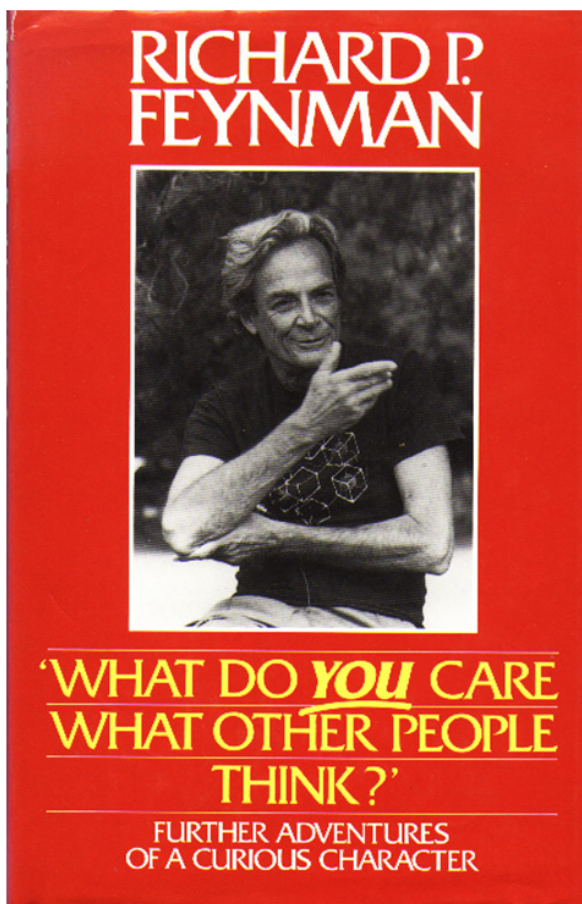


Challenger ulykken¹

Challenger ulykken er en af den nyere tids mest spektakulære teknologiske ulykker: Den 28. januar 1986 forulykkede rumfærgen Challenger kort efter starten på opsendelsen. Tragedien fik en enorm medieomtale bl.a. fordi det var første gang en skolelærer, Christa McAuliffe, skulle med i rumfærgen, det såkaldte 'Teacher in Space'-programme, hvorfor mange amerikanske skoler fulgte opsendelsen intenst.

Ulykken blev undersøgt af en national kommission, der bl.a. udpegede fysikeren og nobelpristageren Richard Feynman som et af sine medlemmer. Det skulle blive Feynman, der løste gåden om den uheldssvangre rumfærges forlis, men meget tyder på at han blev kraftigt støttet af folk fra NASA, der godt kendte problemerne, men ønskede løsningen fremlagt af en kendt fysiker udefra, så opklaringen kunne få den rigtige effekt, såvel eksternt som internt i den store organisation. Feynman selv har beskrevet hændelsesforløbet i bogen: "What do you care what other people think?". Herhjemme har rektor Claus Jensen fra Fåborg Gymnasium udgivet bogen: "Challenger – et teknisk uheld" (Samleren 1993).



Der er masser af stof til tværfagligt samarbejde fx mellem samfundsfag, fysik og matematik. Her vil jeg dog holde mig til matematikken, dvs. specielt statistikken.

Hovedspørgsmålet er da: **Kunne ulykken være undgået?** Havde NASA rent faktisk tilstrækkelig med forhåndsviden til at de burde have udsat flyvningen?

¹ Den amerikanske rumfartsorganisation NASA selv har offentliggjort en stor mængde informativt materiale om ulykken på deres hjemmeside inklusive videoklip fra opsendelsen og selve ulykken, se fx <http://history.nasa.gov/sts51l.html>

