

S0008M – sammanfattning inför tentan

19 okt. 2026, 31 min läsning

UTKAST

#statistik #sannolikhetslära #sammanfattning #tenta

Kurs: S0008M — Sannolikhetslära och statistik

Grundtanken

Del 1 på tentan består av ungefär tolv korta uppgifter, och de kommer i nästan samma ordning varje gång. Den här texten går igenom dem i den ordningen. För varje uppgiftstyp får du veta vad den handlar om, hur du känner igen den, vad du gör steg för steg och vilka fällor som brukar kosta poäng. Texten är skriven för att lyssnas på: orden bär förklaringen, och formlerna står med för att du ska kunna slå upp dem efteråt.

1. Hur tentan ser ut och hur du använder texten

Tentan har två delar. Del 1 har ungefär tolv uppgifter som tillsammans ger tjugofem poäng, och för betyget tre behöver du sjutton av dem. Del 2 rättas bara om del 1 är godkänd. Den här sammanfattningen handlar om del 1, eftersom det är den som avgör om du klarar kursen.

Det viktigaste att veta om del 1 är att den är förutsägbar. Uppgift ett handlar nästan alltid om händelser, uppgift två om Bayes sats, uppgift tre till fem om en enskild fördelning, uppgift sex och sju om centrala gränsvärdessatsen, uppgift åtta till tio om statistisk inferens och uppgift elva och tolv om två slumpvariabler samtidigt. Figuren nedan visar den kartan. Om du vet var du är på kartan vet du också ungefär vilket verktyg som ska fram.

1	händelser: union, snitt, komplement	
2	Bayes sats, total sannolikhet	
3	en fördelning: binomial, Poisson ...	
4	en fördelning: exponential, rektangel ...	sannolikhetslära
5	normal ur kvantiler, väntevärde och varians	
6	centrala gränsvärdessatsen	
7	centrala gränsvärdessatsen	
8	test: kritisk gräns, P-värde	
9	konfidsintervall, oftast stickprov i par	statistisk inferens
10	styrka, teckenintervall, stickprovsstorlek	
11	två variabler: area, kovarians, 2D-normal	två variabler
12	två variabler: area, kovarians, 2D-normal	

Lyssna gärna på ett avsnitt i taget och försök sedan själv med en gammal tentauppgift av samma typ. Exempelen i texten är hopfällda, så att du kan läsa uppgiften först och öppna lösningen efteråt.

2. Lösningssmallen som ger delpoäng

Innan vi går in på uppgiftstyperna: varje lösning i del 1 bör följa samma fem steg. Läraren har sagt att den vanligaste orsaken till att delpoäng uteblir är att slumpvariabeln inte är definierad i ord. Ett rätt svar ger full poäng, men ett fel svar utan uppställning ger noll. Med en tydlig uppställning kan du få poäng även när du räknat fel.

Steg ett är att definiera slumpvariabeln i ord. Skriv till exempel: ξ är antalet trasiga muttrar i urvalet. Eller: ξ är tiden tills nästa kund kommer, mätt i minuter. Meningen ska säga vad som räknas eller mäts och i vilken enhet.

Steg två är att ange fördelningen med parametrar, till exempel $\xi \in \text{Bin}(20; 0,05)$. I kursen skriver man tecknet "tillhör", alltså $\in$, mellan slumpvariabeln och fördelningen. När en parameter är ett decimaltal skiljer man parametrarna med semikolon, så att decimalkommat inte förväxlas med skiljetecknet.

Steg tre är att skriva vad som ska beräknas, till exempel $P(\xi \leq 3)$.

Steg fyra är att räkna. Det räcker att skriva: jag beräknade $P(\xi \leq 3)$ med räknaren och fick 0,02275. Du behöver inte skriva vilket räknarkommando du använde, men uttrycket måste stå där.

Steg fem är att svara exakt så som uppgiften ber om. Står det "svara i procent med två decimaler" ska svaret vara till exempel 55,43 %, inte 55,4 %. Avrunda bara i det sista steget, aldrig i mellanleden.

Lämna alltid in lösningen

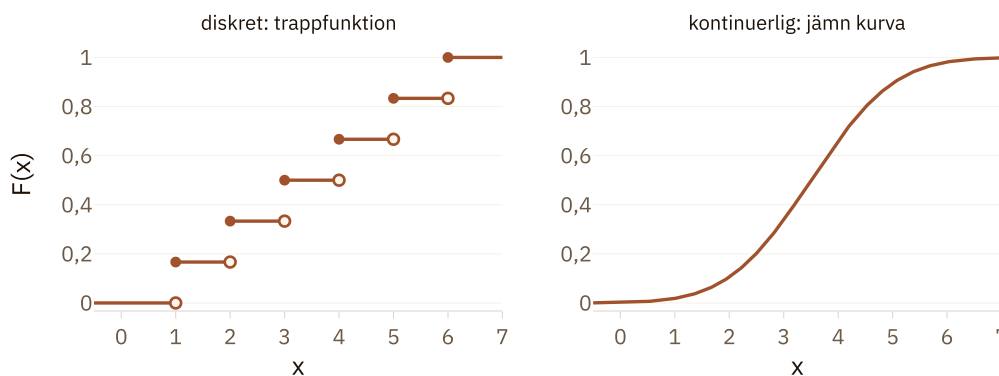
Även i del 1, där bara svaret rättas, ger en uppställning med definierad slumpvariabel chansen till delpoäng.

3. Grundbegreppen som allt vilar på

Innan uppgiftstyperna kommer några begrepp som återkommer i nästan varje uppgift. Om de sitter blir resten mycket lättare.

En slumpvariabel är ett tal som bestäms av ett slumpmässigt försök. Antalet krona vid tio slantsinglingar är en slumpvariabel, liksom tiden tills nästa buss kommer. I kursen betecknas slumpvariabler med grekiska bokstäver, oftast ξ , som uttalas ksi, och η , som uttalas eta. Ett konkret värde som variabeln antar skrivs med vanlig bokstav, till exempel x .

Slumpvariabler är antingen diskreta eller kontinuerliga. En diskret variabel antar enskilda värden som du kan räkna upp, typiskt heltal. Den beskrivs av en sannolikhetsfunktion, som för varje värde anger sannolikheten att variabeln antar just det värdet. Summan av alla sannolikheter är ett. En kontinuerlig variabel kan anta alla värden i ett intervall, och sannolikheten för ett enskilt exakt värde är noll. Den beskrivs i stället av en frekvensfunktion, där sannolikheten för ett intervall är arean under kurvan. Hela arean är ett.



Fördelningsfunktionen $F_\xi(x)$ finns för båda sorterna. Den anger sannolikheten att variabeln är mindre än eller lika med x . Den börjar på noll längst till vänster, växer och slutar på ett. Det är den som räknarens kumulativa kommandon räknar ut, och därför vill räknaren alltid ha "mindre än eller lika med".

Väntevärdet $E(\xi)$ är det värde variabeln antar i genomsnitt i det långa loppet. Föreställ dig att försöket upprepas tusentals gånger och att du tar medelvärde av alla utfall. För en diskret variabel är väntevärdet summan av varje värde gånger dess sannolikhet. För en kontinuerlig variabel är det motsvarande integral.

Variansen $V(\xi)$ mäter hur mycket värdena sprider sig kring väntevärdet. Den är väntevärdet av den kvadrerade avvikelser från väntevärdet. Eftersom variansen har enheten i kvadrat använder man ofta standardavvikelsen $D(\xi)$, som är roten ur variansen och har samma enhet som variabeln själv. I kursen skrivs väntevärdet E , variansen V och standardavvikelsen D .

Två slumpvariabler är oberoende om utfallet av den ena inte säger något om den andra. Oberoende är ett antagande som uppgiften måste ge dig, med ord som “oberoende av varandra” eller ett sammanhang där det är uppenbart, till exempel olika tärningskast. Många av formlerna i kursen gäller bara vid oberoende, och det är en vanlig felkälla att använda dem ändå.

4. Räknaren på tentan

Du får använda TI-Nspire på tentan, och den gör nästan alla beräkningar åt dig. Det viktiga är att veta vilket kommando som hör till vilken fördelning och i vilken ordning argumenten ska skrivas. Fördelningarna finns i menyn Statistics och sedan Distributions.

För binomialfördelningen finns ett kommando för sannolikheten för exakt ett värde, med argumenten antal försök, sannolikhet och värdet. Det finns också ett kumulativt kommando som tar en undre och en övre gräns. Poissonfördelningen fungerar likadant, men med λ i stället för antal försök och sannolikhet. För normalfördelningen anger du undre gräns, övre gräns, väntevärde och standardavvikelse, och oändligheten finns i symbolpaletten. Den inversa normalfördelningen och den inversa t -fördelningen tar arean till vänster, och t -fördelningen dessutom antalet frihetsgrader.

För mätdata matar du in värdena i en lista och kör One-Variable Statistics. Utskriften ger medelvärdet och två standardavvikelser. Den som kallas s_x delar med $n - 1$ och är den som kursen kallar s . Den som kallas σ_x delar med n och ska inte användas. Vid stickprov i par skapar du en tredje kolumn med differenserna och kör statistiken på den.

5. Uppgift 1: händelser

Uppgift ett ger dig några sannolikheter för händelser, oftast två händelser A och B , och frågar efter sannolikheten för en kombination av dem. Det är den enklaste uppgiften på tentan, och den kommer varje gång.

Grundverktygen är tre. Det första är komplementet: sannolikheten att A inte inträffar, som skrivs $P(A^c)$, är ett minus sannolikheten att A inträffar. Det andra är unionen, alltså att minst en av händelserna inträffar. Den är summan av de två sannolikheterna minus snittet, eftersom snittet annars räknas två gånger:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

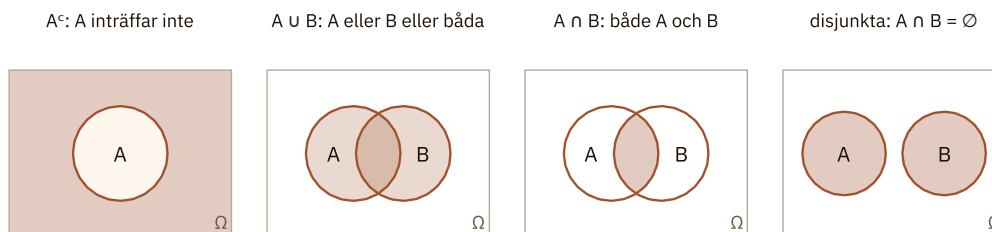
Det tredje är betingning: sannolikheten för A givet att B har inträffat är snittet delat med sannolikheten för B . Man krymper världen till B och frågar hur stor del av den som också är A .

