Nome: Gustavo Estrela de Matos
Nusp: 8536051

Relatório: álbum de figurinhas da copa

São Paulo
2014

Motivação

Em anos de copa do mundo sempre há o lançamento do album de figurinhas com jogadores e seleções que participam do evento. A empresa italiana, Panini, é responsável pela distribuição dos álbuns e pacotes de figurinhas. A empresa de Giuseppe Panini foi fundada em 1961 e no mesmo ano já lançou o álbum de figurinhas do Campeonato Italiano de Futebol. A brincadeira conquistou muitos amantes do esporte e em 1970 a Panini lançou o primeiro álbum de copas do mundo, passando a vender as figurinhas mundialmente.

Neste ano a Panini lançou o álbum da copa do mundo no Brasil, com 649 figurinhas, e, como em todas as ultimas copas, é esperado que muitas pessoas comprem o álbum e tentem completá-lo. Diante de tanto sucesso, é razoável nos perguntar o quão difícil é completar o álbum e o quanto gastaríamos para tal.

Com auxílio do programa R, de um programa que simula o completamento de um álbum e de estatística vamos responder essas perguntas e, também, comparar maneiras de se completar o álbum. Porém, antes, precisamos determinar algumas regras para nosso álbum e simulações.

O álbum da Copa do Mundo de 2014 possui 649 figurinhas e os pacotes vendidos pela Panini possuem 5 figurinhas diferentes, custando 1 real cada pacote. Além disso, vale lembrar que as simulações e cálculos feitos aqui são para situações especiais, por exemplo, completar um álbum sem trocar nenhuma figurinha. Na vida real, a quantidade de dinheiro gasto para se completar um álbum envolveria muitas variáveis e não temos condições de representar todas elas em um computador.

1 - Cálculo justo do tempo médio até completar o album

1.1 – Montando o sistema

Supondo que compramos um álbum e que não estamos fazendo trocas de figurinhas com ninguém, é possível determinar o preço médio para se completar um álbum.

Seja T_i o tempo médio de se completar o album tendo i figurinhas já coletadas, ou seja, quatos pacotes ainda precisam ser comprados, em média, para completar um álbum. Estamos interessados em descobrir T_0 . Seja também, X_i uma hipergeométrica, na qual a população total é 649, a quantidade de figurinhas que nos interessam são 649 – i (figurinhas ainda não tiradas), e as retiradas são feitas 5 a 5.

$$T_0 = 1 + T_5 * P(X_0 = 5)$$

$$T_5 = 1 + T_5 * P(X_5 = 0) + T_6 * P(X_5 = 1) + T_7 * P(X_5 = 2) + T_8 * P(X_5 = 3) + T_9 * P(X_5 = 4) + T_{10} * P(X_5 = 5)$$

$$\vdots$$

$$T_i = 1 + T_i * P(X_i = 0) + T_{i+1} * P(X_i = 1) + T_{i+2} * P(X_5 = 2) + T_{i+3} * P(X_5 = 3) + T_{i+4} * P(X_5 = 4) + T_{i+5} * P(X_5 = 5)$$

$$\vdots$$

$$T_{649} = 0$$

Para encontrar o valor de T_0 utilizaremos o programa R e a Regra de Cramer. Assim, é possível determinar que $T_0 = 913.108$.

Lembrando que cada pacote custa 1 real, desconsiderando o preço do álbum, teríamos gasto, em média, pouco mais de 913 reais para completar o álbum.

1.2 – Montando e resolvendo o sistema no R

Antes de tudo é necessário escrever melhor o sistema acima para termos um formato $Ax = B$ e depois utilizarmos a regra de cramer. Para fazer isso, passaremos tudo o que está na direita, e que não é constante, para o outro lado, deixando o sistema assim:

$$T_i * [1 - P(X_i = 0)] - T_{i+1} * P(X_i = 1) - T_{i+2} * P(X_5 = 2) - T_{i+3} * P(X_5 = 3) - T_{i+4} * P(X_5 = 4) - T_{i+5} * P(X_5 = 5) = 1$$

Assim, A é a matriz formada pelas probabilidades da hipergeométrica, x é o vetor $(T_0, \ldots, T_{649})$ e B é o vetor $(1, \ldots, 1, 0)$.

