

Statistikövning för Excel

© P. Andreas Svensson

Instruktion: Skriv ner dina svar på papper medan du jobbar! Då riskerar du inte att behöva göra om ifall datorn strular. Om du har möjlighet att skriva ut detta dokument så kan du skriva ner svaren direkt här. När du är klar matar du in svaren på MyMoodlesidan under "Statistikövning". Inlämningen rättas automatiskt och du får feedback direkt. Figurer, uträkningar och utdatatabeller som du gör i Excel 365 sparar du direkt i Excelfilen på din egen dator.

Tips anges i blått. Om du har Office ProPlus kan menyer skilja sig något, dessa är då i lila.

Om du använder Excel på engelska anges engelska menyer etc i grönt.

Fråga 1

Du vill undersöka effekten av kalkning på småländska sjöar. Du har en karta med 10 kalkade sjöar och 10 sjöar som inte kalkats. Dessutom har du en pH-mätare.



a) Formulera en arbetshypotes

- "Kalkade sjöar har högre pH än okalkade "
- "Kalkade sjöar har samma pH som okalkade "

b) Formulera en nollhypotes

- "Kalkade sjöar har högre pH än okalkade "
- "Kalkade sjöar har samma pH som okalkade "

Efter mätning visade det sig att de kalkade sjöarna hade *lägre* pH än de okalkade sjöarna.

c) Vad gör du med H_A från a)? ○Förkastar? ○Behåller?

d) Vad gör du med H₀ från b)? ○Förkastar? ○Behåller?

Diskutera med en studiekamrat hur man kanske kan förklara detta resultat.

Fråga 2 Vilken typ av data?

För varje exempel, ange om datan är kategorisk eller numerisk, och om det är nominaldata, ordinaldata, diskret data eller kontinuerlig data.

- a. Aftonbladets filmbetyg som är mellan + och + + + +
 - b. Metalljoner (Cu²⁺, Cd²⁺, Mg²⁺, Mn²⁺)
 - c. Bullermätningar i decibel (dB).
 - d. Antal studenter på Linnéuniversitetets olika kurser
 - e. Ålderskategorierna *pensionär, ungdom, spädbarn,...*
 - f. Antal hjärtslag per minut
 - g. Koldioxidkoncentrationen i luft (i ppm)
-

Fråga 3

Öppna filen **skor.xlsx**

En nyfiken lärare noterade skostorlekarna på ett antal studenter på Miljöanalytikerprogrammet. De Dessa skostorlekar finns i kolumn B

Använd formelfunktionen i Excel för att räkna ut medel, median och typvärde (modalvärde).

Tips, använd formlerna =medel=, =median och =typvärde

(Engelska =average, =median & =mode)

a) Vilken typ av data är skonummer? _____

b) Vad är medelvärdet? _____

c) Vad är medianvärdet? _____

d) Vad är typvärdet? _____



I kolumn F finns skostorlekarna för förstaårsstudenterna på Kemiprogrammet från HT 2013

e) vad är medelvärdet för kemisterna? _____

f) vad är medianvärdet för kemisterna? _____

Tips: om du markerar cellerna med uträkningarna för miljöstudenterna, kan du dra markeringen åt höger (ta tag i lilla gröna kvadraten) och slippa skriva in formlerna igen.

Gör en boxplot över kemisternas skostorlekar

Tips: markera dina data, välj sedan Infoga/Diagram/Statistiskt diagram/låddiagram

(Engelska: Insert/Charts/Insert Statistic Chart/Box and whisker)

Ändra skalan på boxplotens Yaxel så att den börjar på storlek 30

Tips: Dubbelklicka på siffrorna på Yaxeln, i Menyn Formatera axel/Format axis skriv in 30 i rutan "minimum"

Du ser en extrem datapunkt, nämligen kemistudent 10 som har skostorlek 53. Detta låter lite för extremt - kanske är det ett felskrivet "35". Du bestämmer dig för att ta bort detta värde. Gör om boxploten utan det sista värdet. Yaxeln skall börja på storlek 30.

Räkna ut medel och median utan det sista värdet

g) hur påverkades medelvärdet av att du tog bort datapunkten med skostorlek 53?

- Det sjönk mindre än 1 storlek
- Det steg mindre än 1 storlek
- Det var oförändrat
- Det sjönk mer än 1 storlek
- Det steg mer än 1 storlek

h) Hur påverkades medianvärdet?

- Det sjönk mindre än 1 storlek
- Det steg mindre än 1 storlek
- Det var oförändrat
- Det sjönk mer än 1 storlek
- Det steg mer än 1 storlek

i) Vad drar du för slutsats om hur medelvärdet påverkas av extrema värden?

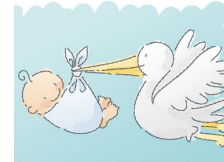
- Medelvärdet är mycket känsligt för extrema värden
- Medelvärdet är mindre känsligt för extrema värden

j) Vad drar du för slutsats om hur medianvärdet påverkas av extrema värden?