Feynman var hård i sin kritik af NASA's håndtering af ulykkesvurderingerne. Specielt hæftede han sig ved den markante forskel mellem vurderinger foretaget af ingeniører henholdsvis administrationen. Ingeniørerne opfattede langt fra rumfæreflyvningerne som sikre. Ved en almindelig rumflyvning siger tommelfingerregler at ca. 4 ud af 100 ender galt, men ingeniørerne var godt klar over at bemandede rumflyvninger blev checket grundigere; Alligevel mente de stadigvæk at risikoen for et alvorligt uheld var af størrelsesordenen 1%. Administrationen satte derimod risikoen for et alvorligt uheld til 10^{-5} . Begrundelsen var en kompliceret udregning, som Feynman fandt totalt meningsløs, idet den byggede på et mylder af forudsætninger, der var rent vilkårlige og tilsyneladende kun tjente til at vise at slutrisikoen var af den ønskede størrelse²! Feynmans konklusion var derfor at rumfærgesendelserne nærmest måtte sammenlignes med russisk roulette med en ikke-forsvindende risiko for at ende i en katastrofe, men hvor man fortsatte flyvningerne nærmest fordi det jo endnu ikke var gået galt. Som Richard Feynman bemærkede det:

“When playing Russian roulette the fact that the first shot got off safely is little comfort for the next.”

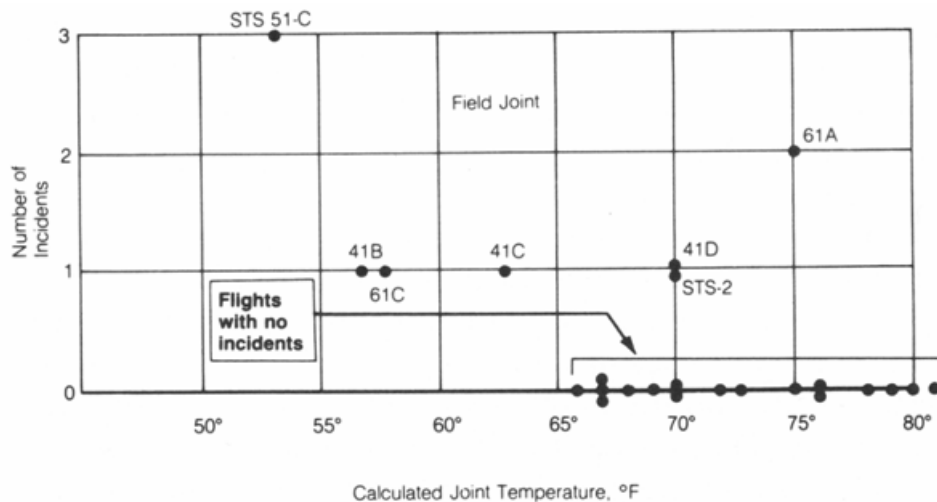
I den konkrete flyvning samlede interessen sig hurtigt om de såkaldte O-ringe. Optagelser fra opsendelsen viste tydelige lække af brændende gasser, og de kunne klart være hovedårsagen til den kraftige eksplosion af hovedtanken. Lækken skulle være sket ved at gas strømmede forbi to O-ringe, fordi disse af en eller anden grund ikke havde sluttet helt tæt. Disse ringe var af gummi og skulle udvide sig hurtigt under opsendelsen: På grund af den kraftige varmeudvikling og de kraftige rystelser udvidede mellemrummet omkring O-ringene sig nemlig under opsendelsen. Tidligere flyvninger og test af sammenslutningerne havde vist at dette kunne føre til erosion af O-ringene. Men jo koldere ringene er jo mere stive er ringene, og dette kunne føre til en problematisk forsinkelse af deres udvidelse. Netop på ulykkesdagen den 28. januar var det betydeligt koldere (lige under frysepunktet) end ved nogen tidligere opsendelse og dette kunne have fået katastrofale konsekvenser.

Inden en offentlig høring lykkedes det Feynman at skille en model af rumfærgen ad og tiltuske sig en O-ring, så han kunne få afprøvet dens manglende fleksibilitet, når ringen blev lagt i et glas med isvand. Derefter blev modellen samlet igen så han på det rette tidspunkt under en tv-transmitteret høring offentligt kunne skille modellen ad og med stor dramatik demonstrere, hvordan ringen først stivnede i isvand og derefter tog nogle sekunder inden den rettede sig ud igen.

