



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 (DOBBELT BOGSTAVJAGT)

- a). **Gør rede for** om der er statistisk belæg for en nulhypotese om, at den forventede ugentlige omsætning af “Dobbelt Bogstavjagt” (*Dobbelt Bogstavjagt (i alt)*) er den samme i alle årets fire kvartaler (*Kvartal*) (brug $\alpha = 5\%$ som signifikansniveau). **Angiv** i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.
- b). **Gør rede for** om antagelsen om normalfordelte data ser ud til at være opfyldt ved at se på histogrammer af fordelingen af den ugentlige omsætning af “Dobbelt Bogstavjagt” indenfor hvert kvartal.

Gør rede for om antagelsen om normalfordelte data ser ud til at være opfyldt ved at se på middelværdi og standardafvigelse af den ugentlige omsætning af “Dobbelt Bogstavjagt” indenfor hvert kvartal og sammenligne histogrammerne ovenfor med den empiriske regel.

EKSEMPEL 2 (MINI QUICK)

- a). **Gør rede for** om der er statistisk belæg for en nulhypotese om, at den forventede ugentlige omsætning af “Mini Quick” (*Mini Quick (i alt)*) er den samme i alle årets fire kvartaler (brug $\alpha = 5\%$ som signifikansniveau). **Angiv** i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.
- b). **Tegn** box plots af fordelingen af den ugentlige omsætning af “Mini Quick” indenfor hvert af årets fire kvartaler.
- Gør rede for** at omsætningen i årets fjerde kvartal ser ud til i et vist omfang at skille sig ud fra omsætningen i de øvrige tre kvartaler.
- c). **Gør rede for** om konklusionen i delspg. a) ændres, såfremt der ses bort fra omsætningen i årets fjerde kvartal i analysen.



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 (UDDANNELSE OG ØKONOMISK VELSTAND)

- a). **Gør rede for** om der er statistisk belæg for en nulhypotese om, at holdningen til spørgsmål 5 (*Spg 5 (grupperet)*)¹ er den samme i hele landet (brug $\alpha = 5\%$ som signifikansniveau). **Angiv** i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.
- b). **Gør rede for** om der er statistisk belæg for en nulhypotese om, at holdningen til spørgsmål 1 (*Spg 1 (grupperet)*)² er den samme i hele landet (brug $\alpha = 5\%$ som signifikansniveau). **Angiv** i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.

¹Spørgsmål 5 lyder "Er du enig eller uenig i, at man ved at tage en videregående uddannelse styrker sine egne forudsætninger for social og økonomisk velstand på lang sigt?"

²Spørgsmål 1 lyder "Er du enig eller uenig i, at det vil øge den økonomiske vækst i samfundet, såfremt flere unge gennemfører en videregående uddannelse?"



VEJLEDENDE LØSNINGER

Eksempel 1 (► Løsning i JMP)

1A):

Test af nulhypotese om at den forventede ugentlige omsætning af "Dobbelt Bogstavjagt" er den samme i alle fire kvartaler:

Antagelser: Observationerne af de enkelte ugers omsætning er indbyrdes uafhængige og approksimativt normalfordelte; normalfordelte indenfor hvert kvartal og med samme standardafvigelse for alle fire kvartaler

Hypoteser: $H_0: \mu_{1. \text{ kvartal}} = \mu_{2. \text{ kvartal}} = \mu_{3. \text{ kvartal}} = \mu_{4. \text{ kvartal}}$
 $H_a: \text{Mindst to af middelværdierne } \mu_{1. \text{ kvartal}}, \mu_{2. \text{ kvartal}}, \mu_{3. \text{ kvartal}}, \mu_{4. \text{ kvartal}} \text{ er ikke ens}$

Signifikanssssh.: 20,68%

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 "Dobbelt Bogstavjagt" er den samme i alle årets fire kvartaler.