- Medianvärdet är mycket känsligt för extrema värden
- Medianvärdet är mindre känsligt för extrema värden

Fråga 4

Fyra nyfödda barn vägde 2,6 , 6,3 , 3,0 och 4,1 kg.
Öppna en ny flik i Exceldokumentet du har öppet och skriv in de 4 värdena i en av kolumnerna.
Beräkna följande med hjälp av Excels formler



a. Medelvärde _____

b. Varians _____

c. Standardavvikelse _____

d. Dessa fyra bebisar är ett stickprov från ett visst sjukhus. Vad är Standard Error i detta stickprov? _____

$$\text{Standard error, SE} = \frac{\text{sd}}{\sqrt{n}}$$

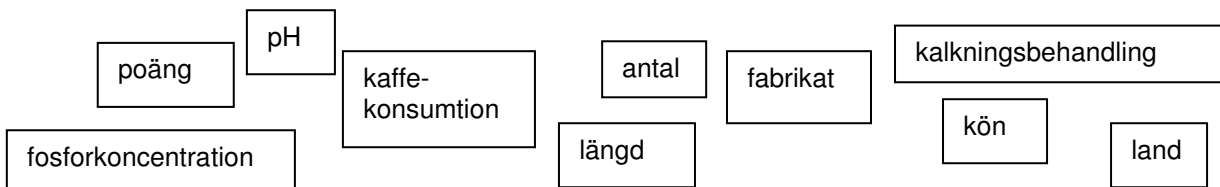
Precisionsmått

Tips: Se hur man räknar ut SE i Excelövningen – "bekanta dig med Excel 365"

Fråga 5

Inom statistiken skiljer man på *förklaringsvariabel* och *responsvariabel*. Ange vilken som är förklaringsvariabel och responsvariabel i undersökningarna nedan.

Välj bland dessa variabler (*varje alternativ skall bara användas en gång*):

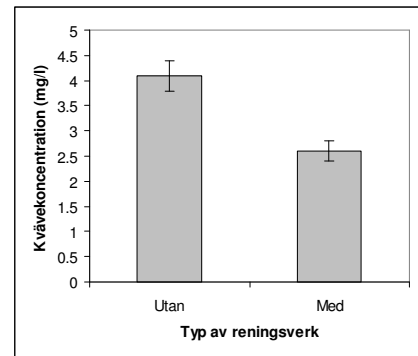


Hypotes	förklaringsvariabel	responsvariabel
a) Svenska och danska reningsverk släpper ut lika mycket fosfor		
b) Kräftmärkena "Ullmo" och "Pandalus" skiljer sig åt i hur många kräftor det är per paket		
c) Grävlingshannar är större än grävlingshonor		
d) Kalkade sjöar är mindre försurade än okalkade		
e) Kaffedrickande studenter får bättre tentaresultat		

Fråga 6

I exemplet på föreläsningen jämförde vi kväveutsläppen från 12 reningsverk med kväverening och 12 utan kväverening

	Utan	Med
medelutsläpp ± SE	4.1 ± 0.3 mg/l	2.6 ± 0.2 mg/l



Ett statistiskt test gav ett P-värde = 0.0004

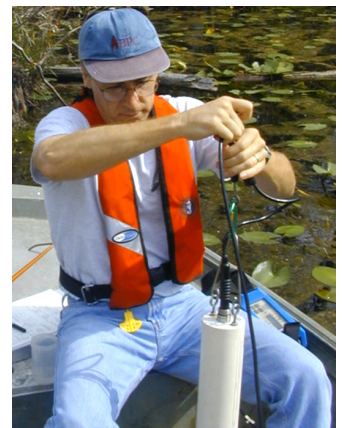
Detta räknas som en "signifikant skillnad". P-et står för "probability" dvs "sannolikhet".

Men vilken sannolikhet är det P-värdet avser?

- sannolikheten att arbetshypotesen är sann, förutsatt att testets antaganden är uppfyllda
- sannolikheten att nollhypotesen är sann, förutsatt att testets antaganden är uppfyllda
- sannolikheten att resultatet uppstått av slumpen (dvs att är ett oturligt sammanträffande)
- sannolikheten att uppnå resultatet, eller ett mer extremt resultat, om nollhypotesen är sann
- sannolikheten att korrekt förkasta nollhypotesen, om det egentligen inte finns någon effekt
- sannolikheten att om man gör experimentet igen så kommer man kunna dra samma slutsats

Fråga 7

En forskare var intresserad av vattnets syremättnad (mätt i procent) i en viss sjö. Han tog ett antal prover på olika platser i sjön med sin syremätare. Han ville veta om dessa data uppfyller antagandena för ett statistisk test.



