MTH3035 Mathematics group project
you need to be in Exeter in person for this module
[redacted]
thursday 10am ossice hour
)
ele choice by today
5% group project plan (written)
5% (2) short group presentations in class
10% final (60 mins) group presentation
75% final written report
5% individual self assessment (January)
Adjustment for low or high contribution
eek 6 reading week

week	date	Monday 10:30	tuesday 1:30
2	27 Sept	Project plan lecture	no class
3	4 Oct	Presentation lecture	no class
4	11 Oct	group presentation 1	group presentation 1

project Management skills, teamwork, Communications

include MTH3035 in 2 subject line
teams ossice hour meeting Tuesday 10-11
doesnt decrease marks for doing bad or being nervous
be understanding of different people & their cultures

already have project plan title
critical path analysis
risk analysis: how can things go wrong? how likely?
getting covid uns an impact more than 1 (because hospital)
room being double book is low prob but high impact
need to interview people involved in this industry like
who makes these predictions professionally
can mitigate for laptop battery while bringing charger

Strategy is making Money? work skills? academic or personal?
are you set up to fail?
blaming others makes you look bad
not responding to communications
making up excuses looks bad
how will you get your team to force yourself to contribute
if someone is not in your team, then advisor, then mark
upload by 12:00 Friday 8th October ele or email
each group only hand in once

nowadays use 16:9 aspect ratio
use transitions sparingly so use it once but not everytime
dont use cliparts
make sure light doesnt wash it out
aim, audience, medium, content, style
keep it short & simple
how confident are you write?
assume less knowledge in the audience
inaudible, invisible, too much material, unclear message, poor explanation, run out of time or finish too early, awful body language (embarrassing silence), poor response to questions
about 1-2 min per slide
what is key message? can you simplify it?
start sentence sos not starts with "in my opinion"
cite sources
dont have to present in order which it happen
use visuals choose most important
avoid irrelevant material attention grabber (one wants to be rich)

dont use comic-sans
use the entire slide - but don't be ridiculous
not ok to right out a complete script
can leave optional extra material to after
make sure not double booked
project voice but don't shout
speak to the audience
drink water before & after speaking
keep an eye on the time - leave enough time for conclusion
explain everything carefully
repeat & even rephrase questions / answer the audience
can paraphrase question to make them open
don't have to answer immediately ask for clarification
conference posters?
be prepared for everything - make a plan
follow up by emails
tuesday 8A 3rd block4 Max 4 Mins

dont produce graphs that produce odd shape in so that does not exist ("chart junk")
pie charts have a place not necessarily in Maths
www.wisecheck.com to check if it works best for everyone
maybe our bar chart should be a standard bar chart or even maybe a pie chart (probably a bad idea)
to publish you need to add labels maybe even labels on the graph instead of a key
can be a key of green heart danger & red rose
no rainbow keys

Aston villa 2-1

United city 2-0

Brentford Norwich 3-0

Chelsea Burnley 0-1

Crystal wolves 0-2

Brighton vs Newcastle 1-1

Arsenal Walsall 2-0

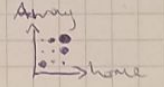
Leed Leicester 3-2

eve tot 0-3

~~Brighton~~ ~~Everton~~

West ham Lives 1-0

MTH Mathematics group project - Statistical forcasting

- Target - who will win
- Lead time - predict one day ahead, one week, whole season before it starts, how long? whole season? all next ~~match~~ lead / horizon
- type of forcasts - exact single value like 23°C is a point forecast
- range/interval forcast str like first goal within 30 to 40 mins
probability/probabilistic forecast - probs for different results happen - like prob of rain is 35%.
probs add upto 100% so win & not win or can do win/lose/draw
represent as pdfs for normal distri as a continuous distri
two teams can affect each other 

discrete 2-Dimensional
poisson is best for number of goal for one team in 1D
pois(2.1)
home advantage - league positioning; current form
keyplayer missing - history of results
covariance and correlation, but better to just assume home and away goals are independent
strong team does not affect weak team
being away on a neutral does affect performance
is Exeter ~ pois(1.3) Warstall(1.1) ∴
Pr(Exeter score=2) = 0.22 Pr(Warstall score 1) = 0.18
if independence $f(x,y) = f(x)f(y)$ is true for all numbers
then independence

Dissertation, thesours

- how many player are still playing - the bookmaker windows - how relevant, so too far back is less relevant and introduces bias

but too small data introduces noise outliers even if data is only the most relevant

- when do you throw away data - do you grow your window or shift it along

measure performance - how to tell which prediction is better?