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Nästan alla varianter av uppgiften är kombinationer av de tre. "Varken A eller B " betyder att ingen av dem inträffar, och det är komplementet till unionen. "Exakt en av dem" är unionen minus snittet, alltså $P(A) + P(B) - 2P(A \cap B)$. "Minst en av n oberoende händelser" är lättast att räkna baklänges: ett minus sannolikheten att ingen inträffar, och om händelserna är oberoende är det produkten av komplementen.

Om händelserna är oberoende får du multiplicera: $P(A \cap B) = P(A)P(B)$. Det gäller bara vid oberoende, och uppgiften säger det i så fall uttryckligen. Om händelserna är beroende måste snittet ges i uppgiften, eller räknas ut via betingning: $P(A \cap B) = P(B)P(A \mid B)$.

Om alla utfall är lika troliga, som för tärningar och mynt, gäller den klassiska definitionen: antalet gynnsamma utfall delat med antalet möjliga. Två tärningar har trettiosex möjliga utfall.



≡ Exempel 1 – varken A eller B (tentan 2025-03-17, uppgift 1)

För händelserna A och B gäller att $P(A) = 0,3$, $P(B) = 0,56$ och $P(A \cap B) = 0,14$. Beräkna $P(A^c \cap B^c)$ i procent med en decimal.

Lösning

"Varken A eller B " är komplementet till "minst en av dem". Först unionen: $P(A \cup B) = 0,3 + 0,56 - 0,14 = 0,72$ Sedan komplementet: $P(A^c \cap B^c) = 1 - 0,72 = 0,28$, alltså 28,0 %.

Fällor i uppgift 1

Glöm inte att dra bort snittet i unionen. Multiplicera inte sannolikheter om uppgiften inte säger att händelserna är oberoende. Och läs noga om frågan gäller "givet att", för då ska du dela med sannolikheten för villkoret.

6. Uppgift 2: Bayes sats och total sannolikhet

Uppgift två handlar om att vända på en betingning. Typiskt vet du hur stor andel av produkterna som kommer från olika fabriker, och hur stor andel av varje fabriks produkter som är defekta. Frågan är sedan: givet att en produkt är defekt, hur troligt är det att den kommer från fabrik A? Samma struktur finns i medicinska test, där du vet hur ofta testet slår ut hos sjuka och friska, och frågan är hur troligt det är att en person med positivt test faktiskt är sjuk.

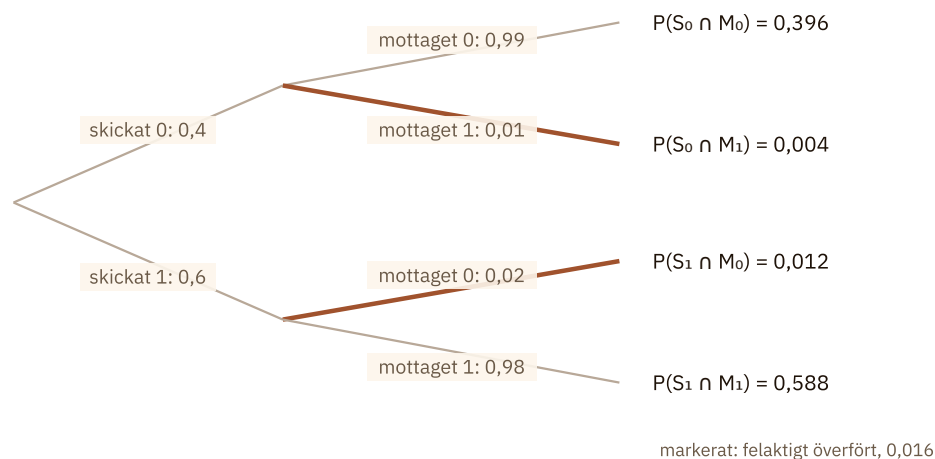
Du löser den i två steg. Först räknar du ut den totala sannolikheten för det du har observerat, till exempel att en produkt är defekt. Det gör du genom att gå igenom varje källa: sannolikheten att produkten kommer från källan gånger sannolikheten att den är defekt givet källan, summerat över alla källor. Det är lagen om total sannolikhet:

$$P(A) = P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2) + \dots + P(B_n)P(A \mid B_n)$$

Sedan använder du Bayes sats: den del av den totala sannolikheten som kommer från just den källa du frågar om, delat med hela den totala sannolikheten.

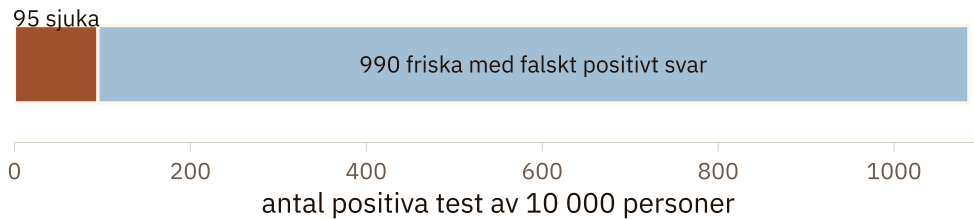
$$P(B_i \mid A) = \frac{P(B_i)P(A \mid B_i)}{P(A)}$$

Ett trädidiagram är det säkraste sättet att hålla ordning. Den första förgreningen är källorna, den andra är utfallet. Varje gren längst ut har en sannolikhet som är produkten längs vägen. Täljaren i Bayes sats är en enda sådan gren, nämnaren är summan av alla grenar som slutar i det observerade utfallet.



Ett annat bra sätt att tänka är med naturliga frekvenser. Föreställ dig tiotusen personer och räkna hur många som hamnar i varje gren. Då blir det uppenbart varför ett test med hög träffsäkerhet ändå kan ge många falska larm när sjukdomen är ovanlig: de friska är så många fler att även en liten andel felaktiga positiva svar blir stor i antal.

$$P(\text{sjuk} \mid +) = 95/1085 \approx 0,088$$



≡ Exempel 2 – vilken signal skickades? (tentan 2025-03-17, uppgift 2)

Ett utsänt tecken 0 tas emot som 1 med sannolikheten 0,02. Ett utsänt tecken 1 tas emot som 0 med sannolikheten 0,035. Sextio procent av alla utsända tecken är ettor. Om en etta tagits emot, vad är sannolikheten att en etta skickades? Svara i procent med två decimaler.

Lösning

Låt S_1 vara att en etta skickades och M_1 att en etta togs emot. Total sannolikhet: $P(M_1) = 0,6 \cdot 0,965 + 0,4 \cdot 0,02 = 0,579 + 0,008 = 0,587$ Bayes sats: $P(S_1 \mid M_1) = \frac{0,579}{0,587} = 0,98637 \approx \boxed{98,64 \%}$

Fällor i uppgift 2

Betinga åt rätt håll: uppgiften ger dig $P(\text{defekt} \mid \text{fabrik})$ men frågar efter $P(\text{fabrik} \mid \text{defekt})$. Kontrollera att källornas sannolikheter summerar till ett. Och räkna inte med tal som inte behövs, uppgiften innehåller ibland överflödig information.

Ibland dyker Bayes upp inne i en fördelning, till exempel: givet att en komponent har hållit i minst fyra år, vad är sannolikheten att den håller i minst sju? Det är samma betingningsformel. Snittet av “minst sju” och “minst fyra” är bara “minst sju”, så kvoten blir sannolikheten för minst sju delat med sannolikheten för minst fyra.

7. Uppgift 3 till 5: välj rätt fördelning

Uppgift tre, fyra och fem handlar oftast om en enda slumpvariabel och en känd fördelning. Det svåra är sällan räkningen, eftersom räknaren gör den. Det svåra är att känna igen vilken fördelning det är. Ställ dig två frågor.

Den första frågan är: räknar jag ett antal, eller mäter jag ett värde? Ett antal är ett heltal som noll, ett, två och så vidare. Ett värde är en mätning som kan anta alla tal i ett intervall, till exempel en tid, en längd eller en vikt.

Om du räknar ett antal finns fyra kandidater. Räknar du hur många av n oberoende försök som lyckas, och sannolikheten är densamma varje gång, är det binomialfördelningen. Drar du utan att lägga tillbaka ur en liten hög, så att sannolikheten ändras efter varje dragning, är det den hypergeometrisk fördelningen. Räknar du hur många händelser som inträffar under en viss tid, till exempel samtal per timme, är det Poissonfördelningen. Räknar du hur många försök som behövs till och med det första lyckade, är det för-första-gången-fördelningen, som i kursen kallas ffg-fördelningen.

Om du mäter ett värde finns också fyra kandidater. En väntetid utan minne, där det inte spelar någon roll hur länge man redan väntat, är exponentialfördelad. Ett värde som är lika troligt överallt på ett intervall är rektangelfördelat. Ett mätvärde som sprider sig symmetriskt kring ett medelvärde är normalfördelat. Och en summa eller ett medelvärde av många oberoende variabler är ungefär normalfördelat enligt centrala gränsvärdessatsen.



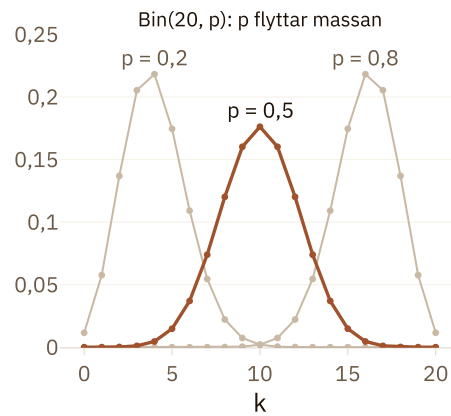
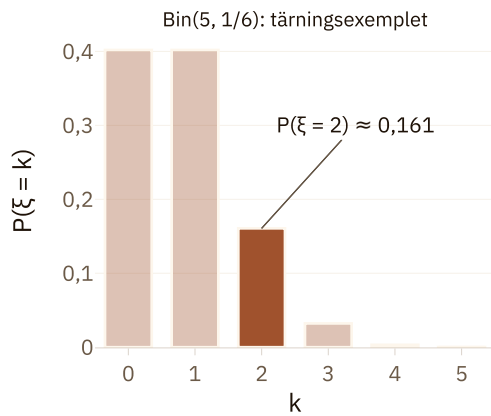
Väntevärde och varians för alla de vanliga fördelningarna står på formelbladet, så dem behöver du inte kunna utantill. Det du behöver kunna är att välja fördelning och ställa upp vad som ska räknas.