Öppna Excelfilen **SyreSjö.xlsx**

Mätvärdena finns i kolumnen DO som står för Dissolved Oxygen. Gör ett histogram över dessa DO-värden.

a) Vad drar du för slutsatser om fördelningen av DO-värdena?

- DO-värdena verkar vara normalfördelade
- DO-värdena verkar inte vara normalfördelade

För att förbättra eventuella normalitetsproblem kan du testa att transformera responsvariabeln (DO). Testa att göra både log-transformering och kvadratrotstransformering.

Tips: Skriv rubriken "logDO" i cell C1 och rubriken "rotDO" i cell D1. En transformering görs genom att skriver in formler i cell C2 och D2. Dessa formler skall använda kolumnen B (dvs DO-värdena) som referens. När du skrivit in formeln i cell C2 och D2 kan du "fylla ned" för att automatiskt räkna ut resten av värdena.

Formeln för logaritm är =log och för kvadratrots är =rot (engelska =sqrt)

Gör ett histogram för de två nya, transformerade, variablerna. Du skall nu ha 3 histogram (DO, logDO och rotDO)

Tips. Om du kopierar ett diagram kan du ändra i den ena vilka data som den skall baseras på.

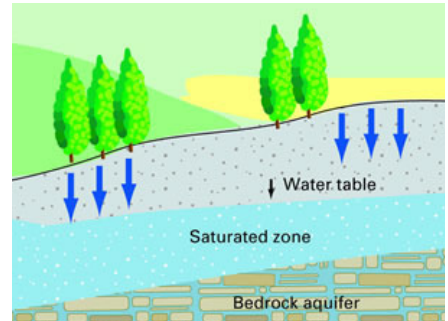
Ändra rubrikerna i histogrammen så att du kan se vilket diagram som baserat på vilken data; DO, logDO eller rotDO.

b) Vilken transformering var bäst på att uppnå normalfördelning för dessa data?

- Log-transformeringen ger mest normal fördelning
- Kvadratrotstransformeringen ger mest normal fördelning
- Båda transformeringarna är sämre än den otransformerade datan

Fråga 8 - Korrelation

En forskare mätte hur djupt grundvattnet befann sig på nio olika platser, först på vintern och sedan på sommaren. Hon var intresserad av om sommarens grundvattendjup var associerat med vinterns grundvattendjup. Dvs, hon ville veta om vinter- och sommarmätningar hängde samman, så att platser med djupt grundvatten på vintern också hade djupt grundvatten på sommaren, osv.



Öppna Excelfilen **Grundvatten.xls**

Börja alltid med att grafiskt plotta dina data! Det enklaste sättet att få en överblick vid korrelation är att göra ett s.k. punktdiagram.

Tips: Markera datan i kolumn B och C. Välj sedan [Infoga/Diagram/punktdiagram utan linje](#).
engelska [Insert/Charts/Scatter no line](#)

Skriv in en vettig diagramrubrik och vettiga namn på axlarna

Tips:

Klicka på Diagramrubriken "Vinter" och ändra till "Grundvatten"

Office 365: [Diagramverktyg/Design/Lägg till diagramelement/Axelrubriker](#).

Office ProPlus: [Diagramdesign/Lägg till diagramelement/Axelrubriker](#)

Engelska: [Chart Tools/Design/Add chart element/Axis Titles](#).

Vågrät axel skall heta "Sommar" och lodrät axel Skall heta "Vinter"

a) Vilka är antagandena för ett korrelationstest?

(Välj tre)

- | |
|--|
| I) Oberoende data
II) Lika många datapunkter i båda variablerna
III) Båda responsvariablerna skall vara normalfördelade
IV) Residualerna skall vara formade som en tratt
V) Datasvärmen skall inte ha någon tydlig kurva
VI) I en Normalsannolikhetskurva skall punkterna forma ett tydligt "S" |
|--|

Antagande 1: _____ Antagande 2: _____ Antagande 3: _____

Kontrollera antagandena genom att göra en boxplot var för de två variablerna Sommar och Vinter.

b) Anser du att antagandena är uppfyllda?

- Nej, boxplottarna är "konsekvent skeva" (dvs att segmenten blir längre och längre). Även med få datapunkter bör man ha helt jämna boxar.

- Ja, visserligen är boxplottarna inte helt symmetriska, men de är inte heller "konsekvent skeva" (dvs att segmenten blir längre och längre). Det är svårt att få helt jämna boxar med så få datapunkter.

Eftersom det är så få datapunkter i detta dataset (bara nio), måste man tolka boxplottarna med viss försiktighet. Forskaren bedömde att antagandet om normalitet var uppfyllt.

Räkna ut korrelationskoefficienten r .

Denna kallas även Pearson's product moment correlation.

