MODELAGEM E SIMULAÇÃO
PPGI/UNINOVE

Henrique Pougy
RA 622150028

GOLF PUTTING

Modelo conceitual

ENTENDIMENTO DO PROBLEMA

O dono de um mini-golfe notou que a taxa de retorno de seus novos clientes é baixa. Após pesquisas qualitativas, percebeu que a principal queixa é devido à **dificuldade** do jogo para clientes novatos.

RESTRIÇÕES

- No entanto, ele sabe que se reduzir excessivamente a dificuldade, perderá também seus clientes regulares.
- É preciso que o jogo permaneça competitivo.
- Ele não tem recursos para criar duas pistas separadas com dificuldades distintas.

OBJETIVO

- Reduzir a dificuldade do jogo para que os jogadores mais inábeis tenham a mesma probabilidade, em média, de fazer pontos (acertar o putting) do que os jogadores com habilidade média.

GOLF PUTTING

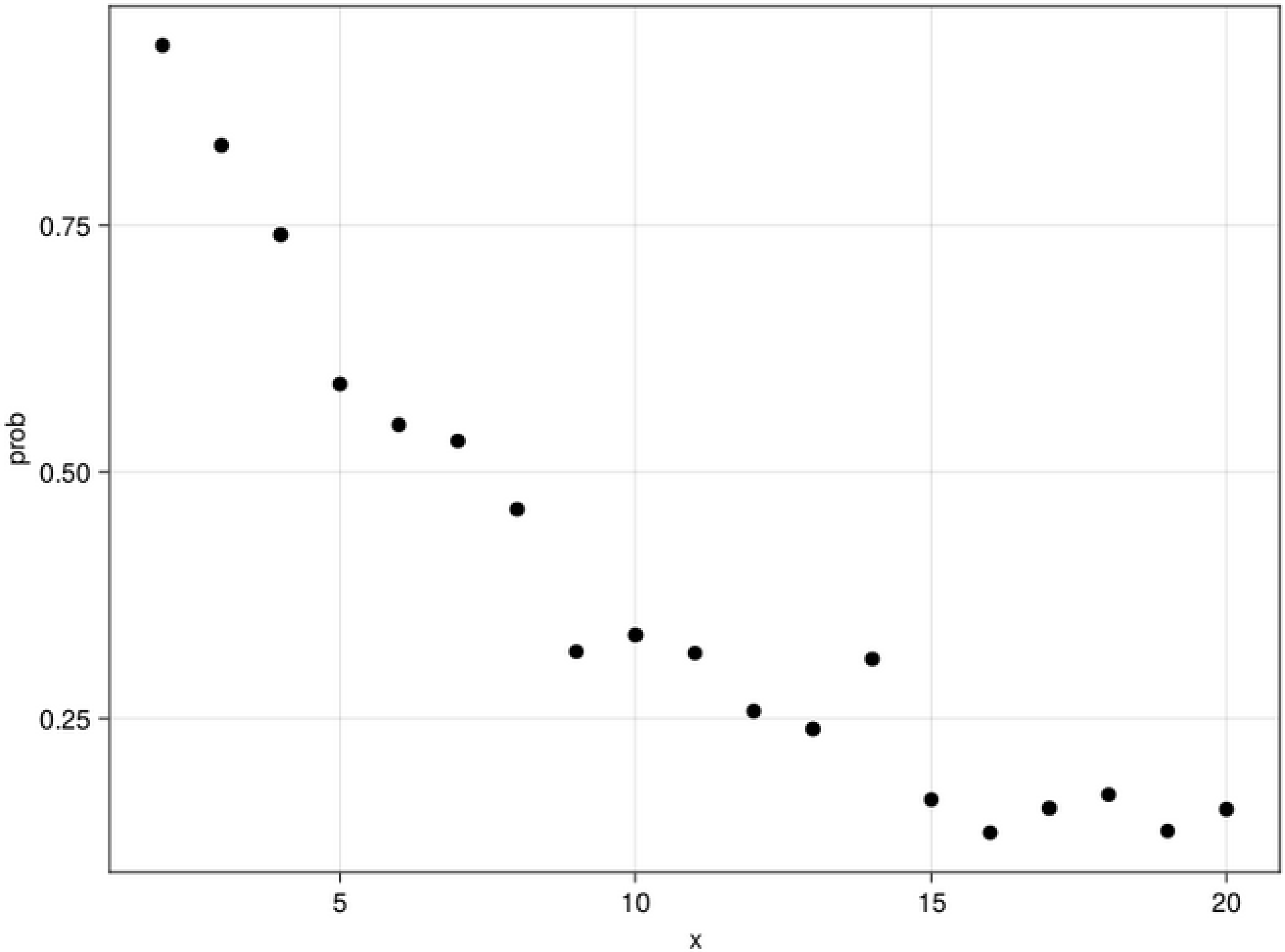
Dados

DADOS DISPONÍVEIS

- Dados sobre putting contendo:
  - X: a distância da bola até o centro do buraco em pés;
  - N: a quantidade total de puttings feita daquela distância;
  - Y: a quantidade de acertos de puttings feitos daquela distância;
- Sabemos também:
  - O diâmetro da bola em polegadas;
  - O diâmetro do buraco em polegadas;

DADOS DISPONÍVEIS

X (distância em pés)	N (nº tentativas)	Y (nº acertos)
2	1443	1346
3	694	577
7	256	136
...	...	...
20	152	24



ACERTO VS DISTÂNCIA

Probabilidade de acerto:

Qtd. Acertos/

Qtd. Tentativas

GOLF PUTTING

Modelagem Conceitual

MODELAGEM CONCEITUAL

- Para uma mesma distância, cada tentativa (putting) é independente das precedentes e das seguintes:
  - P(Acerto [i-1]| distância) ⊥ P(Acerto[i]| distância)

- As amostras são independentes se condicionadas aos parâmetros do modelo (distância X)

$$P(y_i \mid y_{i-1}, D = d) = P(y_i | D = d)$$

MODELAGEM CONCEITUAL

- Para cada distância X[i] e quantidade de tentativas N[i], a quantidade de acertos Y[i] pode ser modelada como uma distribuição binomial.

  - Y[i] ~ Binomial(N[i], P[i])

- Temos que encontrar uma função de ligação que permita modelar P[i] para cada X[i]: ou seja, que defina a probabilidade de acerto em função da distância.

