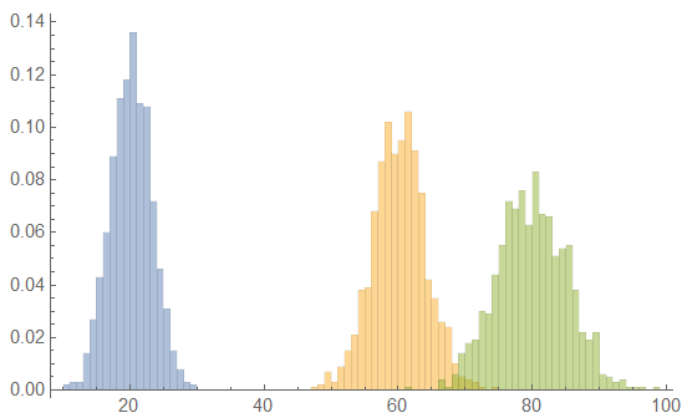


2011年、GDP ランキングにおいて日本は中国に抜かれ3位になった。自分としては、別段何も変わらないのだが、オックスフォードの定宿のB&Bでの出来事。今まで、紅茶と緑茶がセットされていたのが、ウーロン茶に変わった。そういえば、この頃から市街地の観光客で東洋の人は殆ど中国の人となり、日本人観光客を見つけることが難しくなった。夏の時期、サマータウンでホテルを探すことも非常に難しくなってきた。日本の京都でも宿の予約が難しくなった。そして、2020年の今、中国の若い人のように、若い人たちには、できるだけ海外の経験を積んで頂きたいと思う。しかしコロナ禍で渡航できないのは私も学生も同じである。今できることは、足腰と英語力が衰えぬように、鍛錬あるのみ、である。

中心極限定理の前に、「正規分布と正規分布を足し合わせると正規分布になる」という話をする。

問題：

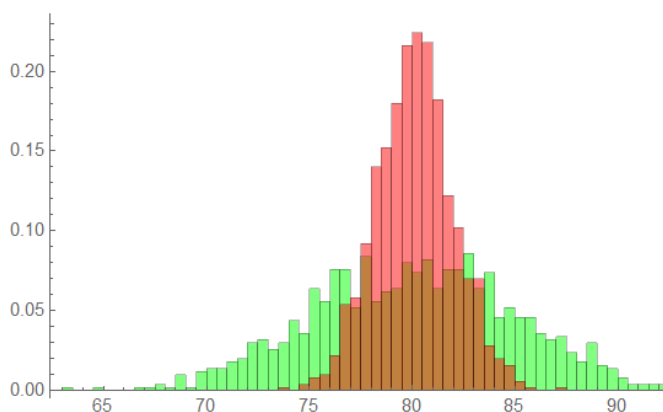
飼い主の体重分布は正規分布 $N(60,16)$ に従っています。犬の体重分布は正規分布 $N(20,9)$ に従っています。飼い主と犬をランダムに抽出してペアを作ったとき、ペアの体重はどのような分布に従いますか？



答えは、"人 $N(60,16)$ 犬 $N(20,9)$ のペア体重は

$N(80,25)$ " である。ペア平均が人平均と犬平均の和となることは予想が付くと思う。そして、分散は、両者の分散の和 $16+9=25$ となる。シミュレーションの様子を上図に示した。青の分布が犬の体重、黄色の分布が人の体重、ペアの体重の分布は緑で示した。分散ではなく標準偏差で考えると、犬分布が3、人分布が4、ペアが5となっている。確かに広がり具合がそのようになっていることが確認できる。

次に、このペア集団(母集団と呼ぶ)の中から、ランダムに7ペアを抽出し、その平均を計算するとする。つまり、7個の値を足して、7で割る。これを標本平均と呼ぶ。サイズ7の標本を抽出して、その標本平均を計算する。標本の抽出を1000回繰り返して、標本平均を1000回求め、その分布を描いた(右図の赤い分布)。母集団 $N(80, 25)$ のペアの分布は緑色で表されている。上図でも、ペア分布の縦軸の確率密度のスケールがトップで0.08程度になっている。さて、母集団と標本平均の分布の分散はどちらが小さいですか？