Tips använd formeln =korrel(sommardata ; vinterdata)

engelska =correl

c) Vad är Korrelationskoefficienten, r ? $r =$ _____

I Excel är det inte så enkelt att få fram t-värdet och P-värdet för att se om korrelationen är signifikant eller inte. Men t- och P-värdet kan man få från en Regression. P-värdet är nämligen identiskt i en enkel korrelation och en enkel linjär regression. Och t-värdet i en korrelation är identiskt med testet för linjens lutning i regression. Så gör en regression med dessa data och ignorera allt utom just t- och P-värdena.

Tips. Välj: Data/Analys/Dataanalys/Analysverktyg :Regression

Engelska Data/Data analysis/Analysis tools :Regression

Markera Sommar-värdena för Y-indataområde (Y range), och Vinter-värdena för X-indataområde (X range). Klicka i "Etiketter" (Labels) om du markerat rubrikerna. I "utdataområde" (output range) väljer du en cell där du vill ha resultaten. Klicka OK. t-värdet och P-värdet anges på understa raden vid "t-kvot" (t Stat) och "p-värde" (p-value).

d) Vad är t-värdet? $t =$ _____

e) Vad är P-värdet? $P =$ _____

f) Vad drar du för biologisk slutsats om grundvattennivån på dessa nio platser?

- Grundvattennivåerna på olika platser var signifikant positivt korrelerade mellan vinter och sommar
- Grundvattennivåerna på olika platser var signifikant negativt korrelerade mellan vinter och sommar
- Grundvattennivåerna på olika platser var inte korrelerade mellan vinter och sommar

Fråga 9

På ett äldreboende ville en sjuksköterska ta reda på om det finns ett samband mellan blodtryck och puls bland personer över 70 år. Hon mätte pulsen och blodtrycket hos trettio patienter.

Öppna Excelfilen **gamlingar.xlsx**

Börja alltid med att grafiskt plotta dina data!

Gör ett punktdiagram med responsvariablerna blodtryck och puls.



Tips. När du gjort diagrammet kan du fixa till axelrubrikerna: markera diagrammet och välj Office 365: Diagramverktyg/ Design/ Lägg till diamelement/Axelrubriker/ primär vertikal och horisontell.

Office ProPlus: Diagramdesign/Lägg till diamelement/Axelrubriker/ primär vertikal och horisontell.

Engelska: Chart tools/ Design/ Add chart element/ Axis titles/ primary vertical & horizontal.

Sedan kan du markera axelrubrikerna och skriva in Puls på X-axeln och Blodtryck på Y-axeln.

Justera in Y-axeln (Blodtryck) så den går från 140 till 230, och X-axeln (Puls) så den går från 40 till 110.

a) Vilken är nollhypotesen (H_0) som testas med en korrelation mellan blodtryck och puls?

- Högre blodtryck leder till högre puls.
- Värdena på blodtryck är inte associerat med värdena på puls.
- Högre blodtryck leder inte till högre puls.
- Högre blodtryck leder till lägre puls.
- Värdena på blodtryck är associerat med värdena på puls.

Gör en boxplot för Puls-värdena och en boxplot för Blodtryck-värden och kolla att de är vara normalfördelade. Vid behov så kan du ändra max/min-värdena på boxplottarnas Yaxlar, så att boxarna inte blir för hoptryckta.

b) Utifrån hur punktdiagrammet ser ut, vilket ungefärligt värde förväntar du dig på korrelationskoefficienten, r ?

- negativt (mellan -1 och 0)
- runt noll
- positivt (mellan 0 och +1)

Räkna ut korrelationskoefficienten r . (Tips: se föregående fråga)

Fundera igenom hur många frihetsgrader du har för detta korrelationstest.

Rapportera följande:

c) $r =$ _____

d) frihetsgrader = _____

Gör en Regression för att få reda på P-värdet på korrelationstestet. (Tips: se föregående fråga)

Var är P-värdet?

e) $P =$ _____

f) Vad gör du med nollhypotesen?

- Behåller den, det finns en statistiskt signifikant association!
- Förkastar den, det finns en statistiskt signifikant association!
- Behåller den, det finns ingen statistiskt signifikant association!
- Förkastar den, det finns ingen statistiskt signifikant association!

g) Vad drar du för medicinsk slutsats?

- Det fanns en signifikant positiv korrelation mellan blodtryck och puls hos personer på äldreboende.
Dvs, patienter med högt blodtryck hade också högre puls, och vice versa.
- Det fanns en signifikant negativ korrelation mellan blodtryck och puls hos personer på äldreboende.
Dvs, patienter med högt blodtryck hade lägre puls, och vice versa.
- Det fanns en signifikant positiv korrelation mellan blodtryck och puls hos personer på äldreboende.
Dvs, patienter med högt blodtryck hade också lägre puls, och vice versa.

Fråga 10

En forskare studerade övergödning i svenska insjöar. Han var intresserad av hur mycket syre som växtplankton (encelliga alger) konsumerade under natten, och om detta påverkades av hur mycket gödningsämnen det fanns i vattnet. Han tog prover i 50 insjöar och mätte dels mängden fosfor (ett gödningsmedel) och dels hur mycket syre algerna i vattnet konsumerade på en timme.



