

Lab 2 - S0001M

19 okt. 2026, 9 min läsning

#S0001M #lab #minitab

Laboration 2 – S0001M

Vad labben handlar om

Fortsättning på Lab 1, med samma data (ditt kropps mått, 18 kvinnor + 18 män, dominant och icke-dominant sida). Du ska:

1. Kolla om datan kan antas vara **normalfördelad** (normalfördelningsdiagram).
2. Jämföra **kvinnor mot män** (två oberoende stickprov), både med och utan normalfördelningsantagande.
3. Jämföra **dominant mot icke-dominant sida** (stickprov i par), både med och utan normalfördelningsantagande.

Varje jämförelse görs med *två* metoder – en som kräver normalfördelning och en som inte gör det – och sedan jämför du slutsatserna. Båda metoderna ska alltid göras, oavsett vad normalfördelningskollen visar.

Litteratur

- Vännman & Jonsson, *Matematisk statistik*, 3:e uppl: **kap 8.4, 9.6, 9.7** (2:a uppl: 8.3, 9.5, 10.1)
- Minitab-handledningen (Canvas), främst **avsnitt 2.4, 3.1, 4.1–4.3**

0. Förberedelser (före labbpasset)

- ☐ Läs kap 8.4, 9.6 och 9.7 i boken
- ☐ Läs hela den här instruktionen + relevanta delar av Minitab-handledningen
- ☐ (Valfritt) Titta på den inspelade genomgången inför Lab 2 i Canvas
- ☐ Ha datafilen från Lab 1 redo (dominant, icke-dominant, kön, differens)

1. Normalfördelningsdiagram

Idé: I ett normalfördelningsdiagram (Minitab: *probability plot*) hamnar punkterna ungefär på en rak linje om datan är normalfördelad. Böjda mönster (S-form, banan-form) tyder på att den inte är det. Lådagrammen från Lab 1 ger en första ledtråd: symmetriska lådor → normalfördelning kan vara rimlig, tydligt sned → troligen inte.

Innan du tittar på din egen data ska du simulera data för att se hur “normal” och “icke-normal” data ser ut i diagrammet.

Minitab-kommandon

- **Simulera:** Calc → Random Data → [fördelning]. Ange antal rader och vilken kolumn de ska sparas i. (Rektangelfördelning = *Uniform*.)
- **Normalfördelningsdiagram:** Graph → Probability Plot → *Single* (en variabel) eller *Multiple* (flera i samma diagram). Normalfördelning är förvald.
- **Döpa kolumn:** skriv namnet i rutan ovanför kolumnen i kalkylbladet.

(a) Litet normalfördelat stickprov

- ☒ Simulera **10 värden från $N(1, 2)$** (väntevärde 1, standardavvikelse 2) till kolumn C1, döp den till **N10**. Gör ett normalfördelningsdiagram över **N10**.

(b) Stora stickprov från tre olika fördelningar

- ☒ Simulera **500 värden från $N(1, 2)$** → kolumn **N500**. Gör normalfördelningsdiagram.
- ☒ Simulera **500 värden från $R(0, 2)$** (rektangel/Uniform, undre gräns 0, övre 2) → kolumn **R500**. Gör normalfördelningsdiagram.
- ☒ Simulera **500 värden från $Exp(1)$** (exponential, mean = 1) → kolumn **E500**. Gör normalfördelningsdiagram.
- ☐ Jämför de tre diagrammen och beskriv **med egna ord** hur de skiljer sig (rak linje? böjd i ändarna? sned åt ena hållet?).

(b½) Övning – ingår INTE i rapporten

Gör detta ändå!

Ditt eget stickprov har bara 18 värden per grupp. Små stickprov ser “spretiga” ut även när de verkligen är normalfördelade. Den här övningen ger dig en känsla för hur mycket avvikelse som är normalt, så att du inte dömer ut din data för snabbt i (c).

- ☒ Simulera **5 stickprov med 18 värden vardera från $N(0, 1)$** (t.ex. C10–C14 på en gång). Gör normalfördelningsdiagram för vart och ett, titta på skärmen och reflektera över hur mycket de avviker från linjen.

(c) Din egen data – kvinnor och män, dominant sida

- ☐ Ange **vilket kroppsmått** du valt och **hur du definierade/mätte det**.
- ☒ Gör normalfördelningsdiagram för den **dominanta sidan** för kvinnor och för män (gärna båda i samma diagram: **Probability Plot → Multiple**, eller *Single* med kön som grupp).
- ☐ Bedöm och **motivera tydligt**: är normalfördelningen en rimlig modell för måttet hos kvinnor respektive män? (Jämför med det du såg i b½.)

