

BAGGRUND OG DATABESKRIVELSE (EKSEMPEL 1-2)

Denne opgave beskæftiger sig med omsætningen af skrabespil på det danske marked. Opgaven tager udgangspunkt i filen DanskeSpil. jmp, hvis indhold er beskrevet på eksempelark 5.

EKSEMPEL 1 (MINI QUICK, SAMLET SALG)

- a). **Tegn** et box plot over fordelingen af den ugentlige omsætning af "Mini Quick" (*Mini Quick* (*i alt*)) for hvert af årene 2014 og 2015 (År).
 - **Estimér** middelværdi og standardafvigelse for den ugentlige omsætning af "Mini Quick" for hvert af årene 2014 og 2015.
 - Gør rede for sammenhængen mellem de beregnede middelværdier og standardafvigelser og de optegnede box plots.
- b). **Beregn** et 95%-konfidensinterval for forskellen mellem den ugentlige omsætning af "Mini Quick" i årene 2014 og 2015 og **giv** en fortolkning af intervallet.
 - **Beregn** et 99%-konfidensinterval for forskellen mellem den ugentlige omsætning af "Mini Quick" i årene 2014 og 2015 og **giv** en fortolkning af intervallet.
 - **Gør rede for** hvorfor det ene af de to beregnede konfidensintervaller er bredere end det andet.
- c). Gør rede for om der er statistisk belæg for en nulhypotese om, at den forventede ugentlige omsætning af "Mini Quick" er den samme i årene 2014 og 2015 (brug $\alpha = 5\%$ som signifikansniveau). Angiv i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.
 - Gør rede for om der er statistisk belæg for en nulhypotese om, at den forventede ugentlige omsætning af "Mini Quick" er den samme i årene 2014 og 2015 (brug $\alpha = 1\%$ som signifikansniveau).
 - Gør rede for sammenhængen med resultaterne i delspg. b).
- d). Gør rede for om der er statistisk belæg for en nulhypotese om, at den forventede ugentlige omsætning af "Mini Quick" er mindst 50.000 kr. større i år 2015 end i år 2014 (brug $\alpha = 5\%$ som signifikansniveau). Angiv i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.



EKSEMPEL 2 (MINI QUICK, SALG I BUTIKSSEGMENTER)

- a). **Tegn** et histogram over fordelingen af den ugentlige omsætning af "Mini Quick" i henholdsvis dagligvarebutikker (*Mini Quick* (*dvh*) og i "convenience"-butikker (*Mini Quick* (*conv*).
 - **Estimér** middelværdi og standardafvigelse af den ugentlige omsætning i hver af de to butikstyper.
- b). Gør rede for om der er statistisk belæg for en nulhypotese om, at den forventede ugentlige omsætning af "Mini Quick" i perioden 2014-2015 er højst 240.000 kr. større i dagligvarebutikker end i "convenience"-butikker (brug $\alpha=5\%$ som signifikansniveau). Angiv i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.
- c). **Gør rede for** om konklusionen i delspg. b) ville ændre sig, såfremt datamaterialet kun bestod af et halvt års observationer (= 26 ugentlige observationer) (og alle andre størrelser antages uændrede).



Baggrund og databeskrivelse (eksempel 3)

Denne opgave beskæftiger sig med befolkningens holdning til en eventuel samfundsøkonomisk gevinst ved at flere unge gennemfører en videregående uddannelse. Opgaven tager udgangspunkt i filen VideregåendeUddannelse. jmp, hvis indhold er beskrevet på eksempelark 4.

EKSEMPEL 3 (STATENS UDDANNELSESSTØTTE (SU))

- a). Beregn et 95%-konfidensinterval for forskellen i andelen, der er enig i spørgsmål 3 $(Spg\ 3\ (grupperet))^1$, blandt mænd og kvinder (Køn) og giv en fortolkning af intervallet.
- b). **Gør rede for** om der er statistisk belæg for en nulhypotese om at andelen, der er enig i spørgsmål 3, er højst 2% større blandt kvinder end blandt mænd (brug $\alpha = 5\%$ som signifikansniveau). **Angiv** i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.
- c). **Gør rede for** om konklusionen i delspg. d) ville ændre sig, såfremt datamaterialet kun bestod af halvt så mange svar (og alle andre størrelser antages uændrede).²

¹Spørgsmål 3 lyder "Er du enig eller uenig i, at Statens Uddannelsesstøtte (5.839 kr. pr. måned før skat for udeboende) er for lav til, at man som studerende kan klare sig økonomisk gennem en videregående uddannelse uden at have et arbejde ved siden af?"

