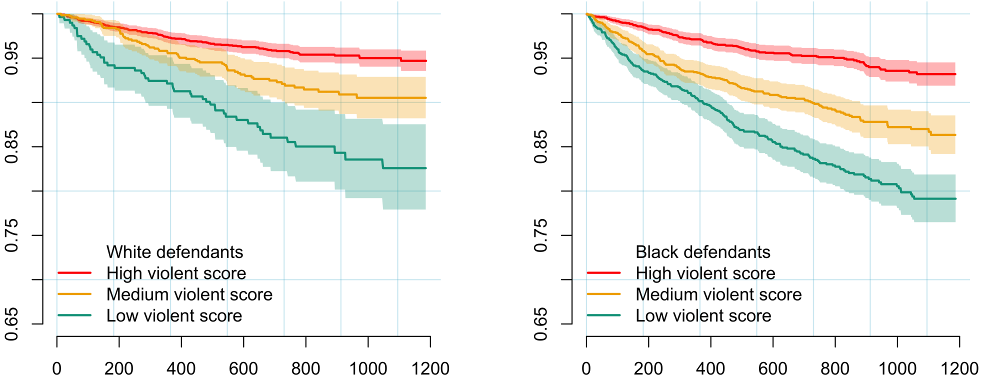
# Motivation (6. Propublica, Actuarial Justice)




Cox Proportional Hazards model, black and white defendant, Larson et al. (2016).

# Motivation (6. Propublica, Actuarial Justice)




Cox Proportional Hazards model, black and white defendant, Larson et al. (2016).

# Motivation (7. Québec)

Au Québec, Charte des droits et libertés de la personne (C-12)

– Article 10 –

Toute personne a droit à la reconnaissance et à l'exercice, en pleine égalité, des droits et libertés de la personne, sans distinction, exclusion ou préférence fondée sur la race, la couleur, le sexe, l'identité ou l'expression de genre, la grossesse, l'orientation sexuelle, l'état civil, l'âge sauf dans la mesure prévue par la loi, la religion, les convictions politiques, la langue, l'origine ethnique ou nationale, la condition sociale, le handicap ou l'utilisation d'un moyen pour pallier ce handicap.

Il y a discrimination lorsqu'une telle distinction, exclusion ou préférence a pour effet de détruire ou de compromettre ce droit.

# Motivation (7. Québec)

Au Québec, Charte des droits et libertés de la personne (C-12)

– Article 20.1 –

Dans un contrat d’assurance ou de rente, un régime d’avantages sociaux, de retraite, de rentes ou d’assurance ou un régime universel de rentes ou d’assurance, une distinction, exclusion ou préférence fondée sur l’âge, le sexe ou l’état civil est réputée non discriminatoire lorsque son utilisation est légitime et que le motif qui la fonde constitue un facteur de détermination de risque, basé sur des données actuarielles.

# Motivation (8. Intention)

En France, Loi n° 2008-496 du 27 mai 2008

– Article 1 –

Constitue une discrimination indirecte une disposition, un critère ou une pratique neutre en apparence, mais susceptible d'entraîner, pour l'un des motifs mentionnés au premier alinéa, un désavantage particulier pour des personnes par rapport à d'autres personnes, à moins que cette disposition, ce critère ou cette pratique ne soit objectivement justifié par un but légitime et que les moyens pour réaliser ce but ne soient nécessaires et appropriés.

Extention de la "Loi n° 72-546 du 1 juillet 1972", qui supprima l'exigence de l'intention spécifique.

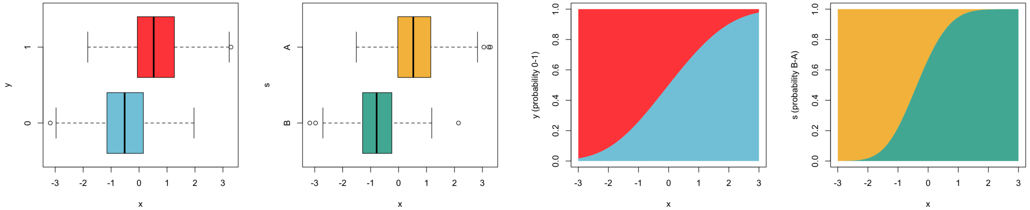
"*Technology is neither good nor bad; nor is it neutral* " , Kranzberg (1986)

# Datasets

- `toydata1`

Consider a confounding Gaussian variable $X_0$, $X_0 \sim \mathcal{N}(0,1)$, and

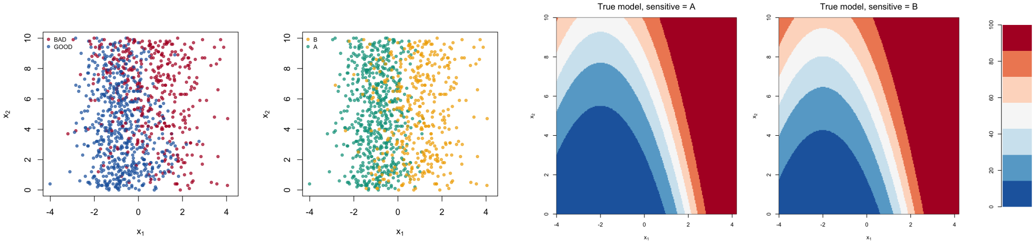
$$\begin{cases} X = X_0 + \epsilon, \;\; \epsilon \sim \mathcal{N}(0, 1/2^2), \\ S = \mathbf{1}(X_0 + \eta > 0), \;\; \eta \sim \mathcal{N}(0, 1/2^2), \;\; s \in \{\mathtt{A}, \mathtt{B}\}, \text{PROBLEME !!!!} \\ Y = \mathbf{1}(X_0 + \nu > 0), \;\; \nu \sim \mathcal{N}(0, 1/2^2), \;\; y \in \{0, 1\}. \end{cases}$$



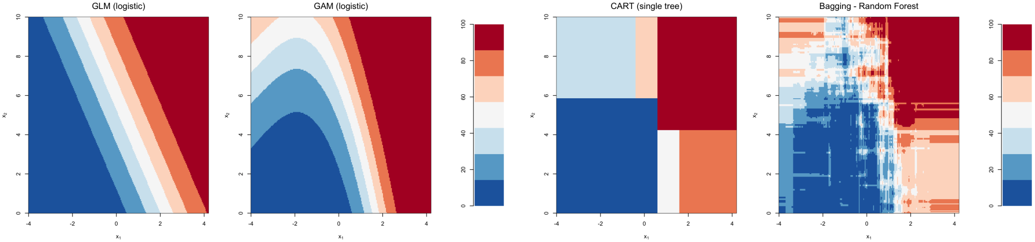

$x \mapsto \mathbb{P}[Y = 0 | X = x]$ (left-hand side) and $x \mapsto \mathbb{P}[S = \mathtt{A} | X = x]$ (right-hand side)

# Datasets

- `toydata2`
  - binary sensitive attribute, $s \in \{\texttt{A}, \texttt{B}\}$, (60% and 40%)
  - $(x_1, x_3) \sim \mathcal{N}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$, $r_{s=\texttt{A}} = 0.4$ and $r_{s=\texttt{B}} = 0.7$
  - $x_2 \sim \mathcal{U}([0, 10])$, independent of $x_1$ and $x_3$
  - $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 \mathbf{1}_{\texttt{B}}(s)$, that does not depend on $x_3$
  - $y \sim \mathcal{B}(p)$ where $p = \exp(\eta)/[1 + \exp(\eta)] = \mu(x_1, x_2, s)$.