2. Jämförelse kvinnor vs. män (två stickprov)

Varför: Lådagrammen i Lab 1 (uppg. g) kanske antydde att kvinnor och män skiljer sig. Nu ska du *påvisa* det statistiskt. Grundantagande: datan kommer från någon kontinuerlig fördelning. Två metoder:

Metod	Kräver normalfördelning?	Jämför
(d) Lådagram med medianintervall	Nej	Medianer
(e) KI för två stickprov (t)	Ja	Väntevärden

(d) Utan normalfördelning – lådagram med medianintervall

Så funkar det: Minitab räknar ut ett **95 % konfidensintervall för fördelningens median** (inte stickprovsmedianen!) för varje grupp, med teckenintervall-metoden (kap 8.2, samma som Lektion 11). Den kräver bara att fördelningen är kontinuerlig.

Beslutsregel

Överlappar intervallen inte → medianerna skiljer sig. Två separata 95 %-intervall ger tillsammans en signifikansnivå på **högst 10 %**.



Graph → Boxplot → One Y / With Groups → Graph variable: dominant sida, Categorical variable: kön → Data View → bocka i **Median confidence interval box**. Den inre lådan är medianintervallet.

- ☒ ~~Gör lådagram med medianintervall för dominant sida, uppdelat på kön.~~
- ☐ Avgör med beslutsregeln om medianerna skiljer sig.
- ☐ Skriv **slutsats + risknivå** (= signifikansnivå, alltså högst 10 %) och motivera tydligt.

(e) Med normalfördelning – konfidensintervall för två stickprov

Under normalfördelning är median = väntevärde, så “genomsnitt” betyder här **väntevärde**. Metoden är effektivare och ger dessutom en uppskattning av **hur stor** skillnaden är. Anta **lika varians** i båda grupperna.



Stat → Basic Statistics → 2-Sample t

- Data i en kolumn med kön i en annan → *Both samples are in one column*; annars *Each sample is in its own column*.
- Options → **Confidence level: 90** och bocka i **Assume equal variances**.

- ☐ Ta fram ett **90 % konfidensintervall** för skillnaden i väntevärde (kvinnor – män eller tvärtom; var tydlig med ordningen).
- ☐ Tolka i ord: **finns det en påvisbar skillnad** (innehåller intervallet 0?) och i så fall **hur stor** är den?

(f) Antaganden + räkna KI:t för hand

- ☐ Skriv upp **modellantagandena för två stickprov** (bok s. 246, gamla s. 216) och **definiera alla beteckningar**. Ungefär:
 - $X_1, \dots, X_{n_1}$ (kvinnor) är oberoende och $N(\mu_1, \sigma)$
 - $Y_1, \dots, Y_{n_2}$ (män) är oberoende och $N(\mu_2, \sigma)$
 - De två stickproven är oberoende av varandra, samma σ i båda.
- ☐ Ta fram medelvärden och standardavvikelser i Minitab (Stat → Basic Statistics → Display Descriptive Statistics, By variables: kön) och **bifoga utskriften**.
- ☐ Räkna intervallet för hand, **redovisa formeln**, och kontrollera att det stämmer med Minitab:

$$\bar{x} - \bar{y} \pm t_{0.05}(n_1 + n_2 - 2) \cdot s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \quad s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

(90 % $\rightarrow \alpha/2 = 0.05$. Med $n_1 = n_2 = 18$ blir frihetsgraderna 34.)

- ☐ Diskutera: **är antagandena rimliga** för din data? Jämför med (c).

(g) Jämför (d) och (e)

- ☐ Ger metoderna **samma** eller **olika** slutsats?
- ☐ Om olika: diskutera vilken slutsats som bör gälla (tänk på vad (c) sa om normalfördelningen och att t-metoden är effektivare om antagandet håller).
- ☐ Slutsats: **kan man på 10 % signifikansnivå påstå att det finns en genomsnittlig skillnad** mellan kvinnor och män – ja eller nej?

3. Jämförelse dominant vs. icke-dominant sida (stickprov i par)

Varför en annan metod? Varje person har *två* mätningar (dominant + icke-dominant), så de två grupperna är inte oberoende – metoden för två stickprov går inte att använda. Istället bildar man **differensen per person** (dominant – icke-dominant, precis som i Lab 1 uppg. h). Då försvinner variationen mellan individer och man analyserar bara differenserna (kap 8.4.1).

Välj grupp

I (h)–(l) väljer du själv **antingen kvinnor eller män**. Använd samma grupp genom hela avsnittet.

Metod	Kräver normalfördelning?
(h) Teckentest (för hand)	Nej
(j) KI för stickprov i par (t)	Ja

(h) Teckentest för hand

Idé: Om det inte finns någon skillnad mellan sidorna är en differens lika sannolikt positiv som negativ. Räkna antal plus-tecken och se om det är orimligt många eller få. (Kap 9.6, se **exempel 9.14**.)