²Herved forstås, at der er svar fra 100 mænd, hvoraf 74 erklærer sig enige i spg. 3, og svar fra 106 kvinder, hvoraf 88 erklærer sig enige i spg. 3.



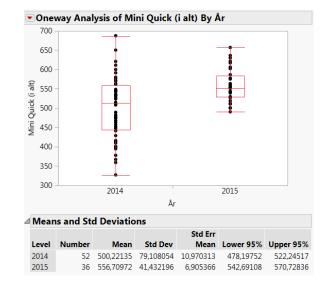
VEJLEDENDE LØSNINGER

EKSEMPEL 1 (► Løsning i JMP)

1a):

Box plots af fordelingen af den ugentlige omsætning af "Mini Quick" for hvert af årene 2014 og 2015 viser, at omsætningen generelt ligger højere og med mindre udsving fra uge til uge i 2015 sammenlignet med 2014.

Den ugentlige omsætning har i 2014 en estimeret middelværdi på 500.221 kr. og en tilhørende standardafvigelse på 79.108 kr. Tilsvarende har den ugentlige omsætning i 2015 en estimeret middelværdi på 556.710 kr. og en tilhørende standardafvigelse på 41.432 kr.



Middelværdien af omsætningen er således ca. 50.000 kr. højere i 2015 end i 2014, hvilket illustreres af at centrum for boxplottet for 2015 ligger ca. 50.000 kr. højere end centrum i boxplottet for 2014. Standardafvigelsen af omsætningen er ca. halvt så stor i 2015 som i 2014, dvs. der er langt mindre variation i omsætningen fra uge til uge i 2015, hvilket illustreres af at bredden af box plottet for 2015 er langt mindre end bredden af box plottet for 2014.

1_B):

Et 95%-konfidensinterval for forskellen mellem den forventede ugentlige omsætning af "Mini Quick" i 2015 og 2014 er givet ved [30.696 kr.; 82.281 kr.].



Med 95% sandsynlighed vil den forventede ugentlige omsætning af "Mini Quick" således ligge mellem 30.696 kr. og 82.281 kr. højere i 2015 end året før.



Med 95% sandsynlighed har der dermed været en vækst i den forventede/gennemsnitlige omsætning på et sted mellem ca. 31.000 kr. og ca. 82.000 kr. fra 2014 til 2015.

(Bemærk: I det omfang at udviklingen i omsætningen over perioden 2014-2015 kan tages som retningsgivende for den fremtidige udvikling af omsætningen, er det således rimeligt at forvente en omsætningsvækst på mellem 31.000 kr. og 82.000 kr. over det kommende år)

Et 99%-konfidensinterval for forskellen mellem den forventede ugentlige omsætning af "Mini Quick" i 2015 og 2014 er givet ved [22.293 kr.; 90.684 kr.].



Med 99% sandsynlighed vil den forventede ugentlige omsætning af "Mini Quick" således ligge mellem 22.293 kr. og 90.684 kr. højere i 2015 end året før.

99%-konfidensintervallet er bredere end 95%-konfidensintervallet, fordi det med større sandsynlighed (99% frem for 95%) indeholder den sande forskel på den forventede omsætning i de to år.

1c):

Test af nulhypotese om at den forventede ugentlige omsætning af "Mini Quick" er den samme i 2014 og 2015:

Antagelser: Observationerne af de enkelte ugers omsætning er

indbyrdes uafhængige indenfor hvert år og også mel-

lem de to år og er approksimativt normalfordelte.

Hypoteser: $H_0: \mu_{2015} - \mu_{2014} = 0 \text{ kr.}$

 $H_a: \mu_{2015} - \mu_{2014} \neq 0 \text{ kr.}$

Signifikansssh.: Mindre end 0,01%

Tekn. konklusion: Signifikanssandsynligheden er mindre end $\alpha = 5\%$

og dermed forkastes H_0 .

Let konklusion: Der er ikke på baggrund af datamaterialet belæg for

en påstand om, at den forventede ugentlige omsætning af "Mini Quick" er den samme i 2014 og 2015. Med andre ord, der er forskel på den forventede om-

sætning i 2014 og 2015.