Perceba que nossa expressão generalizada acima não funciona em um caso especial, quando i vale 649. Por isso, nosso código no R precisará de alguns “retoques *” para representar, de fato, o sistema a que chegamos em 1.1.

Agora podemos construir o sistema e achar T_0 seguindo os seguintes passos:

- Crie uma matriz 650 por 650:
  A = array(1:422550, dim=c(650,650))
  for (i in 1:650)
  for (j in 1:650)
  A[i,j] = 0
- Inicialize essa matriz com zeros
  for (i in 1:650)
  for (j in 1:650)
  A[i,j] = 0
- Preencha a diagonal da matriz
  for (i in 1:650)
  A[i,i] = 1 - dhyper(0, 650 - i, i - 1, 5)
  *A[650,650] = 1 **

- Preencha o restante da matriz

  for (i in 1:650) {
  j = i+1
  while (j < 650 & j < i + 6) {
  A[i,j] = dhyper(j-i, 650 - i, i - 1, 5)-1*
  j = j + 1
  }
  }

- Crie uma outra matriz C igual a A e troque a primeira coluna de C pelo vetor B.

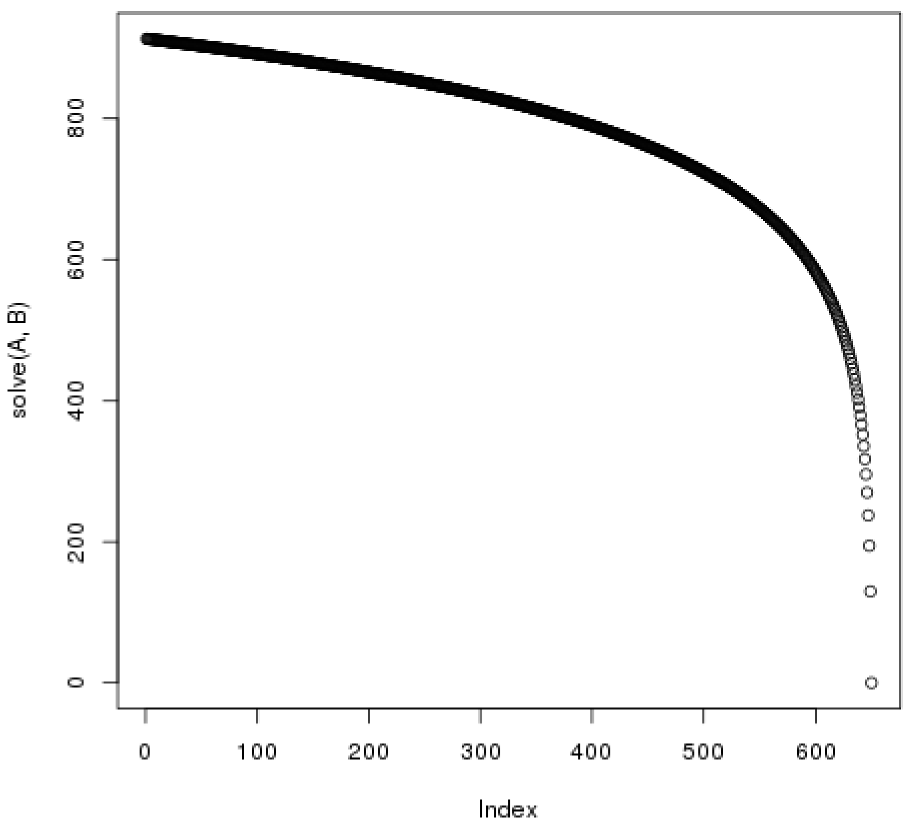
  C = A
  for (i in 1:650)
  C[i,1] = 1
  *C[650,1] = 0 **

Agora, usando a regra de cramer, o valor de T_0 deve ser $det(C)/det(A)$, que vale 913.108. Então, a quantidade média de pacotes, contendo cada um cinco figurinhas diferentes, comprados para se completar um álbum de 649 figurinhas é 913.108.