8. Binomial och hypergeometrisk fördelning

Binomialfördelningen beskriver antalet lyckade försök bland n försök, när varje försök lyckas med samma sannolikhet p och försöken är oberoende. Skriv: ξ är antalet lyckade bland de n försöken, $\xi \in \text{Bin}(n, p)$.

På räknaren finns två kommandon. Det ena ger sannolikheten för exakt ett värde, till exempel exakt tre lyckade. Det andra ger summan av sannolikheterna mellan två gränser. Räknaren räknar bara “mindre än eller lika med”, så “minst k ” måste vändas: sannolikheten för minst tre är ett minus sannolikheten för högst två. Det vanligaste felet är att skriva högst tre i stället för högst två.

$$P(\xi \geq k) = 1 - P(\xi \leq k - 1)$$

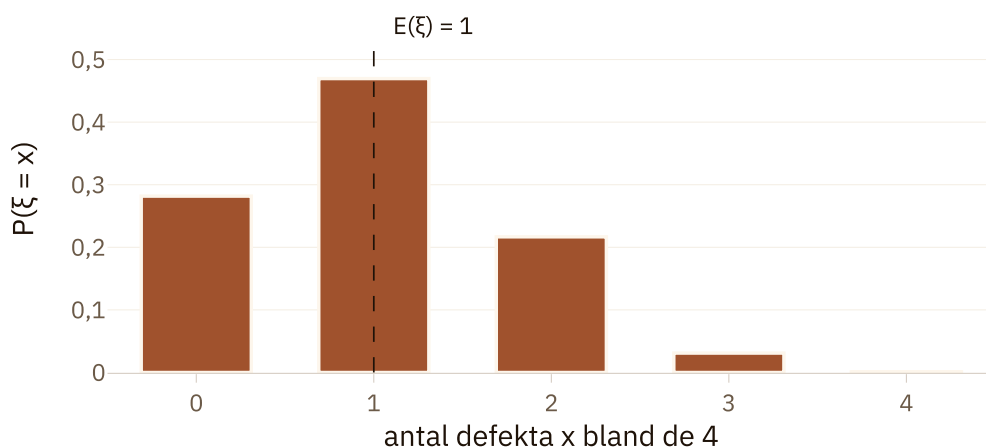


Ibland står inte p direkt i uppgiften. Då räknar du först ut p som sannolikheten att ett enskilt försök lyckas, och använder den sedan i binomialfördelningen. Det kallas två steg. Ett lyckat försök kan till exempel vara att två oberoende egenskaper inträffar samtidigt, och då är p produkten av de två sannolikheterna. Eller så är ett lyckat försök att en exponentialfördelad livslängd överstiger en viss tid, och då är $p = e^{-\lambda t}$.

Det mest sannolika värdet för en binomialfördelad variabel är heltalsdelen av $(n + 1)p$. För Bin(11; 0,15) blir det heltalsdelen av 1,8, alltså ett.

Den hypergeometrisk fördelningen används när du drar utan återläggning ur en ändlig mängd, till exempel kort ur en kortlek eller bollar ur en låda. Den har tre parametrar: hur många som finns totalt, hur många du drar, och andelen lyckade från början. Sannolikhetsfunktionen, väntevärdet och variansen står på formelbladet, och på räknaren räknar du binomialkoefficienterna med kommandot nCr. Om du drar en liten del av en stor mängd, mindre än en tiondel, kan du approximera med binomialfördelningen.

Hyp(20, 4, 0,25): 4 dras ur 20, varav 5 defekta



≡ Exempel 3 — exakt tre av tolv (två steg)

Ett jordbruk odlar ekologiskt med sannolikheten 0,61 och har mjölkkor med sannolikheten 0,87, oberoende av varandra. Man tittar på tolv jordbruk. Vad är sannolikheten att exakt tre av dem både odlar ekologiskt och har mjölkkor? Svara i procent med två decimaler.

 Lösning

Ett lyckat jordbruk gör båda sakerna, så $p = 0,61 \cdot 0,87 = 0,5307$. Låt ξ vara antalet sådana jordbruk av de tolv, $\xi \in \text{Bin}(12; 0,5307)$. $P(\xi = 3) = \binom{12}{3} 0,5307^3 (1 - 0,5307)^9 \approx 0,03631 = \boxed{3,63 \%}$ På räknaren: `binomPdf(12, 0.5307, 3)`. Att de övriga nio "kombinerar på något annat sätt" ingår redan: de är de misslyckade försöken.

≡ Exempel 4 — röda bollar utan återläggning (tentan 2025-03-17, uppgift 5)

I en låda finns 29 bollar, varav 8 är röda. Man drar 4 utan återläggning. Vad är standardavvikelsen för antalet röda bollar? Svara med två decimaler.

 Lösning

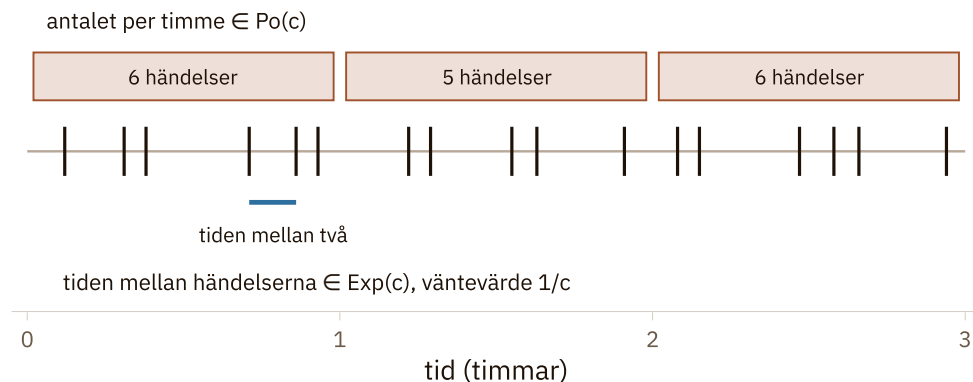
ξ är antalet röda bland de fyra, $\xi \in \text{Hyp}(29; 4; 8/29)$. Formelbladets varians: $V(\xi) = np(1-p) \frac{N-n}{N-1} = 4 \cdot \frac{8}{29} \cdot \frac{21}{29} \cdot \frac{25}{28} \approx 0,713$ Standardavvikelsen är roten: $D(\xi) \approx \boxed{0,84}$.

9. Poisson och exponential: två sidor av samma process

Poissonfördelningen beskriver hur många händelser som inträffar under en viss tid, när händelserna inträffar oberoende av varandra och med en jämn intensitet. Parametern λ är det förväntade antalet under den tid du tittar på, alltså intensiteten gånger tiden. Om det i medel kommer sex kunder per timme och du tittar på en halvtimme är $\lambda = 3$. Väntevärdet och variansen är båda λ , så om uppgiften ger standardavvikelsen är λ standardavvikelsen i kvadrat.

En summa av oberoende Poissonvariabler är också Poissonfördelad, och parametrarna adderas. Det betyder att du ofta kan räkna exakt med räknarens Poissonkommando även när det handlar om en summa av många variabler. När λ är stor går det också att approximera med normalfördelningen, där både väntevärde och varians är λ . Villkoren för approximationen står på formelbladet. Facit har räknat på båda sätten på olika tentor, så räkna helst exakt och skriv gärna att du gjort det.

Exponentialfördelningen beskriver väntetiden till nästa händelse i samma slags process. Det är kopplingen som är lätt att missa. Om händelserna kommer med intensiteten c per timme, så att antalet per timme är $Po(c)$, är tiden mellan två händelser exponentialfördelad med parametern c och väntevärdet $1/c$ timmar. Sex block per timme betyder alltså i genomsnitt tio minuter mellan blocken.



För exponentialfördelningen gäller att sannolikheten att vänta längre än x är $e^{-\lambda x}$. Det är den formel du använder mest. Fördelningsfunktionen är ett minus den. Medianen, alltså den tid som hälften av alla väntetider understiger, är $\ln 2$ delat med λ , och den är alltid kortare än väntevärdet.

Exponentialfördelningen är glömsk: om du redan väntat i fem minuter är sannolikheten att vänta minst fem minuter till precis densamma som från början.

≡ Exempel 5 – från Poisson till exponential

Block skapas i medel sex gånger per timme. Vad är sannolikheten att det dröjer mer än tjugo minuter till nästa block?

Lösning

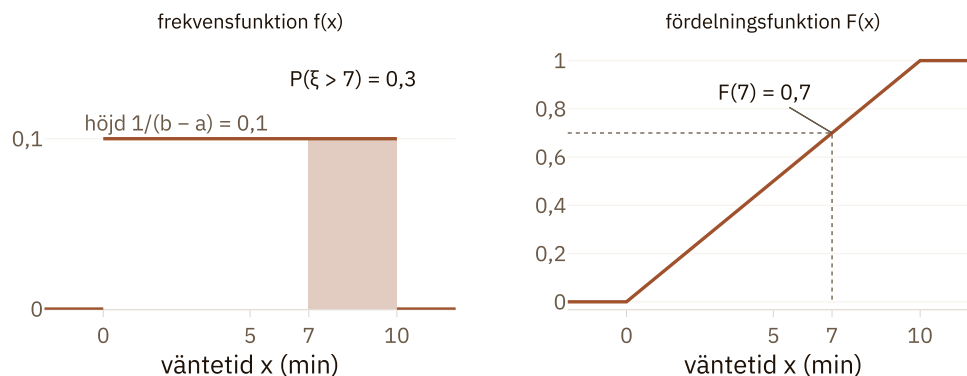
Låt ξ vara tiden till nästa block i timmar. Intensiteten sex per timme ger $\xi \in \text{Exp}(6)$. Tjugo minuter är en tredjedels timme: $P(\xi > \frac{1}{3}) = e^{-6 \cdot 1/3} = e^{-2} \approx \boxed{0,135}$

Fällor med Poisson och exponential

Kontrollera tidsenheten. Om intensiteten anges per timme och frågan gäller minuter måste du räkna om. "Fler än fem" betyder ett minus sannolikheten för högst fem.