Öppna Excelfilen **alger.xlsx**

Datasetet har tre kolumner:

plats	sjöarnas ID-nummer
fosforkonc	fosforkoncentrationen i mg/l
syrekonsum	O ₂ konsumtionen under en timme i mg/l

Utifrån Forskarens frågeställning,

a) vad är förklaringsvariabel (X) i denna undersökning? _____

b) vad är responsvariabel (Y) i denna undersökning? _____

Gör ett punktdiagram. Se till att rätt värden hamnat på rätt axel. Ge axlarna rätt axelrubriker.

c) Titta på datasvärmen i ditt punktdiagram. Baserat på det du kan se såhär långt, anser du att antagandena för en regression verkar uppfyllda? _____

Gör en regression med fosforkoncentration och syrekonsumtion.

Tips. Data/Analys/Dataanalys. Y-indata= kolumn C. X-indata= kolumn B. Välj utdataområde.

Eng: Data/Analys/Data analysis. Y-range= kolumn C. X-range= kolumn B. Välj Output range

Glöm inte att klicka i Residualdiagram (Residual Plot)

och Normalsannolikhetskurva (Normal Probability Plots)!

Till skillnad från en korrelation är det nu viktigt att rätt variabel hamnar som X och rätt variabel hamnar som som Y, *både* i diagrammet och i regressionstestet.

Dra ut diagrammen Residualdiagram och Normalsannolikhetskurva så att de blir ungefär lika höga som breda. Inspektera dessa två diagram

d) Vad drar du för slutsater om antagandena nu?

- Antagandena är uppfyllda.
- Antagandena är inte uppfyllda, dvs denna regression är inte pålitlig.

Skapa en ny variabel genom att log-transformera `syrekonsum`, kalla denna `logsyrekonsum` (Om du inte minns hur man skapar en ny transformerad variabel, se tidigare fråga.)

Gör ett nytt punktdiagram fast med `logsyrekonsum` som responsvariabel

Tips: Om du kopierar det gamla punktdiagrammet och klistrar in en kopia, kan du ändra indata genom att klicka Diagramverktyg/Design/Markera data/

Diagramdesign/Markera data

Chart Tools / Design/Select Data

Ta sedan bort den gamla datan. Klicka sedan "lägg till" (Add) och skriv

Serienamn=`logsyrekonsum`. Markera värden för X-serie (kolumn B) och Värden för Y-serie(Kolumn D). Glöm inte att ändra Yaxelns namn.

Gör om regressionstestet med `logsyrekonsum` som responsvariabel, och inklusive Residualdiagram och Normalsannolikhetskurva. Välj ett nytt utdataområde så du inte skriver över något!

Dra i diagrammen Residualdiagram och Normalsannolikhetskurva så att de blir lika höga som breda. Inspektera dessa två diagram.

e) vad anser du om antagandena för att få göra en regression nu?

- Bättre normalfördelning av residualer och ingen "residualtratt": antagandena är uppfyllda.
- Antagandena är fortfarande inte uppfyllda, dvs denna regression är inte pålitlig.

Inspektera nu resultatet från den nya regressionen. Denna finner du under Koefficienter i utdata-tabellen.

f) Vilka två nollhypoteser (H_{0a} & H_{0b}) är det som testas med de två t-testen nederst i tabellen?

Tips: I Svenska Excel kallas intercept för "konstant" och testas på näst nedersta raden. Lutningen testas på raden längst ner, som av excel kallas "X-värde" alternativt "fosforkonc" .

- H_{0a} :
- Interceptet är noll
 - Interceptet är mindre än ett
 - Då $X=0$ är Y mindre än noll
- H_{0b} :
- När X blir större blir Y större
 - Lutningen är mindre än ett
 - När X blir större blir Y mindre
- Interceptet är större än ett
 - Då $X=0$ är Y större än noll
 - Lutningen är större än ett
 - Lutningen är noll

g) Vad drar du för slutsatser om dessa två nollhypoteser?

- Jag förkastar H_{0a} men behåller H_{0b}
- Jag förkastar både H_{0a} och H_{0b}
- Jag behåller H_{0a} , men förkastar H_{0b}
- Jag behåller både H_{0a} och H_{0b}

h) Vad är estimatet för interceptet? $a =$ _____

i) När fosforconcentrationen är noll mg/l vad är log(syrekonsumtionen) då? _____

j) När fosforconcentrationen är noll mg/l är syrekonsumtionen = _____ mg/l

Tips 1: syrekonsumtionen = $10^{\text{[svaret från i]}}$

Tips 2: i Excel skrivs tiopotenser = 10^{\wedge}

k) Vad är estimatet för lutningen? $b =$ _____

Tips: Eftersom vi har en log-skala på Y-axeln är det lite knepigt att tolka denna siffra. Men det räcker att se att den är positiv, dvs när fosforkoncentrationen ökar, ökar syrekonsumtionen.