GOLF PUTTING

Modelo V.1

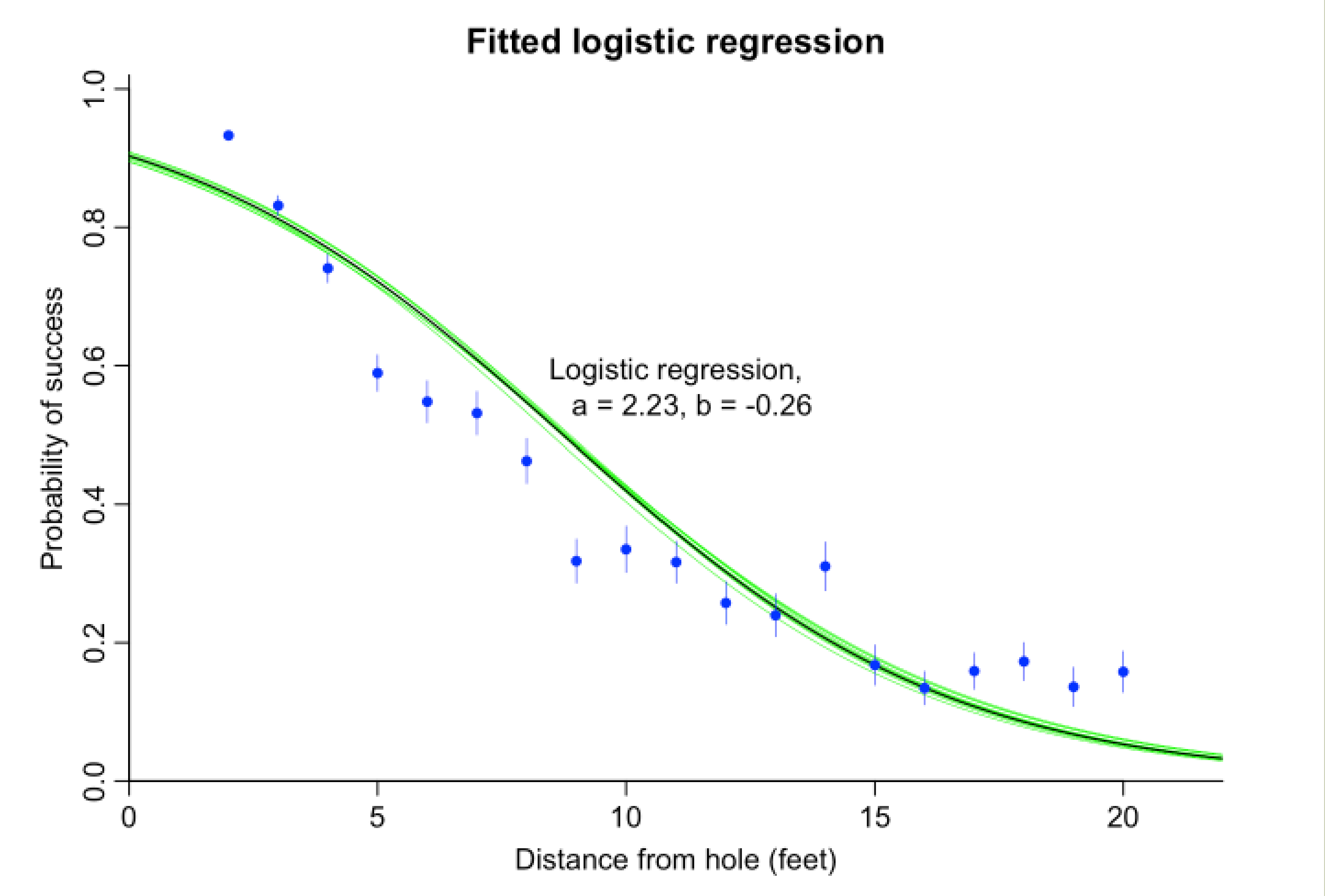
REGRESSÃO LOGÍSTICA

- Y[i] ~ Binomial(N[i], P[i]) onde:
  - P[i] = logit^-1(a + b*X[i])

```
@model function logreg(X, n,  y; predictors=size(X, 2))
    # priors
    α ~ Normal(0, 2.5) # não é alpha, é α
    β = Vector{Float64}(undef, predictors)
    for i in 1:predictors
        β[i] ~ Normal()
    end

    # likelihood
    for i in 1:length(y)
        y[i] ~ BinomialLogit(n[i], α +  X[i, :] ⋅ β) #\cdot TAB (dot product)