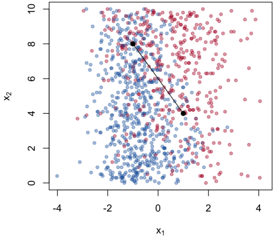


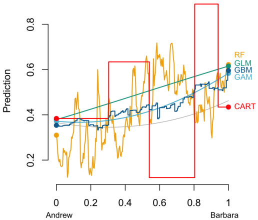
# Datasets




Five models are considered

- plain GLM (logistic)
- GAM (cubic splines)
- CART (classification tree)
- RF (random forest)
- GBM (gradient boosting)



# Datasets

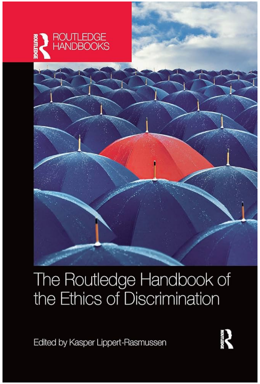
- `GermanCredit`, $m = 1,000$
  - binary sensitive attribute, $s \in \{\mathtt{A}, \mathtt{B}\}$, (64% and 36%) corresponding to gender
  - $y$ denotes a default (30%)
  - $x_1, \cdots, x_k$ denote legitimate credit variables (Duration, Purpose, Credit_amount, Age, Housing, Existing_credits, Foreign_worker, Resident_since, etc)
- `FrenchMotor` (policy observe over one year), $n = 12,437$
  - binary sensitive attribute, $s \in \{\mathtt{A}, \mathtt{B}\}$, (31% and 69%) corresponding to gender
  - $y$ denotes the occurrence of a car accident (8.67%, unbalanced data)
  - $x_1, \cdots, x_k$ denote legitimate credit variables (MariStat, VehAge, SocioCateg, DrivAge, VehBody, VehEnergy, VehMaxSpeed, Garage, VehUsage, etc)

# – Part 1 –

# Insurance

# Discrimination and Insurance

"*What is unique about insurance is that even statistical discrimination which by definition is absent of any malicious intentions, poses significant moral and legal challenges. Why? Because on the one hand, policy makers would like insurers to treat their insureds equally, without discriminating based on race, gender, age, or other characteristics, even if it makes statistical sense to discriminate* (...) *On the other hand, at the core of insurance business lies discrimination between risky and non-risky insureds. But riskiness often statistically correlates with the same characteristics policy makers would like to prohibit insurers from taking into account.*" Avraham (2017)




# Discrimination and Insurance

**Definition 2.2: Mutuality, Wilkie (1997)**

Mutuality is considered as the normal form of commercial private insurance, where participants contribute to the risk pool through a premium that relates to their particular risk at the time of the application, i.e., the higher the risk that they bring to the pool, the higher the premium required.

**Definition 2.3: Solidarity, Wilkie (1997)**

Solidarity is the basis of most national or social insurance schemes. Participation in such state-run schemes is generally compulsory and individuals have no discretion over their level of cover. All participants normally have the same level of cover. In solidarity schemes the contributions are not based on the expected risk of each participant.

# Insurance Pricing and Predictive Modeling

"*Humans think in stories rather than facts, numbers or equations - and the simpler the story, the better*," Harari (2018). For insurers, it is often a mixture of both.

For Glenn (2000), insurer's risk selection process has two sides:
- the one presented to regulators and policyholders (numbers, statistics and objectivity),
- the other presented to underwriters (stories, character and subjective judgment).

The rhetoric of insurance exclusion – numbers, objectivity and statistics – forms what Brian Glenn calls "*the myth of the actuary*," "*a powerful rhetorical situation in which decisions appear to be based on objectively determined criteria when they are also largely based on subjective ones*" or "*the subjective nature of a seemingly objective process*".

Glenn (2003) claimed that there are many ways to rate accurately. Insurers can rate risks in many different ways depending on the stories they tell on which characteristics

# Insurance Pricing and Predictive Modeling

are important and which are not. "*The fact that the selection of risk factors is subjective and contingent upon narratives of risk and responsibility has in the past played a far larger role than whether or not someone with a wood stove is charged higher premiums*." Going further, "*virtually every aspect of the insurance industry is predicated on stories first and then numbers*."

"*all models are wrong but some models are useful*," Box et al. (2011) (in other words, any model is at best a useful fable).

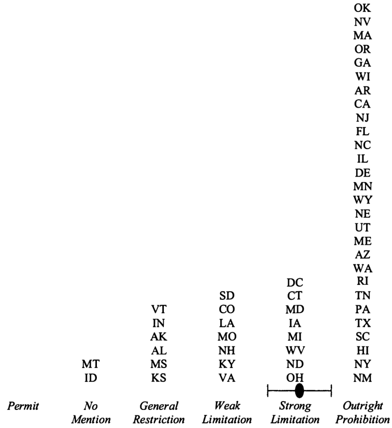
# Insurance Pricing and Predictive Modeling

From Avraham et al. (2013),

- Expressly Permit (-1) - The state has a statute expressly or impliedly permitting insurers to take the characteristic into account.
- No Law on Point (0) - The state laws are silent with respect to the particular characteristic.
- General Restriction (1) - The state has a statute that generally prohibits "unfair discrimination," either across all lines of insurance or in some lines of insurance, but that statute does not provide any explanation as to what constitutes unfair discrimination and does not single out any particular trait for limitation.

(...)

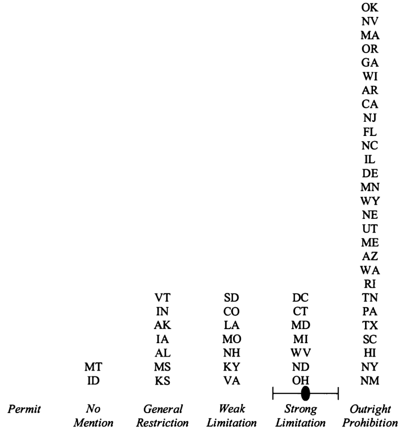
FIGURE 1a. Distribution of States' Scores for Race, in Auto Insurance



# Insurance Pricing and Predictive Modeling

- **Characteristic-Specific Weak Limitation (2)** - The state has a statute that limits the use of a particular characteristic in either issuance, renewal, or cancellation.
- **Characteristic-Specific Strong Limitation (3)** - The state has a statute that prohibits the use of a particular characteristic when the policy is either issued, renewed, or cancelled, or the state has a statute that limits but does not completely prohibit the use of a particular characteristic in rate setting.
- **Characteristic-Specific Prohibition (4)** - The state has a statute the expressly prohibits insurers from taking into account a specific characteristic in setting rates.

