StaM-Bladet

Informationsblad för medlemmar i StaM (Statistisk Metodik), sektion inom SFK, Svenska Förbundet för Kvalitet

Augusti 1998 **årgång 8 nummer 15**

Femtonde numret

Det är något binomialt med Internet: -"Antingen finner man det man söker eller så hittar man något annat". Eftersom detta nummer av StaM-bladet berör binomialfördelningen så tänkte vi att det skulle vara på sin plats med lite länkar till sidor som bl.a. innehåller tillämpningar, teori och simuleringar för binomialfördelningen m.m. Så surfa lugnt!

SISA: Simple Interactive Statistical Analysis Detta är en hemsida där man får förklaringar om, kan enkelt beräkna sannolikheter samt utföra tester. Adress: http://home.clara.net/sisa

SURFSTAT Australia En mycket trevlig "site" är SURFSTAT australia. Då man besvarat några frågor kan man använda de resurser som finns där, allt från sannolikhetskalkylatorer till sannolikhetslära och inferens. Prova!! Adress: http://surfstat.newcastle.edu.au/surfstat

HyperStat Online Denna "site" är en elektronisk uppslagsbok m.m. inom statistik. Den innehåller även länkar till sk Java-applets som simulerar fördelningar m.m. Väl värd ett besök.

Adress: http://www.ruf.rice.edu/~lane/hyperstat/contents.html

Ordförandens ruta

Hej alla nya och gamla medlemmar i SFK-StaM. Jag som skriver heter Mats Franzén och är ny ordförande i SFK-StaM sedan årskiftet. Jag vill påminna om det årliga StaM-seminariet vilket är planerat till 6:e november 1998 i Stockholm. Detta berör bl.a. prognoser och hur presentera data. Dessa är viktiga instrument i dagens "Just In Time" värld. I anslutning till seminariet hålls även årsmöte för SFK-StaM. Boka i kalendern och kom! Mer information om seminariet finns i StaM-Bladet.

Vi i styrelsen håller som bäst på att utforma en ny websida för StaM med ett betydligt rikligare innehåll än tidigare. Den beräknas komma ut på nätet i augusti/september. Inom SFK centralt håller man på att förnya organisationen för att möta framtiden och förändringarna kommer att beröra regioner och sektioner. Vilket även då gäller StaM. Kära StaM medlemmar, tyck till om vår verksamhet!!!

-*Vad behöver vi utveckla för att bli mer attraktiva och erhålla lojala medlemmar?*

Skicka e-post, skriv brev eller ring till oss eller SFK centralt. Tyck till!

Mats Franzén, Ericsson Cables AB

Tänkvärt

The world, created as a result of our thinking thus far, has problems which cannot be resolved by thinking the way we thought when we created them

- Albert Einstein

Förteckning över styrelsen finns på sista sidan

Binomialfördelningen från början

StaM-Bladet brukar inte föra långtgående teoretiska resonemang baserade på formler men ibland får man en ökad förståelse för härledningar och resultat om man kan följa deras väg från ax till limpa. Så i detta nummer släpper vi loss...

Antag att vi studerar en process som tillverkar en grej åt gången och där utfallet bara kan vara av två (bi) slag. Några exempel:

- En förlossningssal där utfallet är 'pojke' respektive 'flicka'.
- Ett löpande band med detaljer där utfallet är 'OK' respektive 'ej OK'.
- Flaskor vars innehåll bedöms 'mer' respektive 'mindre' än 75 cl.
- En kastad tärning där resultatet är 'mindre än eller lika med 2' respektive 'mer än 2'.

Gemensamt för del olika exemplen är att utfallet beskrivs med hjälp av två kategorier eller namn (nom) och att utfallet kan betraktas som slumpmässigt. Detta är naturligtvis inte självklart: det kan ju vara så att först innehåller alla flaskor mer än 75 cl och senare i processen så tenderar det att finnas mindre än 75 cl vätska per flaska.

Antagligen hade vi förväntat oss en slumpmässighet och om vi upptäcker icke-slumpmässig variation kommer vi antagligen att finna någon felkälla som kanske kan elimineras.

Låt oss beteckna det ena utfallet (t.ex. 'pojke', 'ej OK', 'mindre än eller lika med 2') med siffran "1" och det andra möjliga utfallet med siffran "0". Om vi tittar in i bilarna som lämnade sjukhuset med den lyckliga och förväntasfulla mamman och hennes nyfödda barn skulle vi kunna få följande serie:

```
0 0 1 0 1 0 0 1 1 0 0 1 1 0...
```

dvs 'flicka', 'flicka', 'pojke', 'flicka', 'pojke' osv.

Och eftersom vi är intresserade av ämnesområdet statistik skulle vi antagligen beräkna medelvärdet och standardavvikelsen över en sådan mätserie.

Vad blir det teoretiska medelvärdet (väntevärdet μ) i en sådan process? Vad blir den teoretiska standardavvikelsen, σ? Innan vi skriver ned resultatet skall vi definiera en parameter som vi kallar p (observera att vi använder liten bokstav).

