

Kvantitativ Biologi

GRUPPDISKUSSIONER

08:15-10:00 (samt rast på valfri tid)

Diskussion 1: Veckans övningar

Diskutera igenom de olika uppgifterna som ni arbetade med under veckan. Se till att alla i gruppen känner sig bekväma med vad ni svarade på de olika frågorna och varför ni svarade på det sättet. Identifiera sedan de tre svåraste frågorna på övningarna, och motivera varför ni valt dem. En person är sekreterare och lämnar in detta.

Om ni känner att *ni har koll* på veckans uppgifter, gå vidare och diskutera frågorna nedan som ytterligare övning.

Diskussion 2: Seminariefrågorna

Varje vecka delas ett antal frågor ut, som fungerar som uppföljning till veckans övningar. Vrid och vänd på dessa, för att förbereda er så bra som möjligt inför clickersutmaningen. Målet med aktiviteten är att alla ska ha en så djup förståelse som möjligt för alla delmoment. Ta tid och förklara för varandra hur ni tänker kring det ni diskuterar, och låt stämningen i gruppen vara sådan att alla känner att de kan ställa precis vilka frågor de vill!

Diskussion 3: Kreativ design

Designa gemensamt en repetitionsfråga eller en tentafråga (eller en av varje), baserat på blockets innehåll. Observera att vi använder flervalsfrågor för övning och repetition, men inte för tentor, på denna kurs. En person skriver ner en läsbar version av frågan, inklusive eventuella svarsalternativ och fullständig lösning. Om uppgiften innebär att man använder R eller Maxima, ska nödvändig kod vara med.

CLICKERSUTMANINGEN

10:15-12:00 (samt rast)

Med denna aktivitet summerar vi veckan och stämmer av vad ni (och vi) lärt oss. Vi kommer att ställa ett antal frågor, som besvaras individuellt och anonymt med clickers. Efter varje fråga diskuterar vi innehållet efter behov, och dessutom har vi här ett utmärkt tillfälle både för er att ställa frågor och för oss att lägga in förtydliganden kring saker som dykt upp under veckan.

Seminarium

På GM kan du ladda ner de fyra dataset som behövs för att kunna svara på frågorna nedan, de ligger alla i filen **Block4_sem.RData**. För att få ett rent utgångsläge: starta R-Studio och välj *Session – Clear workspace...* i menyn. Om du sen dubbelklickar på filen **Block4_sem.RData** så läses alla data in automatiskt i R-minnet och du kommer se dem i miljöfönstret.

1. Du har mätt relativa ljusnivåer vid markytan i ett regnskogsparti i Xishuangbanna i södra Yunnan, Kina. Ljuskivån är uttryckt på en relativ skala där 1 motsvarar ljuskivån på ett öppet fält. De översta fem raderna av data är inmatade enligt följande:

Tabell 1. Relativa ljuskivåer för 40 mätplatser i Xishuangbanna (Yunnan, Kina)

kronhojd	ljus
låg	0,036982
hög	0,015924
hög	0,021982
låg	0,046734
låg	0,041631
.... OSV...	... OSV ...

Givet att tabellen ovan motsvaras av en dataframe med namnet *yunnan* i R, hur ser koden ut för att analysera detta med ett oberoende t-test mellan två grupper? (Om ni vill testa, finns data i dataframen **Yunnan**.)

2. Du har mätt aktiviteten av ett enzym som förekommer i två former, en kort och en lång. Dina data är inmatade enligt nedanstående tabell som bara visar de första 5 mätningarna. Aktiviteten uppskattades på en skala 1 till 10.

Tabell 2. Rankade aktiviteter uppmätta hos lång respektive kort form hos ett enzym.

Lång	Kort
6	2
4	3
5	1
4	1
6	3
.... OSV...	... OSV ...

- a) Hur ser koden ut för att analysera detta med Wilcoxon-Mann-Whitneys test om varje mätning gjordes på olika celler?
 - b) Hur ser koden ut om dina aktivitetsvärden på varje rad uppskattades från samma cell?
3. Ett forskarlag är intresserat av att få veta om längden på kvinnors menstruationscykel skiljer sig från månens cykel som är ca. 27,5 dagar. De har frågat 15 kvinnor om perioden mellan menstruationscyklerna och deras svar återfinns i vektorn **period** i den laddade RData-filen. Diskutera nu följande frågor:
 - a. Vilka är hypoteserna?
 - b. Vilket test lämpar sig för att testa hypoteserna?
 - c. Hur tolkar du p-värdet i förhållande till hypoteserna?
 - d. Vilken slutsats drar du från analysen?
 - e. Vilken nytta har vi av denna kunskap?

