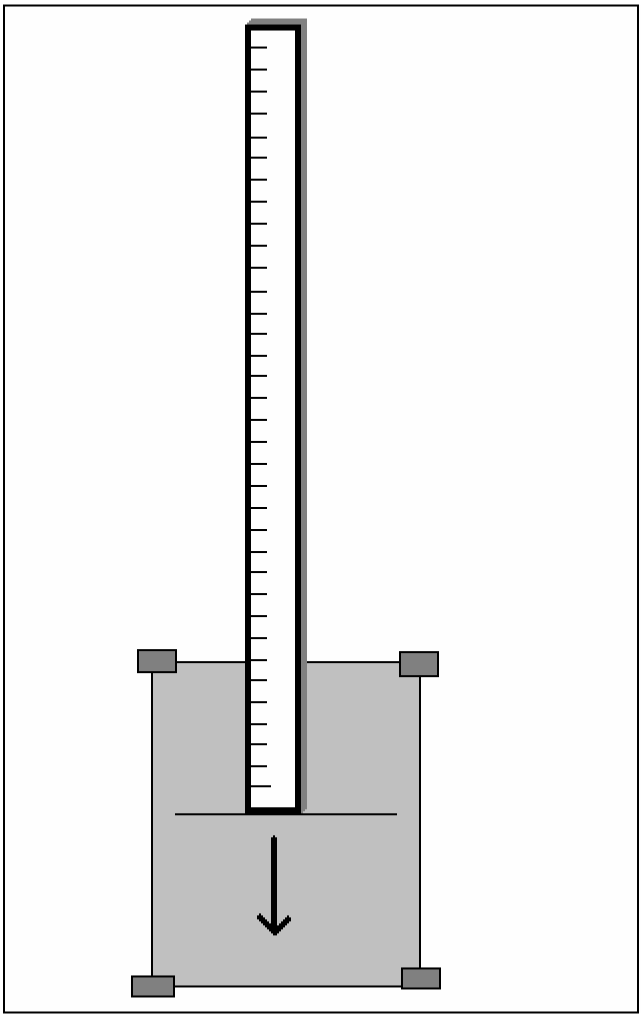
Sammenligning af to måleserier

En af de mest grundlæggende problemstillinger i statistik består i at undersøge om to forskellige måleserier er signifikant forskellige eller om forskellen på de to serier lige så godt kan tilskrives tilfældige variationer. Der findes selvfølgelig utallige simple eksperimenter, der giver anledning til en sådan sammenligning af måleserier, men her vil vi se på et simpelt eksperiment, der nemt kan udføres fra starten af 1g: Måling af reaktionstiden.



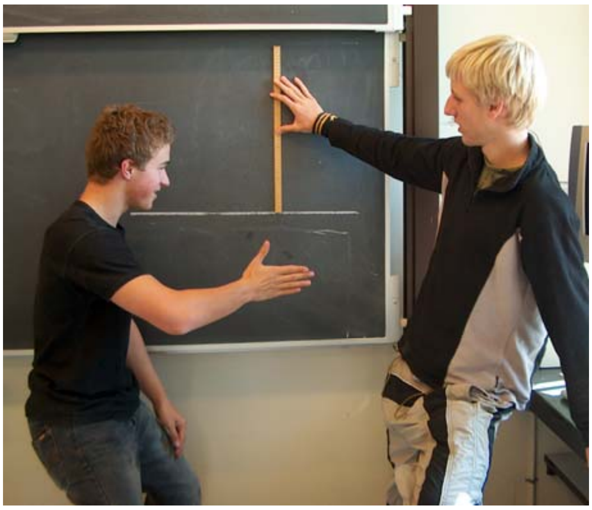
Det sker som vist ved at man lader en lineal falde og derefter ser, hvor langt den falder, og dernæst bruger denne faldlængde til at udregne reaktionstiden. Det forudsætter altså lidt teori om faldloven:

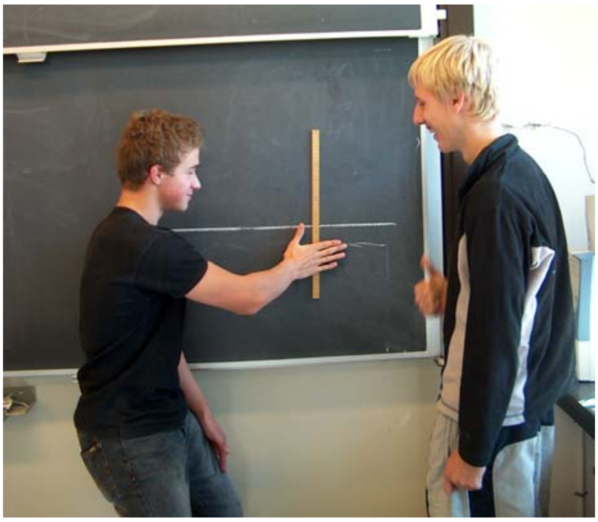
Ifølge Galileis faldlov, som er en af hjørnestenene i den moderne naturvidenskab, er sammenhængen mellem faldtiden t og faldlængden s givet ved formlen:

$$s = \frac{1}{2} \cdot g \cdot t^2$$

hvor g er tyngdeaccelerationen, der varierer en lille smule fra sted til sted. Her vil vi ignorere de lokale variationer og bruge den såkaldte **standardacceleration**, der har værdien 9.80665 m/s^2. Vi kan derfor omsætte en faldlængde til en faldtid ved at isolere t i formlen:

$$t = \sqrt{\frac{2s}{g}}$$





Eleverne arbejder sammen to og to, hvor den ene skal slippe linealen uventet, mens den anden skal forsøge at stande linealen i dens fald så hurtigt som muligt. Derefter bytter de rolle. I hver runde har de netop 10 forsøg. Resultatet er altså en måleserie for hver af eleverne på netop 10 faldlængder, som omregnes til 10 reaktionstider.

I det følgende benytter vi autentiske data fra to elever Janus og Vincent (mens billederne viser to andre elever i færd med at udføre eksperimentet!).

Til selve databehandlingen kan vi benytte et hvilket som helst databehandlingsprogram. Her illustrerer vi udregningerne med DataMeter som er det databehandlingsprogram vi er gået over til at anvende på Haslev Gymnasium, hvor vi tidligere benyttede Excel.

Vi opretter derfor en tabel med to variable, en for eleven og en for faldlængden, så man for hvert af eksperimenterne kan se, hvem der udførte det og hvad resultatet blev. Dernæst udregner vi faldtiderne, hvilket sker ved hjælp af Galileis faldlov. DataMeter har også indbygget alle de gængse naturkonstanter, så DataMeter kender godt standardtyngdeacceleration.

Det giver anledning til den følgende tabel over navne, faldlængder og faldtider. Det er disse data vi vil diskutere i det følgende.