10. Rektangel, ffg och frekvensfunktioner

En rektangelfördelad variabel är lika trolig överallt på ett intervall från a till b . Sannolikheten att den hamnar i ett delintervall är delintervallets längd delat med hela intervallets längd. Det enda du måste komma ihåg är att först klippa delintervallet mot $[a, b]$: den del som ligger utanför har sannolikheten noll.



Ffg-fördelningen räknar antalet försök till och med det första lyckade. Sannolikheten att det behövs fler än k försök är att de k första misslyckas, alltså $(1 - p)^k$. "Minst fyra försök" betyder att de tre första misslyckas. Väntevärdet är ett delat med p .

≡ Exempel 6 – minst fyra försök

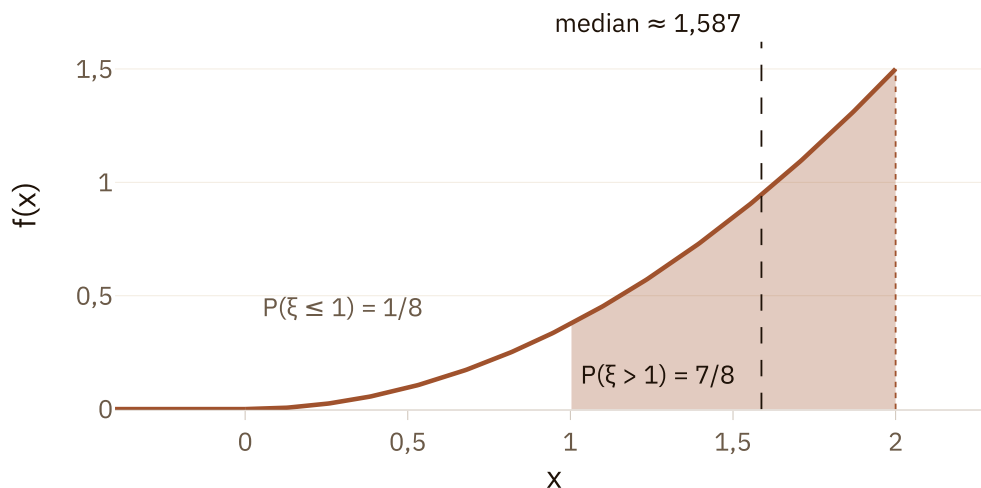
Ett försök lyckas med sannolikheten 0,3, oberoende av tidigare försök. Vad är sannolikheten att det behövs minst fyra försök?

Lösning

Låt ξ vara antalet försök till och med det första lyckade, $\xi \in \text{ffg}(0,3)$. Minst fyra försök betyder att de tre första misslyckas: $P(\xi \geq 4) = P(\xi > 3) = 0,7^3 = \boxed{0,343}$

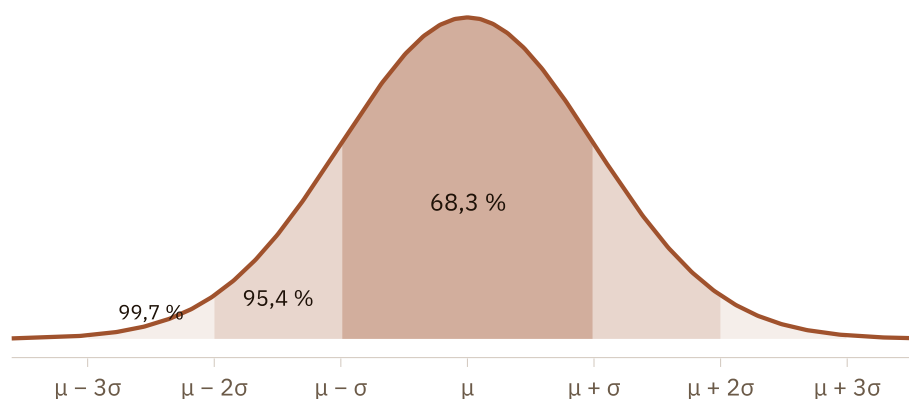
En del uppgifter ger dig en frekvensfunktion, alltså en funktion $f_\xi(x)$ som beskriver hur tätt sannolikheten ligger. Sannolikheten för ett intervall är arean under kurvan över intervallet, alltså en integral. Om funktionen innehåller en okänd konstant bestämmer du den genom att hela arean ska vara ett. Väntevärdet är integralen av x gånger frekvensfunktionen. Medianen är det värde där fördelningsfunktionen, alltså arean till vänster, är en halv.

$f(x) = 3x^2/8$ på $[0, 2]$: sannolikheten är arean



11. Normalfördelningen och kvantiler

Normalfördelningen beskrivs av sitt väntevärde μ och sin standardavvikelse σ . I kursen skrivs den $N(\mu, \sigma)$ med standardavvikelsen, inte variansen, som andra parameter. Ungefär två tredjedelar av värdena ligger inom en standardavvikelse från medelvärdet, och ungefär nittiofem procent inom två.



Allt räknas genom att standardisera: dra bort väntevärdet och dela med standardavvikelsen. Då får du ett värde på standardnormalfördelningen, och sannolikheten får du från Φ , eller direkt med räknarens kommando för normalfördelningen.

En vanlig tentauppgift går åt andra hållet: uppgiften ger två andelar, till exempel att tio procent väger under ett visst värde och fem procent över ett annat, och du ska bestämma μ och σ . Du gör det i två steg. Först översätter du varje andel till ett standardiserat värde med räknarens inversa normalfördelning. Då får du två ekvationer av formen värdet minus μ delat med σ är lika med det standardiserade värdet. Sedan subtraherar du ekvationerna, så att μ försvinner, och löser ut σ . Därefter får du μ ur någon av ekvationerna.

Räknarens inversa normalfördelning vill ha arean till vänster. Om uppgiften säger "fem procent väger mer än" är arean till vänster nittiofem procent. Det är den vanligaste felkällan i den här uppgiften.

≡ Exempel 7 – normalfördelning ur två andelar

Tio procent av alla förpackningar väger under 4,8 kg och fem procent väger över 5,6 kg. Vikten är normalfördelad. Bestäm μ och σ .

Lösning

Arean till vänster är 0,10 respektive 0,95, så de standardiserade värdena är $z_1 = -1,2816$ och $z_2 = 1,6449$: $\frac{4,8 - \mu}{\sigma} = -1,2816$, $\frac{5,6 - \mu}{\sigma} = 1,6449$ Subtrahera: $\sigma = \frac{5,6 - 4,8}{1,6449 + 1,2816} \approx 0,273$. Sedan $\mu = 5,6 - 1,6449 \cdot 0,273 \approx 5,15$. Svar: $\mu \approx 5,15, \sigma \approx 0,273$.

12. Väntevärde och varians av en funktion

Många uppgifter gäller en storhet som beror på en slumpvariabel, till exempel en vinst som är tre kronor per såld enhet minus en fast kostnad. Väntevärdet följer med linjära funktioner: väntevärdet av $a\xi + b$ är a gånger väntevärdet plus b . Standardavvikelsen påverkas bara av faktorn, inte av tillägget, eftersom att flytta alla värden lika mycket inte ändrar spridningen.

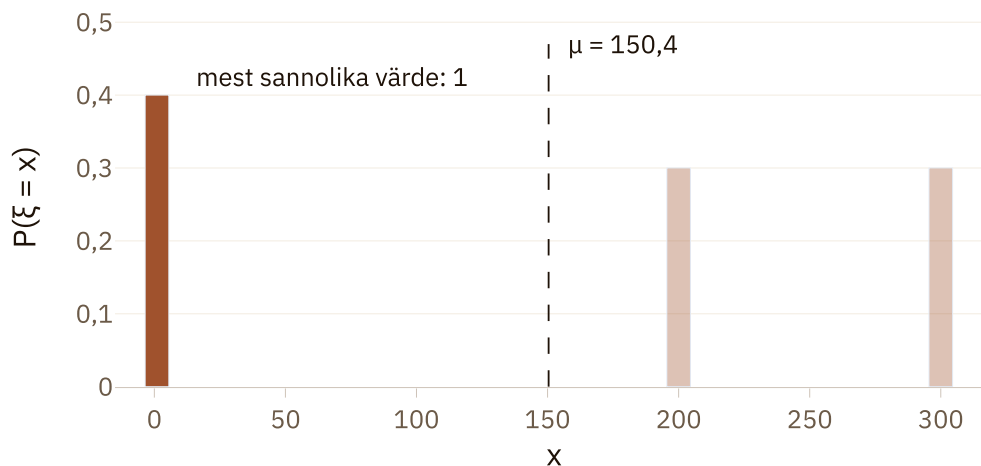
$$E(a\xi + b) = aE(\xi) + b, \quad D(a\xi + b) = |a| D(\xi)$$

För en summa av två variabler adderas variansen, men om variablerna är beroende tillkommer en term med kovariansen:

$$V(a\xi + b\eta) = a^2V(\xi) + b^2V(\eta) + 2ab \operatorname{Cov}(\xi, \eta)$$

Är variablerna oberoende är kovariansen noll och den sista termen försvinner.

Om funktionen inte är linjär, till exempel en energi som beror på hastigheten i kvadrat, fungerar inte reglerna ovan. Då behöver du väntevärdet av kvadraten, och det får du ur variansen: väntevärdet av ξ^2 är variansen plus väntevärdet i kvadrat. För en diskret variabel som ges i en tabell gör du en ny tabell för den nya storheten och räknar väntevärdet direkt.



⚠ Fällor med väntevärde och varians

Variansen kan aldrig vara negativ. Om uppgiften frågar efter standardavvikelsen, glöm inte att dra roten. Och räkna varianser, inte standardavvikelser, när du adderar.