Oneway Anova

Summary of Fit

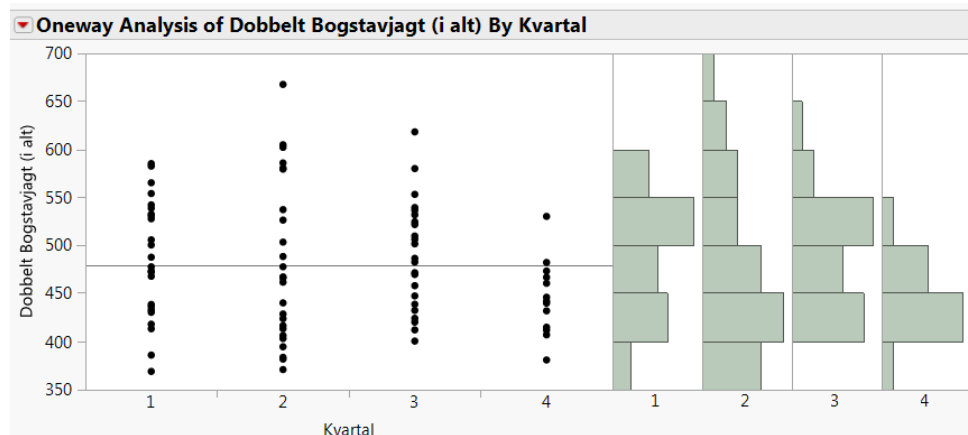
Rsquare	0,052558
Adj Rsquare	0,018721
Root Mean Square Error	64,9886
Mean of Response	478,6664
Observations (or Sum Wgts)	88

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Kvartal	3	19680,86	6560,29	1,5533	0,2068
Error	84	354775,53	4223,52		
C. Total	87	374456,39			

1B):

Histogrammer af fordelingen af den ugentlige omsætning af "Dobbelt Bogstavjagt" indenfor hvert kvartal viser, at fordelingen ser ud til at være pænt klokkeformet og dermed i overensstemmelse med normalfordelingen for alle fire kvartalers vedkommende.



På grund af den pæne klokkeform giver det fin mening at anvende den empiriske regel. Ifølge den vil vi forvente, at næsten alle observationer ligger indenfor ± 3 standardafvigelser af middelværdien. På baggrund af de beregnede middelværdier og standardafvigelser af omsætningen, vil vi derfor forvente at omsætningen i...

- 1. kvartal vil ligge i intervallet (i 1.000 kr.) $[488 - 3 \cdot 59; 488 + 3 \cdot 59] = [311, 665]$
- 2. kvartal vil ligge i intervallet (i 1.000 kr.) $[477 - 3 \cdot 84; 477 + 3 \cdot 84] = [225, 729]$
- 3. kvartal vil ligge i intervallet (i 1.000 kr.) $[490 - 3 \cdot 57; 490 + 3 \cdot 57] = [319, 661]$
- 4. kvartal vil ligge i intervallet (i 1.000 kr.) $[445 - 3 \cdot 39; 445 + 3 \cdot 39] = [328, 562]$

For alle fire kvartalers vedkommende ligger observationerne/histogrammer fint indenfor de beregnede intervaller, og der er således heller ikke ud fra denne vurdering tegn på, at data-materialet ikke er i overensstemmelse med normalfordelingsantagelsen. Dog er standardafvigelsen for omsætningen i 4. kvartal lidt lavere end for de øvrige kvartaler, hvilket ikke er helt i overensstemmelse med antagelsen om ens standardafvigelse i alle fire kvartaler (jf. delspg. a) ovenfor).

Means and Std Deviations						
Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
1	26	487,63538	59,410972	11,65145	463,63877	511,632
2	26	476,63731	84,344696	16,541356	442,56975	510,70487
3	23	489,78739	56,703528	11,823503	465,26695	514,30784
4	13	445,11077	38,672621	10,725855	421,74114	468,4804

²Det er også helt fint, hvis man i stedet argumenterer for, at ca. 95% af observationer ifølge den empiriske regel vil ligge indenfor ± 2 standardafvigelsr af middelværdien.