Använd de två estimaten för att formulera den rätta linjens ekvation i formen:

$$\log Y = b \times X + a$$

Där du ersätter Y med namnet på responsvariabeln, ersätter X med namnet på förklaringsvariabeln, ersätter a med estimatet för interceptet och b med estimatet för lutningen.

l) $\log(\text{_____}) = \text{_____} \times \text{_____} - \text{_____}$

m) Hur stor del av variationen i log(syrekonsumtion) kan förklaras av variationen i fosforkoncentration (ange svaret i %)? _____%

Resultatet av regressionen presenteras även som en ANOVA-tabell

n) Ange antal frihetsgrader för residualerna: $Df =$ _____

o) Ange F-värdet: $F =$ _____ p) Ange P-värdet: $P =$ _____

q) Vad drar du för biologisk slutsats om hur fosfor påverkar syrekonsumtionen i vattnet?

- Det finns ett statistiskt signifikant positivt linjärt samband mellan fosforkoncentration och syrekonsumtion
- Det finns ett statistiskt signifikant positivt linjärt samband mellan fosforkoncentration och log(syrekonsumtion)
- Det finns inget statistiskt signifikant samband mellan fosforkoncentration och syrekonsumtion
- Det finns ett statistiskt signifikant linjärt negativt samband mellan fosforkoncentration och syrekonsumtion

Slutligen skall du lägga till en regressionlinje i ditt punktdiagram med logsyrekonsumtion och fosforkoncentration.

Tips: Markera punktdiagrammet och välj sedan

Diagramverktyg/ Design/ Lägg till diagramelement/ Trendlinje / Linjär.

Diagramdesign/ Lägg till diagramelement/ Trendlinje / Linjär.

Engelska: Chart Tools/ Design/Add chart element/ Trend line / Linear

Fråga 11

Flera typer av båtbottnfärger påverkar marina djurs fortplantning.

Två miljöanalytiker undersökte skillnader i fekunditeten (dvs antalet ägg)

hos en rovlevande marin snäcka (*Lepsiella vinosa*) i två typer av havsvikar:

Havsvikar MED småbåthamn och *Havsvikar UTAN småbåthamn*.



a) Vilken typ av statistiskt test är lämpligt för denna undersökning?

Svar: _____

Öppna Excel-filen **Ward2.xlsx**

b) Vad var medelfekunditeten för de två populationerna *Lepsiella*-snäckor?

Tips: använd formeln medel= eller använd Dataanalys/ Beskrivande Statistik

Engelska: average=

Data analysis /Descriptive Statistics

Medelfekunditet i vikar Med småbåtshamn: _____

Medelfekunditet i vikar Utan småbåtshamn: _____

Gör en boxplot över fekunditetsvärdena, uppdelat på de två typerna av vikar.

Tips: Markera de 2 kolumnerna, välj Infoga/Diagram/låddiagram

Insert/Box and whisker

För att få en förklaring till vilken box som är vilken, markera diagrammet, klicka

Diagramverktyg/Design/Diagramelement/ Förklaring/Höger

Diagramdesign/Lägg till Diagramelement/ Förklaring/Höger

Chart Tools/Design/Add Chart element/Legend/Right

Justera Y-axelns max och min om du tycker det behövs

c) Vad drar du för slutsatser om antagandet om normalitet? ○ Uppfyllt ○ Ej uppfyllt

d) Vad drar du för slutsatser om antagandet om lika varianser? ○ Uppfyllt ○ Ej uppfyllt

Utför det test du valde i a)

Tips: [Data/Analys/Dataanalys](#)

"Antagen medelvärdeskillnad" är hur mycket du antar att de två grupperna skiljer sig åt om nollhypotesen är korrekt.

e) Hur många frihetsgrader hade du i testet? _____

f) Vad var t-värdet (t-värde kallas "t-kvot" i Excel)? _____

g) Vad var P-värdet (detta kallas "P(T<=t) tvåsidig" i Excel)?

- P>0.05 P<0.001 P=5.43 P>76

Vilka statistiska och miljövetenskapliga slutsatser kan du dra?

h) Statistisk slutsats:

- Det var en signifikant skillnad i fekunditet, nollhypotesen förkastas
- Det var ingen signifikant skillnad i fekunditet, vi behåller nollhypotesen

i) Miljövetenskaplig slutsats:

- Lepsiella*-snäckor har signifikant högre fekunditet i vikar med småbåtshamnar jämfört med snäckor utan småbåtshamnar
- Lepsiella*-snäckor har signifikant högre fekunditet i vikar utan småbåtshamnar jämfört med snäckor med småbåtshamnar.
- Lepsiella*-snäckornas fekunditet skiljer sig inte mellan de två typerna av havsvikar.

Avsluta att med hjälp av formler räkna ut Medelvärde och SE för de två grupperna. Gör sedan ett stapeldiagram med de två medelvärdena som har med SE som felstaplar.