Dermed havde man fundet årsagen til ulykken, men det besvarer jo ikke problemet om hvorvidt NASA burde have forudset det. NASA havde godt nok siden 1977 været klar over at O-ringene var et ekstremt svagt led, og de havde derfor skærpet testrutinerne omkring deres samling gennem årene, men ingen havde for alvor testet temperaturafhængigheden. Man samlede derfor nu alle tilgængelige oplysninger om opsendelsestemperaturer og problemer omkring O-ringene fra tidligere flyvninger med rumfærger. I en berømt telefonsamtale mellem Thiokols ingeniører og Nasas ledelse aftenen før den fatale flyvning forsøgte ingeniørerne at stoppe flyvningen netop på grund af den mulige sammenhæng mellem den lave opsendelsestemperatur og så problemerne med O-ringene stabilitet, der kunne gå hen og blive skæbnesvangre. Men det lykkedes ikke for ingeniørerne at påvise en klar sammenhæng og Nasa afviste derfor at udsætte flyvningen.

² Udregningen af den sammensatte sandsynlighed for et alvorligt uheld fulgte standardreglerne for risikovurderinger. Men risikovurderinger er selvfølgelig også underkastet den almene regel: Garbage in – Garbage out. Konklusionen giver kun mening, hvis alle forudsætningerne giver mening.

Episoden er senere blevet kendt som 'The Greatest Missed Opportunity in Statistics'³. Det er denne historiske diskussion af problemerne med O-ringene vi vil prøve at belyse i det følgende. Der er forskellige kilder til de data Thiokols ingeniører havde til rådighed, og tolkningen af NASA's egne rapporteringer omkring tidlige problemer er ikke helt klar, men vi vælger her at følge Feynmans egen forenklede oversigt. Den er blevet standardkilden i de utallige diskussioner man fx finder på Internettet



Hvilke konklusioner kan vi så drage af disse data? Det er nu på tide at vi åbner for et databehandlingsprogram, fx et regneark for at kigge nærmere på dataene. De følgende skærbilleder stammer fra databehandlingsprogrammet **DataMeter** men andre programmer kan også bruges. Dog skal de fleste programmer have lidt assistance med omrøring af variable, idet tilfældige permutationer af data ikke nødvendigvis er inkluderet som standard.

Her er vist et eksempel på en programstump, der kan udføre en tilfældig permutation:

```
randperm(liste)
Func
    n:=dim(liste)                Find længden af listen
    For k = 1 to n
        indeks:=randint(1,n-k+1)  Træk et tilfældigt indeks
        NyListe[k]:=liste[indeks]  Flyt tallet over i ny liste
        For j = indeks to n-k
            liste[j]:=liste[j+1]  Opdater gammel liste
        EndFor j
    EndFor k
    Return NyListe
EndFunc
```

Du behøver ikke forstå koden i detaljer, men kan måske få hjælp fra din lærer til at implementere den i et passende program til din lommeregner, TI-Interactive eller lignende.

³ Se den også i andre henseender interessante hjemmeside

Gallery of Data Visualization: <http://www.math.yorku.ca/SCS/Gallery/>

De tilhørende data ser således ud:

Opsendelser

	Flyvning	Temperatur	Fejl	Dato	Rumfærge	O_ringe
1	STS-1	66	NEJ	81-04-12	Columbia	0
2	STS-2	70	JA	81-11-12	Columbia	1
3	STS-3	69	NEJ	82-03-22	Columbia	0
4	STS-4	80	NEJ	82-06-27	Columbia	0
5	STS-5	68	NEJ	82-11-11	Columbia	0
6	STS-6	67	NEJ	83-04-04	Challenger	0
7	STS-7	72	NEJ	83-06-18	Challenger	0
8	STS-8	73	NEJ	83-08-30	Challenger	0
9	STS-9	70	NEJ	83-11-28	Columbia	0
10	STS 41-B	57	JA	84-02-03	Challenger	1
11	STS 41-C	63	JA	84-04-06	Challenger	1
12	STS 41-D	70	JA	84-08-30	Discovery	1
13	STS 41-G	78	NEJ	84-10-05	Challenger	0
14	STS 51-A	67	NEJ	84-11-08	Discovery	0
15	STS 51-C	53	JA	85-01-24	Discovery	3
16	STS 51-D	67	NEJ	85-04-12	Discovery	0
17	STS 51-B	75	NEJ	85-04-29	Challenger	0
18	STS 51-G	70	NEJ	85-06-17	Discovery	0
19	STS 51-F	81	NEJ	85-07-29	Challenger	0
20	STS 51-I	76	NEJ	85-08-27	Discovery	0
21	STS 51-J	79	NEJ	85-10-03	Atlantis	0
22	STS 61-A	75	JA	85-10-30	Challenger	2
23	STS 61-B	76	NEJ	85-11-26	Atlantis	0
24	STS 61-C	58	JA	86-01-12	Columbia	1

Tabellen rummer data fra de foregående 24 flyvninger med rumfærgerne ordnet kronologisk. Datasættet rummer 6 variable:

Flyvning: Den tekniske betegnelse for den pågældende mission.

Temperatur: Luftens gennemsnitstemperatur omkring rumfærgerne under opsendelsen (målt i Fahrenheit!).

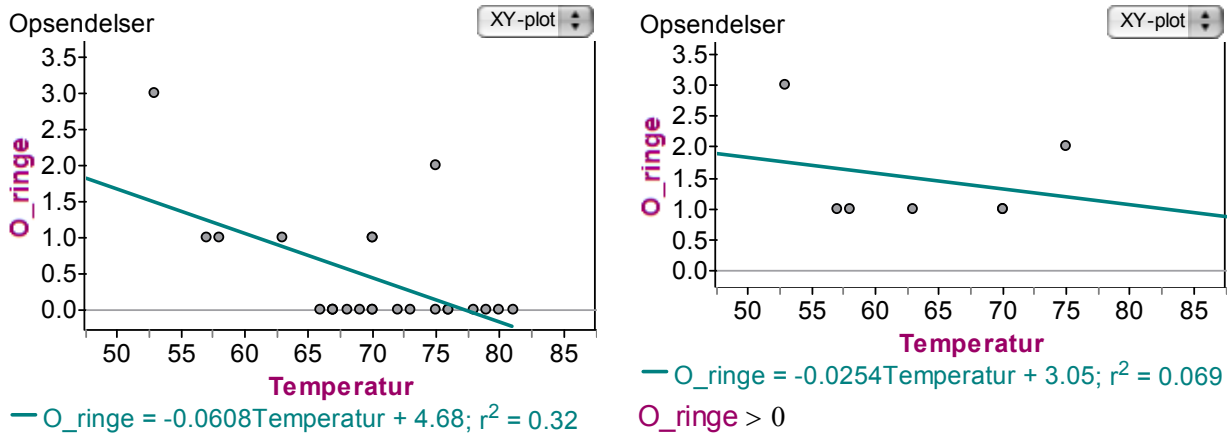
Fejl: Om der efterfølgende var konstateret alvorlige fejl på en eller flere af O_ringene.

Dato: Datoen for opsendelsen i formatet: år-måned-dag.

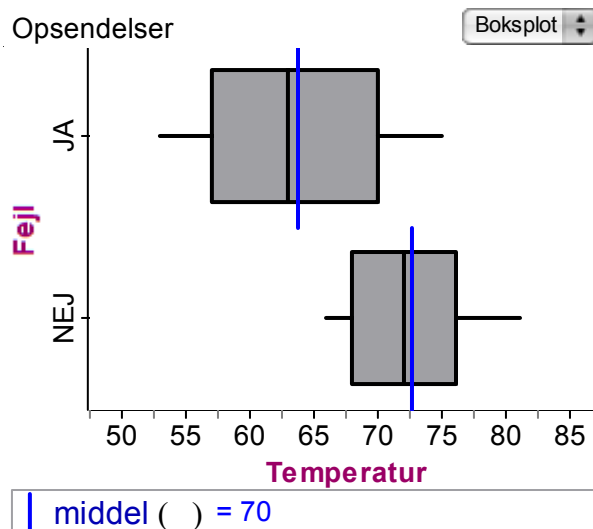
Rumfærge: Hvilken af de fire rumfærger der var tale om.