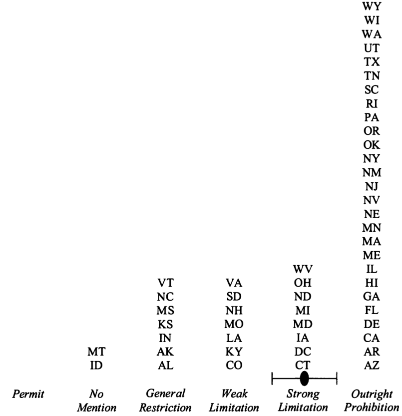
FIGURE 1b. Distribution of States' Scores for National Origin, in Auto Insurance



# Insurance Pricing and Predictive Modeling

"*Race, national origin, and religion have a special place in this country's history; and, as discussed above, discrimination on the basis of these three characteristics has been subject to stricter scrutiny in American law than have other characteristics*," Avraham et al. (2013)

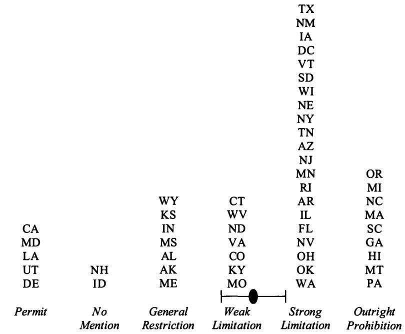
FIGURE 1c. Distribution of States' Scores for Religion, in Auto Insurance



# Insurance Pricing and Predictive Modeling

"*Gender-based discrimination in insurance has long been controversial. And differential treatment on the basis of gender is, of course, in many contexts widely considered unacceptable or illegal. Nevertheless, there does not seem to be the same level of agreement-as there is for race, religion, and national origin-that drawing gender-based distinctions is always wrong. Federal constitutional law treats gender as only a quasi-suspect classification; as a result, laws that discriminate on the basis of gender are subject to an intermediate level of scrutiny.*" Avraham et al. (2013)

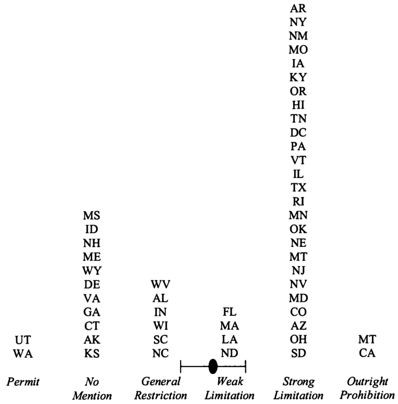
FIGURE 3a. Distribution of States' Scores for Gender, in Auto Insurance



# Insurance Pricing and Predictive Modeling

"*With respect to life insurance, we predict that the laws regulating gender discrimination will be on average relatively weak, since adverse selection in the life insurance market is especially problematic.*" Avraham et al. (2013)

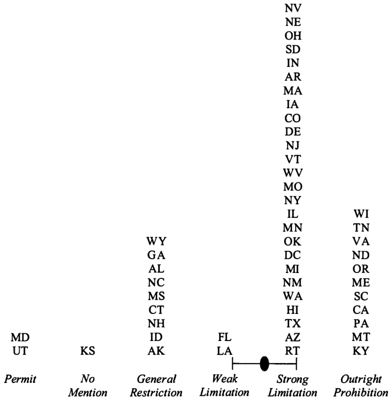
FIGURE 3c. Distribution of States' Scores for Gender, in Disability Insurance



# Insurance Pricing and Predictive Modeling

"*Regarding property/casualty insurance, as there seems to be no conceivable correlation between those risks and gender, we predict either states will cluster around no regulation, or, alternatively, states will cluster around forbidding the use of gender in property/casualty insurance on symbolic or expressive grounds.*" Avraham et al. (2013)

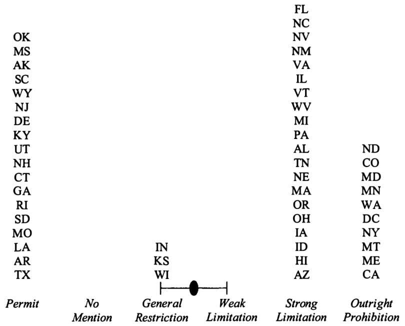
FIGURE 3d. Distribution of States' Scores for Gender, in Property/Casualty Insurance



# Insurance Pricing and Predictive Modeling

"*The gender discrimination will be more strictly regulated on average for health insurance (where gender-rated policies often result in higher premiums for women) than for auto insurance (where gender-rated policies result in higher premiums for men).*" Avraham et al. (2013)

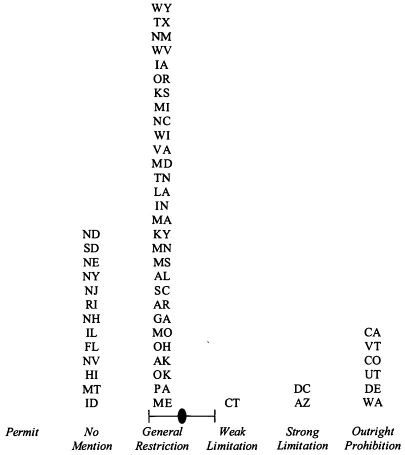
FIGURE 3e. Distribution of States' Scores for Gender, in Health Insurance



# Insurance Pricing and Predictive Modeling

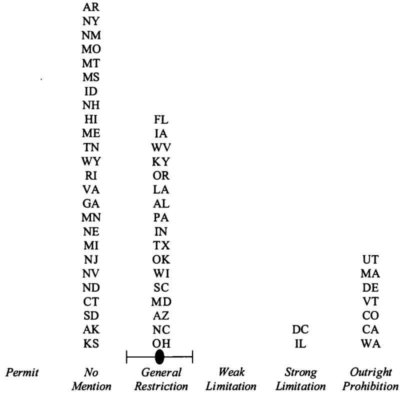
"*Unlike with race, national origin, religion, and gender, legal classifications on the basis of an individual's sexual orientation have not clearly been identified by the Supreme Court as deserving special scrutiny. In addition, unlike race, national origin, and gender, there are no federal laws forbidding discrimination on the basis of sexual orientation in employment.*" Avraham et al. (2013)

FIGURE 4a. Distribution of States' Scores for Sexual Orientation, in Auto Insurance



"*However, there are state laws that forbid discrimination on the basis of sexual orientation, and some lower courts have held that sexual orientation should be a suspect or quasi-suspect characterisation.*" Avraham et al. (2013)