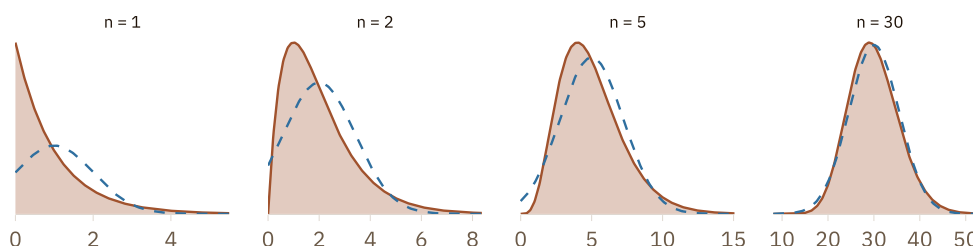
13. Uppgift 6 och 7: centrala gränsvärdessatsen

Uppgift sex och sju handlar nästan alltid om centrala gränsvärdessatsen. Satsen säger att en summa av många oberoende och likafördelade variabler är ungefär normalfördelad, oavsett vilken fördelning varje enskild variabel har. Detsamma gäller för medelvärdet av dem.

Arbetsgången har tre steg. Först tar du fram väntevärdet μ och standardavvikelsen σ för en enda variabel, ofta från formelbladet. Sedan skriver du upp summans eller medelvärdets fördelning. Summans väntevärde är $n\mu$ och dess standardavvikelse är $\sigma\sqrt{n}$. Medelvärdets väntevärde är μ och dess standardavvikelse är $\sigma/\sqrt{n}$. Lägg märke till att spridningen växer med roten ur n , inte med n . Till sist räknar du sannolikheten med normalfördelningen.

$$\xi_1 + \dots + \xi_n \approx N(n\mu, \sigma\sqrt{n}), \quad \bar{\xi} \approx N(\mu, \sigma/\sqrt{n})$$

Summan av n oberoende Exp(1): exakt fördelning (fylld) mot $N(n, \sqrt{n})$ (streckad)



En vanlig variant frågar efter ett gränsvärde i stället för en sannolikhet, till exempel: hur länge räcker ett lager med nittiofem procents sannolikhet? Då söker du det värde T som summan överstiger med sannolikheten nittiofem procent, och det ligger 1,6449 standardavvikelser under

summans väntevärde.

Exempel 8 – hur länge räcker lagret? (tentan 2025-03-17, uppgift 4)

Livslängden för ett elektronrör är exponentialfördelad med väntevärdet 220 timmar. Lagret innehåller 45 rör, och ett trasigt rör byts genast. Bestäm en tid T som lagret räcker åtminstone med 95 procents sannolikhet. Svara med en decimal.

Lösning

Låt ξ_i vara livslängden för rör nummer i . För exponentialfördelningen är $\mu = \sigma = 220$. Den totala tiden är summan av 45 livslängder: $\eta \approx N(45 \cdot 220, 220\sqrt{45}) = N(9900; 1475,8)$. Vi söker T med $P(\eta \geq T) = 0,95$, alltså $T = 9900 - 1,6449 \cdot 1475,8 \approx \boxed{7472,5 \text{ h}}$.

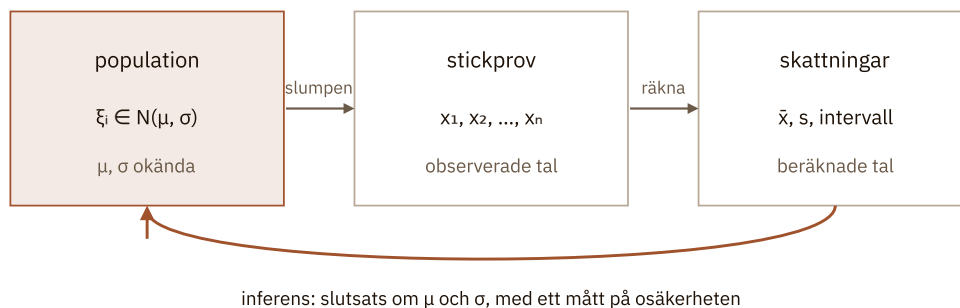
Fällor med centrala gränsvärdessatsen

Använd μ och σ för en enda variabel, inte för summan, när du börjar. Facit använder ingen halvkorrektion om uppgiften inte säger det. Och om variablerna är Poissonfördelade går det ofta att räkna exakt i stället, eftersom summan då själv är Poissonfördelad.

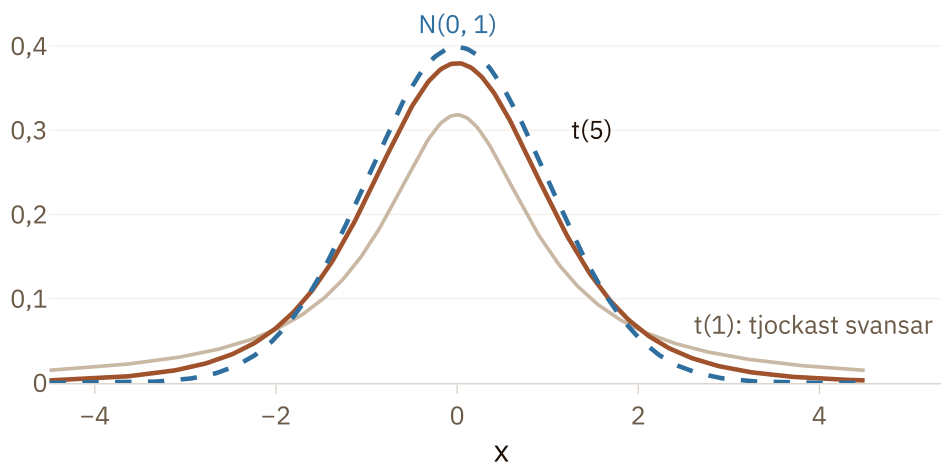
14. Från sannolikhetslära till statistik

Fram till nu har vi utgått från att vi känner fördelningen och frågat hur troliga olika utfall är. Statistisk inferens vänder på frågan. Nu har vi observerat ett stickprov, och vi vill dra slutsatser om den okända fördelningen som stickprovet kommer ifrån, oftast om dess väntevärde.

Stickprovets medelvärde $\bar{x}$ är den naturliga skattningen av väntevärdet μ , och stickprovets standardavvikelse s är skattningen av σ . Medelvärdet är självt en slumpvariabel: tar man ett nytt stickprov blir det ett annat medelvärde. Dess spridning är $\sigma/\sqrt{n}$. Det är därför större stickprov ger säkrare slutsatser: spridningen hos medelvärdet minskar med roten ur stickprovets storlek. För att halvera osäkerheten behöver du fyra gånger så många observationer.

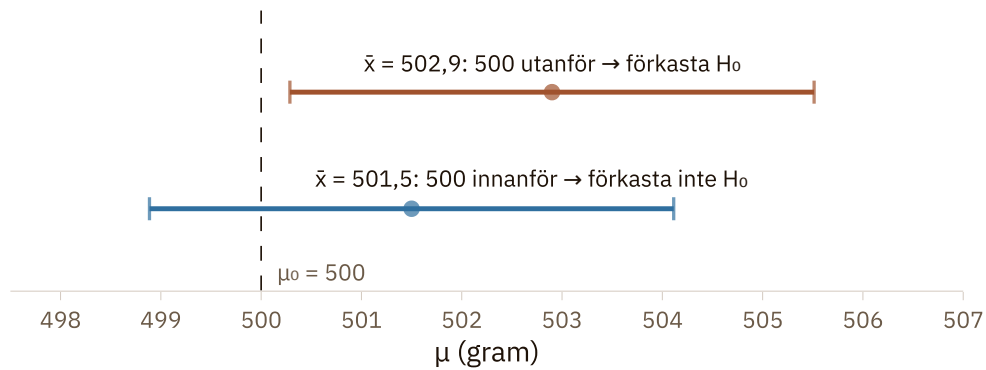


När σ är känd standardiserar man med den, och då gäller normalfördelningen. När σ är okänd, vilket är det vanliga, skattar man den med s . Det tillför extra osäkerhet, och därför används t -fördelningen i stället. Den liknar normalfördelningen men har tjockare svansar, och tjockleken beror på antalet frihetsgrader. För en mätserie är antalet frihetsgrader $n - 1$. Med många observationer blir t -fördelningen nästan identisk med normalfördelningen.



Vid hypotesprövning kan man göra två sorters fel. Det första är att förkasta nollhypotesen fast den är sann. Sannolikheten för det felet är signifikansnivån α , och den väljer man själv, ofta fem procent. Det andra felet är att inte förkasta nollhypotesen fast den är falsk. Sannolikheten för det är β , och styrkan är ett minus β . Man kan inte minska båda felen samtidigt utan att öka stickprovets storlek. Sänker man signifikansnivån flyttas den kritiska gränsen längre ut, och då blir det svårare att upptäcka en verklig skillnad.

Konfidensintervall och hypotesprövning hänger ihop. Ett tvåsidigt test på nivån fem procent förkastar nollhypotesen precis när värdet μ_0 ligger utanför det nittiofemprocentiga konfidensintervallet. Det är samma räkning sedd från två håll.



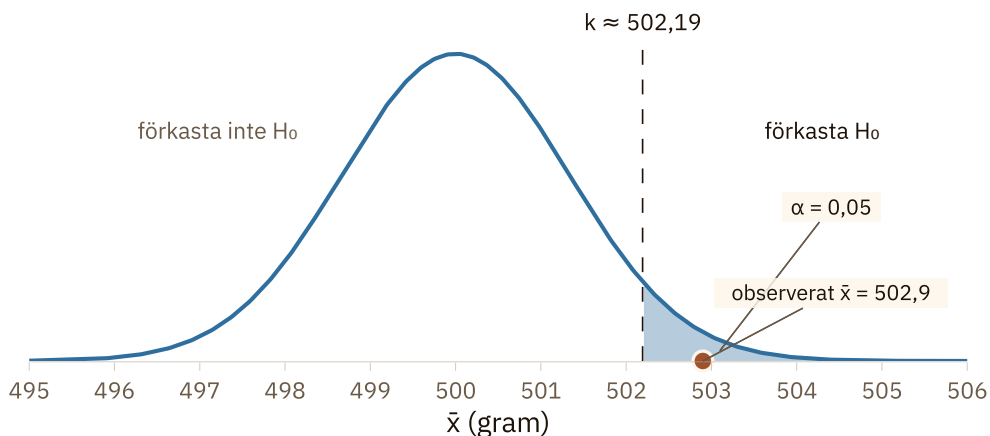
15. Uppgift 8: hypotesprövning, kritisk gräns och P-värde

Från uppgift åtta handlar tentan om statistisk inferens: att dra slutsatser om en okänd parameter utifrån ett stickprov. Hypotesprövning börjar med två hypoteser. Nollhypotesen H_0 är det man utgår från, oftast att ingenting har ändrats. Mothypotesen H_1 är det man vill visa. Tumregeln är att det man vill visa ska stå i H_1 .