- what do other people do to forecast - professionals like betting companies

- keep things really simple at the start

use whole season so accurate for whole years seasons

good at fitting data but not good at predicting

scoring rules on hypothesis tests to tests its fitting & prediction

need more than Chi-Squared

data should come with estimated standard error like a confidence interval

min said predict most likely outcome

we need to do reading of what other people have done

ability score - how well they've done in the last 5 games

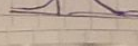
dixon and colls statistical model of football results

just because it fits well doesn't mean it predicts real : don't over fit

: use cross-validation

proper scoring rules → Brier score ← simplest → logarithmic → ranked probab score

Brier score - Binary do $(f-x)^2$ for each forecast then find average

log arithmic score : $-\log[f(x)]$  we need to do more than one

marked probable score & logarithmic score for football use logarithmic
use ranked stats like Z median but not Z mean

can compare true ranking scored by hypothesis test t-test
can check for covariance with Dixon & Coles (1995)
dixon coles on reading list
want to be fighting about 20 sources

first line of raw poisson code use away averages

very final part of SP code weights average and both with 50-50 weighting but we can change the weighting to 30-70 then do a logarithmic or Brier score to compare which is best

southampton only has 2 game away in reset both but arsenal has 3

in weighting calculation it means home games doesn't Models expected away average

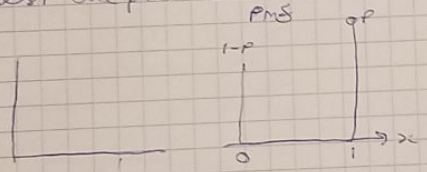
would be give 50% weighting for away-Models expected for 3 home & only 2 away games

for Fulham home vs Man city Away
Model predicts: most probable score 0-1 ∴ away win most probable
with probs of home win Draw Away win:
H: 0.1012704 D: 0.2014161 A: 0.6972599
True result: Away win ∴ & True Results: (H:0, D:0, A:1)
∴ $-(0\times\log_2(0.1012704)+0\times\log_2(0.2014161)+1\times\log_2(0.6972599)=$
$-\log_2(0.6972599)=0.5202\approx 0.52$ is Log Scoring
∴ do that for every game and add up all the log scoring values
for that Model to get the Logarithmic Scoring Results
then just find out which one has the lowest one

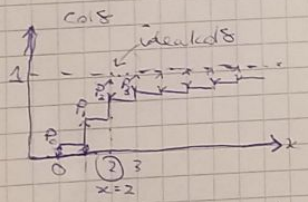
Southampton	SOU
Arsenal	ARS
Aston Villa	AST
Brighton	BHA
Burnley	BUR
Chelsea	CHE
Crystal Palace	CRY
Everton	EVE
Leeds	LEE
Leicestar	LEI
Liverpool	LIV
Man City	MCI
Man United	MUN
Newcastle	NEW
Tottenham	TOT
West Brom	WBA
West Ham	WHU
Wolves	WOL
Fulham	FUL
Sheffield United	SHU
Norwich	NOR
Brentford	BRE

2 good models are kept very sensitive so cross validation is hard

with test the pro Models on our data to compare to our Models



Brier score is $(P-x)^2$



asymptotic to one

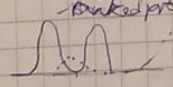
$\therefore \sum_{j \geq 0} (P_j - I(x=j))^2 = rps$

still works is: home-draw-away can be ordered

sensitive to dissonance

	H	D	A
F_1	0.4	0.2	0.4
F_2	0.4	0.3	0.3

So log score is not sensitive to dissonance so would weight equal but this is sensetive to dissonance so says F_2 may better



HT ~ Poi(k)

AT ~ Poi(p)

Scored Arsenal, Barca

Conceded Arsenal, Barca

home win	Draw	Away W	H?	D?	A?
α	β	γ	r	s	t

i) $\alpha, \beta, \gamma \in \mathbb{R}_{\geq 0}$, $\alpha+\beta+\gamma = 1$ $\quad 0 \leq \alpha, \beta, \gamma, \leq 1$