- ☐ Välj grupp (kvinnor/män) och **bifoga utskrift av differensvärdena**.
- ☐ **Stryk alla differenser som är exakt 0**. Antalet kvar kallas n (blir alltså färre än 18).
- ☐ Ställ upp **hypoteserna** (tvåsidig mothypotes):
- H_0 : medianen för differensen är 0 (ingen skillnad mellan sidorna)
 - H_1 : medianen för differensen $\neq 0$

- ☐ Ange **testvariabel**: Y = antal positiva differenser. Under H_0 är $Y \sim \text{Bin}(n, 0.5)$.
- ☐ Ange **beslutsregel** på 10 % nivå: förkasta H_0 om $Y \leq c$ eller $Y \geq n - c$, där c är största talet så att $P(Y \leq c) \leq 0.05$ (5 % i varje svans). Alternativt: räkna p-värdet $2 \cdot P(Y \leq y_{\text{obs}})$ (eller övre svansen) och jämför med 0.10.
- ☐ Räkna ut, dra **slutsats** och motivera tydligt.

Om din data

Mätningarna är avrundade till mm, så det finns troligen **många nollor** bland differenserna. Det gör att n kan bli ganska litet – kom ihåg att använda det nya n i binomialfördelningen.

(i) Kolla normalfördelning

Instruktionen gör förenklingen: *om måttet på båda sidorna är normalfördelat antar vi att differensen också är det* (i praktiken stämmer det oftast, även om sidorna inte är oberoende).

- ☐ Gör **ett normalfördelningsdiagram för icke-dominant sida och ett för dominant sida** (för din valda grupp).
- ☐ Bedöm om normalfördelning är en rimlig modell för differenserna.

(j) Konfidensintervall för stickprov i par (Minitab)

Ska göras även om (i) visade att normalfördelning inte är rimlig. Ger både svar på *om* det finns en skillnad och *hur stor* den är.

Minitab

Stat → Basic Statistics → Paired t → välj kolumnen för dominant och icke-dominant sida (de måste ligga i två separata kolumner, och bara din valda grupp – använd t.ex. Data Options → Subset med Kön="K" eller "M"). Options → **Confidence level: 90**. Alternativ: 1-Sample t direkt på differenskolumnen ger samma intervall.

- ☐ Ta fram ett **90 % konfidensintervall** för den genomsnittliga differensen.
- ☐ Tolka i ord: **finns det en påvisbar skillnad** (innehåller intervallet 0?) och i så fall **hur stor**? Se exempel 8.6 (s. 238) och 8.9 (s. 244) för hur en tolkning kan låta.

(k) Antaganden + räkna KI:t för hand

- ☐ Skriv upp **modellantagandena för stickprov i par** (bok s. 241) och **definiera alla beteckningar**. Ungefär:
 - $D_i = X_i - Y_i$ = differensen dominant – icke-dominant för person i , $i = 1, \dots, n$
 - $D_1, \dots, D_n$ är oberoende och $N(\Delta, \sigma)$, där Δ är den genomsnittliga skillnaden

- ☐ Ta fram medelvärde och standardavvikelse för differenserna i Minitab (som i Lab 1) och **bifoga utskriften**.
- ☐ Räkna intervallet för hand, **redovisa formeln**, och kontrollera mot Minitab:

$$\bar{d} \pm t_{0.05}(n-1) \cdot \frac{s_d}{\sqrt{n}}$$

(Här är $n = 18$ – nollorna stryks **inte** i den här metoden.)

(1) Jämför (h) och (j)

- ☐ Ger teckentestet och KI:t **samma** eller **olika** slutsats?
- ☐ Om olika: diskutera vilken slutsats som bör gälla.
- ☐ Slutsats: **kan man på 10 % signifikansnivå påstå att det finns en genomsnittlig skillnad** mellan dominant och icke-dominant sida?

4. Redovisning

Ingen formell rapport behövs (ingen inledning/metod/teori). Klistra in Minitab-utskrifter, tabeller och diagram i ett dokument och exportera till PDF.

- ☐ **Försättsblad** med namn på alla i gruppen
- ☐ Uppgifterna numrerade **(a)–(l)** som i instruktionen
- ☐ Varje utskrift/diagram är tydligt kopplat till sin uppgift
- ☐ Alla beteckningar **definierade** i (f), (h) och (k)
- ☐ **Beräkningar redovisade** i (f), (h) och (k)
- ☐ Alla slutsatser tydligt **motiverade**
- ☐ Snyggt och läsbart (en slarvig redogörelse måste göras om helt)
- ☐ Inlämnad som **PDF i Canvas** senast 2 veckor efter labbpasset