Se quisermos achar todos os T_i, podemos criar o vetor B e, depois, usar a função *solve(A,B)* do R:

- *B = array(1:650, dim=c(650,1))*
- *for (i in 1:650)*
- *B[i,1] = 1*
- *B[650,1] = 0*
- *solve(A,B)*

Veja abaixo o gráfico do vetor T_i, onde o eixo x é o valorde i – 1 e o eixo y é T_x



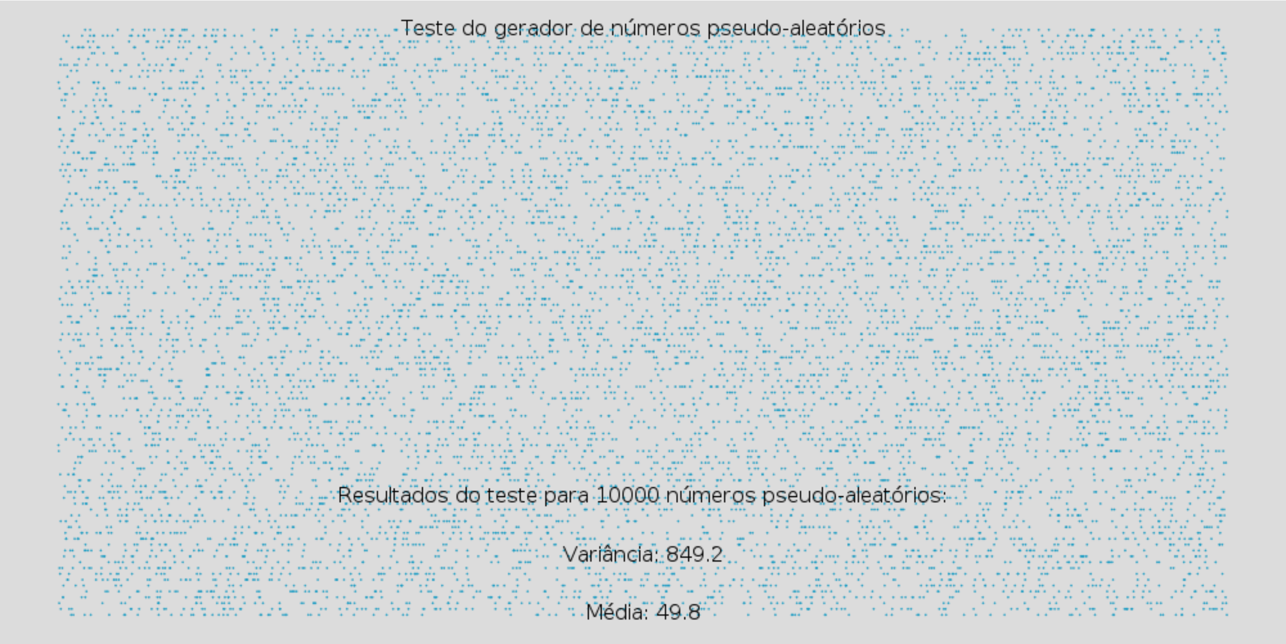
2 - Simulações

2.1 – Gerador de números pseudo-aleatórios

Em um computador não é possível gerar uma sequência de números verdadeiramente aleatória, pois, parttindo de uma mesma semente, sempre poderemos reconstruir essa sequência. Utilizaremos então, números pseudo-aleatórios, que funcionam bem para os nossos testes.

O gerador utilizado foi o gerador de números aleatórios do Java, da classe matemática do java.

Veja abaixo um teste do gerador para números aleatórios entre 0 e 100:



