Chapter 1 - Examples

Conor Houghton

reading Gelman

(2022-01-27) bayesianreadinggroup.github.io

Football data

http://www.stat.columbia.edu/∼gelman/book/data/

home	favorite	underdog	spread	name1	name2	week
1	21	13	2.0	TB	MIN	1
1	27	0	9.5	ATL	NO	1
1	31	0	4.0	BUF	NYJ	1
1	9	16	4.0	CHI	GB	1
1	27	21	4.5	CIN	SEA	1
0	26	10	2.0	DAL	WAS	1
1	24	17	5.0	DET	SF	1
1	20	27	6.0	LAN	HOU	1
0	20	7	1.0	MIA	PHX	1

(where name1 and name2 have been changed to fit)

Football data - some sed-foo

```
sed 's/\s\+/,/g' football_data.txt
 | sed 's/,//'> football_data.csv
```

giving:

```
home,favorite,underdog,spread,name1,name2,week
1,21,13,2.0,TB,MIN,1
1,27,0,9.5,ATL,NO,1
1,31,0,4.0,BUF,NYJ,1
1,9,16,4.0,CHI,GB,1
1,27,21,4.5,CIN,SEA,1
0,26,10,2.0,DAL,WAS,1
1,24,17,5.0,DET,SF,1
1,20,27,6.0,LAN,HOU,1
0,20,7,1.0,MIA,PHX,1
```

Load DataFrame

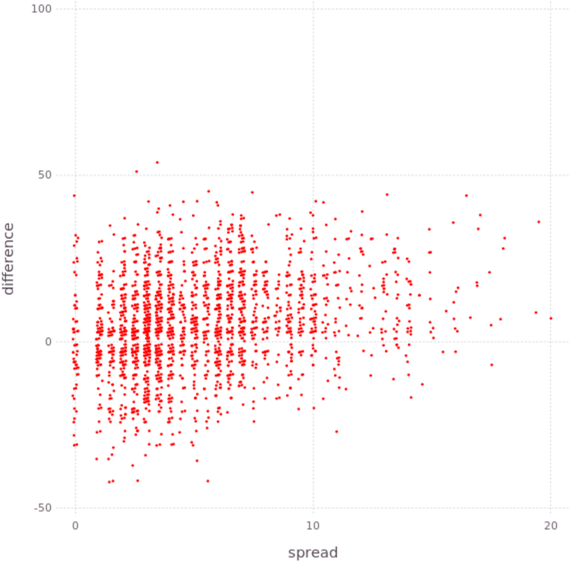
```
using CSV, DataFrames

fD = DataFrame(CSV.File("football_data.csv"))

fD.diff  = fD.favorite - fD.underdog
fD.error = fD.spread   - fD.diff
```

(variable names shortened to fit; long names ftw)

Make the scatter plot

```
using Gadfly,Cairo,Fontconfig

plt=plot(footballData,x=:spread,y=:difference,
         Theme(default_color="red",
               point_size=1pt,
               background_color="white",
               highlight_width=0pt),
         Stat.x_jitter(range=0.5),
         Stat.y_jitter(range=0.5),
         Geom.point,
         Coord.Cartesian(xmin=-0.5,xmax=20.5)
         )
```

Make the scatter plot



Are integer and half-integer spreads the same

```
fD.integerSpread=round.(fD.spread).==fD.spread

mD = groupby(fD, :integerSpread)
mD = combine(mD, nrow, :error => mean => :mean)
```

Are integer and half-integer spreads the same

```
fD.integerSpread=round.(fD.spread).==fD.spread

mD = groupby(fD, :integerSpread)
mD = combine(mD, nrow, :error => mean => :mean)
```

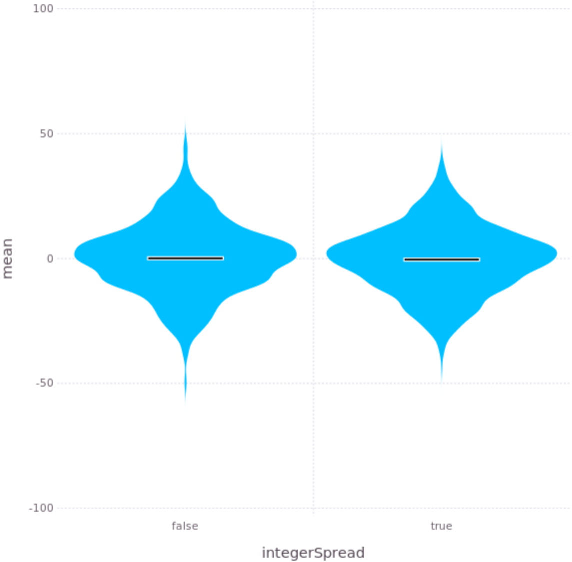
Are integer and half-integer spreads the same

```
fD.intSpread=round.(fD.spread).==fD.spread

mD = groupby(fD, :intSpread)
mD = combine(mD, nrow, :error => mean => :mean)
```