Definition av *p*. p är sannolikheten att vid en slumpmässig dragning av en grej ('människa', 'detalj', flaska', tärningskast) få det ena utfallet ('pojke', 'ej OK', 'mindre än 75 cl', 'mindre eller lika med 2'). Följaktligen är sannolikheten att få det andra utfallet lika med (1 - p).

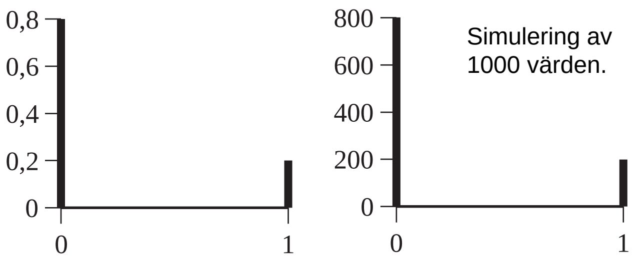
När vi vet p kan vi lätt ange både väntevärde och standardavvikelsen. (Vi anger här bara resultatet men härledningen är lätt som en plätt och återfinns i varje statistiklärobok med självaktning. Titta i index för information om "Bernoulli"):

$$\mu = p \qquad \sigma = \sqrt{p \cdot (1-p)}$$

Obs att då $p = 0$ och då $p = 1$ är standardavvikelsen också noll (antingen är inga grejor felaktiga eller så är alla grejorna felaktiga), ingen variation alltså. Observera också att standardavvikelsen är som högst då $p = 0.5$.

Sammanfattning sidan 2

Vi har tagit minsta möjliga stickprov (dvs 1 detalj) upprepade gånger ur en process och avsynat resultatet. Värdet på variabeln (avsyningsresultatet) var "1" eller "0". En sådan slumpvariabel ger väntevärdet och standardavvikelsen enligt ovan.



Sannolikhetsfördelning och en simulering för en process enligt beskrivning på sidan 2 då n =1 och p = 0.2.

Utvidgning till ett större stickprov

Nu fortsätter vi vidare i vårt teoretiska resonemang och låter varje stickprov bestå av 2 detaljer (obs att vi nu övergår till att diskutera 'detaljer', 'ej OK', 'OK', 'felkvot', m.m.) Resultatet kan ju då skrivas

resultat = *res. avsyn detalj* 1 + *res. avsyn detalj* 2

Detta beteckningssätt blir dock ganska otympligt och långrandigt. Vi skriver det istället som

$$Y = X_1 + X_2$$

De möjliga utfallet på Y är 0, 1 eller 2 'ej OK' detaljer. En dataserie skulle ju kunna bli

```
1, 2, 0, 0, 1, 1, 0, 2, 0, 2...
```

Vad blir väntevärdet och den teoretiska standardavvikelsen för Y? Här utnyttjar vi teorin om linjärkombinationer av variabler. (Linjärkombinationer, dvs summor av variabler, är en vanlig företeelse då man studerar data. Teorin om detta finns beskriven i de flesta elementära böcker i statistik.) Vi utgår ifrån följande:

$$\begin{cases} \mu_{X_1} = p_1 \\ \sigma_{X_1} = \sqrt{p_1 \cdot (1 - p_1)} \end{cases} \qquad \begin{cases} \mu_{X_2} = p_2 \\ \sigma_{X_2} = \sqrt{p_2 \cdot (1 - p_2)} \end{cases}$$

Teorin om linjärkombinationer ger oss för Y:

$$\begin{cases} \mu_Y = \mu_{X_1} + \mu_{X_2} \\ \sigma_Y = \sqrt{\sigma_{X_1}^2 + \sigma_{X_2}^2} \end{cases}$$

Eftersom X_1 och X_2 kommer från samma process är $p_1 = p_2$ vilket vi skriver bara som p. Detta, tillsammans med lite förenklingar ger:

$$\begin{cases} \mu_Y = 2 \cdot p \\ \sigma_Y = \sqrt{2 \cdot p \cdot (1 - p)} \end{cases}$$

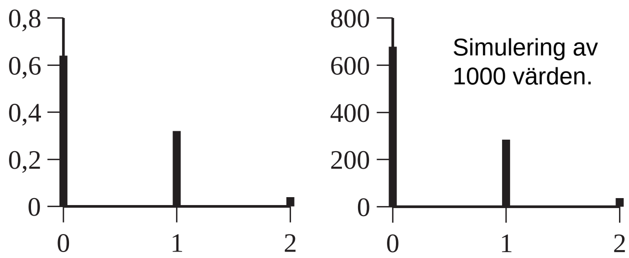
Ett stickprov om *n* detaljer

Vi låter nu stickprovet omfatta n detaljer så att vi får:

$$Y = X_1 + X_2 + \ldots + X_n$$

Vi når till slut vårt mål, dvs väntevärde och standardavvikelse för antal 'ej OK' i ett stickprov om n detaljer ur en process med den konstanta felkvoten p:

$$\begin{cases} \mu_Y = n \cdot p \\ \sigma_Y = \sqrt{n \cdot p \cdot (1 - p)} \end{cases}$$



Sannolikhetsfördelning och en simulering för en process då n = 2 och p = 0.2.