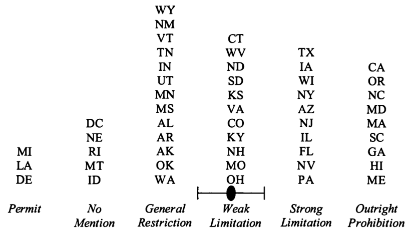
FIGURE 4c. Distribution of States' Scores for Sexual Orientation, in Disability Insurance



# Insurance Pricing and Predictive Modeling

"*We expect that* age *will have the lowest average regulatory score of all the risk characteristics we are studying. First, age is not a suspect classification, at least not by constitutional standards. Second, age tends to correlate causally with several important areas of risk (mortality, health, and perhaps disability risks), thereby increasing the perceived fairness of rating on that basis.*" Avraham et al. (2013)

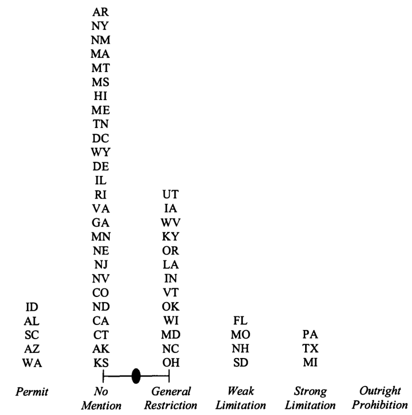
FIGURE 5a. Distribution of States' Scores for Age, in Auto Insurance



# Insurance Pricing and Predictive Modeling

*"Third, age can present serious adverse selection problems for insurers if they are forbidden from taking it into account, since individual insureds know their own age and the associated risks. Fourth, social solidarity arguments with respect to age are relatively weak, since individuals can spread risk over their lifetime through various income smoothing products."* Avraham et al. (2013)

FIGURE 5c. Distribution of States' Scores for Age, in Disability Insurance



# Insurance Pricing and Predictive Modeling

Avraham et al. (2013) suggested to visualize the distribution of scores (Expressly Permit (-1) / No Law on Point (0) / General Restriction (1) / $\cdots$ / Characteristic-Specific Prohibition (4))

FIGURE 6. Distribution of States' Scores for Age, by Line of Insurance

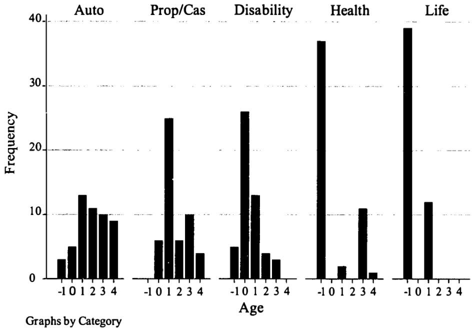
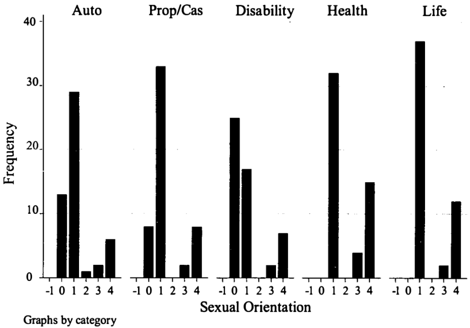



FIGURE 7. Distribution of States' Scores for Sexual Orientation, by Line of Insurance



# Insurance Pricing and Predictive Modeling

FIGURE 8. Distribution of States' Scores for Zip Code, by Line of Insurance

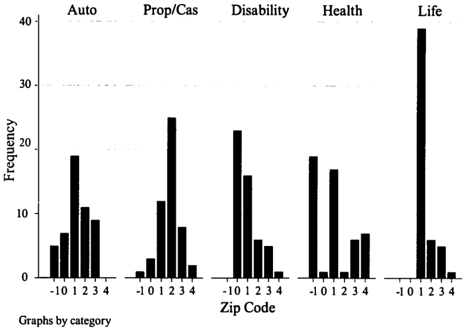
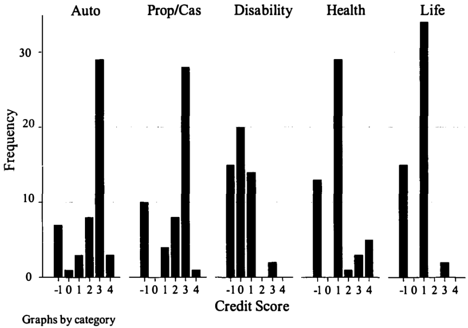



FIGURE 9. Distribution of States' Scores for Credit Score, by Line of Insurance




"*Credit score and zip code are not, by themselves, socially suspect characteristics. However, some commentators have argued that credit score and zip code are used by auto and home insurers as proxies for potentially socially suspect characteristics.*"

**Definition 3.1: Pure premium (homogeneous risks)**

Let $Y$ be the non-negative random variable corresponding to the total annual loss associated with a given policy, then the pure premium is $\mathbb{E}[Y]$.

**Proposition 3.1: Law of Large Numbers (2)**

Consider an infinite collection of i.i.d. random variables $Y, Y_1, Y_2, \cdots, Y_n, \cdots$ in a probabilistic space $(\Omega, \mathcal{F}, \mathbb{P})$, with finite expected value, then

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} Y_i}_{\text{(empirical) average}} \xrightarrow{\text{a.s.}} \underbrace{\mathbb{E}(Y)}_{\text{expected value}}, \text{ as } n \to \infty.$$

# Insurance Pricing and Predictive Modeling

More realistically, population is heterogeneous (with respect to risks), with some covariates $\boldsymbol{x}$ (legitimate, or not).

**Definition 3.2: Pure premium (heterogeneous risks)**

Let $Y$ be the non-negative random variable corresponding to the total annual loss associated with a given policy, with covariates $\boldsymbol{x}$, then the pure premium is $\mu(\boldsymbol{x}) = \mathbb{E}[Y|\boldsymbol{X} = \boldsymbol{x}]$.

In this general setting, $\boldsymbol{x}$ consist in numeric or categorical variables.

**Proposition 3.2: Law of Large Numbers (2')**

Consider an infinite collection of i.i.d. random pairs $(\boldsymbol{X}, Y)$, $(\boldsymbol{X}_1, Y_1)$, $(\boldsymbol{X}_2, Y_2), \cdots, (\boldsymbol{X}_n, Y_n), \cdots$ in a probabilistic space $(\Omega, \mathcal{F}, \mathbb{P})$, with finite expected value, then for any $\mathcal{A} \subset \mathcal{X}$ such that $\mathbb{P}[\boldsymbol{X} \in \mathcal{A}]$,

$$\frac{\displaystyle\sum_{i=1}^{n} Y_i \mathbf{1}(\boldsymbol{X}_i \in \mathcal{A})}{\displaystyle\sum_{i=1}^{n} \mathbf{1}(\boldsymbol{X}_i \in \mathcal{A})} = \underbrace{\frac{1}{n_{\mathcal{A}}} \sum_{i \in \mathcal{I}_n(\mathcal{A})} Y_i}_{\text{conditional average}} \xrightarrow{\text{a.s.}} \underbrace{\mathbb{E}(Y | \boldsymbol{X} \in \mathcal{A})}_{\text{conditional expected value}}, \text{ as } n \to \infty,$$

where $\mathcal{I}_n(\mathcal{A}) = \{i : \boldsymbol{X}_i \in \mathcal{A}\} \subset \{1, 2, \cdots, n\}$ and $n_{\mathcal{A}} = \text{Card}(\mathcal{I}_n(\mathcal{A}))$.

