Opgave 4.6

En salgsvirksomhed med egne forretninger har delt Danmark op i to stort set lige store segmenter – kaldet øst og vest. For at undersøge om der er forskel i omsætningen i de to segmenter, har man gennem en periode registreret salget i én forretning i øst og i én forretning i vest. De to forretninger er stort set lige store mht. indretning, sortiment og areal. Ligeledes har de begge en beliggenhed i en by med ca. 60.000 indbyggere.
Dette har resulteret i følgende, hvor omsætningen er i 1.000 kr.:

Måned	Øst	Vest
Januar	200	240
Februar	210	248
Marts	190	228
April	280	340
Maj	240	288
Juni	230	280
Juli	200	240
August	218	270
September	214	257
Oktober	264	320
November	325	390
December	400	490

1. Undersøg om der er forskel i variansen i omsætningen mellem øst og vest

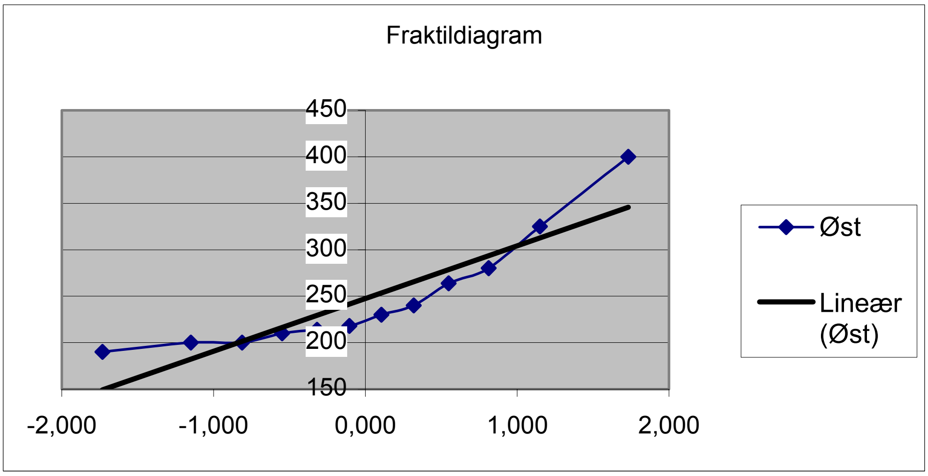
Hypotese:

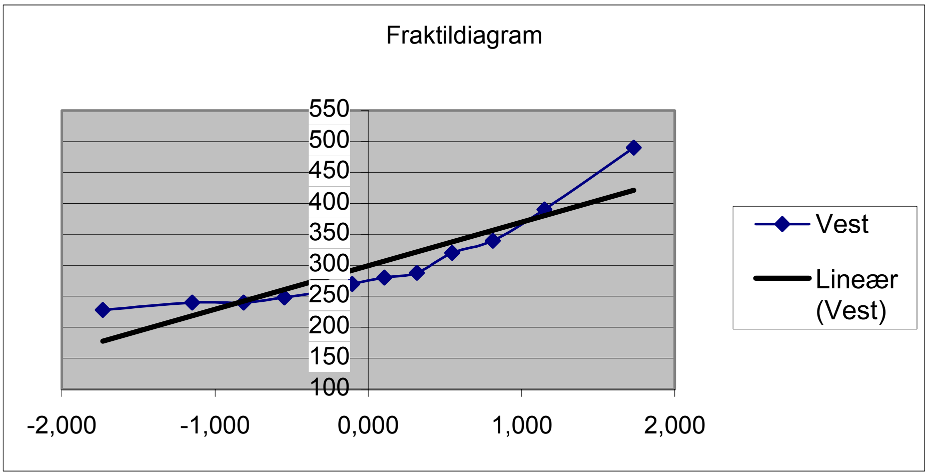
$$H_0 : \sigma^2_{øst} = \sigma^2_{vest}$$

$$H_1 : \sigma^2_{øst} \neq \sigma^2_{vest}$$

Forudsætninger:

De to populationer er normalfordelte.
Dette kan undersøges ved at lave et fraktildiagram for hver af de to populationer. Disse ses nedenfor, og det fremgår, at der ikke er særlig god overensstemmelse mellem tendenslinien og punkterne. Det er især galt i enderne.
Da kravet om normalfordelte population er et skarpt krav ved test for varianshomogenitet skal de efterfølgende konklusioner tages med forbehold!