    end
end;
```

REGRESSÃO LOGÍSTICA



AVALIAÇÃO

- Otimização para o problema de negócio depende da redução da distância da bola em relação aos buracos, mas jogadores piores tendem a fazer mais de um putting, de modo que eles mesmos já realizam esse ajuste naturalmente.

- Modelo não permite entender a relação entre a probabilidade de acerto e a habilidade do jogador.

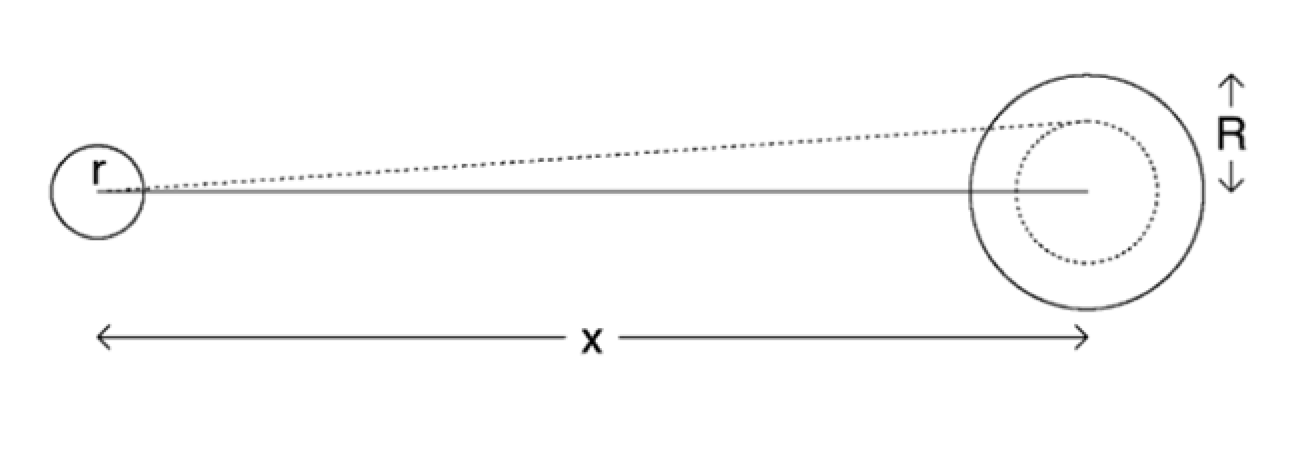
- O ajuste é satisfatório, mas não resolve o problema.