Testet bygger på en kritisk gräns k . Om medelvärdet i stickprovet hamnar bortom gränsen förkastar man nollhypotesen. Gränsen väljs så att risken att förkasta en sann nollhypotes, signifikansnivån α , blir till exempel fem procent. När standardavvikelsen är känd ligger gränsen λ_α standardavvikelser för medelvärdet ovanför μ_0 , och medelvärdets standardavvikelse är $\sigma/\sqrt{n}$.

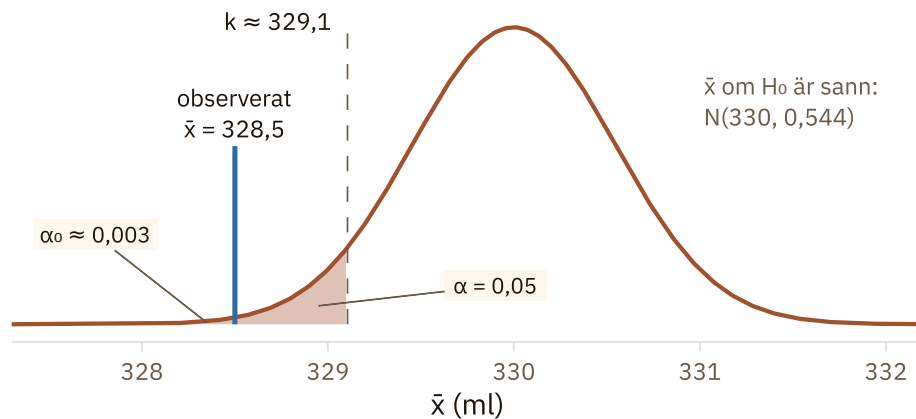
$$k = \mu_0 + \lambda_\alpha \frac{\sigma}{\sqrt{n}}$$

Fördelningen för $\bar{x}$ om $H_0: \mu = 500$ är sann ($\sigma = 4$, $n = 9$)



Om standardavvikelsen är okänd skattas den med stickprovets standardavvikelse s , och då används t -fördelningen med $n - 1$ frihetsgrader i stället för normalfördelningen. På räknaren får du gränsen med den inversa t -fördelningen.

Ett annat sätt att testa är direktmetoden. Där räknar du P-värdet, som skrivs α_0 : sannolikheten att få ett resultat som är minst lika extremt som det observerade, om nollhypotesen är sann. Är P-värdet mindre än signifikansnivån förkastar du nollhypotesen. Vid ett tvåsidigt test, där mothypotesen bara säger att parametern är skild från μ_0 , ska P-värdet multipliceras med två.



Exempel 9 – kritisk gräns

Ett stickprov med 16 observationer tas från $N(\mu; 2)$. Man testar $H_0 : \mu = 10$ mot $H_1 : \mu > 10$ på nivå fem procent. Var ligger den kritiska gränsen?

Lösning

Medelvärdets standardavvikelse är $2/\sqrt{16} = 0,5$, och $\lambda_{0,05} = 1,6449$: $k = 10 + 1,6449 \cdot 0,5 \approx 10,82$. Nollhypotesen förkastas om medelvärdet i stickprovet är större än 10,82.

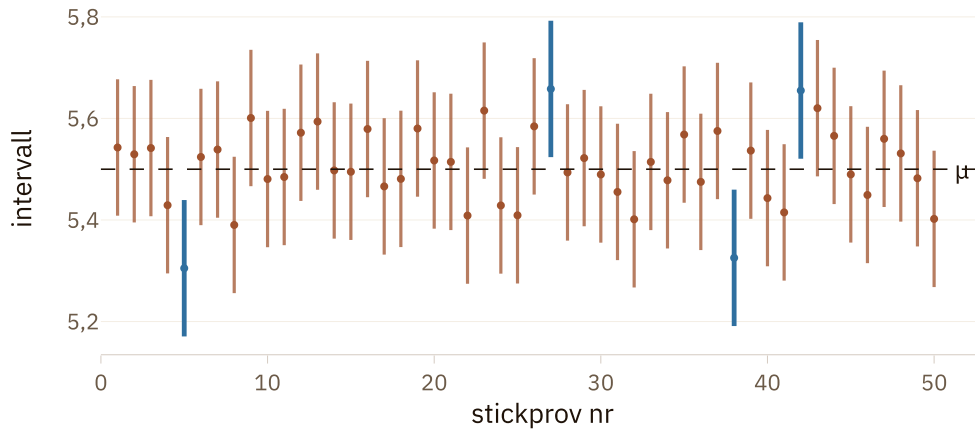
Fällor i uppgift 8

Ett vänstersidigt test, där mothypotesen säger att μ är mindre än μ_0 , ger en gräns under μ_0 . Räknarens inversa funktioner vill ha arean till vänster: för ett högersidigt test på fem procent är det nittiofem procent.

16. Uppgift 9: konfidensintervall

Ett konfidensintervall är ett intervall som med en viss sannolikhet, konfidensgraden, täcker den okända parametern. Ett nittiofemprocentigt intervall för ett väntevärde är medelvärdet plus minus ett t -värde gånger stickprovets standardavvikelse delat med roten ur n . För nittiofem procent används $t_{0,025}$, eftersom de fem procenten delas lika på båda sidorna.

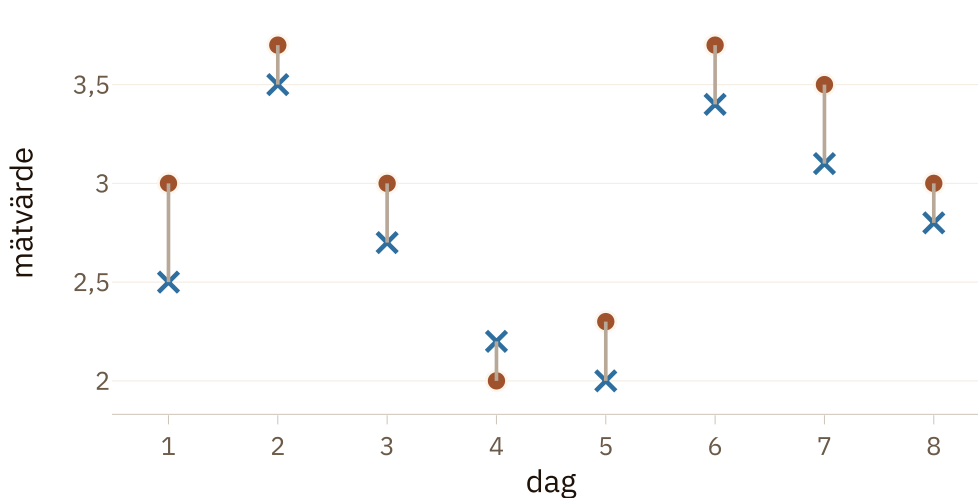
50 stickprov, 90 %-intervall: 46 täcker μ , 4 missar



Den viktigaste frågan i uppgift nio är om det är stickprov i par eller två oberoende stickprov. Stickprov i par betyder att mätningarna hör ihop två och två: samma bil mätt före och efter, samma dag på två platser, samma person med två metoder. Då bildar du differenserna för varje par, och sedan är det ett vanligt intervall för en enda mätserie med $n - 1$ frihetsgrader, där n är antalet par.

$$I_{\Delta} = \bar{z} \mp t_{\alpha/2}(n-1) \frac{s_z}{\sqrt{n}}$$

× metod A ● metod B



Två oberoende stickprov betyder att grupperna inte hör ihop, till exempel två laboratorier som mäter flera gånger var. Då används en gemensam standardavvikelse, som står på formelbladet, och $n_1 + n_2 - 2$ frihetsgrader. En klassisk fälla är två laboratorier som mäter samma prov flera gånger: det är ändå två oberoende stickprov, eftersom mätning nummer ett i det ena labbet inte hör ihop med mätning nummer ett i det andra.

Om noll ligger i ett intervall för en differens kan nollhypotesen att det inte finns någon skillnad inte förkastas. Ligger noll utanför kan den förkastas.

Exempel 10 – stickprov i par

Sex bilar mäts före och efter en justering. Differenserna har medelvärdet 1,2 och standardavvikelsen 0,8. Bestäm ett nittiofemprocentigt konfidensintervall för den genomsnittliga förändringen.

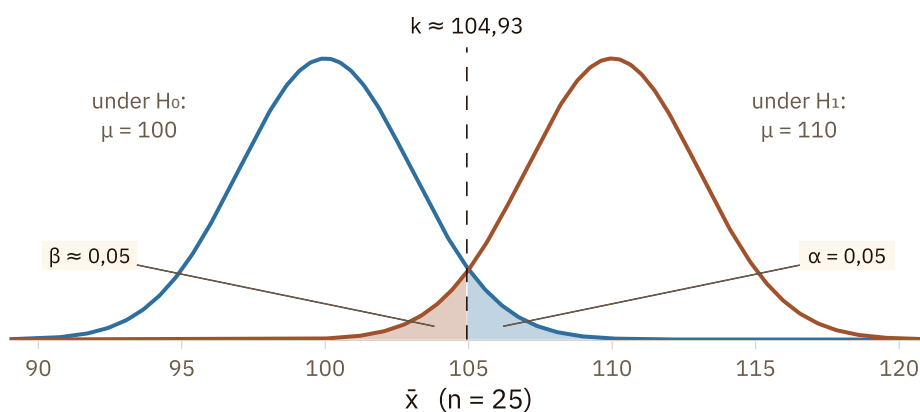
Lösning

Det är stickprov i par med sex par, så $f = 5$ och $t_{0,025}(5) = 2,5706$: $1,2 \mp 2,5706 \cdot \frac{0,8}{\sqrt{6}} = 1,2 \mp 0,84$ Intervallet blir $[0,36; 2,04]$. Noll ligger utanför, så förändringen är signifikant.