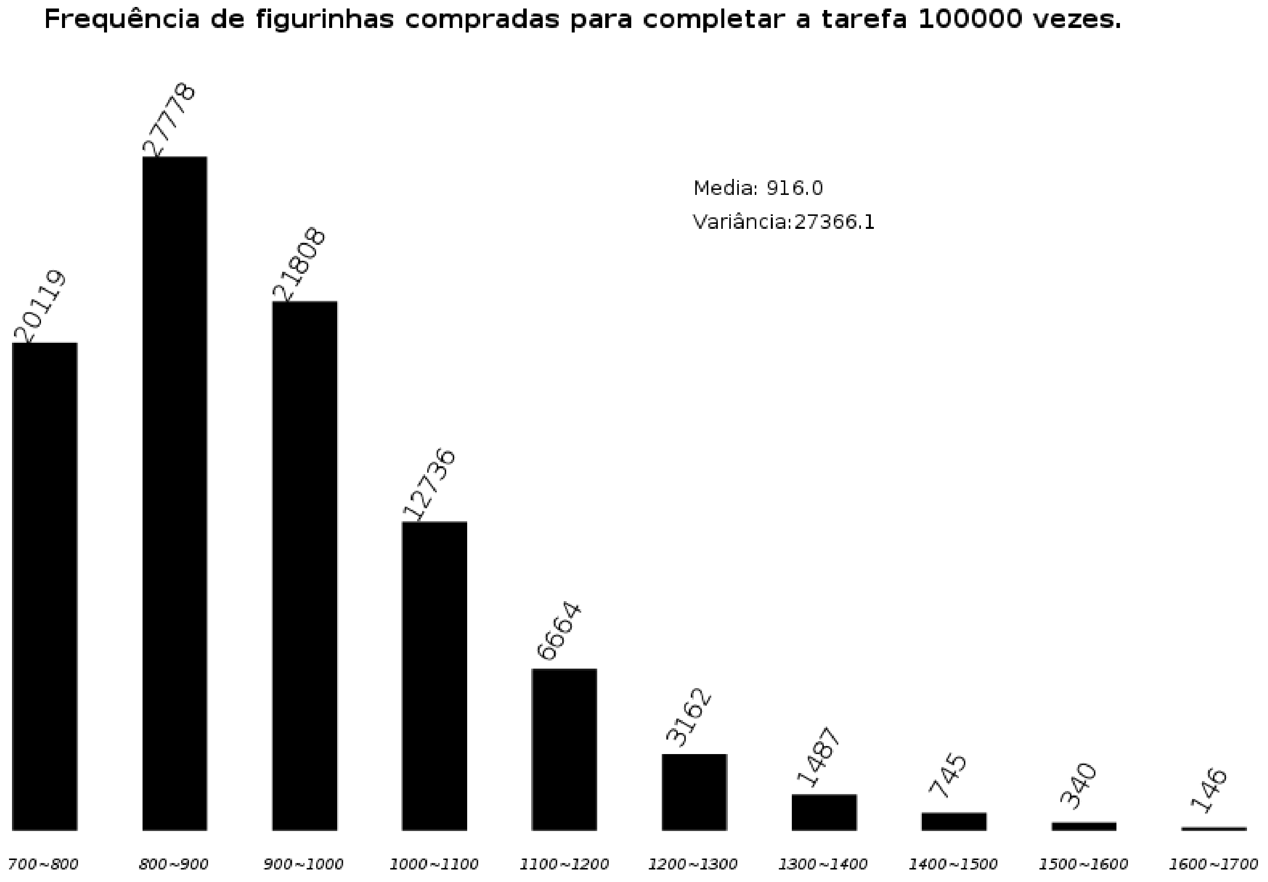
A documentação da classe geradora de números pseudo-aleatórios pode ser consultada no site da Oracle.

2.2 – Completando um álbum com pacotes de n figurinhas diferentes

Nessa simulação, o programa recebe como parâmetros o número de figurinhas em cada pacote e o número de vezes que a simulação será feita. O programa dá como resposta a média de pacotes comprados e a variância.