GOLF PUTTING

Modelo V.2

PRIMEIROS PRINCÍPIOS

- O que faz a bola de golfe entrar?



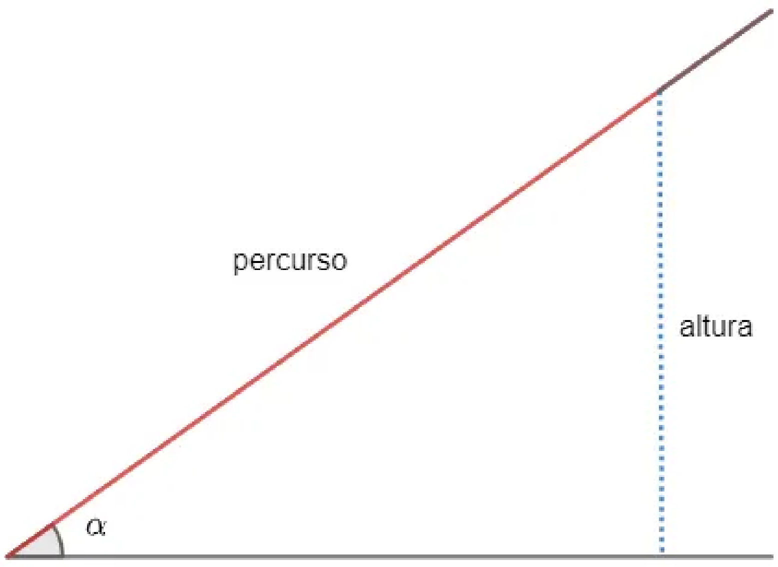
PRIMEIROS PRINCÍPIOS

- A altura máxima será (R-r).

- Ou seja, consideramos o raio do buraco (R) e subtraímos dele o raio da bola (r) pois é preciso haver espaço para a bola entrar.

- Se não subtrairmos o raio da bola e considerarmos a altura máxima como o Raio do buraco, a bola de golfe poderá passar por ele no limite, apenas bordejando o buraco (e logo, não entrando).

PRIMEIROS PRINCÍPIOS



$$\text{sen}\ \alpha = \frac{\text{altura}}{\text{percurso}}$$

PRIMEIROS PRINCÍPIOS

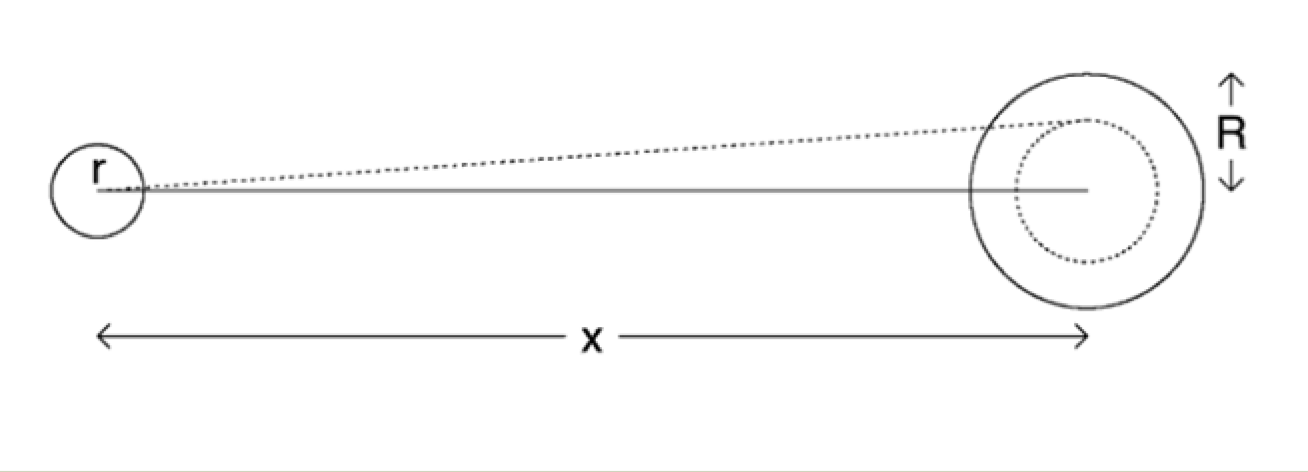
- A função inversa do Seno é o Arcoseno
- Se o ângulo for zero, a altura também será zero (logo, a bola passará com seu centro exatamente no centro do buraco).
- Se o ângulo for negativo, passará “abaixo” do centro, se for positivo, passará “acima” do centro.

