

当ててみよう！水槽の中に○●はいくつある??

BB弾でサンプリング大実験！

解説

付録：実験協力者のお名前

2012年8月15日

©統計数理研究所 2012

※この文書に書かれたデータを夏休みの自由研究や学校の授業等で使用される場合は、ism_news@ism.ac.jp宛にひとことお知らせいただき、統計数理研究所で実施した実験の結果である旨をご明記いただければ幸いです。

「水槽の中の黒玉の数はいくつ？」

本年 2012 年 8 月 4 日（土）の統計数理研究所子ども見学デーでは、BB 弾のサンプリング実験に御協力頂き、どうもありがとうございました。この実験の目的は水槽の中の黒玉の数を予測することでしたが、皆様の予測はうまくいきましたでしょうか？実は水槽の中には黒玉と白玉を合わせて 10 万個の BB 弾がはいっており、そのうち黒玉は 2 万 5 千個入っていました。つまり黒玉の割合は 0.25 です。皆様にはそのうち 300 個を“ランダムに”サンプリングして頂き、その中の黒玉の数をかぞえて頂きました。結果は以下の通りです（一生懸命かぞえてくださいました皆様に改めてお礼を申し上げます）。

76 63 69 66 76 86 50 75 90 72 83 79 71 76 88 73 83 75 85 75 79 71 83 80 84 77 80 76 80 78 74 84 67
76 76 79 78 66 74 78 76 78 76 74 86 55 70 77 79 60 75 74 78 56 68 65 68 68 75 80 60 95 55 61 70 72
75 88 77 72 70 54 77 84 75

計 75 回の計測結果です。これをグラフにしたものが図 1 です。75 個から 80 個の場合が多いことが分かります（灰色のグラフ）。赤線は 2 項分布から計算した理論的な結果です。中心の赤の破線が BB 弾 300 個中の黒玉の理論的な平均値です（ $300 \times 0.25 = 75$ ）。これは真値¹とも呼ばれます。灰色の図と赤色の曲線がうまく合致していることも見て取れると思います。

さて一回のサンプリングでどれだけ正確な予測ができるかを見てみましょう。上記の計測結果の一番目の数は 76 です。理論的な平均値 75 にかなり近く、かなり予測がうまくいっているように思えますが、ただこの結果はあまりあてになりません。この一回だけのサンプリングから求まる 95% 信頼区間（※参照）を図 1 中に緑の線で描きました。×（緑のバツ印）が今回の予測値 76 に対応しています。信頼区間が 62 から 91 の広範囲におよびあまり正確なことは言えません。それでは 5 回のサンプリング実験の結果（76 63 69 66 76）を使った場合はどうでしょうか？この結果の 95% 信頼区間をオレンジ色の線で描きました。予測値（上記の 5 個の計測結果の平均値）は 70 となりましたが信頼区間は以前の緑色のものよりかなり狭くなったことが分かります。それでは上記の 75 回の計測結果を全部使うとどうなるでしょうか？その結果は青の線で描いています。予測値が 74.3 となり信頼区間もかなり狭くなったことが分かります（72.6, 76.0）。

さて今回の最後の結果を使い、水槽中の黒玉の個数を予測致しましょう。水槽の中には全部で 10 万個（100000 個）の BB 弾が入っていました。今回そのうち 300 個だけサンプリングしましたので上記の予測値 74.3 と 95% 信頼区間（72.6, 76.0）にそれぞれ 100000/300 をかけてみます。