Are integer and half-integer spreads the same

```
layer1=layer(fD,x=:intSpread,y=:error,Geom.violin)
layer2=[stuff with mD to get mean bars]

plt=plot(layer2,layer1,
        Theme(background_color="white")
        )
```

Are integer and half-integer spreads the same



Model

$$\text{difference} = \text{score} + \xi$$

where

$$\xi \sim \text{Normal}(0, \sigma^2)$$

and $\sigma = 14$. This implies that a win happens with probability

$$\Pr(\text{difference} > 0) = \Pr(\xi < \text{score})$$

which can be easily found by integrating the Gaussian and gives an error function up to the usual messing with the root of two.

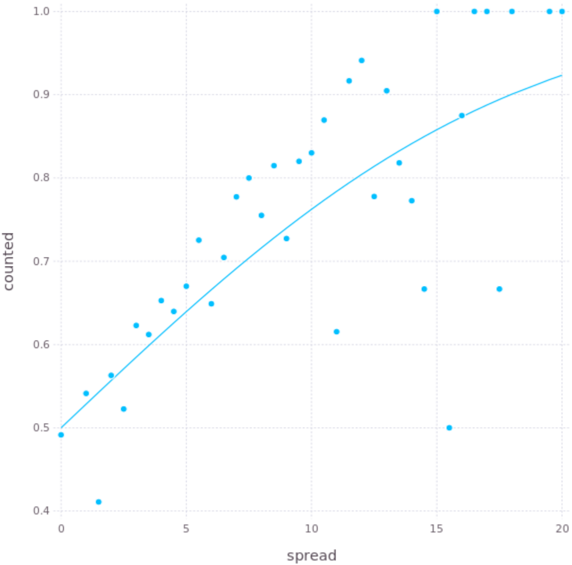
Check predictions

```
model(spread,s) = 0.5+0.5erf(spread /(sqrt(2)*s))

w(a,b)= if(a>b) 1.0 elseif(a<b) 0.0 else 0.5 end
```

Check predictions

```
fD.win = w.(fD.favorite,fD.underdog)

wins = groupby(fD, :spread)
wins = combine(wins,nrow,:win => sum => :totalWin)

s=14.0

wins.counted = wins.totalWin ./ wins.nrow
wins.predicted = modelResult.(wins.spread,s)
```