4. Betaendorfin är ett opium-liknande ämne som kroppen producerar, ofta i respons till stress, smärta, glädje och motion. Upplevelsen av betaendorfin liknar den som morfin ger och är positiv och åtminstone för en del personer så kan detta leda till ett beroende av stress, smärta eller träning. De senare går under termen "träningssnarkomaner". Vi är nu intresserade av att undersöka om träning kan leda till en känsla av välbefinnande även under perioder då man inte tränar. Tio otränade män ställde upp i denna studie (dessa gick innan studien under termen "soffpotatisar"). Deras halter av betaendorfin mättes under vila i början av januari varefter de sattes på ett regelbundet träningsschema. Halterna av betaendorfin mättes sedan igen under vila i maj, minst en vecka efter senaste träningspasset. Du hittar data i dataramen **endo**. Vi vill nu veta om träningen har höjt halten betaendorfin vid vila även under den period då personen inte aktivt tränar.
- Formulera korrekta hypoteser. Ser du något problem mellan hypoteserna och formuleringen av frågan?
 - Vilket test lämpar sig för att undersöka hypoteserna?
 - p-värdet innebär en sannolikhet, sannolikheten för vad exakt?
 - Vilken slutsats kan vi dra från analysen?
 - För en person hade halterna ökat med 5 pg/ml. Hur kan vi hantera detta om vår slutsats är att halterna endorfin minskar?
5. Det finns många studier som visar på fysiologiska effekter hos personer som röker. För gravida, kan dessa effekter på något sätt inverka på det växande embryot? Vi letade upp 15 rökande, minst ett paket om dagen, kvinnor och 15 icke-rökande kvinnor som nyligen blivit gravida. Därefter väntade vi i ca. nio månader och mätte sedan de nyfödda barnens vikt. Vi vill då veta om födelsevikten påverkats av om modern är rökare eller inte. En av kvinnorna hoppade av studien så för gruppen av icke-rökare har vi bara 14 replikat. Du hittar data i dataramen **rokning**.
- Nollhypotesen är att rökning inte påverkar födelsevikten, men vilken är den alternativa hypotesen?
 - Vilket statistiskt test kan användas för att analysera skillnaderna i födelsevikt?
 - Vad blir utfallet av testet, och vilken slutsats drar ni om födelsevikten?
 - Om vi finner att rökning minskar födelsevikten hos barnen kan tobaksbolagen komma med några av följande kommentarer. Hur ställer ni er till dem?
 - Kvinnor vill föda mindre barn så detta är bara positivt
 - Det var en för liten studie så vi kan inte lita på resultaten
 - Det är inte rökning som är problemet utan något annat hos de rökande kvinnorna som ni inte tog hänsyn till
 - Vi har idag så näringsrik mat att en lite mindre födelsevikt inte spelar någon roll utan de minde barnen kommer snart växa ifatt
 - Det var egentligen inga effekter på födelsevikten som sådan utan bara att rökning leder till tidigare födsel innan barnen vuxit till sig

6. Precis som för p-värdet finns en hel del svårigheter kring att förstå begreppet konfidensintervall. För att undersöka detta har vi tagit en population med 65 individer (detta är hela populationen, som därmed alltså är känd), och sedan gjort 20 stickprov om vardera 10 individer. För varje stickprov har sedan ett konfidensintervall bildats, och alla 20 visas i figuren nedan. Den gröna linjen visar populationens väntevärde för medelvärde. Innan ni svarar på frågorna kan ni, om ni vill använda koden i filen Konfidensintervall.R. Kör hela koden ett antal gånger, och titta på de figurer som genereras.

- Förklara den variation som syns i figuren.
- Varför ligger det röda intervallet inte innehåller väntevärdet (gröna linjen).
- Om vi gjorde 1000 stickprov, hur många skulle vara röda?
- Stämmer följande påstående:

"Om jag har ett 95%-igt konfidensintervall för ett stickprov, är sannolikheten att det innehåller populationens sanna medelvärde 95 %."