O_ringe: Antallet af O-ringe, hvor der efterfølgende blev konstateret alvorlige fejl.

Det gør det helle ikke meget mere overbevisende at tilføje en tendenslinje. Selv om den har negativ hældning, og derfor påpeger en mulig sammenhæng mellem mange uheld og lave temperaturer er forklaringsgraden helt ned på 32% (eller endnu værre helt nede på 7% hvis vi udelukker de fejlfrie flyvninger), så det er bestemt tvivlsomt hvor stærk en sådan sammenhæng kan regnes for at være.



Vi skifter derfor strategi og kigger i stedet på boksplottene for de to grupper af opsendelser, dem med fejl og dem uden fejl. Samtidigt med tilføjer vi middeltemperaturerne:



Her synes der at være en langt tydeligere sammenhæng med temperaturen: Der er klart flest fejl, når temperaturen er lav. Men igen er der vist overlap, så hvor stærk er sammenhængen? Kan den motivere en udsættelse?

Trods alt var den Amerikanske præsident Ronald Reagan klar med sin årlige tale til kongressen, hvori rumfærgens Teacher in Space program blev omtalt som en af årets spændende nyheder, ligesom der var fastsat en direkte telefonsamtale mellem besætningen og præsidenten. Der var således et betydeligt politisk pres på NASA om ikke at foretage unødige udsættelser af flyvningen.

Hypotesetestet udført eksperimentelt

Vi kommer ikke udenom at foretage et hypotesetest for at kunne vurdere styrken i den observerede mønster. I den såkaldte **retsags-metafor** for hypotesetest skal vi altså fælde dom i en sag om to påstande:

- a) Ingeniørerne hævder at der er en systematisk sammenhæng, og at man derfor skal være yderst varsom med opsende rumfærgen ved kolde temperaturer på grund af den forøgede risiko (den såkaldt alternative hypotese H_a).
- b) Administrationen hævder at de observerede variationer i antallet af uheld ligeså godt kan tilskrives tilfældigheder, og at der derfor ikke er nogen grund til at udskyde opsættelsen (den såkaldte nulhypotese).

Der er ikke noget fældende bevis, så afgørelsen skal træffes på grundlag af indicier. Vi skal da på forhånd blive enige om to forhold: Hvilket signifikansniveau vil vi lægge til grund for domfældelsen (dvs. hvor stærke skal indicierne være) og hvilken teststørrelse vil vi benytte til at afgøre det?

Først **signifikansniveauet**: Typisk vil man benytte et signifikansniveau på 5%: Hvis sandsynligheden for, at den observerede variation kan forklares med rene tilfældigheder, kommer under 5%, vil vi følge ingeniørerne og dømme systematisk variation på grundlag af de fremlagte indicier. Hvis sandsynligheden for at forklare den observerede variation derimod kommer over 5% vil vi i stedet følge administrationens påstand og dømme tilfældig variation fordi indicierne ikke er stærke nok.

Dernæst **teststørrelsen**: Her er der også et stort spillerum, men igen skal vi blive enige på forhånd. Fx kunne vi kigge på middeltemperaturen i de to grupper eller på medianen for temperaturen i de to grupper og så udregne deres forskel. Her vælger vi middeltemperaturen, for det vil give os mulighed for at sammenligne den eksperimentelle metode med den traditionelle teoretiske metode. Som teststørrelse vælger vi altså forskellen på middeltemperaturerne i de to grupper og vil nu finde sandsynligheden for at denne forskel kan tilskrives rene tilfældigheder!