Data til fraktildiagrammerne:

Øst Omsætning	Obs.nr.			Vest Omsætning	Obs.nr.		
190	1	0,042	-1,732	228	1	0,042	-1,732
200	2	0,125	-1,150	240	2	0,125	-1,150
200	3	0,208	-0,812	240	3	0,208	-0,812
210	4	0,292	-0,549	248	4	0,292	-0,549
214	5	0,375	-0,319	257	5	0,375	-0,319
218	6	0,458	-0,105	270	6	0,458	-0,105
230	7	0,542	0,105	280	7	0,542	0,105
240	8	0,625	0,319	288	8	0,625	0,319
264	9	0,708	0,549	320	9	0,708	0,549
280	10	0,792	0,812	340	10	0,792	0,812
325	11	0,875	1,150	390	11	0,875	1,150
400	12	0,958	1,732	490	12	0,958	1,732

Teststørrelse:

$$F^{test}_{n_{vest}-1;n_{øst}-1} = \frac{s^2_{vest}}{s^2_{øst}} = \frac{5870,3862}{3833,7224} = 1,5312$$

Da det største variansestimat er placeret i tælleren, er det kun i højre hale, der er en chance for at forkaste H_0. Derfor behøver vi kun at finde den kritiske værdi i højre side.

PaceXL Ungrouped Data:

Sample Data:	Øst
Number of Data Points	12
Minimum	190,0
Maximum	400,0
Total	2971,0
Arithmetic mean	**247,5833**
Median	224,0
Mode	200,0
First Quartile	205,0
Third Quartile	272,0
Range	210,0
Inter Quartile Range	67,0
Variance (Sample)	3833,7224
Standard Deviation (Sample)	61,9171
Coefficient of Variation (Sample)	0,2501
Pearson's Skewness Coeff (Sample)	1,1427
Mean Absolute Deviation	46,4444
Standard Error of Mean	17,8739

PaceXL Ungrouped Data:

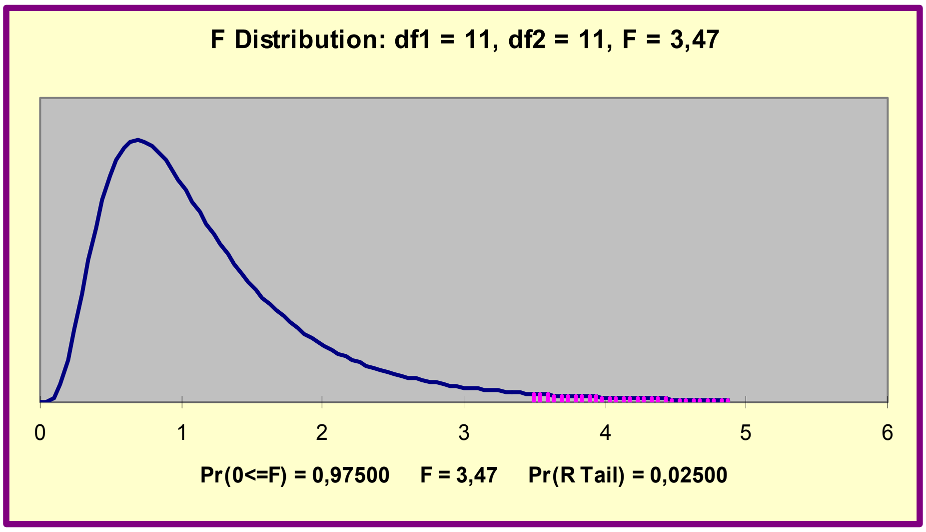
Sample Data:	Vest
Number of Data Points	12
Minimum	228,0
Maximum	490,0
Total	3591,0
Arithmetic mean	**299,25**
Median	275,0
Mode	240,0
First Quartile	244,0
Third Quartile	330,0
Range	262,0
Inter Quartile Range	86,0
Variance (Sample)	5870,3862
Standard Deviation (Sample)	76,6184
Coefficient of Variation (Sample)	0,256
Pearson's Skewness Coeff (Sample)	0,9495
Mean Absolute Deviation	57,1667
Standard Error of Mean	22,1178

Kritisk værdi:

$$F_{n_{vest}-1;n_{øst}-1}^{kritisk,øvre(p=1-\alpha/2)} = F_{11;11}^{kritisk,øvre(p=0,975)} = 3,47$$