**Eksempel 2** (► Løsning i JMP)**2A):**

Test af nulhypotese om at den forventede ugentlige omsætning af “Mini Quick” er den samme i alle fire kvartaler:

Antagelser: Observationerne af de enkelte ugers omsætning er indbyrdes uafhængige og approksimativt normalfordelte; normalfordelte indenfor hvert kvartal og med samme standardafvigelse for alle fire kvartaler

Hypoteser: $H_0 : \mu_{1. \text{ kvartal}} = \mu_{2. \text{ kvartal}} = \mu_{3. \text{ kvartal}} = \mu_{4. \text{ kvartal}}$
 $H_a : \text{Mindst to af middelværdierne } \mu_{1. \text{ kvartal}}, \mu_{2. \text{ kvartal}}, \mu_{3. \text{ kvartal}}, \mu_{4. \text{ kvartal}} \text{ er ikke ens}$

Signifikanssssh.: 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 alle årets fire kvartaler.

Oneway Anova**Summary of Fit**

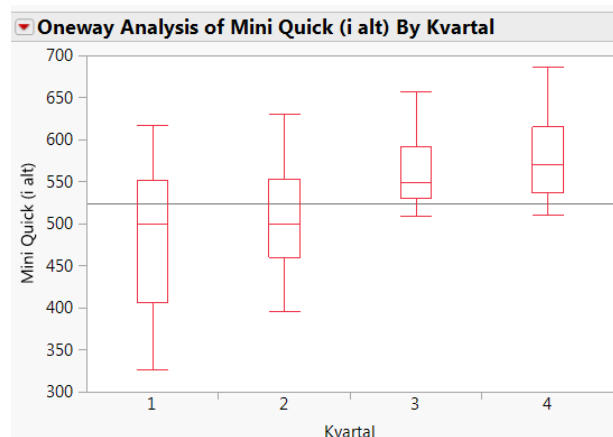
Rsquare	0,269723
Adj Rsquare	0,243642
Root Mean Square Error	62,34738
Mean of Response	523,3302
Observations (or Sum Wgts)	88

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Kvartal	3	120599,72	40199,9	10,3416	<,0001*
Error	84	326524,45	3887,2		
C. Total	87	447124,16			

2B):

Box plots af omsætningen af “Mini Quick” i de fire kvartaler viser, at omsætningen generelt er lavere i 1. og 2. kvartal sammenlignet med 3. og 4. kvartal, mens variationen i omsætningen fra uge til uge til gengæld er lidt større i 1. og 2. kvartal. I 4. kvartal er omsætningen generelt lidt højere end i 3. kvartal, og dermed er omsætningen set over hele året højest i 4. kvartal.



2c):

Test af nulhypotese om at den forventede ugentlige omsætning af “Mini Quick” er den samme i årets første tre kvartaler:

Antagelser: Observationerne af de enkelte ugers omsætning er indbyrdes uafhængige og approksimativt normalfordelte; normalfordelte indenfor hvert af de første tre kvartaler og med samme standardafvigelse for de første tre kvartaler

Hypoteser: $H_0: \mu_{1. \text{kvartal}} = \mu_{2. \text{kvartal}} = \mu_{3. \text{kvartal}}$
 $H_a: \text{Mindst to af middelværdierne } \mu_{1. \text{kvartal}}, \mu_{2. \text{kvartal}}, \mu_{3. \text{kvartal}} \text{ er ikke ens}$

Signifikanssssh.: 0,02%

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 årets første tre kvartaler. På baggrund af box plots i delspg. b) er konklusionen ikke overraskende, idet figurerne viste, at omsætningen generelt er højere i 3. kvartal sammenlignet med 1. og 2. kvartal.