En kommentar om den 'konstanta felkvoten' är kanske på sin plats. Vi menar att felkvoten inte förändras mellan observationerna; varje gång vi avsynar en enhet förväntas p vara densamma.

Om felkvoten varierar, något som vi utreder lite senare, har vi naturligtvis inte en binomialfördelning utan något annat. En av analysens uppgifter är att försöka förstå huruvida en given datamängd kommer från en process med konstant p.

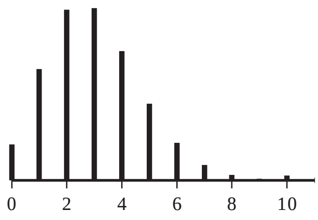
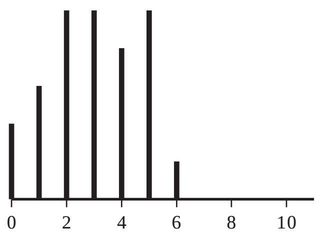
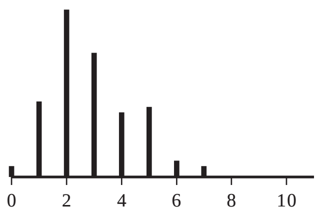
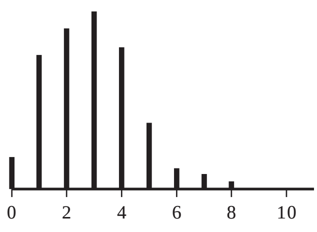
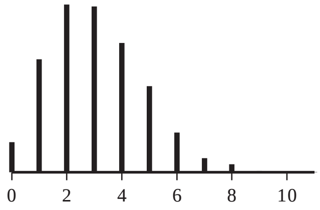
Sammanfattning sidan 3

Vi har ökat stickprovsstorleken till *n* detaljer och värdet på varje detalj var "1" eller "0". Därefter summerade vi resultatet. En sådan summa är en slumpvariabel med väntevärde och standardavvikelse enligt ovan.

Några simuleringar

På dennas sida har vi simulerat binomialfördelningen n = 100 och p = 0.03. De två översta raderna innehåller några olika egenskaper hos fördelningen. Övriga rader innehåller information efter simulareing av 25, 100, 400 och 1600 partier om 100 detaljer. De teoretiska värdena skattas väl av alla simuleringarna.

Det är anmärkningsvärt att trots att den största simuleringen innehåller 64 gånger fler värden än den minsta simuleringen, tycks skattningen av väntevärde och standardavvikelse inte givit speciellt mycket noggrannare resultat.

Parametrar n, p $0 \le x \le n$	$P(X=x)=\binom{n}{x}\cdot p^x\cdot(1-p)^{n-x}$	$\mu = n\cdot p$	$\sigma=\sqrt{n\cdot p\cdot(1-p)}$
$n=100$ $p=0.03$		$\mu = 100\cdot 0.03$ $= 3$	$\sigma=\sqrt{100\cdot 0.03\cdot(1-0.03)}$ ≈ 1.71
N = 25 (dvs 25 sampel om 100 värden)		$\bar{x} = 3.00$	$s = 1.68$
N = 100		$\bar{x} = 2.90$	$s = 1.51$
N = 400		$\bar{x} = 2.88$	$s = 1.63$
N = 1600		$\bar{x} = 3.06$	$s = 1.72$

Hur uttrycks *p*?

Binomialfördelningen innehåller två parametrar, n och p. Parametern n är egentligen inte något problem: n uttrycker stickprovets storlek. Observera att vi tänker oss att vi drar n ur en process och inte ur någon ändlig population och att de n värdena behöver inte 'ligga intill' varandra eller något liknande.

För nybörjaren, som lär sig om binomialfördelningen genom t.ex. simulering, är det till en början en aning förvirrande: "Simulera 200 värden ur en binomialfördelning med p = 0.02 och n =120". Vilket är egentligen antalet som jag använder t.ex. för att beräkna sannolikheter i fördelningen?

Uppmaningen skiljer sig ju inte i sak från "Simulera 200 värden ur en normalfördelning med μ = 10.20 och σ =1.30".

Parametern p brukar dock förekomma i ett antal olika skepnader. Ofta skrivs p snyggt och prydligt som p = 0.021 (mest i läroböcker).

Ibland redovisas p som "4.2 per 1000", eller "250 per 100 000 nyfödda har...". De senare sätten förekommer ofta då man i massmedier redovisar t.ex. hur vanlig en viss sjukdom eller åkomma är hos människor.

Här gäller det dock att inse att "7.5 liter per 100 km" inte är ett mått på p i en binomialfördelning.

Ibland uttrycks p som som en procentsats typ "25% av alla bilar har..." eller som "55 ppm..."