PaceXL: Probabilities

F Distribution:	df1	df2	Mean	StDev	F
	11,	11,	1,2222	0,8809	3,47
	Pr(0<=F)	Pr(R Tail)			
	0,97500	0,02500			



Beslutningsregel :

Nulhypotesen forkastes såfremt $F^{test}_{n_{vest}-1;n_{øst}-1} > F^{kritisk,øvre(p=1-\alpha/2)}_{n_{vest}-1;n_{øst}-1}$, dvs.

såfremt $F^{test}_{n_{vest}-1;n_{øst}-1} > 3{,}47$

Da $F^{test}_{n_{vest}-1;n_{øst}-1} = 1{,}5312$ fastholdes H_0. Vi kan altså ikke påvise, at der er forskel i variansen i omsætningen mellem øst og vest. Men igen – konklusionen skal tages med forbehold pga. afvigelsen fra normaliteten.

2. Undersøg, om der er forskel i den gennemsnitlige omsætning mellem øst og vest
   Der er tale om uafhængige stikprøver

Hypotese:

$H_0 : \mu_{øst} = \mu_{vest} \Leftrightarrow \mu_{øst} - \mu_{vest} = 0 \Leftrightarrow \mu_D = 0$

$H_1 : \mu_{øst} \neq \mu_{vest} \Leftrightarrow \mu_{øst} - \mu_{vest} \neq 0 \Leftrightarrow \mu_D \neq 0$

Forudsætninger:

$\overline{X}_{øst} - \overline{X}_{vest}$ normalfordelt. Vi så i forrige spørgsmål, at populationerne ikke med rimelighed kan antages at være normalfordelte. Teststørrelsen er dog her robust overfor afvigelser, hvis stikprøvestørrelserne er større end eller lig med 30. Dette er imidlertid heller ikke tilfældet her. Da de to stikprøvestørrelser er ens, er konklusionen her dog mere pålidelig end ved testen for varianshomogenitet. Og hvis populationen er nogenlunde symmetrisk er approksimationen i CGV acceptabel selv ved små værdier af n.

Hvis vi antager, der er varianshomogenitet, skal vi benytte følgende teststørrelse:

$$T_{n_{øst}+n_{vest}-2} = \frac{\overline{X}_{øst} - \overline{X}_{vest} - \left(\mu_{øst} - \mu_{vest}\right)_0}{s_{pooled} \cdot \sqrt{\frac{1}{n_{øst}} + \frac{1}{n_{vest}}}} \sim T\left(n_{øst} + n_{vest} - 2\right)$$, hvor

$$s^2_{pooled} = \frac{\left(n_{øst}-1\right) \cdot s^2_{øst} + \left(n_{vest}-1\right) \cdot s^2_{vest}}{n_{øst} + n_{vest} - 2} = \frac{s^2_{øst} + s^2_{vest}}{2}$$

hvor det sidste lighedstegn kun gælder fordi $n_{øst} = n_{vest} = 12$

$$s^2_{pooled} = \frac{3833{,}7224 + 5870{,}3862}{2} = 4852{,}0543$$

$$t^{test}_{n_{øst}+n_{vest}-2} = \frac{\overline{x}_{øst} - \overline{x}_{vest} - 0}{s_{pooled} \cdot \sqrt{\frac{1}{n_{øst}} + \frac{1}{n_{vest}}}} = \frac{247{,}5833 - 299{,}25}{\sqrt{4852{,}0543} \cdot \sqrt{\frac{1}{12} + \frac{1}{12}}} = -1{,}8169$$

Kritisk værdi:

$$t^{kritisk}_{DF=12+12-2;1-\alpha} = t^{kritisk}_{DF=22;0,975} = 2{,}074$$

PaceXL: Probabilities

t Distribution:	T	df	Mean		StDev	
	2,074	22,		0,	1,0488	
	Pr(t <= T)	Pr(t > T)	Pr(-T<t<+T)		Pr(0<..<+T)	Pr(Tails)
	0,97500	0,02500		0,95000	0,47500	0,05000

Beslutningsregel og konklusion:

Nulhypotesen forkastes såfremt $t^{test} < -t^{kritisk}$ eller $t^{test} > t^{kritisk}$, dvs. såfremt $t^{test} < -2{,}074$ eller $t^{test} > 2{,}074$

Da $t^{test} = -1{,}8169$ fastholdes nulhypotesen. Der kan dermed ikke påvises en signifikant forskel i den gennemsnitlige omsætning mellem øst og vest.