# Insurance Pricing and Predictive Modeling

- Excerpt from the Men and Women life tables in 1720 (source: Struyck (1912)). Mortality, as a function of the age and the gender of the individual.

Table des Hommes.

| Années | Personnes | Années | Personnes | Années | Personnes | Années | Personnes | Années | Personnes | Années | Personnes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 710 | 20 | 607 | 35 | 474 | 50 | 313 | 65 | 142 | 80 | 33 |
| 6 | 697 | 21 | 599 | 36 | 464 | 51 | 301 | 66 | 132 | 81 | 29 |
| 7 | 688 | 22 | 591 | 37 | 454 | 52 | 289 | 67 | 123 | 82 | 25 |
| 8 | 681 | 23 | 583 | 38 | 444 | 53 | 277 | 68 | 114 | 83 | 22 |
| 9 | 675 | 24 | 575 | 39 | 434 | 54 | 265 | 69 | 105 | 84 | 19 |
| 10 | 670 | 25 | 567 | 40 | 424 | 55 | 253 | 70 | 97 | 85 | 16 |
| 11 | 665 | 26 | 558 | 41 | 414 | 56 | 241 | 71 | 89 | 86 | 13 |
| 12 | 660 | 27 | 549 | 42 | 404 | 57 | 229 | 72 | 82 | 87 | 10 |
| 13 | 654 | 28 | 540 | 43 | 393 | 58 | 217 | 73 | 75 | 88 | 8 |
| 14 | 648 | 29 | 531 | 44 | 382 | 59 | 206 | 74 | 68 | 89 | 6 |
| 15 | 642 | 30 | 522 | 45 | 371 | 60 | 195 | 75 | 61 | 90 | 4 |
| 16 | 635 | 31 | 513 | 46 | 360 | 61 | 184 | 76 | 54 | 91 | 3 |
| 17 | 628 | 32 | 504 | 47 | 349 | 62 | 173 | 77 | 48 | 92 | 2 |
| 18 | 621 | 33 | 494 | 48 | 337 | 63 | 162 | 78 | 43 | 93 | 1 |
| 19 | 614 | 34 | 484 | 49 | 325 | 64 | 152 | 79 | 38 | 94 | |

Table des femmes.

| Années | Personnes | Années | Personnes | Années | Personnes | Années | Personnes | Années | Personnes | Années | Personnes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 711 | 20 | 624 | 35 | 508 | 50 | 373 | 65 | 205 | 80 | 55 |
| 6 | 700 | 21 | 617 | 36 | 500 | 51 | 362 | 66 | 194 | 81 | 47 |
| 7 | 692 | 22 | 610 | 37 | 492 | 52 | 351 | 67 | 183 | 82 | 40 |
| 8 | 685 | 23 | 603 | 38 | 484 | 53 | 340 | 68 | 172 | 83 | 34 |
| 9 | 679 | 24 | 596 | 39 | 476 | 54 | 329 | 69 | 161 | 84 | 29 |
| 10 | 674 | 25 | 588 | 40 | 468 | 55 | 318 | 70 | 150 | 85 | 24 |
| 11 | 669 | 26 | 580 | 41 | 459 | 56 | 306 | 71 | 140 | 86 | 20 |
| 12 | 664 | 27 | 572 | 42 | 450 | 57 | 294 | 72 | 130 | 87 | 17 |
| 13 | 660 | 28 | 564 | 43 | 441 | 58 | 282 | 73 | 120 | 88 | 14 |
| 14 | 656 | 29 | 556 | 44 | 432 | 59 | 271 | 74 | 110 | 89 | 11 |
| 15 | 652 | 30 | 548 | 45 | 423 | 60 | 260 | 75 | 100 | 90 | 8 |
| 16 | 647 | 31 | 540 | 46 | 414 | 61 | 249 | 76 | 90 | 91 | 6 |
| 17 | 642 | 32 | 532 | 47 | 404 | 62 | 238 | 77 | 81 | 92 | 4 |
| 18 | 636 | 33 | 524 | 48 | 394 | 63 | 227 | 78 | 72 | 93 | 2 |
| 19 | 630 | 34 | 516 | 49 | 384 | 64 | 216 | 79 | 63 | 94 | 1 |

# Insurance Pricing and Predictive Modeling

- Excerpt from the Men and Women life tables in 1720 (source: Struyck (1912))

Mortality, as a function of the age and the gender of the individual.

| men | | | | | | women | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $x$ | $L_x$ | ${}_5p_x$ | $x$ | $L_x$ | ${}_5p_x$ | $x$ | $L_x$ | ${}_5p_x$ | $x$ | $L_x$ | ${}_5p_x$ |
| 0 | 1000 | 29.0% | 45 | 371 | 16.6% | 0 | 1000 | 28.9% | 45 | 423 | 11.8% |
| 5 | 710 | 5.6% | 50 | 313 | 19.2% | 5 | 711 | 5.2% | 50 | 373 | 14.7% |
| 10 | 670 | 4.2% | 55 | 253 | 22.9% | 10 | 674 | 3.3% | 55 | 318 | 18.2% |
| 15 | 642 | 5.5% | 60 | 195 | 27.2% | 15 | 652 | 4.3% | 60 | 260 | 21.2% |
| 20 | 607 | 6.6% | 65 | 142 | 31.7% | 20 | 624 | 5.8% | 65 | 205 | 26.8% |
| 25 | 567 | 7.9% | 70 | 97 | 37.1% | 25 | 588 | 6.8% | 70 | 150 | 33.3% |
| 30 | 522 | 9.2% | 75 | 61 | 45.9% | 30 | 548 | 7.3% | 75 | 100 | 45.0% |
| 35 | 474 | 10.5% | 80 | 33 | 51.5% | 35 | 508 | 7.9% | 80 | 55 | 56.4% |
| 40 | 424 | 12.5% | 85 | 16 | | 40 | 468 | 9.6% | 85 | 24 | |