Assuming unequal variances								
56,4884	t Ratio	4,35776						
12,9627	DF	80,91175						
82,2805	Prob > t	<,0001*						
30,6962	Prob > t	<,0001*						
0,95	Prob < t	1,0000						
	56,4884 12,9627 82,2805 30,6962	equal variances 56,4884 t Ratio 12,9627 DF 82,2805 Prob > t 30,6962 Prob > t 0,95 Prob < t						

Hvis vi i stedet for $\alpha=5\%$ bruger $\alpha=1\%$ som signifikansniveau, så ændrer det ikke på konklusionen, eftersom signifikanssandsynligheden også er mindre end $\alpha=1\%$.



Sammenhængen mellem de ovenstående hypotesetest og konfidensintervallerne i delspg. b) er, at nulhypotesen $H_0: \mu_{2015} - \mu_{2014} = \mu_0$ forkastes ved test på signifikansniveau α netop hvis μ_0 IKKE tilhører et $1-\alpha$ konfidensinterval for $\mu_{2015} - \mu_{2014}$. I tilfældet her forkastes $H_0: \mu_{2015} - \mu_{2014} = 0$ ved test på signifikansniveau $\alpha = 5\%$, hvilket er identisk med at værdien 0 IKKE er en del af $1-\alpha = 95\%$ -konfidensintervallet [30.696 kr.; 82.281 kr.] for $\mu_{2015} - \mu_{2014}$. Præcis det samme gør sig gældende i tilfældet $\alpha = 1\%$.

1_D):

Test af nulhypotese om at den forventede ugentlige omsætning af "Mini Quick" er mindst 50.000 kr. større i 2015 end i 2014:

Antagelser: Observationerne af de enkelte ugers omsætning er indbyrdes uafhængige in-

denfor hvert år og også mellem de to år og er approksimativt normalfordelte.

Hypoteser: $H_0: \mu_{2015} - \mu_{2014} \ge 50.000 \text{ kr.}$

 $H_a: \mu_{2015} - \mu_{2014} < 50.000 \text{ kr.}$

Signifikansssh.: 69,10%

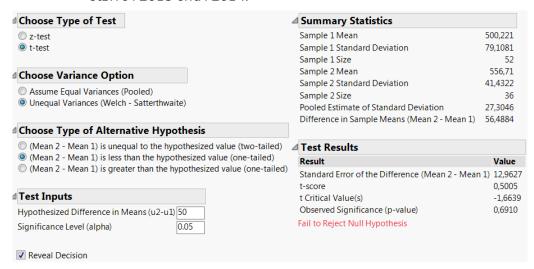
Tekn. konklusion: Signifikanssandsynligheden er større end $\alpha = 5\%$ og dermed forkastes H_0

ikke.

Let konklusion: Der er ikke på baggrund af datamaterialet belæg for afvise en påstand om,

at den forventede ugentlige omsætning af "Mini Quick" er mindst 50.000 kr.

større i 2015 end i 2014.



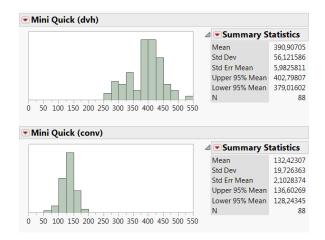


EKSEMPEL 2 (► Løsning i JMP)

2a):

Histogrammerne af fordelingen af den ugentlige omsætning af "Mini Quick" i hhv. dagligvare- og "convenience"-butikker viser, at omsætningen i dagligvarebutikker generelt ligger højere og med større udsving sammenlignet med omsætningen i "convenience"-butikker.

Den ugentlige omsætning i dagligvarebutikker har en estimeret middelværdi på 390.907 kr. og en tilhørende standardafvigelse på 56.122 kr. Tilsvarende har den



ugentlige omsætning i "convenience"-butikker en estimeret middelværdi på 132.423 kr. og en tilhørende standardafvigelse på 19.726 kr.



2в):

Test af nulhypotese om at den forventede ugentlige omsætning af "Mini Quick" er højst 240.000 kr. større i dagligvare- end i "convenience"-butikker:

Antagelser: Observationerne af de enkelte ugers omsætning er indbyrdes uafhængige

indenfor hver butikstype og også mellem de butikstyper og er approksimativt

normalfordelte.

Hypoteser: $H_0: \mu_{dvh} - \mu_{conv} \le 240.000 \text{ kr.}$

 $H_a: \mu_{dvh} - \mu_{conv} > 240.000 \text{ kr.}$

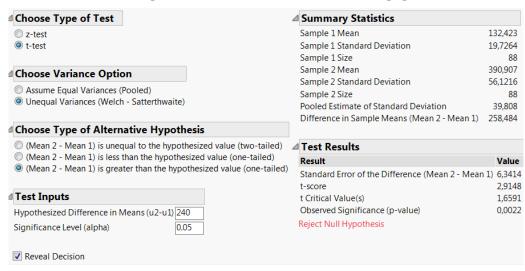
Signifikansssh.: 0,22%

Tekn. konklusion: Signifikanssandsynligheden er mindre end $\alpha = 5\%$ og dermed forkastes H_0 .