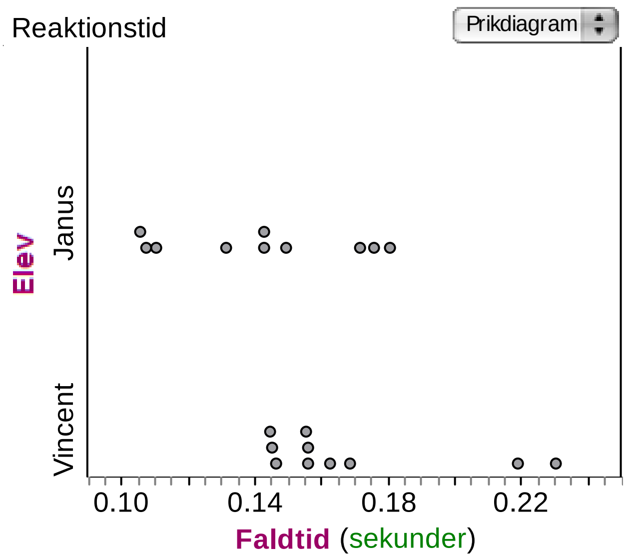
Reaktionstid

	Elev	Faldlængde
enhed		centimeter
=		
1	Janus	14.5 cm
2	Janus	11.0 cm
3	Janus	10.0 cm
4	Janus	5.5 cm
5	Janus	8.5 cm
6	Janus	16.0 cm
7	Janus	6.0 cm
8	Janus	15.2 cm
9	Janus	10.0 cm
10	Janus	5.7 cm
11	Vincent	23.5 cm
12	Vincent	26.0 cm
13	Vincent	12.0 cm
14	Vincent	10.5 cm
15	Vincent	13.0 cm
16	Vincent	10.4 cm
17	Vincent	11.9 cm
18	Vincent	14.0 cm
19	Vincent	10.3 cm
20	Vincent	12.0 cm

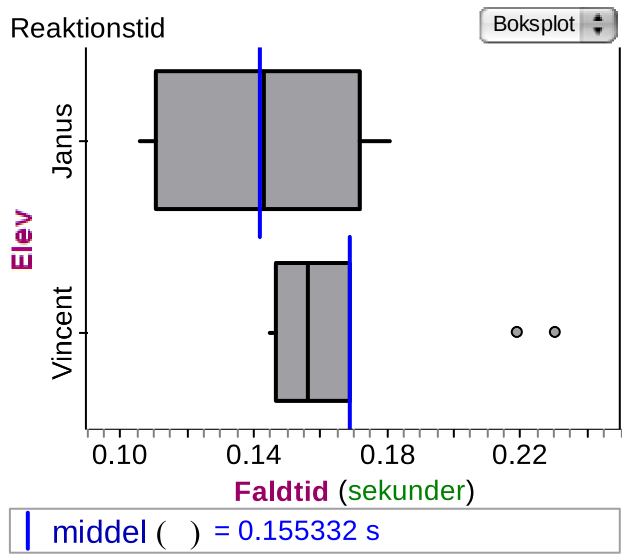
Reaktionstid			
	Elev	**Faldlængde**	**Faldtid**
enhed		centimeter	sekunder
=			$\sqrt{\frac{2\,\text{Faldlængde}}{g}}$
1	Janus	14.5 cm	0.171964 s
2	Janus	11.0 cm	0.149779 s
3	Janus	10.0 cm	0.142809 s
4	Janus	5.5 cm	0.10591 s
5	Janus	8.5 cm	0.131663 s
6	Janus	16.0 cm	0.18064 s
7	Janus	6.0 cm	0.110619 s
8	Janus	15.2 cm	0.176066 s
9	Janus	10.0 cm	0.142809 s
10	Janus	5.7 cm	0.107818 s
11	Vincent	23.5 cm	0.218922 s
12	Vincent	26.0 cm	0.230272 s
13	Vincent	12.0 cm	0.156439 s
14	Vincent	10.5 cm	0.146335 s
15	Vincent	13.0 cm	0.162827 s
16	Vincent	10.4 cm	0.145637 s
17	Vincent	11.9 cm	0.155786 s
18	Vincent	14.0 cm	0.168974 s
19	Vincent	10.3 cm	0.144935 s
20	Vincent	12.0 cm	0.156439 s

For at få en visuel fornemmelse for dataene trækker vi en graf ind i dokumentet og trækker dernæst den *numeriske variabel* **Reaktionstid** ind på førsteaksen og den *kategoriserede variabel* **Elev** ind på andenaksen.

Allerede med prikdiagrammet får vi da en klar fornemmelse af forskellen på de to drenges reaktionsmønster: De har begge en stor variation i reaktionstiderne, men Vincent har mere samling på sine data med kun to reaktionstider, der halter lidt bagefter, mens Janus' reaktionstider falder i tre adskilte klumper. Samtidigt ser det ud som om Janus har en tendens til at være lidt hurtigere på aftrækkeren end Vincent.



Skifter vi til boksplot og afsætter middelværdien (**Plot værdi**) fås et tilsvarende mønster:



Janus' reaktionstider er rimeligt symmetrisk fordelt med en forholdsvis stor kvartilbredde, mens Vincents data er højreskæve med to perifere værdier samt en betragtelig forskel på medianen og middelværdien. Det kunne også godt se ud som om Janus er hurtigere end Vincent: Faktisk ligger hele den nederste halvdel af måleserien for Janus under måleserien for Vincent.

De ovenstående forhold er typiske for mange måleserier: De er ikke specielt stabile og rummer tit nogle kraftige skæverter, også i form af ekstremt små reaktionstider i de tilfælde, hvor det lykkes eleven at gætte sig til hvornår linealen slippes. Vi anbefaler derfor eleverne at de bruger den robuste median som et mål for den *typiske reaktionstid!*

Trækker vi en beregningsboks ind i dokumentet kan vi nemt finde medianerne ved at trække variablene **Faldtid** og **Elev** ind i oversigtstabellen og efterfølgende rette i formlen, så den udregner medianen. I dette tilfælde opnår de to elever altså følgende typiske reaktionstider:

Reaktionstid

		Faldtid
Elev	Janus	0.1428087 s
	Vincent	0.15643909 s
	Søjle total	0.15278246 s

R1 = median ( )

På basis af denne foreløbige numeriske og grafiske analyse af dataene (en såkaldt **Explorative Data Analysis**) synes konklusionen altså at være:

Janus er hurtigere end Vincent.

Men det er selvfølgelig en konklusion Vincent ikke kan være helt tilfreds med, så vi vil give ham lov til at udfordre den i en statistisk test!

Er Janus så virkeligt hurtigere end Vincent?

Vi har set at Janus har en median, der ligger under Vincents median, og at mange af Janus reaktionstider ligger under Vincents reaktionstider. På den anden side er der kun 10 observationer fra hver, så hvor overbevisende er Janus sejr egentlig? Kunne den ikke lige så godt tilskrives tilfældigheder? Det vil vi nu udføre en statistisk test på og for bedre at kunne leve sig ind i argumentationen vil vi benytte en meget udbredt metafor for en sådan statistisk test i form af en retssag:

Sagen Janus versus Vincent

De får hver lov til at fremlægge deres påstand:

Janus: Jeg har vundet fordi jeg er den bedste!
Jeg vinder med overvældende sandsynlighed også næste gang!

Vincent: Du har alene vundet fordi du er heldig!
Næste gang kan det lige så godt være mig, der vinder!

Teknisk bemærkning: Kernen i enhver statistisk test er en undersøgelse af om en observeret forskel skyldes systematiske variationer i data eller tilfældige variationer i data. De to påstande knyttet til den statistiske test kaldes hypoteser:

- Påstanden om at forskellen skyldes tilfældige variationer i data kaldes nulhypotesen H_0. Den udsiger altså at der ingen systematisk forskel er på de to måleserier, dvs. den systematiske forskellen er nul.
- Påstanden om at forskellen skyldes systematiske variationer i data kaldes den alternative hypotese H_a.

Strengt taget skal disse hypoteser foreligge før målingerne udføres. Man må altså forestille sig at Janus kigger Vincent dybt i øjnene og siger: "Jeg er bare hurtigere end dig" og at Vincent efterfølgende udfordrer ham på en reaktionstest for at få sat påstanden fra Janus på prøve. Det er så resultaterne fra denne test, der skal afgøre om Janus kan opretholde sin påstand.

Spørgsmålet er nu hvordan vi kan finde en rimelig afgørelse på tvisten? Vi må altså blive enige om nogle spilleregler. Der findes nu forskellige strategier for hvordan man udfører detaljerne i testen, men det er de samme overordnede rammer man anvender i de forskellige typer test. I alle tilfælde går man ud fra Vincents påstand om at der i virkeligheden ikke er nogen forskel og at de observerede forskelle alene skyldes tilfældigheder (den såkaldte nulhypotese) og antager altså midlertidigt at han har ret:

Retten bygger sin dom på en analyse af den følgende **midlertidige antagelse**: Det er alene tilfældigheder, der har afgjort konkurrencen mellem de to.