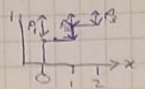
~~only one~~ $r, s, t = (0 \text{ or } 1)$ only one of r, s, t can be ~~1~~ at once. ∴

if $r=1 \Rightarrow s, t = 0$

if $s=1 \Rightarrow r, t = 0$

if $t=1 \Rightarrow r, s = 0$

eg if $r=1$: ~~$rps = (1-\alpha)^2$~~



$\therefore \sum_{j=0} (p_j - I(x=j))^2 = rps$

∴ $p_0 = 0.54$ $\quad \alpha = 0.54$ $\quad \beta = 0.24$ $\quad \gamma = 0.21$

i) ∴ $(1-\alpha) = 0.46$ $\quad (1-\beta) = 0.76$ $\quad (1-\gamma) = 0.79$

$\alpha+\beta = 0.78$ $\quad (1-0.78) = 0.22$

$\alpha+\beta+\gamma = 1$ $\quad 1-(\alpha+\beta+\gamma) = 0$

$I(x=j)$ is 0 or 1 ∴ for $r=1$ ∴ $x=0$ ∴

$I(0=0) = 1$ and $I(0=1) = 1$, $I(0=2) = 1$ ∴ do

$(0.46-1)^2 + (0.22-1)^2 + (0-1)^2 = rps = 1.9$ want rps to be as low as possible

∴

~~if $\alpha = 0.11$, $\beta = 0.22$, $\gamma = 0.65$, $t=1$ ∴ $x=2$ ∴~~

~~$1-1=0$ ∴ $I(x=0) = 0$ $I(x=1) = 0$ $I(x=2) = 1$ ∴~~

~~$(0.11-0)^2 + (0.22-0)^2 + (0-1)^2 = rps = 1.0605$~~

if $\alpha = 0.35$, $\beta = 0.28$, $\gamma = 0.35$, $t=1$ ∴ $x=2$

∴ $1-1=0$ $\quad I(x=0) = 0$ $\quad I(x=1) = 0$ $\quad I(x=2) = 1$ ∴

~~$(0.35)^2 + (0.28-0)^2 + (0$~~ $\alpha+\beta = 0.63$ ∴

$(0.35-0)^2 + (0.63-0)^2 + (0-1)^2 = rps = 1.5194$

if $\alpha = 0.11$, $\beta = 0.22$, $\gamma = 0.65$, $t=1$ ∴ $x=2$ ∴

$I(x=0) = 0$, $I(x=1) = 0$, $I(x=2) = 1$, $1-1 = 0$, $\alpha+\beta = 0.33$ ∴

$(0.11-0)^2 + (0.33-0)^2 + (0-1)^2 = rps = 1.121$

$1.121 < 1.5194$ which is expected for a better model

if $\alpha = 0.40$, $\beta = 0.26$, $\gamma = 0.35$, $S = 1$ $\therefore$ $x = 1$ $\therefore$

$I(x \le 0) = 0$, $I(x \le 1) = 1$, $I(x \le 2) = 1$ $\therefore$ $1 - 1 = 0$

$\alpha + \beta = 0.40 + 0.26 = 0.66$ $\quad 1 - 0.66 = 0.34$ $\therefore$

$(0.40 - 0)^2 + (0.34 - 1)^2 + (0 - 1)^2 = rps = 1.5756$ is high as expected

Equivalent: $(0.40 - 0)^2 + (0.66)^2 + 1 = 1.5756$

Equivalent: $(0.40 - I(x \le 0))^2 + (\alpha + \beta)^2 + 1$ for $S = 1$: $(\alpha - H)^2 + (\alpha + \beta)^2 + 1$

for $r = 1$: ~~or~~ if: $\alpha = 0.54$, $\beta = 0.24$, $\gamma = 0.21$, $r = 1$ $\therefore$ $x = 0$ $\therefore$

$I(x \le 0) = 1$, $I(x \le 1) = 1$, $I(x \le 2) = 1$ $\therefore$ $1 - 1 = 0$

$1 - \alpha = 0.46$, $\alpha + \beta = 0.54 + 0.24 = 0.78$ $\therefore$ $1 - (\alpha + \beta) = 1 - 0.78 = 0.22$ $\therefore$

$(0.46 - 1)^2 + (1 - (\alpha + \beta) - 1)^2 + (0 - 1)^2 = 1.90$

$(0.46 - 1)^2 + (0.78)^2 + 1 = 1.9$

$(\alpha)^2 + (\alpha + \beta)^2 + 1 = 1.9$

for $S = 1$: $H = 0$ $\therefore$ $(\alpha)^2 + (\alpha + \beta)^2 + 1 = rps$

for $r = 1$ $(\alpha)^2 + (\alpha + \beta)^2 + 1 = rps$

for $t = 1$: $\alpha = 0.11$, $\beta = 0.22$, $\gamma = 0.65$ $\therefore$ $t = 1$ $\therefore$ $x = 2$ $\therefore$