PRIMEIROS PRINCÍPIOS

- Uma boa prior distribution para o ângulo é a distribuição normal centrada em zero, pois:
- Os jogadores miram– e acertam em média – no centro
- Os erros estão centrados nessa média.
- Não há preferência por “errar” mais para cima ou para baixo.

PRIMEIROS PRINCÍPIOS

- Dado que já conhecemos a média, o parâmetro a ser inferido é portanto o desvio padrão desta distribuição normal: ou seja, o grau de controle que os jogadores têm sobre esse ângulo (que deveria ser 0°).



PRIMEIROS PRINCÍPIOS

- A probabilidade de que a bola entre no buraco, para uma dada distância X, é dada pela probabilidade de que o módulo do ângulo seja menor do que o ângulo limite:
  - Pr(|ângulo| < arcoseno((R-r)/x)
- Essa probabilidades será igual a:
  - 2 ϕ (ângulo_limite-média/desvio_padrão)-1
  - Φ é a CDF da Normal
  - A média é 0
  - 2 * ϕ porque é bicaudal
  - -1 porque x – (1-x): p é o "centro" dos limites

MODELO

- Modelo em Turing

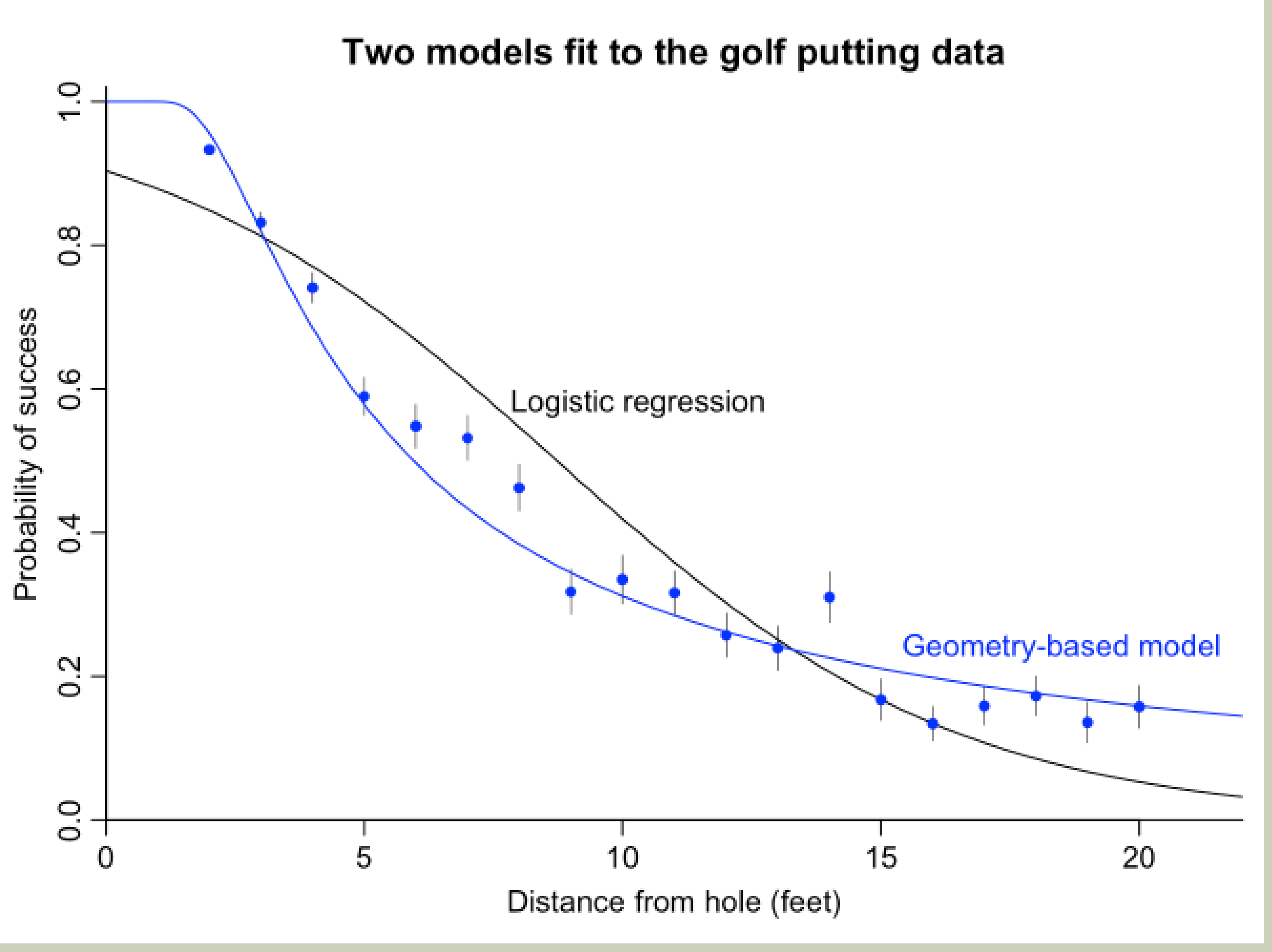
```
function max_angle(r, R, x)

    return asin((R-r)/x)
end
```

```
@model function trigon_reg(x, n,  y, r, R)
    # priors
    σ ~ truncated(Normal(0, 2.5); lower=0)

    # likelihood
    for i in 1:length(y)
        p = 2 * cdf(Normal(0, σ), max_angle(r, R, x[i]))-1
        y[i] ~ BinomialLogit(n[i], p)
    end
end;
```