17. Uppgift 10: styrka, stickprovsstorlek och teckenintervall

Uppgift tio varierar mer än de andra. De tre vanligaste typerna är styrka, stickprovsstorlek och teckenintervall.

Styrkan är sannolikheten att testet förkastar nollhypotesen när den är falsk, alltså när det sanna väntevärdet är μ_1 . Arbetsgången har två steg. Först bestämmer du den kritiska gränsen under nollhypotesen, som i uppgift åtta. Sedan räknar du sannolikheten att hamna bortom gränsen, men nu med det sanna väntevärdet μ_1 . För ett högersidigt test är det ett minus Φ av gränsen minus μ_1 , delat med medelvärdets standardavvikelse. För ett vänstersidigt test är det Φ av samma uttryck. Styrkan ska alltid ligga mellan signifikansnivån och ett.



Uppgiften kan också gå baklänges: du får en önskad styrka och ska bestämma signifikansnivån. Då lägger du gränsen så att den önskade styrkan uppnås under μ_1 , och räknar sedan hur stor del av fördelningen under μ_0 som ligger bortom gränsen.

☰ Exempel 11 – styrka

Fortsätt exempel 9, där gränsen är 10,82 och medelvärdets standardavvikelse 0,5. Vad är styrkan om det sanna väntevärdet är 11?

Lösning

$$S(11) = 1 - \Phi\left(\frac{10,82 - 11}{0,5}\right) = 1 - \Phi(-0,355) \approx \boxed{0,64}$$

☰ Exempel 12 – styrka baklänges (tentan 2025-03-17, uppgift 8)

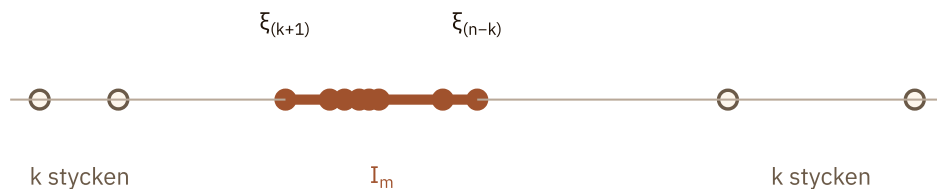
Ett stickprov med 25 observationer tas från $N(\mu; 4,15)$. Man testar $H_0 : \mu = 10$ mot $H_1 : \mu = 12,5$ och vill att styrkan ska vara 90 procent. Vad blir signifikansnivån? Svara i procent med två decimaler.

Lösning

Medelvärdets standardavvikelse är $4,15/5 = 0,83$. Styrkan 90 procent under $\mu_1 = 12,5$ ger gränsen $k = 12,5 - 1,2816 \cdot 0,83 \approx 11,436$. Signifikansnivån är sannolikheten att hamna över gränsen när $\mu = 10$: $\alpha = 1 - \Phi\left(\frac{11,436 - 10}{0,83}\right) \approx \boxed{4,18 \%}$

Stickprovsstorleken bestäms antingen av hur brett ett intervall får vara eller av vilken styrka testet ska ha. Formlerna löser ut n ur uttrycket för bredden eller för styrkan, och svaret avrundas alltid uppåt, eftersom ett mindre stickprov inte räcker.

Teckenintervallet används när man vill skatta medianen och inte kan anta normalfördelning. Idén är enkel: varje observation hamnar under medianen med sannolikheten en halv, så antalet observationer under medianen är binomialfördelat med p lika med en halv. Om du sorterar observationerna och tar intervallet från den $(k + 1)$:a minsta till den $(k + 1)$:a största, är konfidensgraden ett minus två gånger sannolikheten att högst k observationer hamnar under medianen.



≡ Exempel 13 – konfidsensgrad för ett teckenintervall (tentan 2025-03-17, uppgift 7)

Ett stickprov på åtta observationer kommer från en fördelning som inte är normal. Beräkna konfidsensgraden för intervallet från den näst minsta till den näst största observationen. Svara i procent med två decimaler.

Lösning

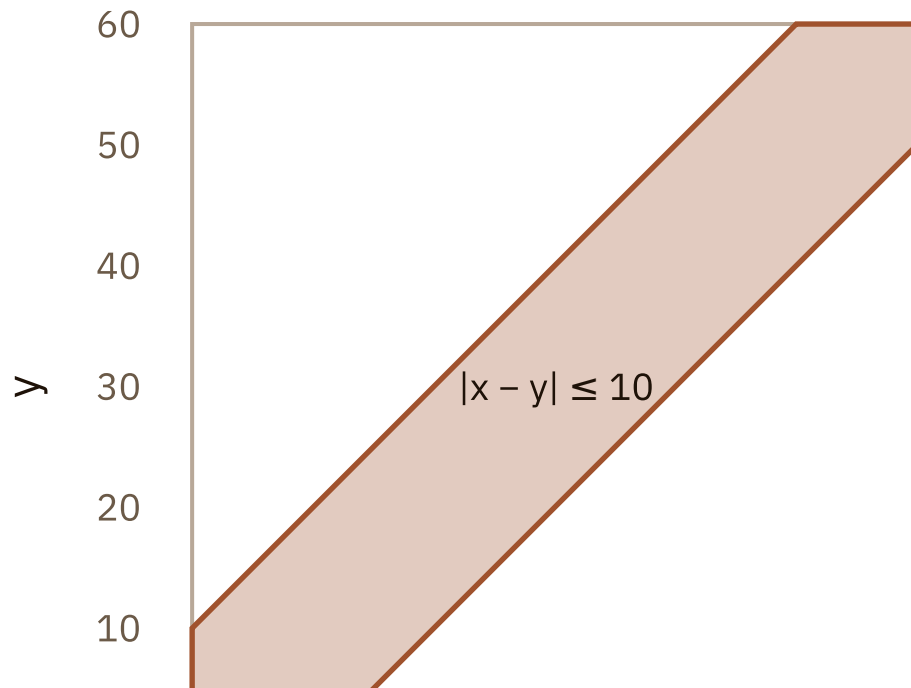
Låt η vara antalet observationer under medianen, $\eta \in \text{Bin}(8; 0,5)$. Intervallet missar medianen om högst en observation hamnar på ena sidan: $1 - 2P(\eta \leq 1) = 1 - 2 \cdot$

$$\frac{1 + 8}{256} = \frac{238}{256} \approx \boxed{92,97 \%}$$

18. Uppgift 11 och 12: två slumpvariabler

De två sista uppgifterna handlar om två slumpvariabler samtidigt. Fyra typer återkommer: area, kovarians, tvådimensionell normalfördelning och en blandning av simultana fördelningar, slumpvandring och max eller min.

Areauppgifterna är ofta de mest tacksamma. Om ett par (ξ, η) är likformigt fördelat över ett område är sannolikheten för en händelse arean av händelsens del av området delat med hela områdets area. Rita alltid. Ett klassiskt exempel är två personer som kommer slumpmässigt under en timme och väntar en viss tid på varandra: de träffas om skillnaden mellan ankomsttiderna är mindre än väntetiden, och det är ett band längs diagonalen i en kvadrat. Det enklaste är att räkna arean av de två triangelarna utanför bandet och dra bort den från ett.



Exempel 14 – möte inom tio minuter

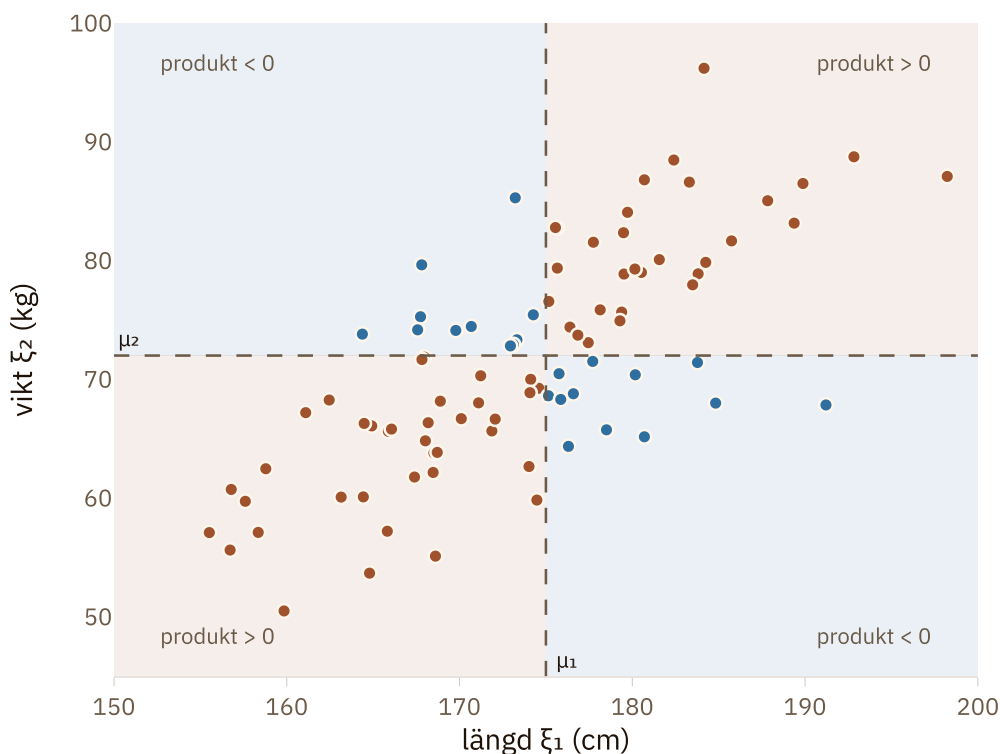
Två personer kommer oberoende av varandra vid en slumpmässig tidpunkt under samma timme, och var och en väntar tio minuter. Vad är sannolikheten att de träffas?