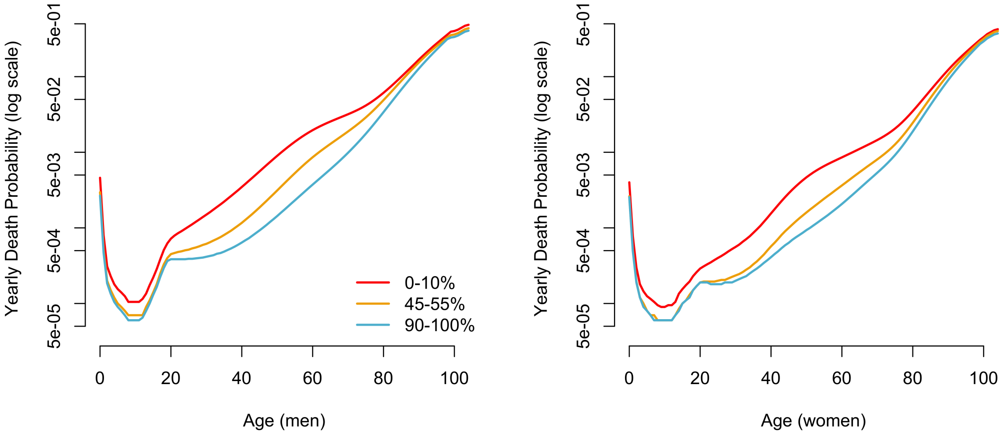
# Insurance Pricing and Predictive Modeling

- Excerpt from the Men and Women life tables in 2016 (source: Blanpain (2018))

Mortality, as a function of the age, the gender and the wealth of the individual.

| men | | | | women | | | |
|---|---|---|---|---|---|---|---|
| $x$ | 0-5% | 45-50% | 95-100% | $x$ | 0-5% | 45-50% | 95-100% |
| 0 | 100000 | 100000 | 100000 | 0 | 100000 | 100000 | 100000 |
| 10 | 99299 | 99566 | 99619 | 10 | 99385 | 99608 | 99623 |
| 20 | 99024 | 99396 | 99469 | 20 | 99227 | 99506 | 99526 |
| 30 | 97930 | 98878 | 99094 | 30 | 98814 | 99302 | 99340 |
| 40 | 95595 | 98058 | 98627 | 40 | 97893 | 98960 | 99074 |
| 50 | 90031 | 96172 | 97757 | 50 | 95021 | 97959 | 98472 |
| 60 | 77943 | 91050 | 95649 | 60 | 88786 | 95543 | 97192 |
| 70 | 59824 | 79805 | 90399 | 70 | 79037 | 90408 | 94146 |
| 80 | 38548 | 59103 | 76115 | 80 | 63224 | 79117 | 85825 |
| 90 | 13337 | 23526 | 38837 | 90 | 31190 | 45750 | 55918 |
| 100 | 530 | 1308 | 3231 | 100 | 2935 | 5433 | 8717 |

# Insurance Pricing and Predictive Modeling




Force of mortality (log scale) for various income quantile, in France, Blanpain (2018).

# Insurance Pricing and Predictive Modeling

U.S. DECENNIAL LIFE TABLES FOR 1969-71

Volume I, Number 1

NCHS

## United States Life Tables: 1969-71


1. Life table for the total population: United States, 1969-71 ---- 6
2. Life table for males: United States, 1969-71 ---- 8
3. Life table for females: United States, 1969-71 ---- 10
4. Life table for the white population: United States, 1969-71 ---- 12
5. Life table for white males: United States, 1969-71 ---- 14
6. Life table for white females: United States, 1969-71 ---- 16
7. Life table for the population other than white: United States, 1969-71 ---- 18
8. Life table for males other than white: United States, 1969-71 ---- 20
9. Life table for females other than white: United States, 1969-71 ---- 22
10. Life table for the Negro population: United States, 1969-71 ---- 24


TABLE 10. LIFE TABLE FOR THE NEGRO POPULATION: UNITED STATES, 1969-71

| AGE INTERVAL | PROPORTION DYING | OF 100,000 BORN ALIVE | | STATIONARY POPULATION | | AVERAGE REMAINING LIFETIME |
|---|---|---|---|---|---|---|
| PERIOD OF LIFE BETWEEN TWO AGES | PROPORTION OF PERSONS ALIVE AT BEGINNING OF AGE INTERVAL DYING DURING INTERVAL | NUMBER LIVING AT BEGINNING OF AGE INTERVAL | NUMBER DYING DURING AGE INTERVAL | IN THE AGE INTERVAL | IN THIS AND ALL SUBSEQUENT AGE INTERVALS | AVERAGE NUMBER OF YEARS OF LIFE REMAINING AT BEGINNING OF AGE INTERVAL |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| $x$ to $x+t$ | ${}_tq_x$ | $l_x$ | ${}_td_x$ | ${}_tL_x$ | $T_x$ | $\mathring{e}_x$ |
| DAYS | | | | | | |
| 0-1 | 0.01348 | 100,000 | 1,348 | 272 | 6,411,264 | 64.11 |
| 1-7 | .00668 | 98,652 | 659 | 1,616 | 6,410,992 | 64.99 |
| 7-28 | .00254 | 97,993 | 249 | 5,631 | 6,409,376 | 65.41 |
| 28-365 | .01037 | 97,744 | 1,013 | 89,778 | 6,403,745 | 65.52 |

Mortality, gender and “race”

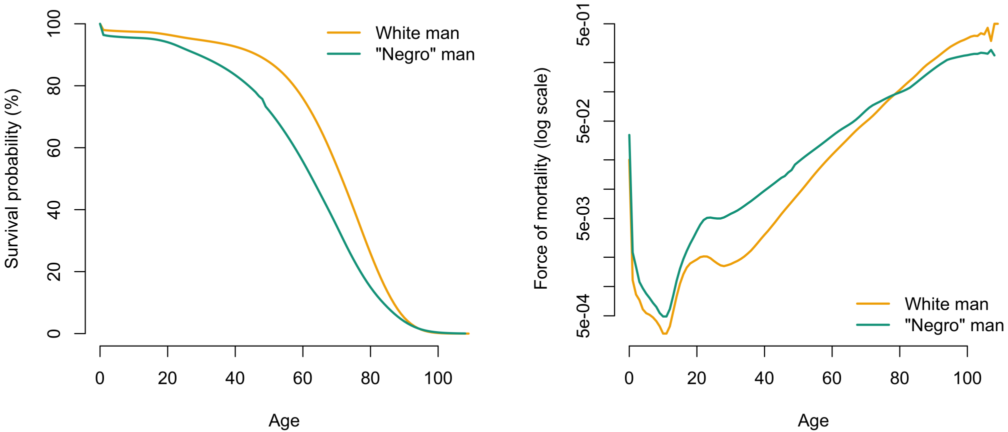
Frederick L. Hoffman

Hoffman (1896, 1918, 1931)