Ett något mer diffust sätt är "i detta stickprov om 235 människor borde vi ha funnit 13 personer med denna åkomma...". Här kan vi beräkna att man anser att p är lika med 13/235 = 0.055.

En icke-statistisk aspekt av problemet är vad som definieras som 'felaktig', 'OK' eller vad som nu studeras. Inom elektronikindustrin har man ofta svårigheter med vad som t.ex. är en dålig lödning, en 'ej korrekt lackning' etc. Med fotografier, likare, normaler, etc försöker man lösa problemet.

Ett iterativt sätt att beräkna sannolikheter

Den som inte har ett lämpligt statistikprogram kanske själv vill beräkna sannolikheter i binomialfördelningen. Men att med hjälp av kalkylator eller dator beräkna sannolikheter direkt med hjälp av ursprungsformeln brukar ge upphov till svårigheter eftersom n-fakultet, $n!$, ganska snart blir väldigt stort, större än vad som kalkylatorn eller datorn kan klara av.

Istället kan vi försöka utnyttja ett iterativt förfarande. Först beräknar vi sannolikheten att x = 0, med hjälp av detta resultatet beräknar vi sannolikheten att x = 1 osv.

Vi skapar en formel genom att skriva $P(X = x)$ och $P(X = x+1)$ som ett förhållande:

$$\frac{P(X=x+1)}{P(X=x)} = \frac{\binom{n}{x+1}\cdot p\cdot(1-p)^{n-(x+1)}}{\binom{n}{x}\cdot p\cdot(1-p)^{n-x}} = \frac{a-b\cdot x}{x+1}$$

Vi kan skriva om detta uttryck på följande sätt:

$$P(X=x+1) = \frac{a-b\cdot x}{x+1}\cdot P(X=x)$$

där $a = \dfrac{n\cdot p}{1-p}$ och $b = \dfrac{p}{1-p}$

Vi behöver nu ett startvärde för $P(X = 0)$. Vi kan välja vilket värde som helst t.ex. 100.

Antag nu att n = 20 och p = 0.12. Med hjälp av ovanstående formel får vi följande tabell:

x	0	1	2		
$P(X=x)$	100	272.727	353.306	...	0.005

Raden $P(X = x)$ beräknas tills resultatet är lågt, t.ex. 0.005 varefter raden summeras. Därefter divideras varje värde med denna summa varvid vi till slut får de önskade sannolikheterna.

Vad händer om *p* inte är konstant?

Vad händer med väntevärde och standardavvikelse om *p* inte är konstant? Vi skall här studera några enkla men grundläggande fall.

Fall A. Antag att vi har en maskin med tre spindlar som monterar komponenter på ett kretskort. De tre spindlarna monterar 15, 8 respektive 7 komponenter av exakt samma typ på kortet. Om vi räknar antal felaktigt monterade komponenter så kommer vi sannolikt att tro att detta antal följer en binomialfördelning med $n = 30$ och något p.

Men vad händer om de tre spindlarna har olika p? Vad blir då väntevärdet och standardavvikelsen för antal felaktigt monterade komponenter? Vi använder teorin om linjärkombinationer igen. Precis som tidigare kallar vi summan 'ej OK' för Y:

$$Y = \underbrace{X_1 + X_1 + \ldots}_{\text{15 komponenter}} + \underbrace{X_2 + X_2 + \ldots}_{\text{8 komponenter}} + \underbrace{X_3 + X_3 + \ldots}_{\text{7 komponenter}}$$

Detta ger oss ett uttryck för väntevärdet.

$$\mu_Y = \underbrace{p_1 + \ldots}_{\text{15 komp}} + \underbrace{p_2 + \ldots}_{\text{8 komp}} + \underbrace{p_3 + \ldots}_{\text{7 komp}} = \sum p_1 = n \cdot \overline{p}$$

Och för standardavvikelsen får vi följande:

$$\sigma_Y = \sqrt{\underbrace{p_1 \cdot (1 - p_1) + \ldots}_{\text{15 komponenter}} + \underbrace{p_2 \cdot (1 - p_2) + \ldots}_{\text{8 komponenter}}}$$

$$= \sqrt{n \cdot \left[\overline{p} \cdot (1 - \overline{p}) - \sigma_p^2\right]}$$

Här är $\overline{p}$ och σ_p medelvärdet respektive standardavvikelsen över de 30 felkvoterna.

En simulering av fall A

Om vi simulerar 1000 kretskort med komponenter monterade enligt fall A:

$$p_1 = 0.08 \qquad p_2 = 0.15 \qquad p_3 = 0.22$$
$$\overline{p} = 0.131 \qquad \sigma_p = 0.0579$$
$$\mu = 3.940 \qquad \sigma = 1.823$$
$$\overline{x} = 3.938 \qquad s = 1.795$$

Som synes överensstämmer det simulerade resultatet väl med den teoretiska härledningen.

Fall B. Antag att vi i stället har fyra maskiner som monterar komponenter på kretskort. Varje dag väljer vi slumpmässigt ut en maskin och avsynar ett kretskort. Antag att kortet innehåller 40 komponenter av en viss sort. Om vi räknar antal felaktigt monterade komponenter så kommer vi sannolikt att tro att detta antal följer en binomialfördelning med $n = 40$ och något p.