Lösning

Ankomsttiderna är likformigt fördelade på kvadraten $[0, 60] \times [0, 60]$. De träffas om skillnaden är högst tio minuter. Utanför bandet ligger två trianglar med kateten 50: $P = 1 - \left(\frac{50}{60}\right)^2 = 1 - \frac{25}{36} = \frac{11}{36} \approx \boxed{30,6\%}$

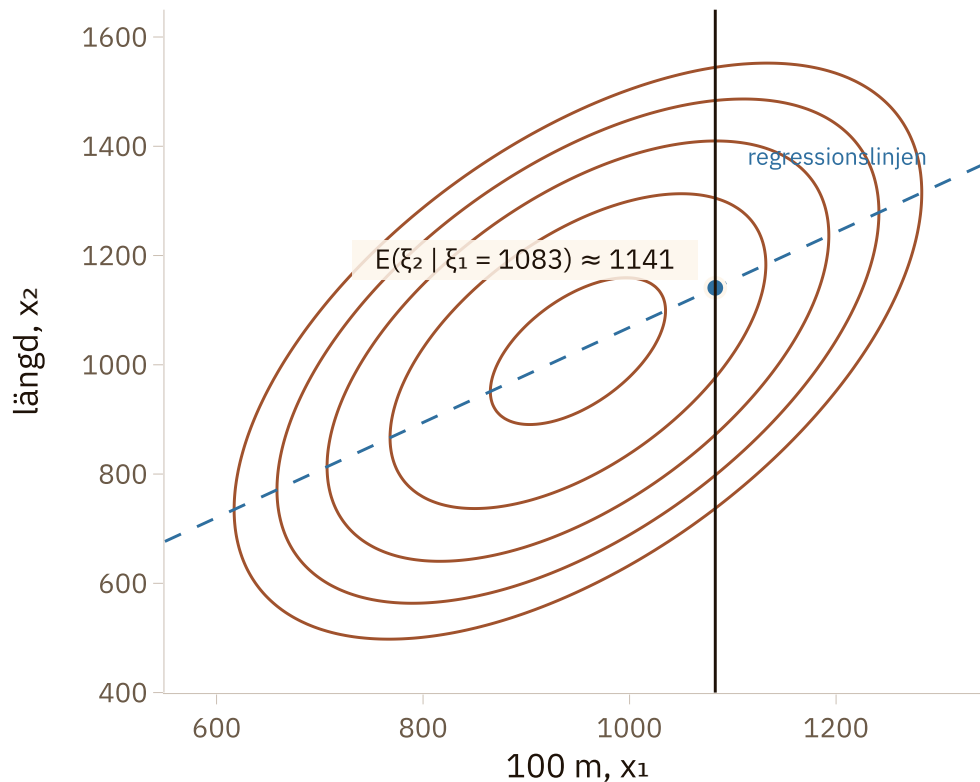
Kovariansen mäter hur två variabler samvarierar. Den är positiv när stora värden på den ena brukar gå ihop med stora värden på den andra. Korrelationskoefficienten ρ är kovariansen delad med produkten av standardavvikelseerna, och den ligger alltid mellan minus ett och ett. Oberoende variabler har kovariansen noll. Kovariansen kan delas upp över summor, och en faktor framför en variabel kommer med som en faktor framför kovariansen.

$$\text{Cov}(\xi, \eta) = E(\xi\eta) - E(\xi)E(\eta) = \rho D(\xi) D(\eta)$$



Den tvådimensionella normalfördelningen dyker upp i uppgifter som gäller två mått som hör ihop, till exempel längden hos far och son, eller två idrottares resultat. Tricket är nästan alltid att bilda skillnaden eller summan mellan dem. Skillnaden mellan två normalfördelade variabler är normalfördelad, med väntevärdet lika med skillnaden mellan väntevärdena. Variansen är summan av varianserna minus två gånger kovariansen. Frågan “vem blir längst?” blir då frågan om skillnaden är större än noll.

$$D(\xi_1 - \xi_2) = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$



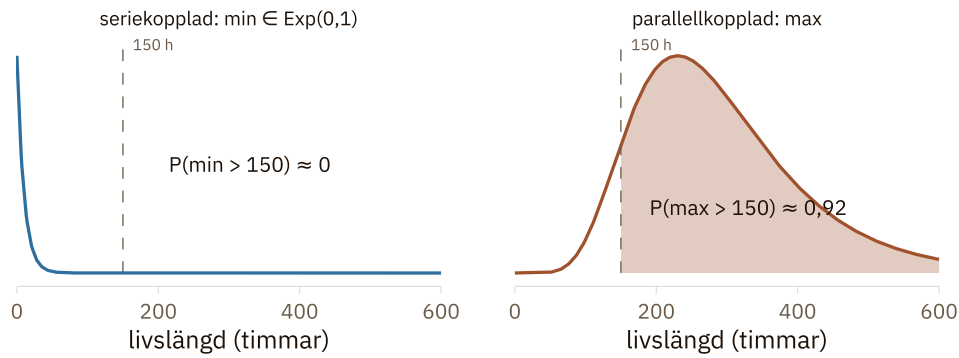
≡ Exempel 15 – blir sonen längre än fadern?

Faderns längd är $N(178; 7)$ och sonens $N(180; 7)$, med korrelationen $\rho = 0,5$. Vad är sannolikheten att fadern är längre än sonen?

Lösning

Låt ζ vara faderns längd minus sonens. Då är $E(\zeta) = -2$ och $D(\zeta) = \sqrt{49 + 49 - 2 \cdot 0,5 \cdot 7 \cdot 7} = 7$. Sannolikheten är $P(\zeta > 0) = 1 - \Phi(2/7) \approx \boxed{0,39}$.

Max och min dyker upp i uppgifter om system av komponenter. Ett seriekopplat system fungerar bara så länge alla komponenter fungerar, så dess livslängd är den kortaste komponentens livslängd, alltså minimum. Sannolikheten att systemet håller längre än en viss tid är produkten av sannolikheterna att varje komponent gör det. Ett parallellkopplat system fungerar så länge minst en komponent fungerar, så dess livslängd är maximum, och sannolikheten att det har gått sönder före en viss tid är produkten av komponenternas fördelningsfunktioner.



En slumpvandring är ett antal steg åt höger eller vänster. Om ξ är antalet steg åt höger av n är ξ binomialfördelad, och positionen efter n steg är $2\xi - n$. En fråga om positionen översätts därmed till en fråga om ξ , som räknaren klarar.

Exempel 16 – slumpvandring

En partikel tar tio steg, var och ett åt höger eller vänster med samma sannolikhet. Vad är sannolikheten att den hamnar minst fyra steg till höger om startpunkten?

Lösning

Låt ξ vara antalet steg åt höger, $\xi \in \text{Bin}(10; 0,5)$. Positionen är $2\xi - 10 \geq 4$, alltså $\xi \geq 7$:

$$P(\xi \geq 7) = 1 - P(\xi \leq 6) = \frac{176}{1024} \approx \boxed{0,172}$$

19. De vanligaste felen, samlade

Den här listan samlar de fel som oftast kostar poäng på del 1. Lyssna på den kvällen före tentan.

Det första felet gäller “minst”. Sannolikheten för minst k är ett minus sannolikheten för högst $k - 1$, inte högst k . Det gäller binomial-, Poisson- och ffg-fördelningen och allt annat som räknar heltal.

Det andra felet gäller arean till vänster. Räknarens inversa funktioner vill alltid ha arean till vänster om värdet. Om uppgiften säger “de fem procent tyngsta” eller “fem procent över” är arean till vänster nittiofem procent.

Det tredje felet gäller standardavvikelse och varians. Kursen skriver normalfördelningen med standardavvikelsen, inte variansen. När du adderar spridningar adderar du varianser och drar roten på slutet, aldrig standardavvikelsena direkt.

Det fjärde felet gäller tidsenheten. En intensitet per timme måste räknas om om frågan gäller minuter, och en exponentialfördelnings λ ska ha samma tidsenhet som frågan.

Det femte felet gäller par eller två stickprov. Stickprov i par kräver att mätningarna hör ihop två och två. Två laboratorier eller två grupper är två oberoende stickprov, med en gemensam standardavvikelse och $n_1 + n_2 - 2$ frihetsgrader.

Det sjätte felet gäller ensidigt och tvåsidigt. Vid ett tvåsidigt test delas signifikansnivån på båda sidorna, och P-värdet multipliceras med två. Vid ett konfidensintervall används alltid $\alpha/2$ i varje svans.

Det sjunde felet gäller avrundning. Avrunda bara i det sista steget, och svara med exakt det antal decimaler som efterfrågas. En stickprovsstorlek avrundas alltid uppåt.

Det åttonde felet gäller slumpvariabeln. Definiera den i ord, med vad den räknar eller mäter och i vilken enhet. Det är den vanligaste orsaken till uteblivna delpoäng.

20. Strategi på tentan

Läs igenom alla tolv uppgifter först och börja med dem du känner igen direkt. Uppgift ett, två och sex brukar vara de säkraste poängen. Kartan i början visar var de brukar ligga.

För varje uppgift: definiera slumpvariabeln i ord, ange fördelningen, skriv vad som ska beräknas, räkna och svara exakt som uppgiften ber om. Kontrollera rimligheten innan du går vidare. En sannolikhet ligger mellan noll och ett. En varians är aldrig negativ. Ett antal är ett heltal mellan noll och n . Medianen i en exponentialfördelning är mindre än väntevärdet. Styrkan ligger mellan signifikansnivån och ett.

Om du kör fast, skriv ändå upp slumpvariabeln och fördelningen. Det kan ge delpoäng, och ofta lossnar uppgiften när uppställningen står på papperet.

Se även

- S0008M
- Betingad sannolikhet · Bayes sats · Lagen om total sannolikhet
- Binomialfördelning · Hypergeometrisk fördelning · Poissonfördelning · Exponentialfördelning · ffg-fördelning · Likformig fördelning
- Normalfördelning · Median och kvantiler · Linjära funktioner av stokastiska variabler
- Centrala gränsvärdessatsen · Normalapproximation
- Hypotesprövning · Direktmetoden · Konfidensintervall · Stickprov i par · Två stickprov
- Signifikansnivå och styrka · Styrkefunktion · Teckenintervall
- Flerdimensionella fördelningar · Kovarians · Korrelation · Tvådimensionell normalfördelning · Max och min av stokastiska variabler