$I(x \le 0) = 0$, $I(x \le 1) = 0$, $I(x \le 2) = 1$ $\therefore$ $1 - 1 = 0$ $\therefore$ $\alpha + \beta = 0.33$

$(\alpha - I(x \le 0))^2 + ((\alpha + \beta) - I(x \le 1))^2 + (0 - 1)^2 =$

$(\alpha - 0)^2 + ((\alpha + \beta) - 0)^2 + 1 =$

$\alpha^2 + (\alpha + \beta)^2 + 1 = 0.11^2 + (0.33)^2 + 1 = 1.121$

if $r = 1$ $\alpha = 0.6$, $\beta = 0.24$, $\gamma = 0.16$

$\alpha^2 + (\alpha + \beta)^2 + 1 = (0.6)^2 + (0.6 + 0.24)^2 + 1 = rps = 2.0656$

$$rps = [Pr(\text{Home W})]^2 + [Pr(\text{Home W}) + Pr(\text{Draw})]^2 + 1$$

$$rps = [Pr(\text{Home W})]^2 + [Pr(\text{Home W}) + Pr(\text{Draw})]^2 + [Pr(\text{Home W}) + Pr(\text{Draw}) + Pr(\text{Away W})]^2$$

Say $Pr(\text{Home win}) = 0.4 = P_1$, $Pr(\text{Draw}) = (0.26) = P_2$, $Pr(\text{Away win}) = 0.34 = P_3$

true result Draw:

$\therefore H=0, D=1, A=0 \therefore 0.4+0.26=0.46$, $0.4+0.26+0.34=1$

$\therefore \frac{1}{2}\left[(P_1 - H)^2 + (P_1 - H + P_2 - D)^2 + (P_1 - H + P_2 - D + P_3 - A)^2\right] =$

$\frac{1}{2}\left[(0.4-0)^2 + (0.4-0+0.26-1)^2 + (0.4-0+0.26-1-0.34-0)^2\right] =$

$\frac{1}{2}\left[(0.4-H)^2 + (0.66-1)^2 + (1-1)^2\right] =$

$\frac{1}{2}\left[(0.4-H)^2 + (0.66-1)^2\right] = \frac{1}{2}\left[(0.4-0)^2 + (0.66-1)^2\right] = 0.1378$

If $D=1$, $Pr(\text{Home W}) = 0.3 = P_1$ $Pr(\text{Draw}) = 0.25 = P_2$ $Pr(\text{Away W}) = 0.45 = P_3$ $\therefore$

$H=0, D=1, A=0 \therefore 0.3+0.25 = 0.55$, $0.3+0.25+0.55 = 1$ $\therefore$

$\left[(P_1 - H)^2 + (P_1 - H + P_2 - D)^2 + (P_1 - H + P_2 - D + P_3 - A)^2\right] =$

$\frac{1}{2}\left[(0.3-0)^2 + (0.3-0+0.25-1)^2 + (0.3-0+0.25-1+0.45-0)^2\right] =$

$\frac{1}{2}\left[(0.3-0)^2 + (0.55-1)^2 + (1-1)^2\right] =$

$\frac{1}{2}\left[(0.3-0)^2 + (0.55-1)^2\right] = 0.1+025$

If $H=1$, $Pr(\text{Home W}) = 0.54 = P_1$, $Pr(\text{Draw}) = 0.24 = P_2$, $Pr(\text{Away W}) = 0.22 = P_3$

$\therefore H=1, D=0, A=0 \therefore 0.54+0.24 = 0.78$ $\therefore 0.54+0.24+0.22 = 1$ $\therefore$

$\frac{1}{2}\left[(P_1 - H)^2 + (P_1 - H + P_2 - D)^2 + (P_1 - H + P_2 - D + P_3 - A)^2\right] =$

$\frac{1}{2}\left[(0.54-1)^2 + (0.54-1+0.24-0)^2 + (0.54-1+0.24-0+0.22-0)^2\right] =$

$\frac{1}{2}\left[(0.54-1)^2 + (0.78-1)^2 + (1-1)^2\right] =$

$\frac{1}{2}\left[(0.54-1)^2 + (0.78-1)^2\right] = 0.13$ {pretty good since H wins with Prob 0.54}

If $A=1$, $Pr(\text{Home W}) = 0.11 = P_1$, $Pr(\text{Draw}) = 0.23 = P_2$, $Pr(\text{Away W}) = 0.66 = P_3$ $\therefore$