Men vad händer om de fyra maskinerna har olika p? Vad blir då väntevärdet och standardavvikelsen för antal felaktigt monterade komponenter?

Observera att vi i denna situation har en blandning av variabler. Vi får följande uttryck:

$$\begin{cases} \mu_Y = n \cdot \overline{p} \\ \sigma_Y = \sqrt{n \cdot \left[\overline{p} \cdot (1 - \overline{p}) + (n-1) \cdot \sigma_p^2\right]} \end{cases}$$

En simulering av fall B

Om vi simulerar 1000 kretskort slumpmässigt dragna från de fyra maskinerna får vi följande:

$$p_1 = 0.05 \qquad p_2 = 0.13 \qquad p_3 = 0.18 \qquad p_4 = 0.22$$
$$\overline{p} = 0.145 \qquad \sigma_p = 0.073$$
$$\mu = 5.80 \qquad \sigma = 3.65$$
$$\overline{x} = 5.85 \qquad s = 3.50$$

Som synes överensstämmer det simulerade resultatet väl med den teoretiska härledningen.

Ytterligare utvidgning

Man kan också tänka sig att p varierar enligt någon kontinuerlig fördelning. En dylik fördelning måste vara begränsad på intervallet 0 till 1 (p finns ju inte utanför dessa gränser). Det vanligaste valet brukar då vara den s.k. *betafördelningen*.

Kombinerar man detta med en binomialfördelning får vi en diskret fördelning som ibland kallas just *betabinomialfördelningen*. Många läroböcker i statistik går igenom denna fördelning.

Hur bedömer man huruvida ett data-material följer en binomialfördelning?

Att svara på rubrikens fråga kräver mer utrymme än vad som gives här. Det första man bör vara klar över är huruvida det rent teoretiskt kan tänkas vara en binomialfördelning. Uppfattningen att "antal dagar per order" är en binomialfördelad variabel bara för att ordet "antal" ingår i beskrivningen är ett villospår.

Om man har en verkligt stor datamängd, vilket man sällan har, kanske det räcker med att jämföra histogrammet med en teoretisk binomialfördelning.

Ett vanligt förslag är att man skall utföra ett så kallat 'Chitvå'-test. Där gör man en matematisk jämförelse mellan histogrammet och en teoretisk fördelning.

Den som vill fördjupa sig i problemet bör studera den statistiska teorin och leta under ord och rubriker såsom

"Chisquare", "Goodness of fit", "deviance" m.m.

Se t.ex. MODELLING BINARY DATA av D. Collett (Chapman & Hall)

THE 1ST INTERNATIONAL SYMPOSIUM ON INDUSTRIAL STATISTICS

UNDERSTANDING VARIATION:
A KEY TO SUCCESSFUL QUALITY IMPROVEMENT

19 - 21 August 1999 Stockholm, Sweden

The theme of this Symposium is that of the Shewhart program, i.e. principles addressed by Shewhart already in his book from 1931:

"...The first step of the engineer in trying to satisfy these wants is, therefore, that of translating as nearly as possible these wants into physical characteristics of the thing manufactured to satisfy these wants. In taking this step intuition and judgement play an important role as well as the broad knowledge of the human element involved in the wants of individuals.

The second step of the engineer is to set up ways and means of obtaining a product which differ from the arbitrarily set standards for these quality characteristics by no more than may be left to chance."

The aim of the conference is to promote research and utilization of statistical tools for quality improvement in industry. To achieve this the promotion of statistical tools has to be adequate and problem oriented and research on new statistical methods fulfilling quality improvement needs has to be institutionalized. Both areas will be addressed in the conference.

The 1st ISIS is held in Stockholm, Sweden in connection with the 52nd Session of the ISI in Helsinki, Finland. The symposium is organized by the Industrial Statistics committee of the ISI and the Division Quality Technology and Management at Linköping University will be acting as the local host.

For further information, please visit the First ISIS homepage at `box.ikp.liu.se/isis`.

Binomialfördelningen i massmedier - ett exempel

För den som vill skärpa sitt sinne för statistikt tänkande är våra massmedier en outsinlig källa. Där presenteras inte bara en massa sifferuppgifter utan även tvärsäkra påståenden. Inte sällan består dessa påståenden av en procentsats i samband med det ena eller det andra och så här i examenstider har betygsättningen granskats.

När massmedier presenterar nyheter inom medicin eller inom statsvetenskap eller inom ekonomi finns ofta ett uttalande eller en referens till en eller annan 'forskare' (om det är i TV finns denne naturligtvis med i studion).

Denne person ger vanligtvis en professionell syn på den aktuella frågan och tillfogar ofta ett eller annat alternativt synsätt på problemet. Det hela ger naturligtvis en viss glans åt reportaget, en vilja att ge saken en allsidig belysning.