RESULTADOS



Two models fit to the golf putting data
Probability of success
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
Distance from hole (feet)
Logistic regression
Geometry-based model

AVALIAÇÃO

- Ajuste é superior
- Agora temos um novo parâmetro: o grau de controle do jogador sobre o ângulo em que ele acerta a bola
  - Jogadores piores terão um desvio padrão maior
  - Jogadores melhores terão um desvio padrão menor
- A distribuição posterior de sigma representa como esse grau de controle se distribui no conjunto de jogadores
- Agora conseguimos resolver o problema!

GOLF PUTTING

Simulação

SIMULAÇÃO – TAMANHO DOS BURACOS

- Podemos identificar na distribuição posterior do parâmetro sigma original o percentil extremo à direita (ou seja, o maior nível de “erro” dos jogadores que conseguiram acertar)

	parameters Symbol	2.5% Float64	25.0% Float64	50.0% Float64	75.0% Float64	97.5% Float64
1	:σ	0.0295844	0.0295844	0.0317637	0.0343794	0.0700404

SIMULAÇÃO – TAMANHO DOS BURACOS

- Podemos gerar vários modelos, um para cada diâmetro de buraco, e ver como isso afeta o parâmetro sigma:

```
function model_factory(R::Float64)

    temp_model = trigon_reg(df[!, "x"], df[!, "n"], df[!, "y"],
                            (1.68/2)/12, (R/2)/12)
    return temp_model

end
```

```
begin
    simulated_data = []
    for R in 4.25 : 1 : 10
        temp_model = model_factory(R)
        corrente = sample(temp_model, MH(), 2_000)
        push!(simulated_data, (R=>mean(corrente["σ"])))
    end

end
```

SIMULAÇÃO – TAMANHO DOS BURACOS

Buracos maiores, tendem a ter um sigma médio superior (ou seja, na média, é possível “errar mais” em relação ao ângulo

Diametro (polegadas)	Sigma médio
4.25	0.0381874
5.25	0.0471095
6.25	0.0624821
7.25	0.0765048
8.25	0.0813176
9.25	0.0999186

SIMULAÇÃO – TAMANHO DOS BURACOS

O diâmetro ideal dos buracos é de aprox. 7 polegadas!

OBRIGADO!

Referências:

Gelman, Andrew (2019). "**Model building and expansion for golf putting**". Disponível em: https://mc-stan.org/users/documentation/case-studies/golf.html

Dados extraídos de:

https://statmodeling.stat.columbia.edu/2019/03/21/new-golf-putting-data-and-a-new-golf-putting-model/