# Insurance Pricing and Predictive Modeling

| White, men | | | | | | "Negro", men | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $x$ | $L_x$ | ${}_5p_x$ | $x$ | $L_x$ | ${}_5p_x$ | $x$ | $L_x$ | ${}_5p_x$ | $x$ | $L_x$ | ${}_5p_x$ |
| 0 | 100000 | 2.3% | 55 | 83001 | 8.5% | 0 | 100000 | 4.2% | 55 | 66101 | 13.1% |
| 5 | 97671 | 0.2% | 60 | 75969 | 12.7% | 5 | 95826 | 0.3% | 60 | 57457 | 17.4% |
| 10 | 97441 | 0.2% | 65 | 66343 | 18.4% | 10 | 95497 | 0.4% | 65 | 47485 | 22.2% |
| 15 | 97208 | 0.7% | 70 | 54138 | 25.5% | 15 | 95161 | 1.2% | 70 | 36925 | 29.8% |
| 20 | 96480 | 1.0% | 75 | 40324 | 35.8% | 20 | 94053 | 2.3% | 75 | 25921 | 36.1% |
| 25 | 95524 | 0.8% | 80 | 25885 | 47.7% | 25 | 91904 | 2.5% | 80 | 16560 | 41.7% |
| 30 | 94716 | 0.9% | 85 | 13527 | 62.1% | 30 | 89584 | 3.0% | 85 | 9648 | 51.3% |
| 35 | 93843 | 1.3% | 90 | 5125 | 75.1% | 35 | 86885 | 4.0% | 90 | 4696 | 63.4% |
| 40 | 92631 | 2.1% | 95 | 1274 | 85.2% | 40 | 83441 | 5.4% | 95 | 1721 | 71.6% |
| 45 | 90725 | 3.3% | 100 | 189 | 90.5% | 45 | 78976 | 7.2% | 100 | 489 | 74.8% |
| 50 | 87690 | 5.3% | 105 | 18 | 100.0% | 50 | 73282 | 9.8% | 105 | 123 | 100.0% |

# Insurance Pricing and Predictive Modeling




Force of mortality (log scale) white men and "Negro" men, 1968-71, U.S.

**Definition 3.3: Balance Property**

A pricing function $m$ satisfies the balance property if $\mathbb{E}_{\boldsymbol{X}}[m(\boldsymbol{X})] = \mathbb{E}_Y[Y]$.

**Proposition 3.3: Law of total expectations**

$$\mathbb{E}_Y[Y] = \mathbb{E}_{\boldsymbol{X}}\big[\mathbb{E}_{Y|\boldsymbol{X}}[Y|\boldsymbol{X}]\big] = \mathbb{E}_{\boldsymbol{X}}\big[\mu(\boldsymbol{X})\big].$$

**Proof** Since $\mathbb{E}(Y) = \displaystyle\int y f_y(y)\mathrm{d}y$ and $\mathbb{E}(Y|\boldsymbol{X}=\boldsymbol{x}) = \displaystyle\int y f_{y|\boldsymbol{x}}(y|\boldsymbol{x})\mathrm{d}y$,

$$\mathbb{E}(\mathbb{E}(X|Y)) = \int\left(\int x\mathbb{P}[X=x|Y=y]\mathrm{d}x\right)\mathbb{P}[Y=y]\mathrm{d}y = \int\int x\mathbb{P}[X=x,Y=y]\mathrm{d}x\mathrm{d}y$$

$$= \int x\left(\int \mathbb{P}[X=x,Y=y]\mathrm{d}y\right)\mathrm{d}x = \int x\mathbb{P}[X=x]\mathrm{d}x = \mathbb{E}(X).$$

# Insurance Pricing and Predictive Modeling

Homogeneous risk sharing

| | Policyholder | Insurer |
|---|---|---|
| Loss | $\mathbb{E}[Y]$ | $Y - \mathbb{E}[Y]$ |
| Average loss | $\mathbb{E}[Y]$ | 0 |
| Variance | 0 | $\text{Var}[Y]$ |

$\mathbb{E}[Y]$ is the premium paid, and $Y$ the total loss,
from De Wit and Van Eeghen (1984) and Denuit and Charpentier (2004)

# Insurance Pricing and Predictive Modeling

## Heterogeneous risk sharing, with perfect information

| | Policyholder | Insurer |
|---|---|---|
| Loss | $\mathbb{E}[Y\|\Theta]$ | $Y-\mathbb{E}[Y\|\Theta]$ |
| Average loss | $\mathbb{E}[Y]$ | $0$ |
| Variance | $\text{Var}[\mathbb{E}[Y\|\Theta]]$ | $\text{Var}[Y-\mathbb{E}[Y\|\Theta]]$ |

where $\Theta$ denotes the heterogeneous risk factor.

The term on the bottom right is $\mathbb{E}[\text{Var}[Y|\Theta]]$, corresponding to the standard variance decomposition (or Pythagoras theorem)

$$\text{Var}[Y]=\text{Var}[\mathbb{E}[Y|\Theta]]+\mathbb{E}[\text{Var}[Y|\Theta]].$$

to go further ➔ (for more details on Lebesgue spaces, and $L^2$)

**Proposition 3.4: Variance decomposition (1)**

For any measurable random variable $Y$ with finite variance

$$\text{Var}[Y] = \underbrace{\mathbb{E}[\text{Var}[Y|\Theta]]}_{\to \text{ insurer}} + \underbrace{\text{Var}[\mathbb{E}[Y|\Theta]]}_{\to \text{ policyholder}}.$$

**Proof**:

$$\begin{aligned}
\text{Var}[Y] &= \mathbb{E}\left[Y^2\right] - \text{E}[Y]^2 = \mathbb{E}\left[\text{Var}[Y|\Theta] + \mathbb{E}[Y|\Theta]^2\right] - \mathbb{E}[\mathbb{E}[Y|\Theta]]^2 \\
&= (\mathbb{E}[\text{Var}[Y|\Theta]]) + \left(\mathbb{E}\left[\mathbb{E}[Y|\Theta]^2\right] - \mathbb{E}[\mathbb{E}[Y|\Theta]]^2\right) = \mathbb{E}[\text{Var}[Y|\Theta]] + \text{Var}[\mathbb{E}[Y|\Theta]].
\end{aligned}$$

# Insurance Pricing and Predictive Modeling

Heterogeneous risk sharing, with imperfect information

| | Policyholder | Insurer |
|---|---|---|
| Loss | $\mathbb{E}[Y\vert\boldsymbol{X}]$ | $Y-\mathbb{E}[Y\vert\boldsymbol{X}]$ |
| Average loss | $\mathbb{E}[Y]$ | $0$ |
| Variance | $\text{Var}[\mathbb{E}[Y\vert\boldsymbol{X}]]$ | $\mathbb{E}[\text{Var}[Y\vert\boldsymbol{X}]]$ |

$$\mathbb{E}[\text{Var}[Y|\boldsymbol{X}]] = \underbrace{\mathbb{E}[\text{Var}[Y|\Theta]]}_{\text{perfect ratemaking}} + \underbrace{\mathbb{E}\{\text{Var}[\mathbb{E}[Y|\Theta]|\boldsymbol{X}]\}}_{\text{misclassification}}$$

This "misclassification" term (on the right) is called "*subsidierende solidariteit*" in De Pril and Dhaene (1996), or "*subsidiary solidarity*", as opposed to "*kanssolidariteit*" or "*random solidarity*" term (on the left).