När det gäller siffror är det naturligtvis mycket enklare. Då är det ju bara att beräkna en procentsats, lätt som en plätt, samt att skriva en dräpande eller syrlig eller ironisk kommentar, alltefter tycke och smak.

(I genomsnitt vart 3-4 år finns det någon professionell person som kommenterar siffrorna på TV-skärmen.)

I massmedierna, och på många av våra företag, finns nämligen inte osäkerhet. Varje beräknad siffra är en otvivelaktig sanning som kan kommenteras och tolkas i det oändliga.

Ett exempel på senare tid är Sveriges framgång i grenen matematik. Där blev vi bäst! Om man tittar på ett diagram över de olika nationernas resultat undrar man dock vad som skulle ha hänt om det hela hade upprepats med andra människor vid ett annat tillfälle. Hade Sverige fortfarande varit bäst? Men att undra detta, då resultatet redovisades, hade väl varit som att svära i kyrkan.

Nu i början på juni får vi ytterligare data som handlar om kunskap. I nyhetssändningar på radion läser man upp resultatet i matematik för niondeklasser. Man angav att alltifrån 0 till 20 % av eleverna var underkända.

Medelnivån låg på 4.9 % enligt radion. Antag att man med ovanstående menar att p = 0.049. Siffrorna 0 - 20 % måste vara observerade värden dvs det vi brukar skriva p-hatt ().

Med hjälp av fördelningsfunktionen för binomialfördelningen kan vi skapa en viss insikt om siffrorna. Vi behöver dock ett n, dvs storleken på en skolklass i nian. Antag att n = 25 respektive 30. (Vänster respektive höger kolumn.)

$$P(X \leq x) = \sum_{0}^{x} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

x	$P(X \leq x)$	$P(X \leq x)$
0	0.2848	0.2215
1	0.6516	0.5639
2	0.8784	0.8198
3	0.9680	0.9428
4	**0.9934**	0.9856
5	0.9989	**0.9970**
6	0.9999	0.9995
7	1.0000	0.9999
8	1.0000	1.0000

Sannoliketen att få 0 % underkända är ganska hög, 0.28 respektive 0.22. Att få 20 % (eller mer) under- kända i en klass krävs åtminstone 5 respektive 6 underkända elever. Vi kan beräkna denna sannolikhet på följande sätt:

$$1 - P(X \leq 4) = 1 - 0.9934 \approx 0.007$$
$$1 - P(X \leq 5) = 1 - 0.9970 \approx 0.003$$

Kommentar

Eftersom vi inte vet exakta storlekar på klasserna i nian, har vi använt n = 25 respektive 30. Om vi skulle beskriva den naturliga variationen av 'ej godkänd' skulle vi antagligen ange 0-5 respektive 0-6 'ej godkända', att döma av sannolikhetsberäkningarna.

Om vi använder den vanliga tumregeln

$$\mu \pm 3 \cdot \sigma$$

för att approximativt beräkna omfattningen av den total variationen får vi

$$25 \cdot 0.049 \pm 3 \cdot \sqrt{25 \cdot 0.049 \cdot (1 - 0.049)}$$

respektive

$$30 \cdot 0.049 \pm 3 \cdot \sqrt{30 \cdot 0.049 \cdot (1 - 0.049)}$$

Detta ger intervallen [0 till 4.46] och [0 till 5.02]. Tumregeln ovan passar bäst för kontinuerliga fördelningar men binomialfördelningen är en diskret fördelning så intervallet borde ju skrivas i heltal t.ex. [0 till 4] och [0 till 5].

Beräkningarna visar alltså att händelsen '20 % underkända i en klass' kan inträffa (även om sannolikheten är liten). I varje fall borde det inte ge upphov till en sådan indignerad rubrik som

"Stor skillnad på betyg i olika skolor"

Ytterligare synpunkter

I en tidningsartikel om just detta problem påstås det att av 95 695 elever undersöktes 66 procent. (95 695 elever låter rimligt. Häromdagen nämnde radion att det dött cirka 93 000 människor förra året så det verkar vara balans.) Om vi använder klasstorleken n = 30 får vi

$$\frac{0.66 \cdot 95695}{30} \approx 2105$$

klasser i landet. Med sannolikheten att få 6 eller fler (dvs 20 %) underkända är 1 - 0.9970 = 0.0030 kan vi använda binomialfördelningen igen.

Vi kan alltså betrakta det hela som ett nytt stickprov ur en process där n = 2105 och p = 0.0030. Vad är sannolikheten att få åtminstone en klass med 20 % 'ej godkända'? Vi använder samma formel som tidigare ($P(X \leq 0)$ är sannolikheten för *ingen* klass med 20 % 'ej godkända'):

$$1 - P(X \leq 0) = 1 - 0.0018 \approx 0.998$$

Det är alltså en ganska stor sannolikhet att få åtminstone en klass med 20 % 'ej godkända'.

I den analys som beskrivs här har vi inte haft tillgång till hela datamaterialet. Vi har antagit att p = 0.049 samt att n = 25 respektive 30 och utnyttjat lite statistisk teori. Om t.ex. klasstorleken varierar blir variationen och därmed tolkningen annorlunda.