Tips: [Se hur man räknar ut gör stapeldiagram med felstaplar i Excelövningen – "bekanta dig med Excel 365"](#)

Fråga 12

Medley och Clemens undersökte effekten av graden av zinkförorening på antalet arter (diversiteten) kiselalger i Klippiga bergens floder. Zinkkoncentrationen i floderna klassificerades i kategorierna: *natural*, *background*, *low*, *medium* eller *high*. Prova på antalet kiselalger togs sedan i olika floder med olika grader av zinkförorening.



Öppne Excel-filen **Medley3.xlsx**

Datasetet innehåller följande information

DIVERSITY Diversiteten, dvs antal arter (medelvärde per μl från varje flod)

ZINC Klassificering av zinkföroreningsgrad: BACKGROUND (lägst), LOW, MEDIUM, HIGH

a) Vad är förklaringsvariabel?

- ZINC DIVERSITY

b) Vilken typ av data innehåller förklaringsvariabeln?

- Numeriska, diskreta Numeriska, kontinuerliga Kategoriska, nominala Kategoriska, ordinala

c) Var är responsvariabel?

- ZINC DIVERSITY

d) Vilken typ av data innehåller responsvariabeln?

- Numeriska, diskreta Numeriska, kontinuerliga Kategoriska, nominala Kategoriska, ordinala

e) Hur många olika grupper finns det i förklaringsvariabeln?

- Tre: ANTAL, ZINC, DIVERSITY
- Fyra: Back, low, medium, high
- Två: ZINC, DIVERSITY
- Lika många som antalet arter kiselalger

f) Vilken typ av statistiskt test är lämpligt för denna undersökning?

- T-test
- Regression
- Korrelation
- ANOVA

g) Formulera en nollhypotes:

- Det är en signifikant effekt av Zinknivå på diversiteten av kiselalger
- Diversiteten är densamma i ZINC och i DIVERSITY
- Diversiteten är densamma i alla 4 olika Zinkkoncentrationerna
- Det är ingen skillnad på Zinknivåerna i olika floder

Plotta dina data grafiskt genom att göra en boxplot.

Tips. Markera rutan med data (C1: F10) och välj

Infoga/Diagram/statistiska diagram/Låddiagram.

Diagramverktyg/Design/Diagramelement/ Förklaring/Höger

Insert/Charts/Box and whisker

Klicka Diagramverktyg/Design/Diagramelement/ Förklaring nederkant

Diagramdesign/Lägg till Diagramelement/ Förklaring nederkant

Chart Tools/Design/Add chart elements/ Legend Bottom

Klicka Diagramverktyg/Design/Diagramelement/Axelrubriker.

Diagramverktyg/Design/Diagramelement/Axelrubriker.

Add chart elements/ Axis titles

Döp Xaxeln till "Zink level" och Yaxeln till "Diversity"

Utifrån denna boxplot, vad är din uppfattning om antagandena för testet du valde i f)?

h) Normalitetskravet?

- Ingen konsekvent assymetri - dvs kravet på normalitet är uppfyllt
- Ingen konsekvent assymetri - dvs kravet på normalitet är inte uppfyllt
- De olika boxarna är olika stora - kravet är inte uppfyllt

i) Kravet på lika varianser?

- Det går inte att se i denna typen av graf
- Ett tydligt samband mellan medelvärdet och variansen - dvs kravet på lika varianser är inte uppfyllt
- Inget samband mellan medelvärdet och variansen - dvs kravet på lika varianser är uppfyllt.

Utför testet du valde i f)

Tips: Data/Dataanalys/välj test /markera indataområde och utdataområde (input/output range).

j) Ange antal frihetsgrader för faktorn zink-nivå ("mellan grupper"): Df= _____

k) Ange antal frihetsgrader för residualerna ("inom grupper"): Df = _____

l) Ange F-värdet: F = _____

m) Ange P-värdet: P = _____

n) Vad drar du för slutsats?

- Det var ingen signifikant effekt av Zinkkoncentration på diversiteten av kiselalger
- Det var en signifikant effekt av Zinkkoncentration på Artantalet (diversiteten) av kiselalger
- Medelvärdena för zinkkoncentrationerna är inte de samma i de fyra olika grupperna
- Medelvärdena för diversiteten kiselalger är de samma i de fyra olika zinkkoncentrationerna

Vi skulle kunna gå vidare här med att ta reda på vilka zinknivåer som skiljer sig från vilka. I så fall behöver vi antingen använda ett annat statistikprogram, eller manuellt göra 6 t-tester och justera p-värdet med Bonferroni-korrektion. Du behöver inte göra det på denna övning.

Fråga 13 - Regression

En lantbrukare var intresserad av hur mängden gödning påverkar hur mycket gräset växer. Hon sådde in ett område med gräsfrön. Sedan applicerade hon olika mängder handelsgödsel på olika 1 m² stora rutor. Två månader senare skördades varje yta för sig och gräset torkades och vägdes.