**Proposition 3.5: Variance decomposition (2)**

For any measurable random variable $Y$ with finite variance

$$\text{Var}[Y] = \underbrace{\mathbb{E}[\text{Var}[Y|\boldsymbol{X}]]}_{\to \text{ insurer}} + \underbrace{\text{Var}[\mathbb{E}[Y|\boldsymbol{X}]]}_{\to \text{ policyholder}},$$

where

$$\begin{aligned} \mathbb{E}[\text{Var}[Y|\boldsymbol{X}]] &= \mathbb{E}[\mathbb{E}[\text{Var}[Y|\Theta]|\boldsymbol{X}]] + \mathbb{E}[\text{Var}[\mathbb{E}[Y|\Theta]|\boldsymbol{X}]] \\ &= \underbrace{\mathbb{E}[\text{Var}[Y|\Theta]]}_{\text{perfect ratemaking}} + \underbrace{\mathbb{E}\{\text{Var}[\mathbb{E}[Y|\Theta]|\boldsymbol{X}]\}}_{\text{misclassification}}. \end{aligned}$$

# References

Andrus, M., Spitzer, E., Brown, J., and Xiang, A. (2021). What we can't measure, we can't understand: Challenges to demographic data procurement in the pursuit of fairness. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 249–260.

Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals and it's biased against blacks. *ProPublica*, May 23.

Avraham, R. (2017). Discrimination and insurance. In Lippert-Rasmussen, K., editor, *Handbook of the Ethics of Discrimination*, pages 335–347. Routledge.

Avraham, R., Logue, K. D., and Schwarcz, D. (2013). Understanding insurance antidiscrimination law. *Southern California Law Review*, 87:195.

Barry, L. and Charpentier, A. (2020). Personalization as a promise: Can big data change the practice of insurance? *Big Data & Society*, 7(1):2053951720935143.

Bickel, P. J., Hammel, E. A., and O'Connell, J. W. (1975). Sex bias in graduate admissions: Data from Berkeley. *Science*, 187(4175):398–404.

Blanpain, N. (2018). L'espérance de vie par niveau de vie-méthode et principaux résultats. *INSEE Document de Travail*, F1801.

Box, G. E., Luceño, A., and del Carmen Paniagua-Quinones, M. (2011). *Statistical control by monitoring and adjustment*, volume 700. John Wiley & Sons.

# References


Charpentier, A. (2023). *Insurance: biases, discrimination and fairness*. Springer Verlag.

Crossney, K. B. (2016). Redlining. *https://philadelphiaencyclopedia.org/essays/redlining/*.

De Baere, G. and Goessens, E. (2011). Gender differentiation in insurance contracts after the judgment in case c-236/09, Association Belge des Consommateurs Test-Achats asbl v. conseil des ministres. *Colum. J. Eur. L.*, 18:339.

de La Fontaine, J. (1668). *Fables*. Barbin.

De Pril, N. and Dhaene, J. (1996). Segmentering in verzekeringen. *DTEW Research Report 9648*, pages 1–56.

De Wit, G. and Van Eeghen, J. (1984). Rate making and society's sense of fairness. *ASTIN Bulletin: The Journal of the IAA*, 14(2):151–163.

Denuit, M. and Charpentier, A. (2004). *Mathématiques de l'assurance non-vie: Tome I Principes fondamentaux de théorie du risque*. Economica.

Dieterich, W., Mendoza, C., and Brennan, T. (2016). Compas risk scales: Demonstrating accuracy equity and predictive parity. *Northpointe Inc*, 7(7.4):1.

Dressel, J. and Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science advances*, 4(1):eaao5580.

# References

Elliott, M. N., Morrison, P. A., Fremont, A., McCaffrey, D. F., Pantoja, P., and Lurie, N. (2009). Using the census bureau's surname list to improve estimates of race/ethnicity and associated disparities. *Health Services and Outcomes Research Methodology*, 9:69–83.

Feeley, M. and Simon, J. (1994). Actuarial justice: The emerging new criminal law. *The futures of criminology*, 173:174.

Feller, A., Pierson, E., Corbett-Davies, S., and Goel, S. (2016). A computer program used for bail and sentencing decisions was labeled biased against blacks. it's actually not that clear. *The Washington Post*, October 17.

Glenn, B. J. (2000). The shifting rhetoric of insurance denial. *Law and Society Review*, pages 779–808.

Glenn, B. J. (2003). Postmodernism: the basis of insurance. *Risk Management and Insurance Review*, 6(2):131–143.

Harari, Y. N. (2018). *21 Lessons for the 21st Century*. Random House.

Hoffman, F. L. (1896). *Race traits and tendencies of the American Negro*, volume 11. American Economic Association.

Hoffman, F. L. (1918). *Mortality from respiratory diseases in dusty trades (inorganic dusts)*. Number 231. US Government Printing Office.

Hoffman, F. L. (1931). Cancer and smoking habits. *Annals of surgery*, 93(1):50.

# References


Kearns, M. and Roth, A. (2019). *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press.

Kranzberg, M. (1986). Technology and history:" kranzberg's laws". *Technology and culture*, 27(3):544–560.

Larson, J., Mattu, S., Kirchner, L., and Angwin, J. (2016). How we analyzed the compas recidivism algorithm. *ProPublica*, 23-05.

Li, K. C.-W. (1996). The private insurance industry's tactics against suspected homosexuals: redlining based on occupation, residence and marital status. *American Journal of Law & Medicine*, 22(4):477–502.

Mcdonald, S. (2015). Indirect gender discrimination and the 'test-achats ruling': an examination of the uk motor insurance market. In *Royal Economic Society Conf., Manchester*.

Merriam-Webster (2022). *Dictionary*. .

Perry, W. L. (2013). *Predictive policing: The role of crime forecasting in law enforcement operations*. Rand Corporation.

Rhynhart, R. (2020). Mapping the legacy of structural racism in philadelphia. *Philadelphia, Office pf the Controller*.

# References


Schanze, E. (2013). Injustice by generalization: notes on the Test-Achats decision of the european court of justice. *German Law Journal*, 14(2):423–433.

Schauer, F. (2006). *Profiles, probabilities, and stereotypes*. Harvard University Press.

Squires, G. D. (2003). Racial profiling, insurance style: Insurance redlining and the uneven development of metropolitan areas. *Journal of Urban Affairs*, 25(4):391–410.

Squires, G. D. and Velez, W. (1988). Insurance redlining and the process of discrimination. *The Review of Black Political Economy*, 16(3):63–75.

Struyck, N. (1912). *Les oeuvres de Nicolas Struyck (1687-1769): qui se rapportent au calcul des chances, à la statistique général, la statistique des décès et aux rentes viagèter*. Société générale néerlandaise d'assurances sur la vie et de rentes viagères.

The Zebra (2022). Car insurance rating factors by state. *https://www.thezebra.com/*.

Van Gerven, G. (1993). Case c-109/91, Gerardus Cornelis Ten Oever v. Stichting bedrijfspensioenfonds voor het glazenwassers-en schoonmaakbedrijf. *EUR-Lex*, 61991CC0109.

Wilkie, D. (1997). Mutuality and solidarity: assessing risks and sharing losses. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1357):1039–1044.