Både Janus og Vincent forpligter sig naturligvis på dette udgangspunkt.

Dernæst vælges en **teststørrelse**: Hvorfor mener Janus han er bedre end Vincent. Det kan han have mange gode grunde til, men han er nødt til at vælge en af dem for at gennemføre testen. Janus kan så fx sige: Jeg er bedre end Vincent, fordi jeg har en bedre median. Vi vedtager altså at testen skal gå på medianforskellen. Igen må begge siderne altså forpligte sig på at lade retten basere sin afgørelse på en analyse af denne teststørrelse. Hvis det alene var tilfældet som rådede, ville man forvente en medianforskel på 0 s : Det ville være lige så sandsynligt at Janus vandt som at Vincent vandt. I den faktiske konkurrence finder man nu en medianforskel på -0.0136 s, dvs. så meget hurtigere er Janus i forhold til Vincent målt på medianen:

Reaktionstid

	-0.013630393 s

R1 = median (Faldtid; Elev = "Janus") − median (Faldtid; Elev = "Vincent")

Janus fastholder nu, at denne forskel virker meget stor, og derfor svær at opnå ved alene at lade tilfældet råde, mens Vincent fastholder, at den virker meget lille, og derfor er nem at forklare alene ved tilfældighedernes spil.

I retten **udregner man nu sandsynligheden** for at konkurrencen kunne ende så skævt, som den faktisk gjorde:

- Hvis denne sandsynlighed er meget stor, står Vincent stærkt: Hvis man alene lader tilfældet råde vil det være nemt at få et udfald som det faktisk observerede.
- Hvis denne sandsynlighed derimod er meget lille, står Vincent dårligt: Der skal da nærmest et mirakel til for at få et så ekstremet udfald blot ved at lade tilfældet råde.

Problemet er selvfølgelig, hvor grænserne ligger for en klar afgørelse. Også disse grænser bør aftales på forhånd! Der er da tradition for at man fastlægger et **kritisk niveau**. Det kan fx være på 1%. Hvis sandsynligheden er under 1% vil man altså højst kunne opnå et så skævt resultat i 1 ud af 100 konkurrencer, hvis det alene var tilfældet. Det svækker Vincent betydeligt, for det gør det højst usandsynligt at de skulle få et så skævt resultat, når de kun prøvede kræfter den ene gang. Igen skal de to kombattanter forpligte sig på det kritiske niveau og respektere rettens afgørelse:

Aftalen: Hvis sandsynligheden for, at det kan ende så skævt – som det rent faktisk gjorde – når vi lader tilfældighederne råde, er under 1% vinder Janus retssagen. I modsat fald vinder Vincent retssagen.

Tilbage er så blot spørgsmålet om hvordan vi finder sandsynligheden! Det kan gøres på to væsentligt forskellige metoder:

- Ved en rent teoretisk beregning af teststørrelsens fordeling.
- Ved en rent eksperimentel undersøgelse af teststørrelsens fordeling.

Bemærkning: Det er den teoretiske analyse af testsandsynligheden, der er meget krævende. Udover et stort matematisk apparat afhænger den meget af en lang række detaljer såsom kendskab til den teoretiske fordeling af reaktionstiderne hos de to kombattanter og det specifikke valg af teststørrelse. I praksis afgør man derfor sådanne tvistigheder ved hjælp af en række **kanoniske tests**, der på forhånd er undersøgt teoretisk i alle detaljer, hvorfor man kan nøjes med at checke om forudsætningen for teorien er opfyldt og derefter slå sandsynligheden op i en tabel eller udregne den via et program. I vores konkurrence, ville man således traditionelt anvende et t-test. Det forudsætter at

1. reaktionstiderne er normalfordelte (herunder symmetriske), hvad vi allerede har set de ikke er!
2. spredningerne er de samme, hvad vi også har set de ikke er!
3. vi vælger forskellen i middelværdi som teststørrelse, hvad vi netop ikke har gjort!

Hvis vi skulle basere rettens afgørelse på en teoretisk funderet kanonisk test ville vi altså være ilde stedt!

Vi vælger derfor den anden mulighed og gennemfører rettens analyse ud fra en eksperimentel undersøgelse af teststørrelsens fordeling. Vi kommer da til at træffe endnu et valg: Hvor præcist vil vi kende fordelingen? Hvis den kritiske sandsynlighed er 1% skal vi i praksis kende fordelingen med en nøjagtighed på nogle få promille. Det må vi altså tage højde for i det følgende!

Tilbage står så blot at **simulere** udfaldene af eksperimentet på tilfældig vis. Det kan gøres på forskellige måder, men her benytter vi følgende synsvinkel:

Hvis Vincent har ret er det i virkeligheden rent tilfældigt hvordan de fremkomne reaktionstider er fordelt på de to kombattanter. Vi kan derfor simulere de tilfældige udfald af eksperimentet ved at **rører rundt i variablene** i datasættet, dvs. vi gennemfører en tilfældig permutation af navnene og bryder dermed enhver sammenhæng mellem personer og reaktionstider. I så fald vil vi få et datasæt, hvor forskellen mellem de to personer med garanti kan tilskrives tilfældigheder.

Spørgsmålet er så i hvor høj grad det omrørte datasæt ligner det originale datasæt. Hvis de ligner hinanden meget taler det for også at opfatte forskellene mellem de to personer i det originale datasæt som værende fremkommet ved tilfældigheder. Hvis de derimod overhovedet ikke ligner hinanden taler det for at opfatte forskellen mellem de to personer i det originale datasæt som en reel systematisk forskel, dvs. at Janus reelt synes at være hurtigere end Vincent.

Retten er sat!

Omrøring i hånden:
Inden vi gennemfører omrøringen med DataMeter vil vi lige påpege at det er nemt at udføre omrøringen i hånden. Det kan derfor være en god ide at gennemføre en omrøring med klassen i hånden så de kan derved bedre kan fange ideen.

Hver elev får da to sæt kort af ti kort med hver sin farve, fx blå og gul, og tildeler Janus 10 blå kort og Vincent 10 gule kort. Reaktionstiderne for Janus skrives op på de blå kort, mens reaktionstiderne for Vincent skrives op på de gule kort. Derefter blandes kortene rigtig godt og grundigt og de tyve kort fordeles nu på Janus og Vincent på en systematisk måde, fx de ti første kort til Janus og de ti sidste kort til Vincent. Eleverne har da netop fået omrørt observationerne godt og grundigt og derved sikret at Janus i det omrørte datasæt får en tilfældig blanding af blå og gule kort og tilsvarende med Vincent.

Eleverne lægger nu de ti omrørte kort for Janus op i rækkefølge efter stigende reaktionstid og noterer hans median i det omrørte sæt. Tilsvarende gøres for Vincent. Til sidst finder eleverne forskellen mellem medianerne for reaktionstiderne for Janus og Vincent. Det er her tydeligvis en fordel at vi arbejder med medianen og ikke fx middelværdien, da vi så stort set slipper for at foretage manuelle udregninger. Der kan højst blive tale om gennemsnittet af to tal!