Öppna Excel-filen **gödningsexperiment.xlsx**

Filen har två kolumner

A) GÖDNING mängd gödning (g/m²)

B) SKÖRD Mängden skördat gräs (g/m² torrsvikt)

13_1. Vilken är den lantbruksvetenskapliga hypotesen av intresse?

- a) "Det finns ett samband mellan mängden gödning och storleken på grässkörden."
Dvs "Gödningens mängd påverkar grässkörden storlek"
- b) "Det finns en korrelation mellan mängden gödning och storleken på grässkörden "
Dvs "Gödningens mängd och grässkörden mängd är associerade"
- c) "Det finns ett samband mellan storleken på grässkörden och mängden gödning."
Dvs " Grässkörden storlek påverkar gödningens mängd "

13_2. Vilken är den statistiska nollhypotesen av störst intresse?

- a) Grässkörden påverkar inte mängden gödning (lutningen i regressionen är noll)
- b) Mängden gödning påverkar inte grässkörden (lutningen i regressionen är noll)
- c) Mängden gödning påverkar inte grässkörden (interceptet i regressionen är noll)

Gör ett punktdiagram med SKÖRD och GÖDNING. Fundera igenom vad som skall vara på X-axeln, och vad som skall vara på Y-axeln i diagrammet.

Tips. Markera datan, Infoga/diagram/punktdiagram. Lägg till Axeltitlar på X och Y-axel.

Eng: Insert/Chart/Scatter

Axis titles

13_3. Är detta ett av antagandena för en regression?

- a) Lutningen på regressionlinjen är noll **ja/nej**
- b) Oberoende data **ja/nej**
- c) Interceptet på regressionslinjen är noll **ja/nej**
- d) Residualerna är normalfördelade **ja/nej**
- e) Inget samband mellan emdelvärde och varians **ja/nej**
- f) Ingen tydlig kurva **ja/nej**

13_4. Så långt, kan du se att några av antagandena för en regression inte är uppfyllda?

- a) Ja b) Nej

Gör nu en regression. Tänk på att rätt variabel hamnar på X och på Y.

Tips. Data/Dataanalys/regression.

Välj Residualdiagram och Normalsannolikhetskurva. Expandera dessa så de blir ungefär lika höga som breda

13_5. Tyder Residualdiagrammet (Residual plot) på att det är problem med dessa data?

- a) Ja, det är ett starkt samband mellan medelvärde och varians
- b) Nej allt tyder på att datan är normalfördelad
- c) Nej, det finns ingen "triangel" dvs inget samband mellan medel och varians

Kolla också Normalsannolikhetskurvan för att säkerställa att punkterna ligger något snånär på en rak linje.

Titta på estimaten för a och b i outdatatabellen. Interceptet "a" kallas "Konstant" i svenska Excel och testas på näst understa raden. Lutningen "b" testas på understa raden

$$\begin{array}{l} \text{Räta linjens ekvation är:} \\ \text{eller} \end{array} \quad \begin{array}{l} \text{SKÖRD} = \text{Intercept} + \text{Lutning} \times \text{GÖDNING} \\ \text{SKÖRD} = a + b \times \text{GÖDNING} \end{array}$$

13_6. Ange följande

- | | |
|---|---------------|
| a) Intercept | a= _____ |
| b) Lutning | b= _____ |
| c) P-värde för H_0 att lutningen är 0 | P= _____ |
| d) F-värde för hela regressionsmodellen | F= _____ |
| e) P-värde för hela regressionsmodellen | P= _____ |
| f) R^2 (ange som proportion) | R^2 = _____ |

13_7. Hur skulle du sammanfatta detta resultat i en vetenskaplig rapport?

- a) Det var ett signifikant positivt linjärt samband mellan gödningsnivå och gräsavkastning (Regression; $P < 0,001$).
- b) Gräsavkastningen påverkade gödningsnivån positivt på ett signifikant sätt (Regression, $R^2=0,92$; $F_{1\&8}=94,04$; $P < 0,001$).
- c) Gödningsnivå och gräsavkastning var signifikant korrelerade (Korrelation; $t=9,697$; $P < 0,001$).
- d) Det var ett signifikant positivt linjärt samband mellan gödningsnivå och gräsavkastning (Regression; $R^2=0,92$; $F_{1\&8}=94,04$; $P < 0,001$).

När man presenterar signifikanta regressioner i rapporter brukar man använda punktdiagram med trendlinje (dvs med regressionslinjen). Lägg till en trendlinje till ditt punktdiagram

Tips: markera diagrammet, klicka plus-symolen och

[Diagramelement/Trendlinje/linjär](#)

[Diagramdesign/lägg till diagramelement/Trendlinje/linjär](#)

[Add Chart Elements/Trend line /linear](#)