Simulering i hånden:

Inden vi gennemfører en simulering med databehandlingsprogrammet vil vi lige påpege at det er nemt at udføre en simulering i hånden.

Hver elev får da 24 kort af ti kort med temperaturerne for de 24 flyvninger. Derefter blandes kortene rigtig godt og grundigt og eleven trækker de første syv af de firetyve kort. De første syv kort repræsenterer da flyvningerne med fejl, mens de resterende 17 repræsenterer de fejlfrie flyvninger. På denne måde sikres det at fejlene fordeles tilfældigt på flyvningerne.

Eleverne udregner nu middeltemperaturen for de syv første kort og tilsvarende for de 17 sidste kort og trækker derefter de to middeltemperaturer fra hinanden. Med en klasse på fx 25 elever har vi altså nu 25 forskellige målinger af forskellen i middeltemperaturer. Gentages det nogle gange kan man rimeligt hurtigt opbygge en samling på 100 målinger af forskellen i middeltemperaturer. Disse middelforskelle 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 netop 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 er reel og at ingeniørerne derfor står stærkt i deres krav om at udsætte flyvningen.

Vi udregner derfor den valgte teststørrelse, dvs. forskellen i middeltemperaturer, som en **måling** for vores datasæt. I DataMeter sker det ved at dobbeltklikke på datasættet **Opsendelser** og vælge fanebladet **Målinger** inde i **datainspektøren**:

Inspicér Opsendelser		
Data	Målinger	Kommentarer Udseende Kategorier
Måling	Værdi	Formel
Middelforsk	-8.87395	$\text{middel}(\text{Temperatur}; \text{Fejl} = \text{"JA"}) - \text{middel}(\text{Temperatur}; \text{Fejl} = \text{"NEJ"})$
<ny>		

Derved sikrer vi os at vi kan finde fordelingen af teststørrelsen, når vi nu om lidt laver om på den! I det rent faktisk tilfælde var der altså en forskel på -8.87395°F .

Så skal vi have udregnet sandsynligheden for at man kan få så stor en forskel ved rene tilfældigheder. Vi bryder derfor en eventuel afhængighed mellem temperaturerne og fejlene ved at **omrøre** fejlene, dvs. røre rundt i dem uafhængigt af temperaturerne. Vi permuterer med andre ord fejlværdierne! Det gøres i Datameter ved at højreklikke på datasættet og vælge kommandoen **Rør rundt i en variabel**. Straks dukker der et afledet datasæt op med de omrørte variable:



Omrøring af Opsendelser

Inspicér Omrøring af Opsendelser

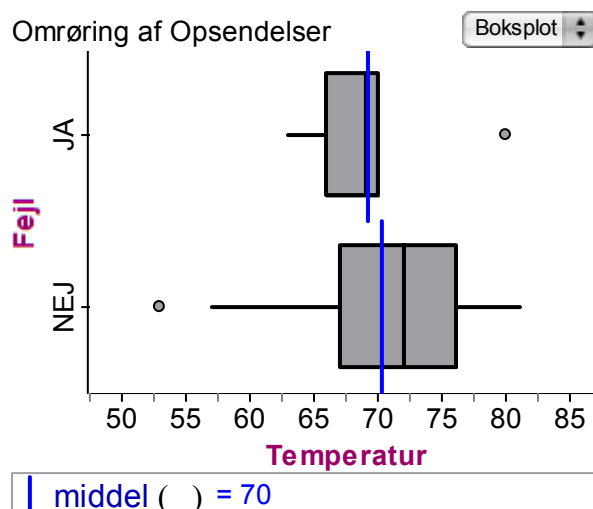
Data Målinger Kommentarer Udseende Kategorier **Omrøring**

