

## 2 Oversigt over $\chi^2$ -test

	2 $\chi^2$ -test af typen "Goodness of fit" (GOF)	2 $\chi^2$ -test for "uafhængighed"
<b>Anvendelse</b>	Vil undersøge om observerede data stemmer overens med forventede data. EKS: planters fordeling på type, hvor man forventer at Mendels arveligheds love gælder. EKS: opinionsundersøgelse, hvor man sammenligner med sidste valgresultat.	Vil undersøge om observerede data af en type er uafhængig af observerede data af en anden type. EKS: er rygning uafhængig af køn (ryger/ikke-ryger contra mand/kvinde)? EKS: er en ny operationsmetode bedre end den nuværende (man anvender så 2 grupper forsøgspersoner, nogen gennemgår ny type operation resten den kendte type operation)?
<b>Givet på forhånd</b>	Der skal være givet <u>både</u> <b>observerede data</b> og <b>forventede data</b> .	Der er <u>kun</u> givet <b>observerede data</b> .
<b>Nulhypotese</b>	At de observerede data stemmer overens med de forventede data (som er kendte).	At de observerede data stemmer overens med de forventede data (som man selv skal beregne).
<b>Datastruktur</b>	Observerede data opskrives i en <b>liste OBS</b> . Forventede data opskrives i en <b>liste FORV</b> . <b>NB: %-tal omregnes til faktiske tal.</b>	Observerede data opskrives i en <b>matrix OBS</b> . Forventede data <b>beregnes</b> og opskrives i en <b>matrix FORV</b> .
<b>Omregning fra %-tal til faktiske tal</b>	Kan evt. anvende en omregningsfunktion $f_{\sim}$ på <b>listen</b> med forventede %-tal.	
<b>Beregning af forventede værdier</b>		$\frac{\text{vandret sum}}{\text{sum ialt}} \cdot \text{lodret sum}$
<b>Antal frihedsgrader <math>v</math> (df =degree of freedom)</b>	$v = df = \text{Antal opdelinger} - 1$	$v = df = (\text{Antal rækker} - 1) \cdot (\text{Antal søjler})$

		- 1)
<b><math>\chi^2</math>-teststørrelsen Q</b> NB: Q afhænger ikke af antal frihedsgrader eller af signifikansniveauet!	$\sum \frac{(obs - forv)^2}{forv}$ (summerer over alle <b>listens</b> elementer).	$\sum \frac{(obs - forv)^2}{forv}$ (summerer over alle <b>matricens</b> elementer).
<b>Signifikansniveau</b> <b>Signifikansniveauet er sandsynligheden for at forkaste en SAND nulhypotese!</b>	Typisk 5% = 0.05 (se teksten i opgaven).	Typisk 5% = 0.05 (se teksten i opgaven).
<b>p-værdi</b> <b>p-værdi = <math>P(X \geq Q)</math></b>	Arealet under $\chi^2$ -fordelingen fra Q til $\infty$ . $pVÆRDI := 1 - CDF(X, Q)$ eller med Gym-pakken: $pVÆRDI := 1 - chicdf(v, Q)$	Arealet under $\chi^2$ -fordelingen fra Q til $\infty$ . $pVÆRDI := 1 - CDF(X, Q)$ eller med Gym-pakken: $pVÆRDI := 1 - chicdf(v, Q)$
<b>Kritiske værdi q</b> <b>Angiver det sted, hvor den kritiske mængde starter.</b>	Arealet under $\chi^2$ -fordelingen fra q til $\infty$ = signifikansniveauet. $q := Quantile(X, 1 - Signifikansniveau)$ Med Gym-pakken: $invchi(v, 1 - Signifikansniveau)$	Arealet under $\chi^2$ -fordelingen fra q til $\infty$ = signifikansniveauet. $q := Quantile(X, 1 - Signifikansniveau)$ Med Gym-pakken: $invchi(v, 1 - Signifikansniveau)$
<b>Kritisk mængde</b> <b>Går fra kritisk værdi q til <math>\infty</math>.</b>	Gråt område under $\chi^2$ -fordelingen (fra q til $\infty$ ).	Gråt område under $\chi^2$ -fordelingen (fra q til $\infty$ ).
<b>Oprettelse af en stokastisk variabel X, som er <math>\chi^2</math>-fordelt.</b>	<i>with(Statistics) :</i> $X := RandomVariable(ChiSquare(v)) :$ <b>Ikke nødvendig med Gym-pakken.</b>	<i>with(Statistics) :</i> $X := RandomVariable(ChiSquare(v)) :$ <b>Ikke nødvendig med Gym-pakken.</b>
<b><math>\chi^2</math>-test</b>	Med Gym-pakken: $ChiKvadratGOFtest(OBS, FORV)$ <b>hvis signifikansniveau = 0.05</b> ellers: $ChiKvadratGOFtest(OBS, FORV, level = Signifikansniveau)$	Med Gym-pakken: $ChiKvadratUtest(OBS)$ <b>hvis signifikansniveau = 0.05</b> ellers: $ChiKvadratUtest(OBS, level = Signifikansniveau)$
<b>Argumenter</b>	ACCEPT: hvis p-værdi > signifikansniveau ACCEPT: hvis $\chi^2$ -teststørrelsen Q < kritiske	ACCEPT: hvis p-værdi > signifikansniveau ACCEPT: hvis $\chi^2$ -teststørrelsen Q < kritiske

<b>(konklusion på <math>\chi^2</math>-test)</b>	værdi $q$ ACCEPT: hvis den blå streg på figuren ligger udenfor det grå område -----	værdi $q$ ACCEPT: hvis den blå streg på figuren ligger udenfor det grå område -----
	----- FORKASTES: hvis $p$ -værdi < signifikansniveau FORKASTES: hvis $\chi^2$ -teststørrelsen $Q >$ kritiske værdi $q$ FORKASTES: hvis den blå streg på figuren ligger indenfor det grå område	----- FORKASTES: hvis $p$ -værdi < signifikansniveau FORKASTES: hvis $\chi^2$ -teststørrelsen $Q >$ kritiske værdi $q$ FORKASTES: hvis den blå streg på figuren ligger indenfor det grå område