Vi vet inte hur materialet fördelat sig. Kanske har det varit många '0-procentare' och lite för många "högprocentare' och ändå givet ett medelvärde på 4.9 procent. Kanske hade en analys avslöjat att utfallet inte följer en enda binomialfördelning utan att det är en blandning av många olika p-värden. Och kanske är det rent av troligt att det är så, det är kanske otroligt om denna 'undervisningsprocess' är lika över hela landet med skiftande kommunala satsningar, olika förutsättningar, varierande kompetens hos lärare, etc etc.

"Olika tolkningar av betygsreglerna"

"Resultatet visar att skolan inte är lika över hela landet"

"Man blir orolig utifrån ett likvärdighetsperspektiv"

"Oavsett var man bor ska man ha lika chanser"

Vi kan åtminstone konstatera att det är alltför lättvindigt att utifrån enkla tabeller med procentsatser dra (och publicera) långtgående slutsatser som inte direkt ökar förståelsen av ett viktigt problem. Observera att vi inte anser att 4.9 % underkända är ett acceptabelt resultat eller något slags normalnivå. Vad vi vänder oss emot är den ytliga analysen av något som angår oss alla.

p som en funktion av väntevärde och varians

Väntevärdet och standardavvikelsen för en binomialfördelning har tidigare beskrivits med följande två uttryck:

$$\begin{cases} \mu = n \cdot p \\ \sigma = \sqrt{n \cdot p \cdot (1-p)} \end{cases}$$

Det behöver ju inte sägas att både μ och σ är positiva tal (dvs inte negativa eller noll) och att n är stickprovets storlek, ett heltal. Med lite manipulering av dessa två formler kan vi skriva parametern p som en funktion av μ och σ:

$$\sigma^2 = \mu \cdot (1-p)$$

$$\sigma^2 = \mu - \mu \cdot p$$

$$p = 1 - \frac{\sigma^2}{\mu}$$

Några slutsatser

1. σ^2 måste vara mindre än μ annars blir ju p negativt
2. om $p = 0.5$ måste σ^2 vara hälften så stort som μ
3. om p är nära 1 måste σ^2 vara litet i förhållande till μ
4. om p är nära 0 måste σ^2 vara nära μ

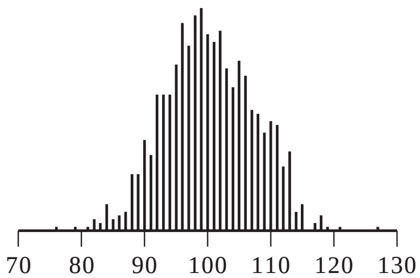
Det fjärde påståendet riktar oss mot en annan oerhört viktig och central del av statistikteorin. Nämligen approximationen av binomialfördelningen med Poissonfördelningen.

Den senare är en annan 'händelsefördelning' av typen 'antal telefonsamtal per 60 minuter', 'antal lackfel per ytenhet', 'antal partiklar per volymsenhet' osv.

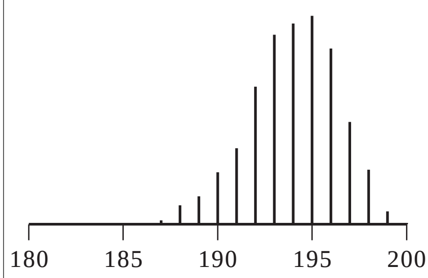
Studerar man Poissonfördelningens egenskaper finner man att $\mu = \sigma^2$ och läroböckerna brukar ange att om p är litet så kan man approximera en binomialfördelning med en Poissonfördelning.

Och det framgår tydligt av slutsats 4 att p är litet endast då σ^2 är nära μ.

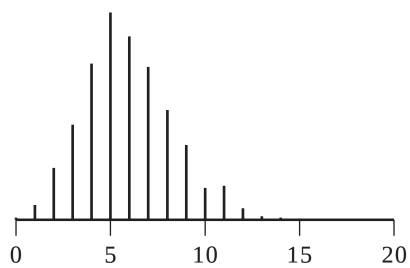
Nedan har vi simulerat tre olika datamängder för att belysa de delar som här diskuterats. Varje histogram innehåller 1000 värden.



$p = 0.5$
$n = 200$
$\bar{x} = 100.3$
$s^2 = 50.8$



$p = 0.97$
$n = 200$
$\bar{x} = 194.0$
$s^2 = 5.7$



$p = 0.03$
$n = 200$
$\bar{x} = 5.9$
$s^2 = 5.7$

Genom att studera siffrorna till höger om histogrammen finner man att påståendena 1-4 stämmer väl.

Ingemar Sjöström
Ericsson Quality Management Institute

Inbjudan till SFK-StaM seminarium 1998
fredagen den 6 november i Stockholm
”Analyser av omvärld och processer”

Förändringstakten i världen har ökat, leverantörer vill delta i kunders värdeskapande processer, kunder kräver kortare och säkra leveranstider från leverantörer. Då behövs systematisk metodik för att analysera omvärlden, kunder samt egna processer. Seminariet kommer att ge inspiration och exempel på olika tillvägagångssätt.