Let konklusion: Der er ikke på baggrund af datamaterialet belæg for en påstand om, at den

forventede ugentlige omsætning af "Mini Quick" er højst 240.000 kr. større i dagligvare- end i "convenience"-butikker. Med andre ord, den forventede

omsætning er mere end 240.000 kr. større i dagligvarebutikker.





2c):

Test af nulhypotese om at den forventede ugentlige omsætning af "Mini Quick" er højst 240.000 kr. større i dagligvare- end i "convenience"-butikker:

Antagelser: Observationerne af de enkelte ugers omsætning er indbyrdes uafhængige

indenfor hver butikstype og også mellem de butikstyper og er approksimativt

normalfordelte.

Hypoteser: $H_0: \mu_{dvh} - \mu_{conv} \le 240.000 \text{ kr.}$

 $H_a: \mu_{dvh} - \mu_{conv} > 240.000 \text{ kr.}$

Signifikansssh.: 6,16%

Tekn. konklusion: Signifikanssandsynligheden er større end $\alpha = 5\%$ og dermed forkastes H_0

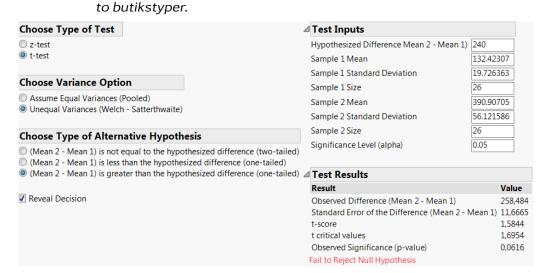
ikke.

Let konklusion: Der er ikke på baggrund af datamaterialet belæg for at afvise en påstand om,

at den forventede ugentlige omsætning af "Mini Quick" er højst 240.000 kr.

større i dagligvare- end i "convenience"-butikker.

Intuitionen er, at når vi kun anvender data for et halvt år (26 uger), så er der så stor usikkerhed forbundet med vores estimater af den forvente-de/gennemsnitlige omsætning, at den observerede forskel i omsætningen mellem de to butikstyper ikke er stor nok til at give os belæg for at afvise en påstand om en forskel i forventet omsætning på højst 240.000 kr. mellem de





EKSEMPEL 3 (► Løsning i JMP)

3a):

Et 95%-konfidensinterval for forskellen mellem andelen af mænd og andelen af kvinder, der er enige i spørgsmål 3, er givet ved [-17,2%; -1,4%].

Two Sample Test for P	ropo	ortions			
Description		portion ference	Lov	ver 95%	Upper 95%
P(Enig Mand)-P(Enig Kvinde)	-0,09387		-	0,17224	-0,01398
Adjusted Wald Test (Null Hypothesis)		P	rob		
P(Enig Mand)-P(Enig Kvinde) ≤ 0	0,98	95		
P(Enig Mand)-P(Enig Kvinde) ≥ 0	0,01	.05*		
P(Enig Mand)-P(Enig Kvinde) = 0	0,02	11*		

Med 95% sandsynlighed vil andelen af mænd, der er enige i spørgsmål 3, således være mellem 1,4% og 17,2% lavere end andelen af kvinder, der er enige i spørgsmål 3.



3в):

Test af nulhypotese om at andelen af kvinder, der er enige i spørgsmål 3, er højst 2% større end andelen af mænd, der er enige i spørgsmål 3:

Antagelser: Observationerne af de enkelte respondenters svar er indbyrdes uafhængi-

ge indenfor hvert køn og også mellem de to køn. Respondenterne har kun to svarmuligheder, "Enig" eller "Uenig"), hvor p_{kvinde} hhv. p_{mand} betegnelser andelen af kvinder hhv. mænd, der er "Enig" (= sandsynligheden for at være "Enig") i spørgsmål 3. Desuden opfylder datamaterialet betingelserne $n_{kvinde} \cdot \hat{p}_{kvinde} = 212 \cdot 0,830 = 176 > 5$ og $n_{kvinde} \cdot (1 - \hat{p}_{kvinde}) = 212 \cdot (1 - 0,830) = 176 > 5$