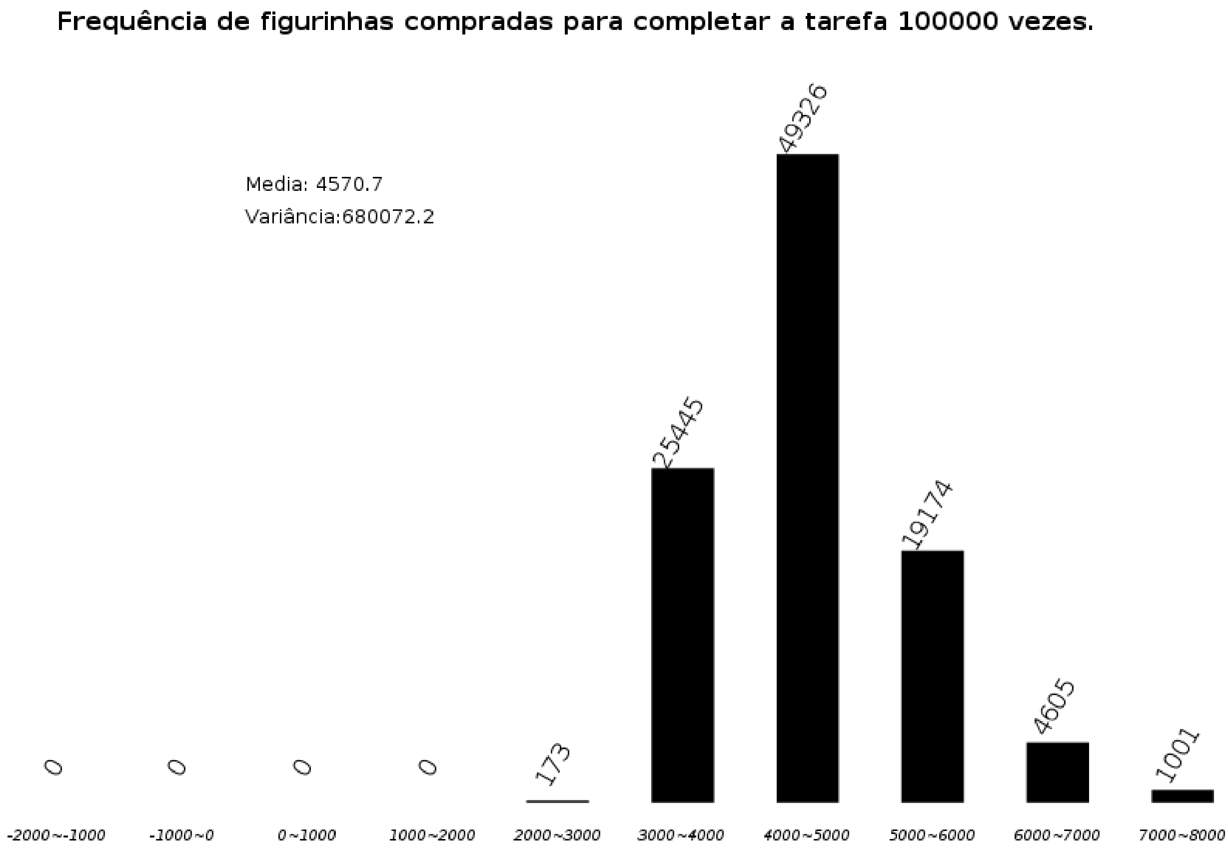
Veja os resultados abaixo para 100 mil testes:

- *para n = 5*



Podemos observar que o valor da simulação ficou bem proximo ao que tinhamos visto em 1.1.

- *para n = 1*



Neste caso também sabemos calcular a média de pacotes comprados. Se já possuimos i figurinhas, a probabilidade de conseguir uma nova figurinha é $\frac{(649-i)}{649}$, então o tempo médio para se conseguir uma nova figurinha é a esperança de uma geométrica de parâmetro $(649-i)/649$. Assim, nosso T_0 valerá $1+\frac{649}{648}+\frac{649}{647}+\frac{649}{646}+...+649$ que pode ser aproximado para $649*\ln(649)$.

