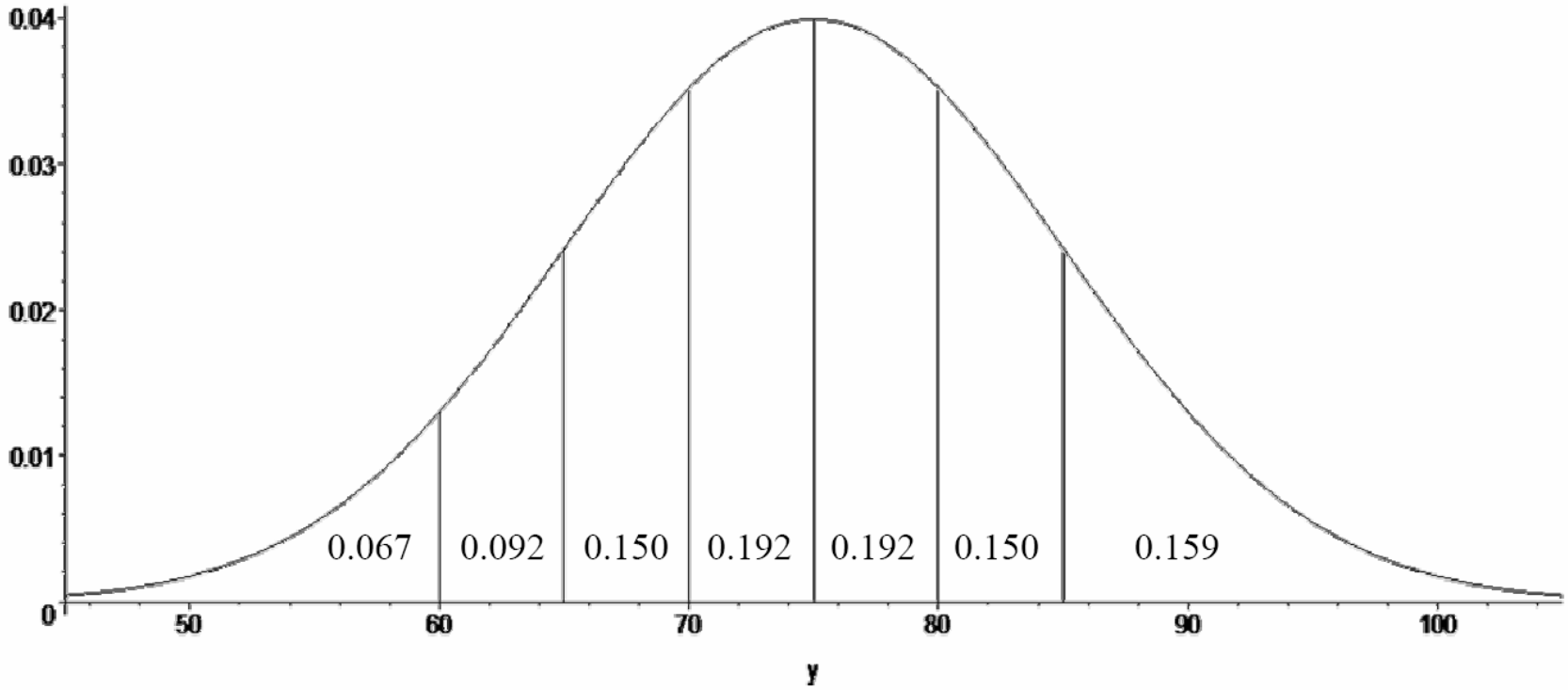
KJIKVADRAT-TEST FOR TESTING AV MODELL

Normalfordelingen N(75,10^2):