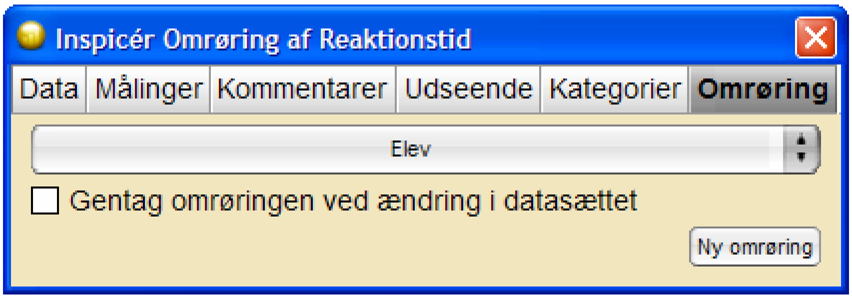
Med en klasse på fx 25 har vi altså nu 25 forskellige målinger af forskellen i medianerne i de scramblede datasæt. Gentages det nogle gange kan man rimeligt hurtigt opbygge en samling på 100 målinger af forskellen mellem medianerne i de omrørte datasæt. Disse medianforskelle kan løbende indtastes i et datasæt, så man hurtigt kan få frembragt et billede af deres fordeling og supplere med de nødvendige udregninger.

Det gør det muligt at se hvor typisk forskellen i det originale datasæt er, idet man kan se om den ligger midt inde i klumpen af de omrørte forskelle, eller tværtimod langt ude i enderne. Derved kan man finde sandsynligheden for at frembringe den observerede forskel ved et rent tilfælde. Hvis denne sandsynlighed er endog meget lav tyder det på at forskellen på Janus og Vincent er reel. Ellers kan vi lige så godt forklare den ved tilfældigheder!

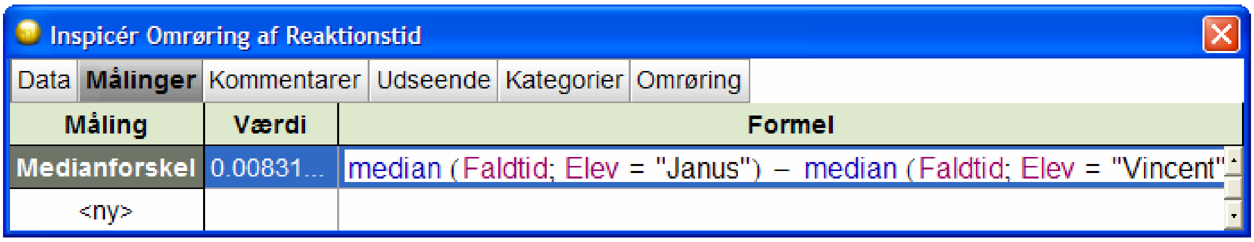
For at udføre en enkelt omrøring højreklikker vi på datasættet for **Reaktionstider** og vælger **Rør rundt i en variabel** fra menuen (eller det tilsvarende menupunkt på **Datasæt**-menuen). Der dukker da et nyt afledet datasæt på skærmen med navnet

Omrøring af Reaktionstid

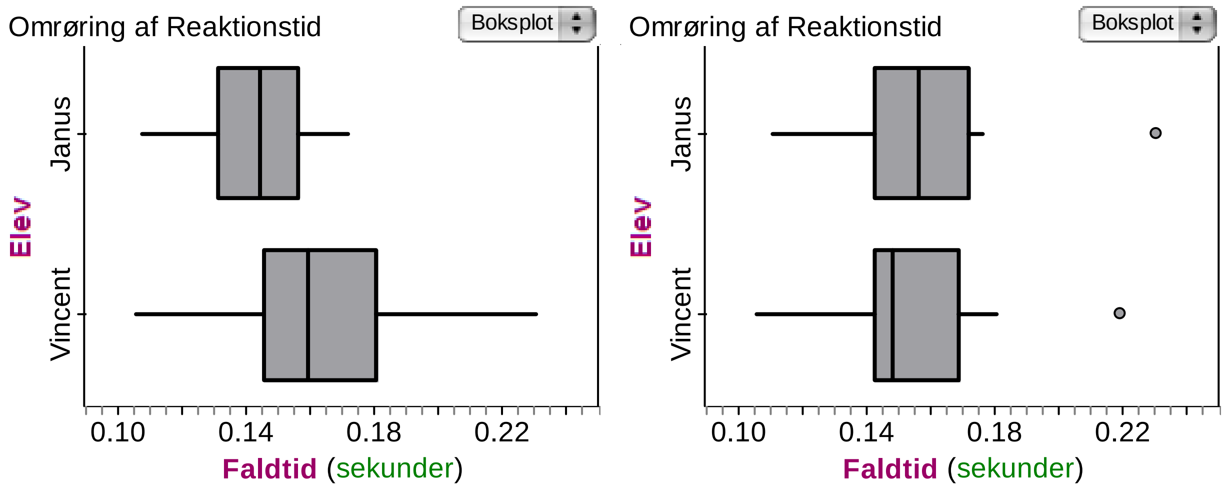
som vi kan inspicere ved at dobbeltklikke på det. Som det sidste faneblad finder vi **Omrøring**. Her kan vi vælge hvilken variabel vi vil omrøre. I dette tilfælde er der tre variable: **Elev**, **Faldlængde** og **Faldtid**. Som udgangspunkt røres der rundt i den første variabel. Det passer fint her, fordi vi så rører rundt i den kategoriserede variabel, dvs. navnet. Men da man kun kan røre rundt i uafhængige variable skal man huske at klippe formlen for **Faldtid** først! Ellers går der kludder i enhederne.



Vi skal nu have oprettet en måling af forskellen i medianer i det omrørte datasæt (hvis vi da ikke på forhånd har indført målingen i det oprindelige datasæt og dermed trukket den med over i det omrørte datasæt). Det sker ved at åbne fanebladet for målinger og så indskrive formlen (eller kopiere den og indsætte den, hvis den allerede er indført i en beregningsboks)

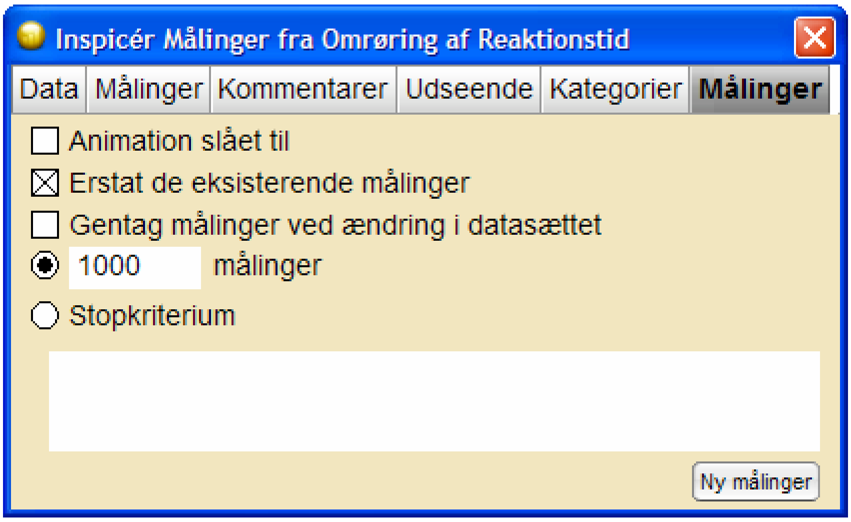


Vi kan nu se hvordan eksperimentet opfører sig, når vi gentager omrøringen. Visuelt kan man fx se på hvordan boksplottene for de to kombattanter opfører sig når vi gentager omrøringen:

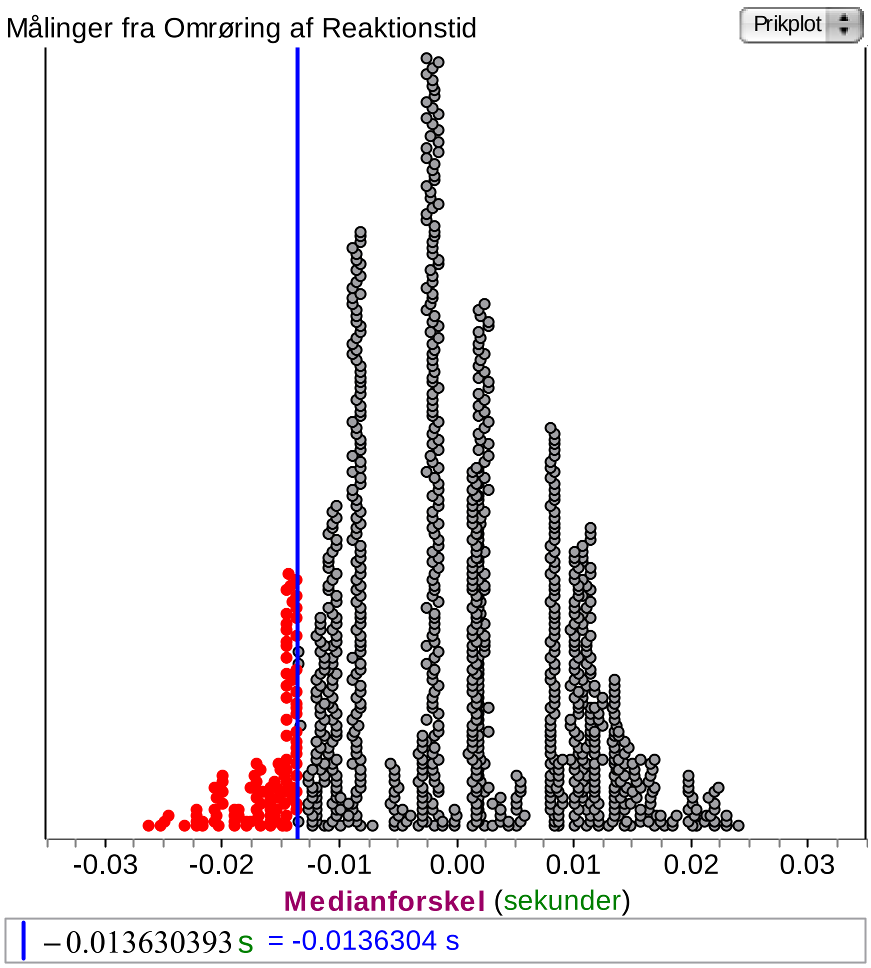


Vi ser da at det sommetider er Janus, der vinder, og sommetider er det Vincent, der vinder. Og i snit vinder de selvfølgelig ca. lige mange gange hver i den omrørte version, fordi det nu netop kun er tilfældighedernes spil, der råder.

Men det interessante er så med hvor meget de vinder, dvs. hvor stor forskellen bliver mellem medianerne. Vi foretager derfor en **gentagen måling** på det omrørte datasæt ved at højreklikke og vælge menupunktet **Udfør gentagne målinger**. Der oprettes da et nyt afledet datasæt med de gentagne målinger og på det sidste faneblad i den tilhørende datainspektør kan vi sætte betingelserne for den gentagne måling:



Her har vi slået animationen fra (som er tidsrøvende idet alle andre grafer osv. hele tiden skal opdateres løbende mens målingerne gentages). Tilsvarende har vi slået **Erstat de eksisterende målinger** til, så vi starter forfra på målingerne. Derved bliver det nemmere at holde styr på det samlede antal målinger. Endelig har vi sat antallet af målinger til 1000, hvilket er mange, men det giver en rimelig høj sikkerhed for at præcist billede af fordelingen:



Som det ses af prikdiagrammet er fordelingen for medianforskellene grynet. Det er karakteristisk for fordelinger, der involverer medianer: Der er kun et forholdsvis lille antal mulige medianer, og dermed også kun et overskueligt antal medianforskelle.

Endvidere kan vi også se at den observerede medianforskel på -0.0136…s slet ikke er så usædvanlig endda. Vi kan nu nemt finde ud af hvilken sandsynlighed den observerede medianforskel svarer til, dvs. hvor mange observationer der går forud for den observerede forskel. Det kan fx ske ved at tælle kugler i det tilhørende Prikdiagram eller ved hjælp af en formel i en **Beregningsboks**.

Målinger fra Omrøring af Reaktionstid

	103

R1 = tæl (Medianforskel ≤ -0.013630393 s)

Vi finder altså at 103 af kuglerne ligger mindst lige så langt ude, dvs. i ca. 1/10 af tilfældene vil vi kunne frembringe en forskel, der er mindst lige så stor. Det tyder på at vi forholdsvis nemt kan forklare den observerede forskel som et resultat af tilfældigheder, så mere overbevisende er Janus sejr altså heller ikke! Vi konkluderer derfor at det blev Vincent, der har vundet retssagen: Janus har ikke kunnet overbevise retten om at han reelt er bedre end Vincent.

Til slut vil vi lige kort se på hvad der sker, hvis vi i stedet for at sammenligne medianer benytter den indbyggede standardtest til at sammenligne middelværdier. Vi trækker derfor **Test-værktøjet** ned og vælger testet for sammenligning af to middelværdier:

Stikprøveresultater — Test af to middelværdier

Første variabel (numerisk): Ikke tildelt
Anden variabel (numerisk eller kategoriseret): Ikke tildelt

Stikprøvens størrelse for **Første variabel** : **20**
Stikprøvens størrelse for **Anden variabel** : **20**
Stikprøvens middelværdi for **Første variabel** : **100**
Stikprøvens middelværdi for **Anden variabel** : **110**
Standardafvigelsen for **Første variabel** : **10**
Standardafvigelsen for **Anden variabel** : **15**
Standardfejlen for middelværdien af **Første variabel** : **2.23607**
Standardfejlen for middelværdien af **Anden variabel** : **3.3541**
Alternativ hypotese: Populationsmiddelværdien af **Første variabel er forskellig fra** den fra **Anden variabel**

Teststørrelsen, Student's t, på grundlag af **ukombinerede varianser** , er **-2.481**.
Der er **33.1031** frihedsgrader .

Hvis det var sandt at populationsmiddelværdien af **Første variabel** var den samme som den fra **Anden variabel** (nulhypotesen), og vi gentog stikprøven mange gange, ville sandsynligheden for at få en værdi for Student's t der var **med en numerisk værdi, der er mindst lige så stor** være **0.018**.

Vi får da en skabelon for testen, hvor vi skal indsætte forskellige oplysninger markeret med blå skrift. Hvis vi som her har de rå data til rådighed kan vi simpelthen trække de to relevante variable ind i skemaet. Ellers må vi selv indskrive nøgletallene: stikprøveantal, stikprøvemiddelværdier samt standardafvigelser. Disse nøgletal kunne fx være oplyst i en opgavetekst.

Men her trækker vi de to variable ind. Første variabel er numerisk og giver direkte anledning til middelværdierne. Det er altså variablen **Faldtid**. Anden variabel er kategorisk med to forskellige værdier, som splitter den første variabel i to kategorier, her **Elev**.

Test fra Reaktionstid — Test af to middelværdier

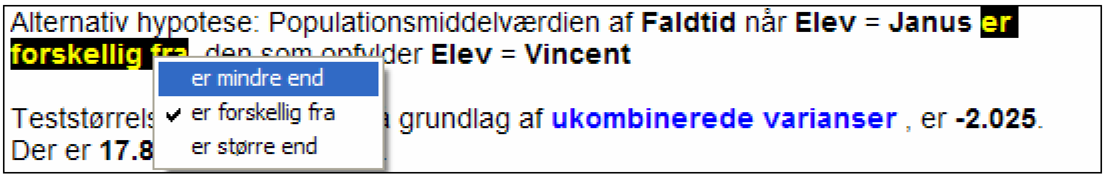
Første variabel (numerisk): Faldtid
Anden variabel (numerisk eller kategoriseret): Elev

Stikprøveantal fra **Elev** = **Janus**: **10**
Stikprøveantal fra **Elev** = **Vincent**: **10**
Stikprøvemiddelværdi fra **Faldtid** når **Elev** = **Janus**: **0.142008 s**
Stikprøvemiddelværdi fra **Faldtid** når **Elev** = **Vincent**: **0.168657 s**
Standardafvigelse af **Faldtid** når **Elev** = **Janus**: **0.0282367 s**
Standardafvigelse af **Faldtid** når **Elev** = **Vincent**: **0.0305726 s**
Standardfejl af **Faldtid** når **Elev** = **Janus**: **0.00892922 s**
Standardfejl af **Faldtid** når **Elev** = **Vincent**: **0.00966792 s**
Alternativ hypotese: Populationsmiddelværdien for **Faldtid** når **Elev** = **Janus** **er forskellig fra** den som opfylder **Elev** = **Vincent**

Teststørrelsen, Student's t, på grundlag af **ukombinerede varianser** , er **-2.025**. Der er **17.8875** frihedsgrader.