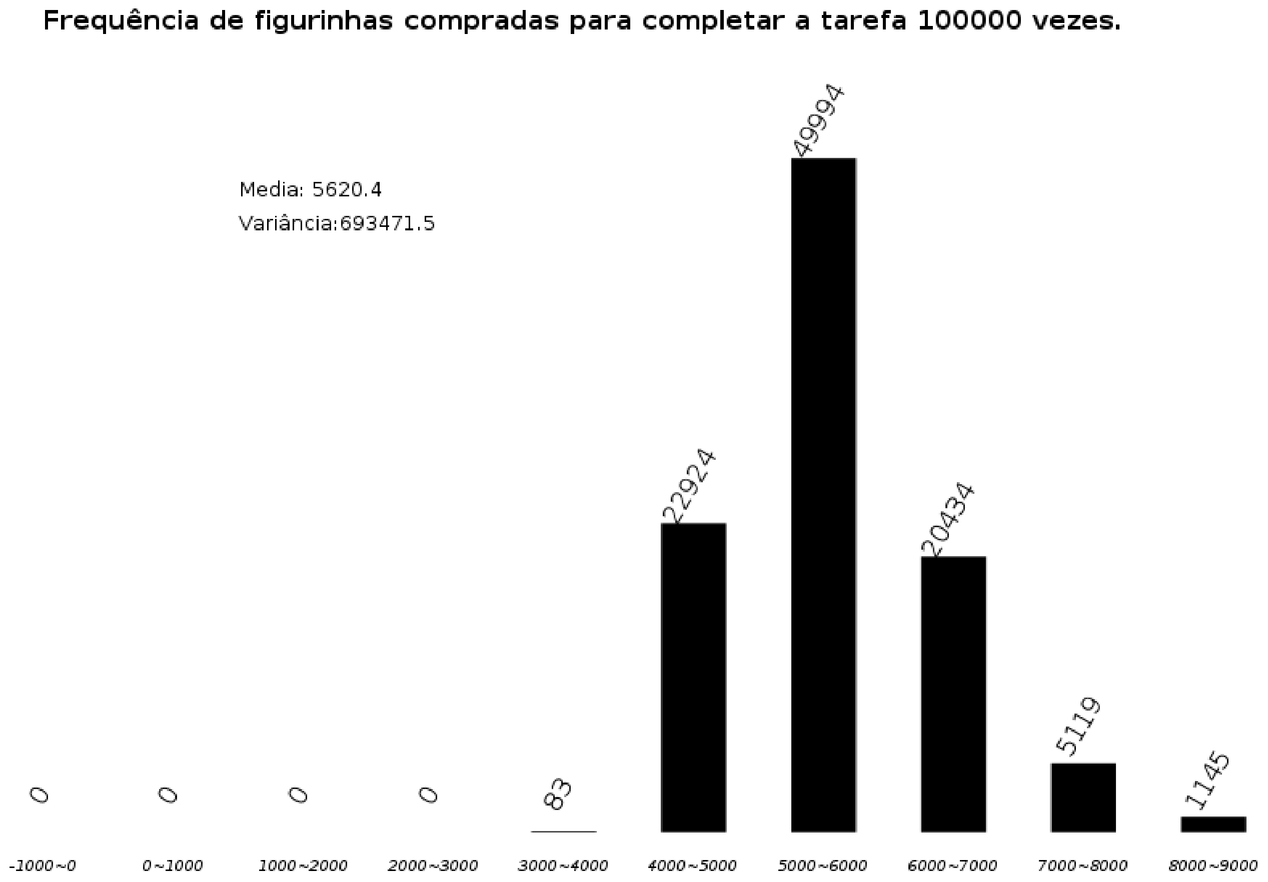
Se fizermos na calculadora, veremos que $649*\ln(649) \simeq 4202$, que é um pouco longe dos 4570, mas isso ocorre porque multiplicamos o erro da aproximação por 649.

2.3 – Completando cinco álbuns com pacotes de cinco figurinhas sem trocas

Nessa simulação, o programa recebe como parâmetro apenas o número de vezes que a simulação será feita. O programa dá como resposta a média de pacotes comprados para completar os cinco albúns e a variância.

Neste teste, a distribuição das figurinhas nos pacotes é feita da seguinte maneira: a cada pacote de figurinha aberta, a i-ésima figurinha vai para o i-ésimo álbum.

Veja o resultado abaixo:



Se dividirmos a média encontrada por cinco, teremos um valor em volta de 1120 pacotes comprados, o que faz dessa simulação não muito próxima do que acontece quando se completa um álbum recebendo pacotes de cinco figurinhas.