Fejl

☐ Gentag omrøringen ved ændring i datasættet

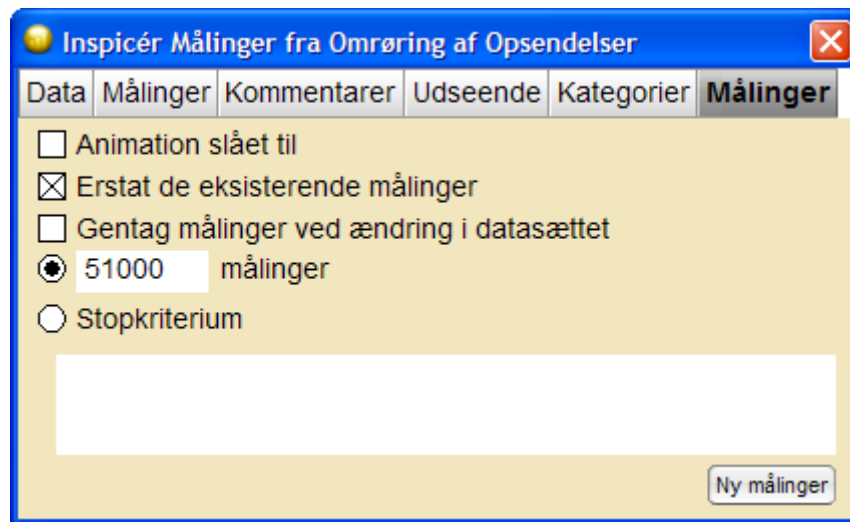
Ny omrøring

I første omgang omrøres den første variabel, dvs. her **Flyvning**. Men det kan vi rette ved at dobbeltklikke på det omrørte datasæt og vælge fanebladet **Omrøring**, hvor vi udskifter variabelen **Flyvning** med **Fejl**: Derefter gentages omrøringen naturligvis! Til illustration fremskaffer vi også en graf over **Fejl** som funktion af **Temperatur** inklusive middelværdierne, der afsættes med **Plot Værdi**:

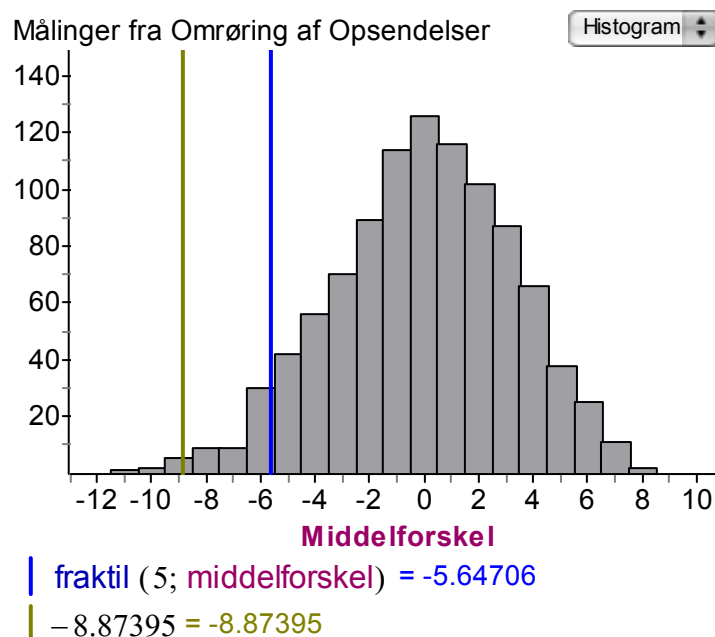


Ikke overraskende ligner de to grupper nu hinanden meget mere og begge middeltemperaturerne ligger tæt ved den fælles middeltemperatur på 70°F .

Vi er nu klar til at udføre målingerne af de omrørte temperaturforskelle, hvor pointen altså er at vi har sikret os at enhver mulig sammenhæng mellem temperatur og fejl er brudt, dvs. at temperaturforskellene nu virkeligt *er* fremkommet ved rene tilfældigheder, fordi fejlen fordeles fuldstændigt tilfældigt på de 24 flyvninger. I DataMeter højreklikker vi derfor på **det omrørte datasæt** og vælger menupunktet **Udfør gentagne målinger**. Under de fem obligatoriske målinger blafrer boksplotterne frem og tilbage, idet somme tider den ene og somme tider den anden lægger sig i spidsen. Ved at dobbeltklikke på det afledte datasæt over målingerne kan vi nu få frembragt 1000 gentagne målinger, idet vi husker at få slukket for animationen, så der ikke spilles tid med hele tiden at få opdateret resultaterne fra den nye måling:



Det tager ikke særlig lang tid at få udført de 1000 omrøringer med tilhørende målinger og dermed få udregnet de 1000 forskelle i middeltemperaturer for de to grupper **under antagelsen** af at de i virkeligheden er uafhængige (den såkaldte nulhypotese)! Dermed kan vi få tegnet et histogram over målingerne sammen med den observerede forskel i middeltemperaturer på -8.87395°F :



Histogrammet viser klart, hvor svært det er at frembringe en så stor forskel alene ved tilfældigheder. Den observerede forskel ligger et godt stykke inde i den kritiske zone afgrænset ved 5%-fraktilen, der slutter ved -5.64706°F .

Vi kan også få udregnet den faktiske (eksperimentelle) sandsynlighed. I DataMeter kan det fx gøres ved at trække datasættet for de gentagne målinger ind i en beregningsboks og derefter udregne formelen

Målinger fra Omrøring af Opsendelser

	5

$R1 = \text{tæl} (\text{Middelforskel} \leq -8.87395)$

Det viser sig da at der kun er 5 ud af 1000 målinger, der ligger lige så langt eller længere ude, så sandsynligheden for at den observerede forskel skyldes rene tilfældigheder er altså ca. ½%.

Dermed har ingeniørerne vundet retssagen (men i den virkelige verden valgte administrationen altså at sidde deres advarsler overhørige!)

Bemærkning: Vi kan også gennemføre testen mere traditionelt. I DataMeter trækker vi da et testværktøj ned i dokumentet og ser hvilke muligheder vi har. Vi vælger **Sammenlign middelværdier** og der dukker nu en skabelon op for testen, som vi skal have udfyldt. Dels skal vi have trukket de to fundamentale variable **Temperatur** (numerisk) og **Fejl** (kategorisk) ind i testskemaet, dels skal vi have rettet i den alternative hypotese: Påstanden fra ingeniørerne var ikke blot at der var en sammenhæng, men at fejlen var hyppigere ved lave temperaturer, så den alternative hypotese skal ændres til at middeltemperaturen for de fejlbehandlede flyvninger **er mindre end** middeltemperaturen for de fejlfri flyvninger:

Test fra Opsendelser Sammenlign middelværdier

Første variabel (numerisk): Temperatur
Anden variabel (numerisk eller kategoriseret): Fejl
Stikprøveantal fra Fejl = JA : 7
Stikprøveantal fra Fejl = NEJ : 17
Stikprøvemiddelværdi fra Temperatur når Fejl = JA : 63.7143
Stikprøvemiddelværdi fra Temperatur når Fejl = NEJ : 72.5882
Standardafvigelse af Temperatur når Fejl = JA : 8.15913
Standardafvigelse af Temperatur når Fejl = NEJ : 5.06284
Standardfejl af Temperatur når Fejl = JA : 3.08386
Standardfejl af Temperatur når Fejl = NEJ : 1.22792
Alternativ hypotese: Populationsmiddelværdien af Temperatur når Fejl = JA er mindre end den som opfylder Fejl = NEJ
Teststørrelsen, Student's t, på grundlag af ukombinerede varianser , er -2.673 . Der er 7.97814 frihedsgrader.
Hvis det var sandt at populationsmiddelværdien af Temperatur når Fejl = JA var den samme som den fra Temperatur når Fejl = NEJ (nulhypotesen), og vi gentog stikprøveudtagningen mange gange, ville sandsynligheden for at få en værdi for Student's t mindst lige så lille være 0.014 .

Vi ser da at sandsynligheden for at forklare den observerede forskel i middeltemperaturerne alene som et resultat af rene tilfældigheder er nede på 1.4%. Det fører altså til samme konklusion som før!