Hvis det var sandt at populationsmiddelværdien af **Faldtid** når **Elev** = **Janus** var den samme som den fra **Faldtid** når **Elev** = **Vincent** (nulhypotesen), og vi gentog stikprøven mange gange, ville sandsynligheden for at få en værdi for Student's t **med en numerisk værdi, der er mindst lige så stor** være **0.058**.

Det giver anledning til den viste udfyldning af skemaet. Men her har vi stadigvæk mulighed for at justere den alternative hypotese (mens nulhypotesen ligger fast: De to middelværdier er ens og enhver observeret forskel skyldes tilfældige variationer). I vores tilfælde var den alternative hypotese at Janus er hurtigere end Vincent og dermed har en mindre middelværdi:



Teknisk bemærkning: Endelig er der en lidt kryptisk bemærkning om at teststørrelsen er beregnet på grundlag af ukombinerede varianser. Forudsætningen for at udføre et kanonisk test er at de to måleserier har samme varians, og at denne fælles værdi derfor kan beregnes ved at samle de to måleserier til en lang måleserie. Men man kunne også give et skøn over den fælles varians ved at kombinere de to stikprøvevarianser.

Det endelige resultat fra testværktøjet giver en testsandsynlighed på 2.9%, så heller ikke her er Janus alt for overbevisende! Vi ligger dog og roder nede i nærheden af 1%, så denne gang ville vi konkludere at testet var niveaufølsomt (og afgørelsen derfor ikke helt klar til fordel for en af hypoteserne).

Test fra Reaktionstid — Test af to middelværdier

Første variabel (numerisk): Faldtid

Anden variabel (numerisk eller kategoriseret): Elev

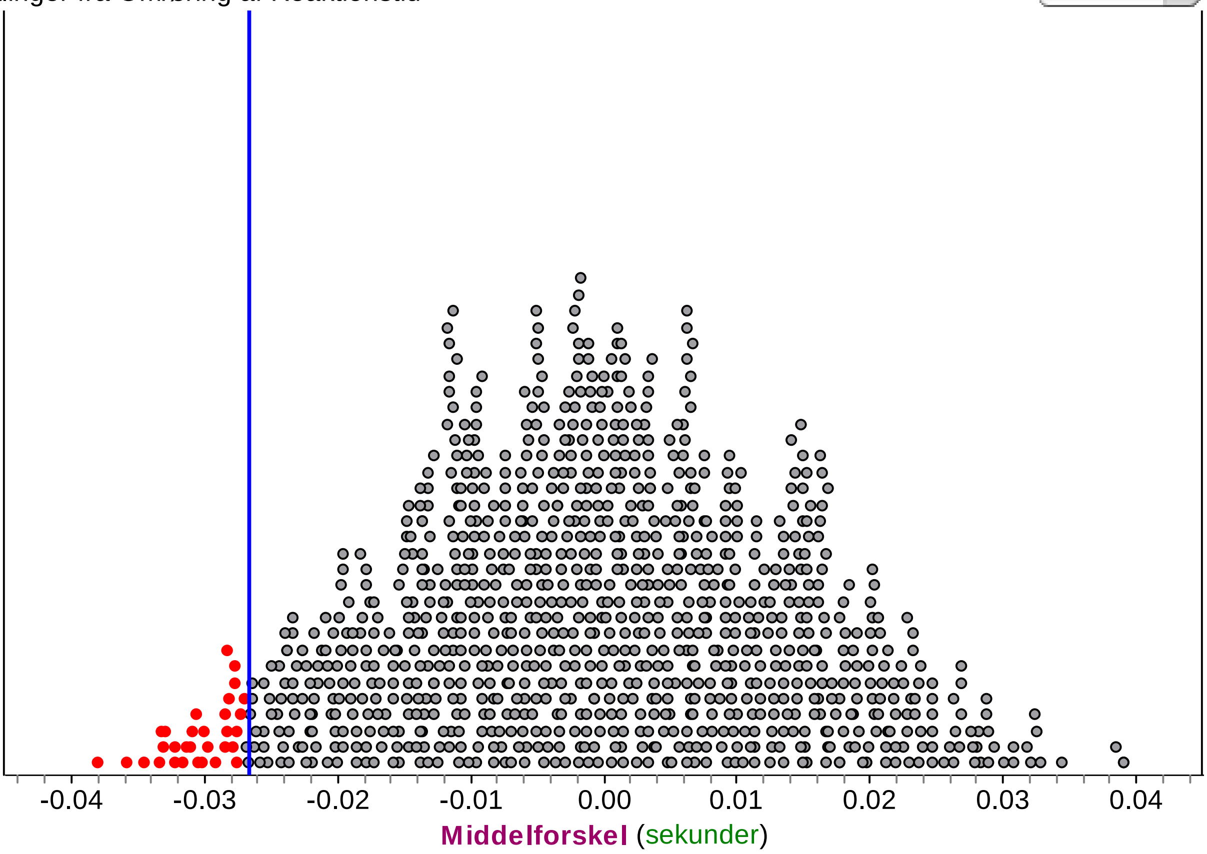
Stikprøveantal fra **Elev** = **Janus**: **10**
Stikprøveantal fra **Elev** = **Vincent**: **10**
Stikprøvemiddelværdi fra **Faldtid** når **Elev** = **Janus**: **0.142008 s**
Stikprøvemiddelværdi fra **Faldtid** når **Elev** = **Vincent**: **0.168657 s**
Standardafvigelse af **Faldtid** når **Elev** = **Janus**: **0.0282367 s**
Standardafvigelse af **Faldtid** når **Elev** = **Vincent**: **0.0305726 s**
Standardfejl af **Faldtid** når **Elev** = **Janus**: **0.00892922 s**
Standardfejl af **Faldtid** når **Elev** = **Vincent**: **0.00966792 s**
Alternativ hypotese: Populationsmiddelværdien for **Faldtid** når **Elev** = **Janus** **er mindre end** den som opfylder **Elev** = **Vincent**

Teststørrelsen, Student's t, på grundlag af **ukombinerede varianser** , er **-2.025**. Der er **17.8875** frihedsgrader.

Hvis det var sandt at populationsmiddelværdien af **Faldtid** når **Elev** = **Janus** var den samme som den fra **Faldtid** når **Elev** = **Vincent** (nulhypotesen), og vi gentog stikprøven mange gange, ville sandsynligheden for at få en værdi for Student's t **mindst lige så lille** være **0.029**.

Igen kunne vi have tilføjet middelforskellen som en måling og lave 1000 gentagne målinger af middelforskellen og derved eksperimentelt fastslå sandsynligheden for at få et resultat, der er mindst lige så skævt som den observerede middelforskel på –0.026648773 s. Ved en prøvekørsel fandt jeg da også 1.9% i rimelig overensstemmelse med *p*-værdien på 2.9% fra den officielle t-test.