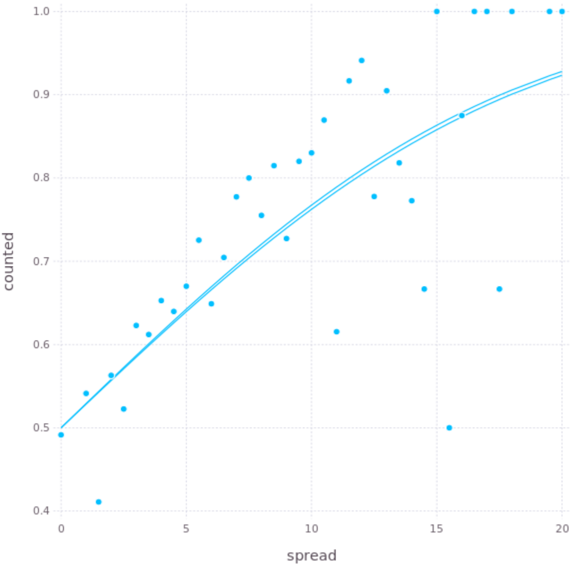
Check predictions



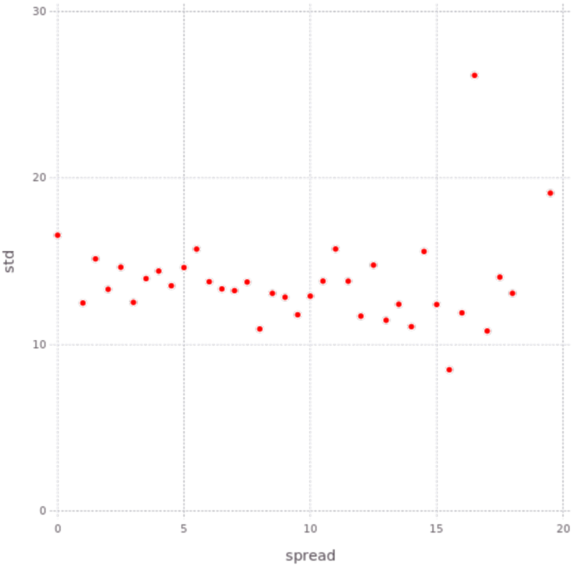
Check predictions

```
fD.win = score.(fD.favorite,fD.underdog)

wins = groupby(fD, :spread)
wins = combine(wins,nrow,:win => sum => :totalWin)

s=std(footballData.error)

wins.counted = wins.totalWin ./ wins.nrow
wins.predicted = modelResult.(wins.spread,s)
```

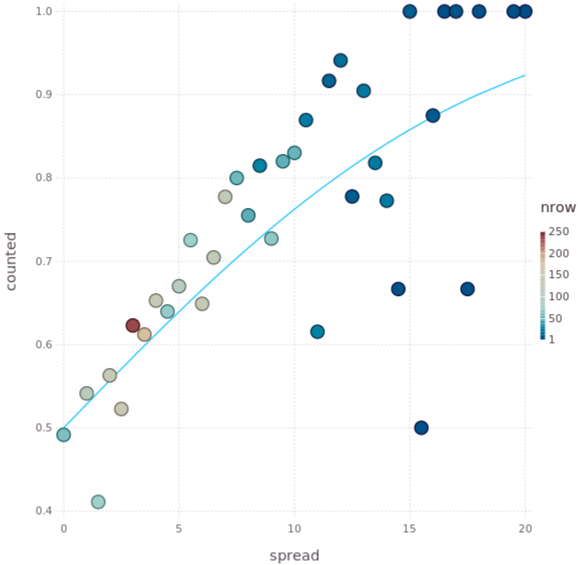
Check predictions



Check predictions - sigma



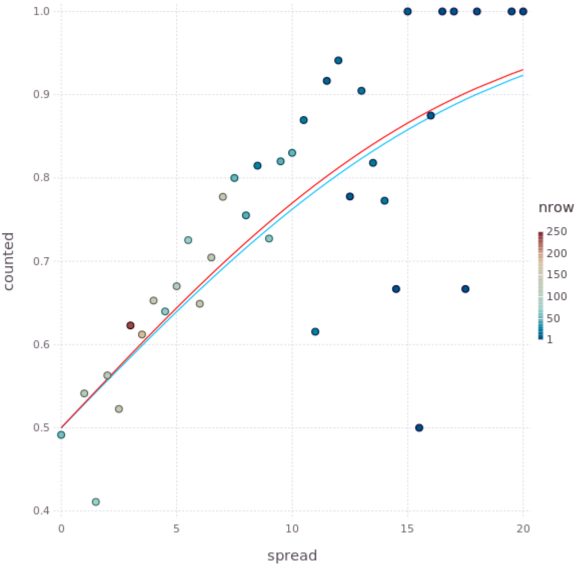
Check predictions - color for data size



Filter for data size

```
cutOff=100
cut=wins.spread[findall(x -> x>cutOff,wins.nrow)]
isGood(x) = x in cut

model = std(filter(:spread => isGood,fD).error)
```

Check predictions - new sigma



Regression

```
using GLM

fm = @formula(difference ~ spread)
linearRegressor = lm(fm, footballData)
```

Regression

```
difference ~ 1 + spread

Coefficients:
-----------------------
                Coef.
-----------------------
(Intercept)  0.152845
spread       1.0139
-----------------------
```

Another question

8. **Subjective probability**: discuss the following statement. 'The probability of event E is considered "subjective" if two rational persons A and B can assign unequal probabilities to E, $P_A(E)$ and $P_B(E)$. These probabilities can also be interpreted as "conditional":

$$P_A(E) = P(E|I_A)$$

and

$$P_B(E) = P(E|I_B)$$

where I_A and I_B represent the knowledge available to persons A and B, respectively.' Apply this idea to the following examples.

a The probability that a '6' appears when a fair die is rolled, where A observes the outcome of the die roll and B does not.

b The probability that Brazil wins the next World Cup, where A is ignorant of soccer and B is a knowledgeable sports fan.