答えは、標本平均の分布の分散の方が小さい、である。この事例では標本サイズが7なので、分散は $25 \div 7$ となる。

ここで中心極限定理である。

中心極限定理：

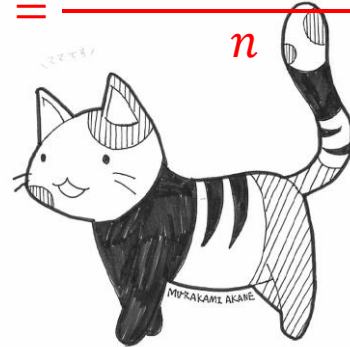
平均 μ ，分散 σ^2 の母集団から無作為抽出したサイズ n の標本平均の分布は， n が大きくなるにつれて，

平均 μ ，分散 $\frac{\sigma^2}{n}$ の正規分布に近づく。

Sample average

本事例では母集団の分散が 25 で，標本サイズが 7 であるので，標本平均の分散は $\frac{\sigma^2}{n} = \frac{25}{7}$ となる。標本抽出された 7 ペアの中には，重いペアも居れば，軽いペアもいるので，相殺されて，分散は母集団分散に比べて小さくなる。標本内で大きいものと小さいものが相殺される，というところがポイントである。経営の例をあげると，1 日の需要分布が $N(1000, 700)$ だったとき，1 週間での標本平均の分布は $N(1000, 100)$ となる。 $700 \div 7 = 100$ という計算である。

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$$



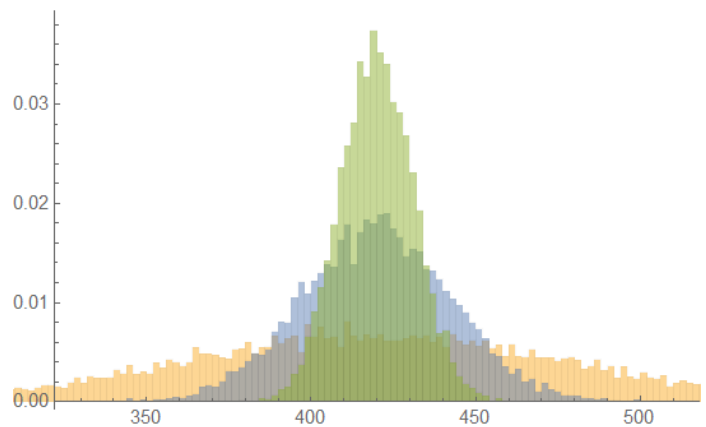
中心極限定理は，標本平均の分布に関する定理である。標本平均であることを強調したいとき，私のテキストでは，右図にあるフランケンシュタイン猫のアイコンを使う。 n 分の 1 ずつ足し合わせた感じが出ているのではないだろうか？この猫は院生の村上朱音氏に描いて頂いた。村上氏に感謝する。

中心極限定理で驚くべきところは，母集団が正規分布でなくても，どのような分布であっても，標本平均分布が正規分布に近づくことである。サイズが 7 位の小さい数だと，まだ正規分布に見えてこないが，30 以上となると，正規分布に近づいてくる。母集団は，2 コブラクダのような分布でもよいし，指数分布でもよい。どのような形状の分布でも，標本平均の分布は正規分布に近づく，という点が全く驚きである。

最後に，もうひとつ例を見てみよう。

白熊ストア目白店の 1 日の売上高分布は，右図の黄色の分布 $N(420, 360)$ だったとする。

1 週間の標本平均の分布は $N(420, \frac{360}{7})$ となる（右図の青の分布）。30 日間の標本平均の分布は $N(420, \frac{360}{30})$ となり，右図の緑の分布となる。標本サイズが大きくなるに従って，正規分布がとがってくる。



終わり