Målinger fra Omrøring af Reaktionstid — Prikdiagram



| -0.026648773 s = -0.0266488 s

Reaktionstider for en hel årgang

Nu har vi længe nok vadet grundigt rundt i et opgør mellem to elever. Vi slutter med at se på den typiske fordeling af reaktionstiderne for en større population, fx fra en bestemt årgang. Her ses fx resultaterne fra den lille årgang 2003 på Haslev Gymnasium. Der er også medtaget en enkelt lærer (mig selv!) for at få det til at gå op med et lige antal:

Reaktionstider fra en hel årgang

	Elev	Klasse	Reaktionstid	Køn
enhed			sekunder	
=				
37	37	1bx	0.189 s	p
38	38	1bx	0.166 s	d
39	39	1bx	0.157 s	d
40	40	1a	0.142 s	p
41	41	1a	0.135 s	p
42	42	1y	0.11 s	d
43	43	1z	0.148 s	p
44	44	1z	0.167 s	p
45	45	1bx	0.186 s	p
46	46	1bx	0.143 s	p
47	47	1z	0.124 s	d
48	48	1z	0.125 s	d
49	49	1a	0.194 s	p
50	50	1a	0.146 s	p
51	51	1bx	0.16 s	p
52	52	1bx	0.141 s	p
53	53	1a	0.132 s	d
54	54	Lærer	0.15 s	d

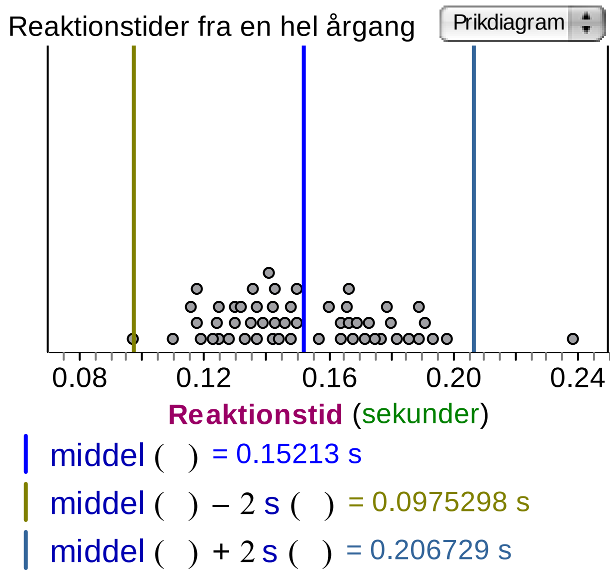
For at danne os et overblik over datasættet tegner vi forskellige grafer over fordelingen, der på mange måder viser det samme fra lidt forskellige synsvinkler:

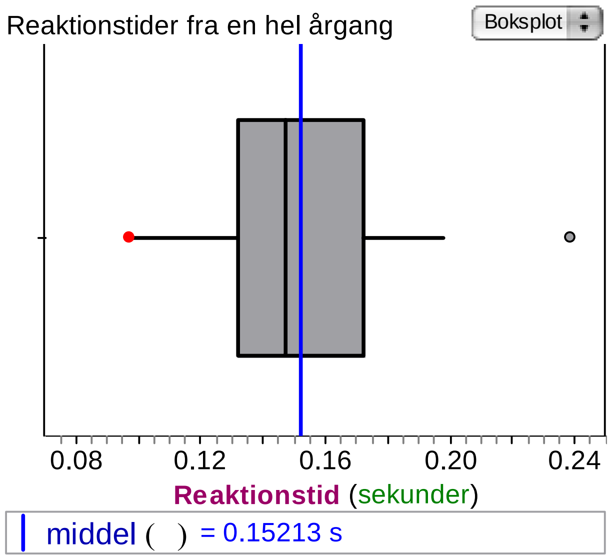
Et prikdiagram, et histogram, et boksplot og et normalfordelingsplot.

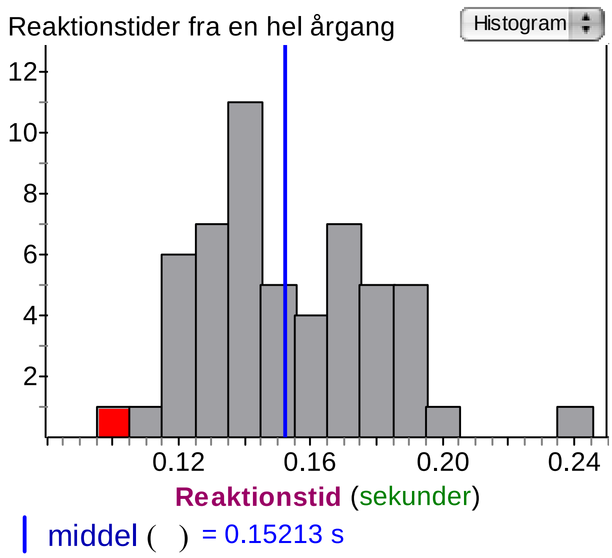
På alle graferne har vi tilføjet middelværdien og på prikdiagrammet og normalfordelingsplottene har vi ydermere tilføjet grænserne for normalområdet, dvs.

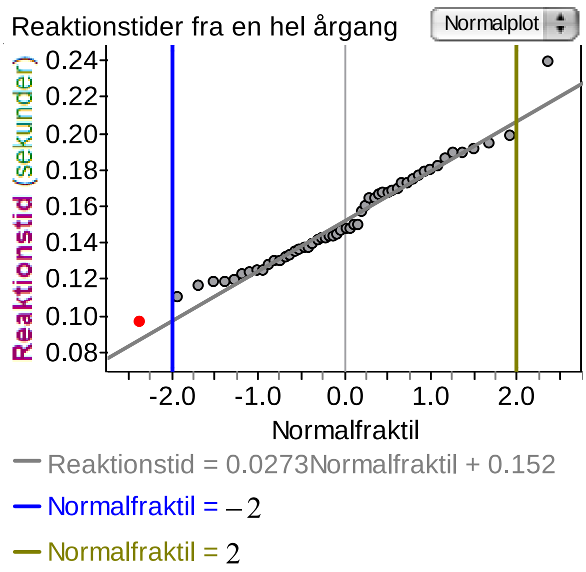
middelværdien ± 2 standardafvigelser.

Det fremgår af graferne at fordelingen er trukket lidt højreskæv af en enkelt perifer værdi, samt at middelværdien er 0.152 s og at alle observationerne bortset fra den hurtigste og den langsomste ligger i normalområdet. Der er altså kun to observationer der skiller sig ud. Fordelingen kan også med rimelighed beskrives ved en normalfordeling med middelværdi 0.152 s og spredning 0.0273 s.