Oneway Anova

Summary of Fit

Rsquare	0,214883
Adj Rsquare	0,193074
Root Mean Square Error	63,71851
Mean of Response	514,3885
Observations (or Sum Wgts)	75

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Kvartal	2	80007,74	40003,9	9,8531	0,0002*
Error	72	292323,52	4060,0		
C. Total	74	372331,26			

**Eksempel 3** (► Løsning i JMP)**3A):**

Test af nulhypotese om at der er uafhængighed mellem holdningen til spørgsmål 5, og hvor i landet man er bosiddende:

- Antagelser:** Observationerne af respondenters svar er indbyrdes uafhængige. Endvidere er det forventede antal observationer i hver af krydstabellens er mindst 5. Sidstnævnte er opfyldt med undtagelse af antallet af uenige respondenter i Region Sjælland (forventet: 3,9) og Region Nordjylland (forventet: 2,6).
- Hypoteser:** H_0 : Uafhængighed mellem Spg. 5 (grupperet) og Region
 H_a : Afhængighed mellem Spg. 5 (grupperet) og Region
- Signifikanssssh.:** 95,51%
- 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 holdningen til spørgsmål 5 (en videregående uddannelse styrker egne forudsætninger for social og økonomisk velstand) er den samme i hele landet.

Contingency Analysis of Spg. 5 (grupperet) By Region
 region REGION

Contingency Table

		Spg. 5 (grupperet)		
	Count	Enig	Uenig	Total
Expected		125	7	132
Hovedstaden	124,253	7,74718		
Region	Sjælland	62	5	67
		63,0677	3,93228	
Syddanmark	87	5	92	
		86,6005	5,39955	
Midtjylland	100	7	107	
		100,72	6,27991	
Nordjylland	43	2	45	
		42,3589	2,64108	
Total		417	26	443

Tests

N	DF	-LogLike	RSquare (U)
443	4	0,33023246	0,0033

Test	ChiSquare	Prob> ChiSq
Likelihood Ratio	0,660	0,9561
Pearson	0,669	0,9551

**3B):**

Test af nulhypotese om at der er uafhængighed mellem holdningen til spørgsmål 1, og hvor i landet man er bosiddende:

- Antagelser:** Observationerne af respondenters svar er indbyrdes uafhængige. Endvidere er det forventede antal observationer i hver af krydstabellens er mindst 5. Sidstnævnte er opfyldt med undtagelse af antallet af uenige respondenter i Region Nordjylland (forventet: 3,8).
- Hypoteser:** H_0 : Uafhængighed mellem *Spg. 1 (grupperet)* og *Region*
 H_a : Afhængighed mellem *Spg. 1 (grupperet)* og *Region*
- Signifikanssssh.:** 1,00%
- 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 holdningen til spørgsmål 1 (øget samfundsøkonomisk vækst såfremt flere unge tager en videregående uddannelse) er den samme i hele landet. Hvis man kigger nærmere på tallene kan man se, at der generelt er noget større andele af befolkningen, der er uenige i spørgsmål 1 i Region Sjælland (16%) og Region Nordjylland (26%) sammenlignet med i de øvrige tre regioner (mindre end 11%).

Contingency Analysis of Spg. 1 (grupperet) By Region

Region REGION

Contingency Table

		Spg. 1 (grupperet)		
		Enig	Uenig	Total
Region	Count			
	Row %			
	Expected			
	Hovedstaden	102	9	111
		91,89	8,11	
		99,0551	11,9449	
Sjælland		48	9	57
		84,21	15,79	
		50,8661	6,13386	
Syddanmark		76	9	85
		89,41	10,59	
		75,853	9,14698	
Midtjylland		88	5	93
		94,62	5,38	
		82,9921	10,0079	
Nordjylland		26	9	35
		74,29	25,71	
		31,2336	3,7664	
Total		340	41	381

Tests

	N	DF	-LogLike	RSquare (U)
	381	4	5,8665641	0,0451

Test	ChiSquare	Prob> ChiSq
Likelihood Ratio	11,733	0,0195*
Pearson	13,274	0,0100*