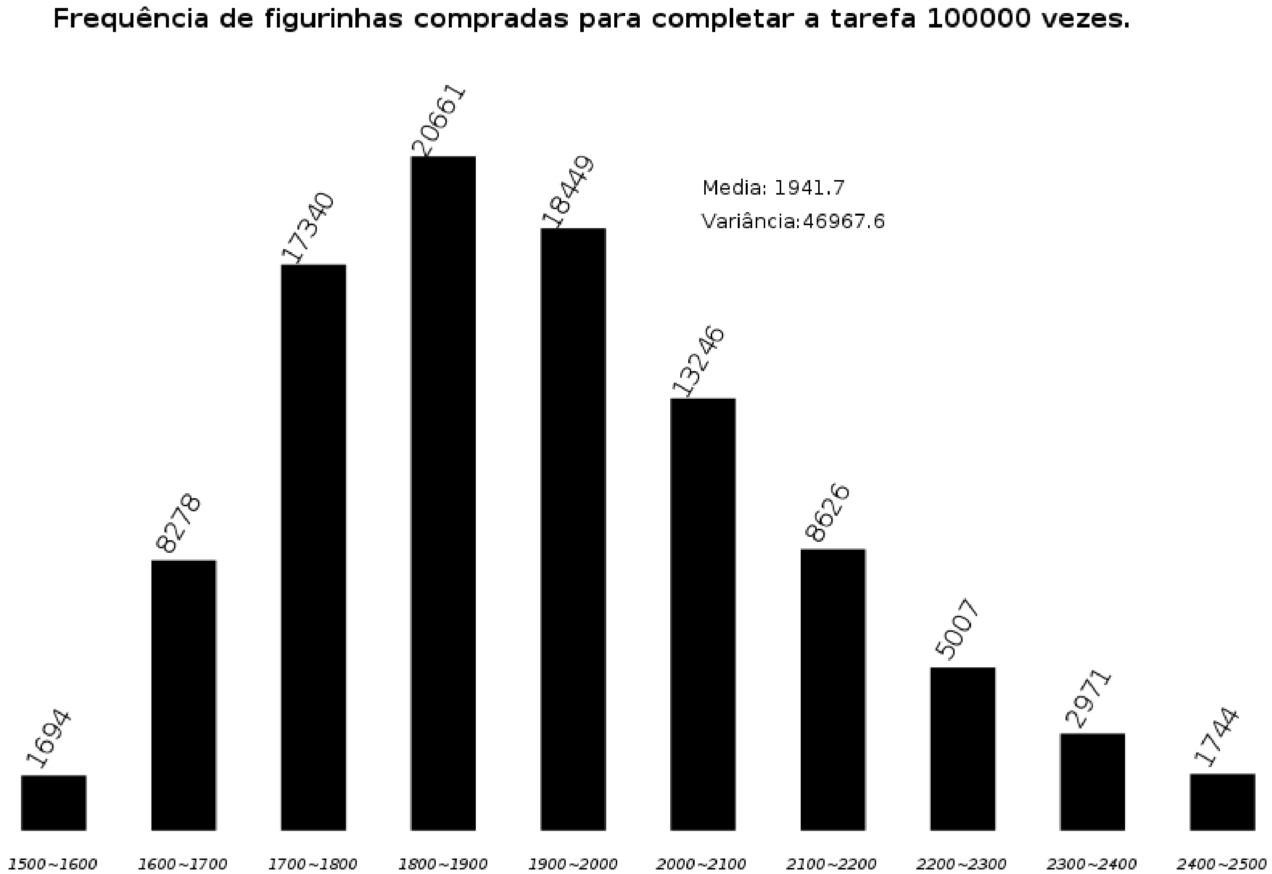
Esse aumento na média ocorre porque precisamos comprar pacotes até completar todos os álbuns, logo, pegaríamos o maior valor dos T_0 para cada álbum.

2.4 – Completando cinco álbuns com pacotes de cinco figurinhas sem trocas

Nessa simulação, o programa recebe como parâmetro apenas o número de vezes que a simulação será feita. O programa dá como resposta a média de pacotes comprados para completar os cinco albúns e a variância.

Neste teste, a distribuição das figurinhas nos pacotes é feita da seguinte maneira: uma figurinha sorteada irá para o primeiro álbum se ela não estiver em nenhum; irá para o segundo se ela não estiver no primeiro; irá para o terceiro se não estiver nem no primeiro e nem no segundo, etc.

Veja o resultado abaixo:



Vemos que, permitindo troca de figurinhas, a média de pacotes comprados para se completar os cinco álbuns diminuiu consideravelmente comparado ao mesmo caso, sem troca de figurinhas.

Dividindo a média encontrada por cinco, teremos um valor em volta de 388 pacotes, bem abaixo dos 915 encontrados na primeira simulação, o que mostra que a troca de figurinhas diminui bastante a quantidade média de pacotes comprados para completar um álbum.

Observações finais

O código das simulações, com manual, está disponível no site: https://linux.ime.usp.br//~gustavoem/figurinhas/