Risk Weighting

The best thing since One Hot Encoding?

Laurel Ruhlen
Data Science @ Braintree
github.com/lruhlen

Agenda

- What is risk-weighting?
- Simple example
- Code
- Pros, cons, and algebra

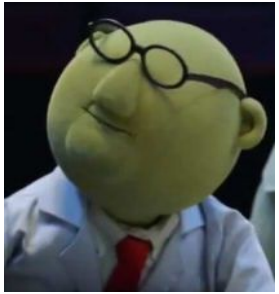
What is this & why do we care?

- Because: categorical variables
  - With high ordinality

What is this & why do we care?

```
Kermit = [“green”, 64, 4]
```

For example









Monsters?





Monsters?

Monsters	Total
2	4
2	4
3	4

Monsters?

Color	Monsters	Total	Risk of being a monster
Green	2	4	0.5
Red	2	4	0.5
Blue	3	4	0.75

Monsters?

~~Kermit = [“green”, 64, 4]~~

Kermit = [0.5, 64, 4]

As Code

```
data = pd.read_csv("example_data.csv").set_index("name")
```

name	color	age	weight	is_monster
Kermit The Frog	green	64	4	False
Dr. Bunsen Honeydew	green	47	8	False
Karli	green	2	3	True
Oscar The Grouch	green	57	12	True
Beaker	red	40	4	False
Pepe The King Prawn	red	30	1	False
Animal	red	60	7	True
Elmo	red	35	4	True
Sam The Eagle	blue	70	12	False
Gonzo	blue	67	5	True
Grover	blue	55	5	True
Cookie Monster	blue	63	20	True

As Code

```
feature_names=["color",]
target_name="is_monster"

rwt = RiskWeightTransformer(feature_names, target_name=target_name)
rwt.fit(data)

data["color_monster_risk"] = rwt.transform(data)
```

name	color_monster_risk	age	weight	is_monster
Kermit The Frog	0.50	64	4	False
Dr. Bunsen Honeydew	0.50	47	8	False
Karli	0.50	2	3	True
Oscar The Grouch	0.50	57	12	True
Beaker	0.50	40	4	False
Pepe The King Prawn	0.50	30	1	False
Animal	0.50	60	7	True
Elmo	0.50	35	4	True
Sam The Eagle	0.75	70	12	False
Gonzo	0.75	67	5	True
Grover	0.75	55	5	True
Cookie Monster	0.75	63	20	True

Why Not OHE?

	Is green	Is blue	Is purple	Is orange	Is red	Is yellow	Is pink	Is beige	...
	0	0	0	1	0	0	0	0	0
	0	1	0	0	0	0	0	0	0

- 2 equations (rows)
- 8 unknowns (columns)
- Can’t solve!

Why Not OHE?

$$y = 2x + 1$$

When To Use

Risk Weighting	**One Hot Encoding**
High ordinality	Low ordinality
Supervised learning	Unsupervised learning
True/False outcomes	3+ outcomes

Let's be friends!

- https://github.com/lruhlen
- @yelling_at_computers on ChiPy Slack!