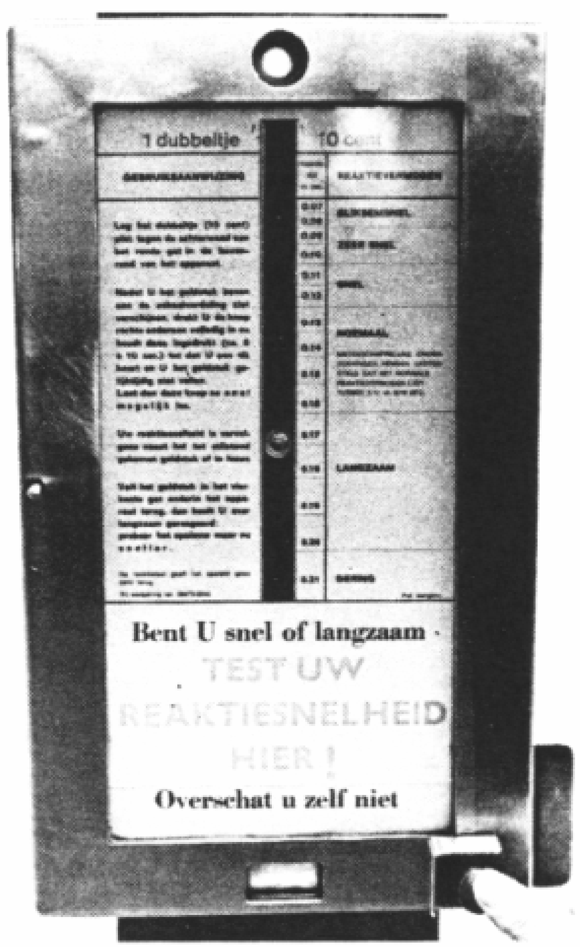






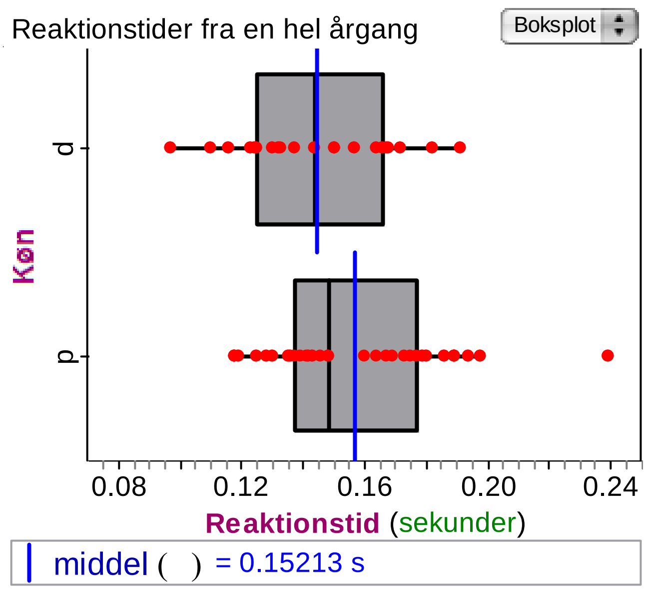
Er det så hvad man ville forvente for en undersøgelse af reaktionstider? For at besvare dette spørgsmål må vi kende lidt til mere omfattende undersøgelser af reaktionstider. Sådanne undersøgelser ligger bl.a. til grund for nogle simple maskiner, der tidligere blev brugt til at måle reaktionstiden. Her er vist et billede af en hollandsk maskine, men tilsvarende maskiner har også tidligere været stillet op i Danmark. Maskinen fungerer ved at man skubber en mønt ind, som på et tilfældigt tidspunkt efterfølgende falder frit. Ved at trykke på en tast kan man stoppe faldet og aflæse på skalaen, hvor lang tid man har brugt og hvor hurtig man har været.

Af det tilhørende skema fremgår at en normal reaktionstid går fra 0.13 s til 0.17 s med en midterværdi på 0.15 s. Det passer fint med en observeret middelværdi på 0.152 s. Vores observerede normalområde, der går fra 0.10 s til 0.20 s passer også pænt med de afgrænsende områder SNEL og ZER SNEL (der rækker ned til 0.09 s) på den ene side og LANGZAAM (der rækker helt op til 0.21 s) på den anden side. Vi har ingen elever, der er BLIKSEMSNEL og kun en enkelt der er GERING.

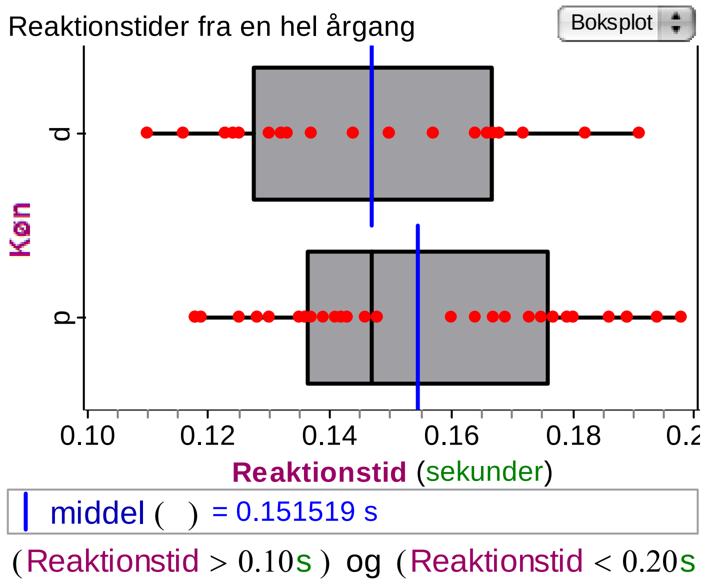


Reaktie-tijd in sec.	REAKTIE-VERMOGEN
0.07	BLIKSEMSNEL
0.08	
0.09	ZEER SNEL
0.10	
0.11	SNEL
0.12	
0.13	NORMAAL wetenschappelijke onderzoekingen hebben vastgesteld dat het normale reaktievermogen ligt tussen 0,12 en 0,18 sec.
0.14	
0.15	
0.16	
0.17	LANGZAAM
0.18	
0.19	
0.20	
0.21	GERING

Når man har en hel årgang til rådighed kan man også sammenligne fx drenge og piger. Vi trækker derfor variablen **Køn** ind på andenaksen i et boksplot:



Vi ser da at drengenes fordeling er pænt symmetrisk, mens pigernes trækkes skævt af den perifere langsomme pige. Vi ser også at drengene tilsyneladende i middel er lidt hurtigere end pigerne ligesom pigernes fordeling falder i to tydelige halvdele. Men fjerner vi den hurtigste dreng og den langsomste pige fx ved at sætte et passende filter på er forskellen knap så markant (om end pigernes fordeling stadigvæk er trukket lidt skæv til højre):



Som i eksemplet med Janus og Vincent kan vi teste om forskellen er signifikant. Her nøjes vi med at se på udfaldet af en kanonisk t-test med tilhørende grafisk fremstilling af testfordelingen:

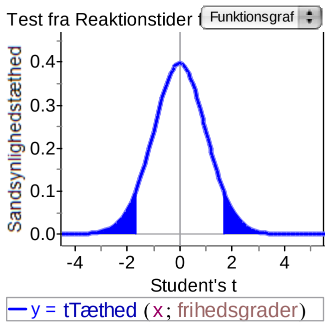
Test fra Reaktionstider fra en hel årgang — Test af to middelværdier

Første variabel (numerisk): Reaktionstid
Anden variabel (numerisk eller kategoriseret): Køn

Stikprøveantal fra **Køn** = **d**: **21**
Stikprøveantal fra **Køn** = **p**: **33**
Stikprøvemiddelværdi fra **Reaktionstid** når **Køn** = **d**: **0.144667 s**
Stikprøvemiddelværdi fra **Reaktionstid** når **Køn** = **p**: **0.156879 s**
Standardafvigelse af **Reaktionstid** når **Køn** = **d**: **0.0250187 s**
Standardafvigelse af **Reaktionstid** når **Køn** = **p**: **0.0279886 s**
Standardfejl af **Reaktionstid** når **Køn** = **d**: **0.00545952 s**
Standardfejl af **Reaktionstid** når **Køn** = **p**: **0.00487218 s**
Alternativ hypotese: Populationsmiddelværdien for **Reaktionstid** når **Køn** = **d** **er forskellig fra** den som opfylder **Køn** = **p**

Teststørrelsen, Student's t, på grundlag af **ukombinerede varianser**, er **-1.669**. Der er **46.2196** frihedsgrader.

Hvis det var sandt at populationsmiddelværdien af **Reaktionstid** når **Køn** = **d** var den samme som den fra **Reaktionstid** når **Køn** = **p** (nulhypotesen), og vi gentog stikprøven mange gange, ville sandsynligheden for at få en værdi for Student's t **med en numerisk værdi, der er mindst lige så stor** være **0.1**.



Der synes altså ikke at være signifikant forskel på drengenes og pigernes reaktionstider (målt på middelværdierne af de to stikprøver).