$$72.6 \times 100000 / 300 = 24200$$

$$74.3 \times 100000 / 300 = 24767$$

$$76.0 \times 100000 / 300 = 25333$$

結局今回のサンプリング実験で、水槽の中の黒玉の数の予測値は 24767 でその 95% 信頼区間は（24200,

¹ 今回のサンプリング実験で予測対象となる値。黒玉の割合 0.25、またはそれに対応する黒玉の数。

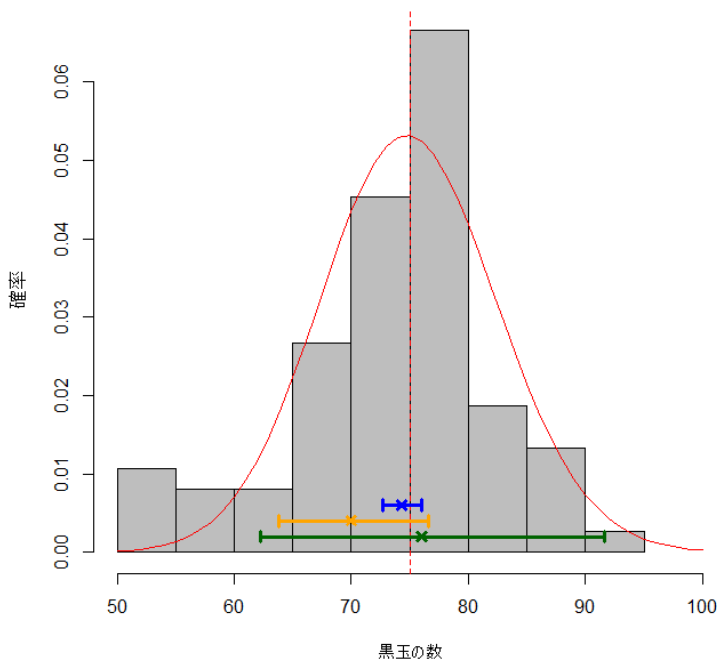
25333)であることが分かりました。

今回 300 個の BB 弾のサンプリングを 75 回して頂きましたので、合計で 22500 個の BB 弾をかぞえて頂きました。それでは黒玉の割合 0.25 (または水槽中の黒玉の数 2 万 5 千個) をある程度正確に予測するには一般的にどれだけの BB 弾をかぞえる必要があるのでしょうか？もし 3000 個をかぞえて頂いた場合、確率 0.95 で黒玉の割合の予測値のずれを 0.016 以下に抑えることができます。ただこのずれを 0.01 に縮めるためには約 2 倍の 7000 個を数える必要があり、労力の割にはそれほど予測の精度は向上しません。今回 22500 個を数えて頂きましたのでこのずれは 0.006 以下に抑えることができました(実際の黒玉の割合の予測値は 0.247(=上記の 75 回の計測値の平均/300)となり 0.25 とのずれは 0.003 です)。

今回のサンプリング実験と同様なことは広く世の中で行われています。世論調査や選挙の当選予想などでは、全ての日本人(10 万個の BB 弾)に対して調べるのではなく、数千人(数千個の BB 弾)に対して調査を行えば、かなり正確に知りたい情報(水槽中の黒玉の数)を知ることができます。これも統計学の手法を用いることによってはじめて可能となるのです。

分析及び執筆²: 統計数理研究所 統計思考院 小森 理

図 1



※ 95%信頼区間とは？

確率 0.95 で予測したい本当の黒玉の数(真値)が含まれるように作った区間。サンプル数を多くすると、この区間は狭くなる(予測の精度が上がる.)。

ランダム性はこの区間自体が持ち、予測したい真値にはランダム性がないことに注意。

² 記述を簡明にするため、数値計算の細かい端数は省略しております。

実験協力者方のお名前（公開を承諾して下さった方のみ）

ひかりさん	東京都青梅市	しゅうちゃんさん	東京都東大和市
ゆうかさん	東京都青梅市	たつくんさん	東京都東大和市
ちひろさん	埼玉県滑川町	みーちゃんさん	東京都東大和市
ゆきのさん	埼玉県滑川町	つばささん	東京都日野市
ほのかさん	埼玉県東松山市	りょうまさん	東京都武蔵村山市
はるかさん	東京都八王子市	けいなさん	
げつきさん	東京都三鷹市	おとかさ	東京都立川市
ゆうまさん	東京都立川市	さきさん	東京都立川市
ゆうきさん	東京都多摩市	徳久さん	埼玉県日高市
ゆうさくさん	東京都国立市	柊さん	埼玉県
あかねさん	東京都国立市	はなさん	東京都立川市
こじろうさん		遼さん	神奈川県川崎市
あおとさん		ここなさん	東京都小平市
りょうさん	東京都小平市	けんたろうさん	東京都国立市
みきさん	東京都国立市	雄大さん	東京都立川市
みおさん	東京都国立市	大空さん	東京都立川市
ふうかさん	長野県上田市	杏果さん	東京都武蔵村山市
はるきさん	千葉県習志野市	弘さん	Taipei, Taiwan
ひなさん	千葉県習志野市	遥人さん	東京都立川市
たいちさん	東京都小金井市	百合菜さん	東京都立川市
しょうたさん	東京都小金井市	樹さん	東京都立川市
みちとしさん	東京都調布市	裕作さん	東京都立川市
かずしさん	東京都立川市	美涼さん	神奈川県厚木市
あさひさん	東京都大田区	涼太郎さん	神奈川県厚木市
りょうたさん	東京都立川市	一泰さん	東京都調布市
あいりさん	東京都立川市	大造さん	東京都調布市
みづきさん	東京都立川市	健史さん	東京都調布市
ゆきさん	東京都小平市	航太郎さん	東京都三鷹市
あかりさん	東京都昭島市	隼太郎さん	東京都三鷹市
れいさん	東京都昭島市	ゆりさん	東京都小平市
颯太さん	東京都小平市	玲奈さん	東京都立川市
しおのさん	東京都小平市	吉造さん	東京都立川市

※方が一お名前が間違っている場合は申し訳ございません。お詫び申し上げます。

上記以外の方でご協力いただいた方も含め、ご協力ありがとうございました！