$H=0, D=0, A=1 \therefore 0.11+0.23 = 0.34$, $0.11+0.23+0.66 = 1$ $\therefore$

$\frac{1}{2}\left[(P_1 - H)^2 + (P_1 - H + P_2 - D)^2 + (P_1 - H + P_2 - D + P_3 - A)^2\right] =$

$\frac{1}{2}\left[(0.11-0)^2 + (0.11-0+0.23-0)^2 + (0.11-0+0.23-0+0.66-1)^2\right] =$

$\frac{1}{2}\left[(0.11-0)^2 + (0.34-0)^2 + (1-1)^2\right] =$

$\frac{1}{2}\left[(0.11-0)^2 + (0.34-0)^2\right] = 0.06385$ {very good; A with prob 0.66}

$\therefore$

$\therefore$ Formula: $\frac{1}{2}\left[(P_1 - H)^2 + (P_1 - H + P_2 - D)^2\right]$

We tried to improve Z first model by factoring in recent form, that is weighting Z average goals in Z previous 5 games more heavily than less recent results. we calculated recent form as a teams average goals from Z last 5 games, ignoring whether they were played at home or away. We tried weighting form at 25, 50 and 75% however all 3 attempts had worse results than Z original model (where w=0). we believe this is due to Z fact that ignoring whether Z team was playing home or away when calculating recent form was adding more noise than accuracy to Z predictions.

we next tried calculating recent form separately for home & away games, using a formula as before, but considering only home or away games. we again tested for 25, 50 and 75% weighting of recent form, & found that 25% gave Z best predictions, & was an improvement on Z first model. Calculating form seperately for home & away Z noise induced when we calculated it as a single variable.

we tried again to improve our model further by including conceded goals by Z opponent to our third model. We attempted this by adding goals conceded by Z away teams to Z home teams expected goal number, & vice versa. this would then mean our model took into account Z opponent Z teams were facing. we added two factors for conceded goals, (conceded average goals, & conceded form, both relative to home & away matches.

as previously stated, we added Z conceded goals factor to our previous model, leading to our expected goal numbers. Calculated like so. we weighted all 4 factors equally, & also 12.5% weightings on forms, as that worked well with model 3. Z results showed that although both weightings were good, adding this factor didn't change things too much.

we tried to improve Z first model by factoring in recent form, that is weighting Z average goals in Z previous 5 games more heavily than less recent results. we calculated recent form as a teams average goals from Z last 5 games, ignoring whether they were played at home or away. We tried weighting form at 25, 50 & 75% however all 3 attempts had worse results than Z original model (where w=0). we believe this is due to Z fact that ignoring whether Z team was playing home or away when calculating recent form was adding more noise than accuracy to Z predictions.

we next tried calculating recent form separately for home & away games, using a similar formula as before, but considering only home or away games. we again tested for 25, 50 & 75% weightings of recent form, & found that 25% gave Z best predictions, & was an improvement on Z first model. Calculating form separately for home & away removes Z noise induced when we calculated it as a single variable.

we tried again to improve our model further by including conceded goals by Z opponent to our 3rd model. we attempted this by adding goals conceded by Z away team into Z home teams expected goal number & vice versa. This would then mean Z opponent Z teams were would matter & be taken into account by Z model. we added in two factors for conceded goals, conceded average goals, & conceded form, both relative to home & away matches.

as previously stated, we added Z conceded goals factors to our previous model, leading to our expected goal numbers Calculated like so, we weighted all 4 factors equally, & also for 12.5% weightings on forms, as that worked well with model 3. Z results showed that although both weightings good, adding this factor didn't change this too much

- we tried to improve Model 3 by including conceded goals by the opponent
- do this by adding goal conceded by away team to home teams expected and vice versa.
- this means our Model takes into account the opponents
- Conceded goals was made as two factors: Conceded away goals, & conceded home
- both relative to home and away

~~not the improved~~ → how we tried to improve →
how we implemented this improvement →
why this is an improvement →
Specifics of the improvement
extra Specifics

we tried to improve our Model further by adding conceded goals. that is adding the conceded goals of the away team to the home teams expected goals and vice versa. this means the Model takes into account what the opponent is. we did this relative to home and away

as previously stated we added the conceded goals factors to our previous Model, leading to our expected goal numbers. Calculated like so, we weighted ~~both~~ factors equally, & also for 12.5% weightings on terms, as that worked well with Model 3. Results showed that although both weightings were good, adding this factor didn't change things too much