Välkomna!

09:00 **Registrering och kaffe**

09:40 **Inledning**
Ordförande Mats Franzén

09:45 **Omvärldsanalys eller business intelligence - vad är det?**
Niklas Lundblad, forskare och projektledare vid IPF i Uppsala
Omvärlden är inte längre vad den brukade vara. Från att ha framstått som relativt stabil och överblickbar tycks omvärlden bli allt mer föränderlig och framtiden allt mer osäker. Mot den bakgrunden framstår spaning allt tydligare som ett överlevnadskriterium.

10:30 **Paus med frukt**

10:45 **Omvärldsanalys i praktiken**.
Anders Baudin, rektor vid Handelshögskolan i Umeå, Universitetslektor och docent i statistik
Här diskuteras vilka faktorer man bör ta upp i bevakning av omvärlden till ett företag

11:30 **Lunch**

12:15 **Sektionsmöte med sedvanliga årsmötesförhandlingar**

13:00 **Hur ska man presentera fakta?**
Ej klart med föreläsare

13:45 **Kund-, medarbetar- och verksamhetsanalys i ett företag.**
Mats Franzén, kvalitetsäkringschef hos Ericsson Cables AB, Telekabel Divisionen
Metoder i praktiken för att utveckla sin verksamhet med fokus på kunden, medarbetarna och verksamheten.

14:30 **Paus med frukt**

14:45 **Samband mellan kundkrav, produktionsförhållanden och ekonomi**
Olle Carlsson, docent i statistik vid Handelshögskolan i Göteborg och vid högskolan i Örebro

15:30 **Avrundning och diskussion**

16:00 **Slut**

Plats: ERICSSON, Konferensanläggningen i Älvsjö, Götalandsvägen 230, Stockholm

Avgift: Utgår med 2500kr, studerade med giltigt kårlegitimation ges 1000kr rabatt. Lunch och kaffe samt seminariedokumentation ingår i kostnaden. Avgiften betalas via inbetalningskort som bifogas bekräftelsen.

Anmälan görs snarast till: Mats Franzén,
tel 0650-36161
fax 0650-36064
E-post mats.franzen@eca.ericsson.se
eller till någon annan styrelsemedlem i SFK-StaM

OBS! Anmälan är bindande

SFK-StaMs styrelse

Ordförande
Mats Franzén
Ericsson Cables AB
824 82 Hudiksvall
tel 0650-361 61
fax 0650-360 64
mats.franzen@eca.ericsson.se

Sekreterare
Sture Rydin
Stora Cell AB
814 81 Skutskär
tel 026-851 07
fax 026-854 00
srn@swipnet.se

Kassör
Per Persson
Linköpings Universitet
Kvalitetsteknik/IKP
581 83 Linköping
tel 013-28 23 93
fax 013-28 27 42
perp2@ikp.liu.se

Ledamöter

Lennart Nilsson
Umeå Universitet
Inst för mat stat
901 87 Umeå
tel 090-786 60 77
fax 090-786 76 58
ln@matstat.umu.se

Ann Brännström-Stenberg
Högskolan i Karlstad
Inst för teknik
651 88 Karlstad
tel 054-83 82 36
fax 054-83 84 49
ann.brannstrom-stenberg@hks.se

Johannes Forkman
Pharmacia-Upjohn Diagnostics
F35-2
751 82 Uppsala
tel 018-16 36 78
fax 018-16 63 90
johannes.forkman@eu.pnu.com

Valberedning

Susanna Weinberger
Ovako Steel AB
813 82 Hofors
tel 0290-252 96
fax 0290-259 87
seskfws5@ibmmail.com

Göran Lande
Ericsson Radio Systems AB
824 82 Stockholm
tel 08-
fax 08-
goran.lande@era.ericsson.se

Redaktör
Ingemar Sjöström
Ericsson Quality Management Institute
125 82 Stockholm
tel 08-721 75 05
fax 08-726 31 21
ingemar.sjostrom@eqmi.ericsson.se

Revisiorer

Auli Sarlin Gilljan
Kreativ Kvalitet
Flugsvampsvägen 101
141 60 Huddinge
tel 070-482 89 93
auli@komvux.tyreso.se

Anders Hynén
ABB Corporate Research
Avd R
721 78 Västerås
tel 021-32 32 24
fax 021-14 21 90
anders.hynen@secrrc.abb.se

Redaktionskommitté
Mats Franzén
Lennart Nilsson
Ingemar Sjöström

Bidrag accepteras gärna via 3.5"-diskett med textmängden i format WordPerfect, Word e.d.

Man blir medlem i SFK–StaM genom att kontakta Svenska Förbundet för Kvalitet telefon 08 – 783 82 54 eller 783 01 71 (fax 661 1967). Kanslisekreterare är Berit Winnsjö.

Några hemsidor

http://www.sfkvalitet.se

http://www.sfkvalitet.se/sektion/sfk-stam.htm