 $201 \cdot (1 - 0,736) = 53 > 5.$

Hypoteser: $H_0: p_{kvinde} - p_{mand} \le 2\%$

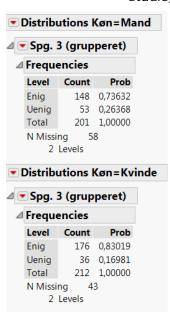
 $H_a: p_{kvinde} - p_{mand} > 2\%$

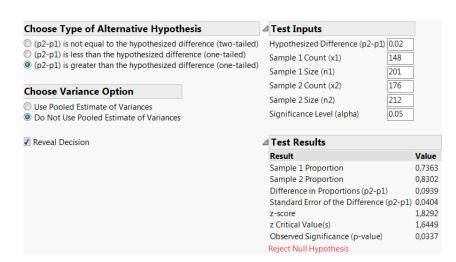
Signifikansssh.: 3,37%

Tekn. konklusion: Signifikanssandsynligheden er mindre end $\alpha = 5\%$ og dermed forkastes H_0 .

Let konklusion: Der er ikke på baggrund af datamaterialet belæg for en påstand om, at an-

delen af kvinder, der er enige i spørgsmål 3, er højst 2% højere end andelen af mænd, der er enige i spørgsmål 3. Med andre ord, andelen af kvinder, der er enige i at SU'en er for lav til at man kan undgå at arbejde ved siden at sit studie, er mere end 2% højere end andelen af mænd, der mener det samme.







3c):

Test af nulhypotese om at andelen af kvinder, der er enige i spørgsmål 3, er højst 2% større end andelen af mænd, der er enige i spørgsmål 3:

Antagelser: Observationerne af de enkelte respondenters svar er indbyrdes uafhængi-

ge indenfor hvert køn og også mellem de to køn. Respondenterne har kun to svarmuligheder, "Enig" eller "Uenig"), hvor p_{kvinde} hhv. p_{mand} betegnelser andelen af kvinder hhv. mænd, der er "Enig" (= sandsynligheden for at være "Enig") i spørgsmål 3. Desuden opfylder datamaterialet betingelserne

 $n_{kvinde} \cdot \hat{p}_{kvinde} = 88 > 5 \text{ og } n_{kvinde} \cdot (1 - \hat{p}_{kvinde}) = 18 > 5 \text{ og } n_{mand} \cdot \hat{p}_{mand} = 74 > 5$

og $n_{mand} \cdot (1 - \hat{p}_{mand}) = 26 > 5$.

Hypoteser: $H_0: p_{kvinde} - p_{mand} \le 2\%$

 $H_a: p_{kvinde} - p_{mand} > 2\%$

Signifikansssh.: 10,03%

Tekn. konklusion: Signifikanssandsynligheden er større end $\alpha = 5\%$ og dermed forkastes H_0

ikke.

Let konklusion: Der er ikke på baggrund af datamaterialet belæg for afvise en påstand om,

at andelen af kvinder, der er enige i spørgsmål 3 , er højst 2% højere end

andelen af mænd, der er enige i spørgsmål 3.

Intuitionen er, at når vi kun anvender svar fra halvt så mange personer, så er der så stor usikkerhed forbundet med vores estimater af andelen af hvert køn, der er enige i spørgsmål 3, at den observerede forskel mellem andelene af de to køn ikke er stor nok til at give os belæg for at afvise en påstand om, at andelen af kvinder er højst 2% større end andelen af mænd.

Choose Type of Alternative Hypoth	esis	△ Test Inputs		
 (p2-p1) is not equal to the hypothesized difference (two-tailed) (p2-p1) is less than the hypothesized difference (one-tailed) (p2-p1) is greater than the hypothesized difference (one-tailed) 		Hypothesized Difference (p2-p1) 0,02		
		Sample 1 Count (x1)	74	
		Sample 1 Size (n1)	100	
Choose Variance Option		Sample 2 Count (x2)	88	
Use Pooled Estimate of Variances		Sample 2 Size (n2)	106	
Do Not Use Pooled Estimate of Variances		Significance Level (alpha)	0.05	
✓ Reveal Decision		△ Test Results		
		Result		
		Sample 1 Proportion	0,74	
		Sample 2 Proportion	0,830	
		Difference in Proportions (p2-p1)		
		Standard Error of the Difference	e (p2-p1) 0,057	
		z-score	1,230	
		z Critical Value(s)	1,644	
		Observed Significance (p-value	e) 0,109	
		Observed Significance (p. value	0,100	