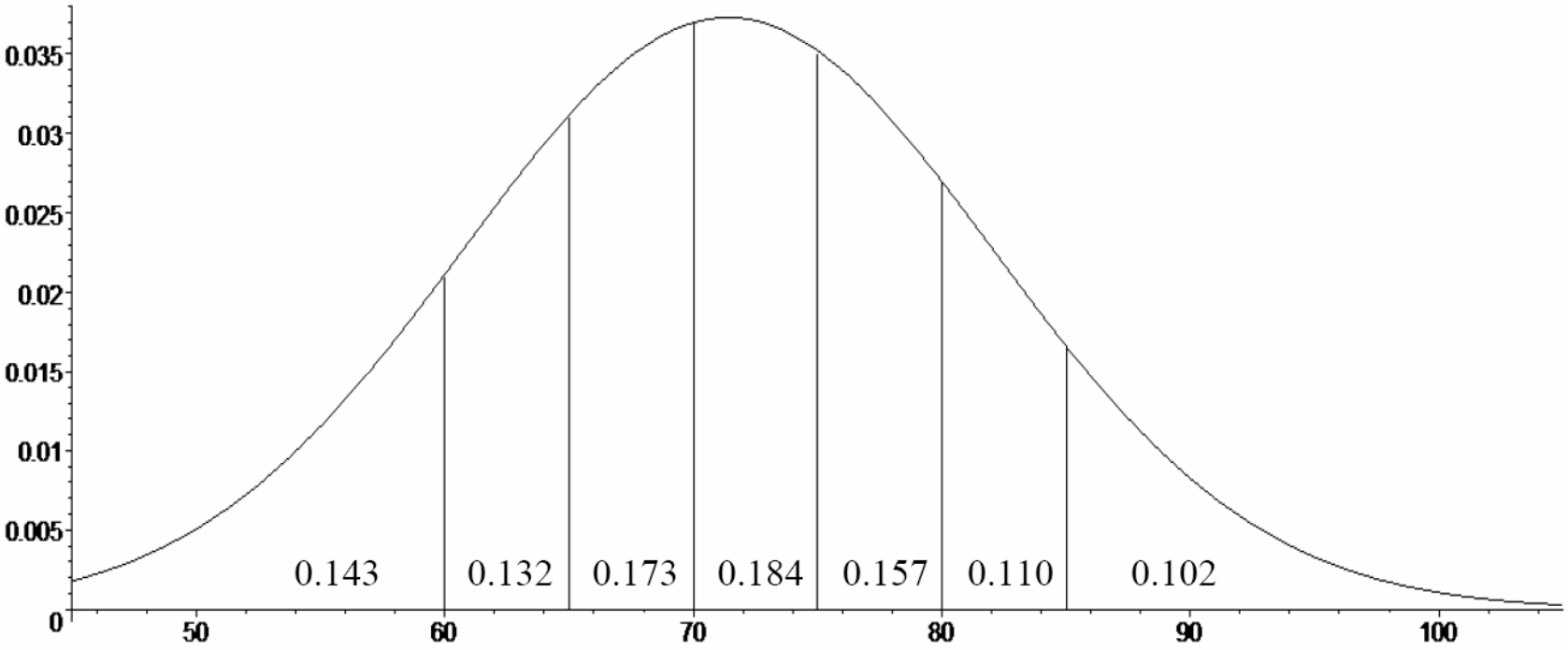
Forventet (n=60):	4.0	5.5	9.0	11.5	11.5	9.0	9.5
Observert	7	11	16	7	8	5	6

Kjikvadratobservator: 19.1

Fordeling under *H0: N(75,10^2)*: Kjikvadrat med df=7-1=6. Kritisk verdi ved 5%: 12.59

P-verdi: P(kjikvadr6 >= 19.1) = 0.004

Forventning og varians for Y estimert fra de 60 observasjonene viser seg å være 71.4 og 10.7^2. For å teste nullhypotesen *H0: Y er normalfordelt* brukes disse for å beregne kjivadrat-testobservatoren:



Forventet (n=60)	8.6	7.9	10.4	11.0	9.4	6.6	6.1
Observert	7	11	16	7	8	5	6

Kjikvadratobservator: 6.6

Fordeling under *H0: Y normalfordelt)*: Kjikvadrat med df=7-1-2=4. Kritisk verdi ved 5%: 9.49

P-verdi: P(kjikvadr4 >= 6.6) = 0.16

10.15 Test for independence

Eye color: 1=blue, 2=brown
Hair color: light blond, dark blond, dark

MINITAB:

Chi-Square Test: l.blond; d.blond; dark

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

	l.blond	d.blond	dark	Total
1	29	12	7	48
	23,40	14,40	10,20	
	1,340	0,400	1,004	
2	10	12	10	32
	15,60	9,60	6,80	
	2,010	0,600	1,506	
Total	39	24	17	80

Chi-Sq = 6,860; DF = 2; P-Value = 0,032

10.16 Test for homogeneity

Blood types: A, B, AB, O
Populations: 1, 2, 3

MINITAB:

Chi-Square Test: A; B; AB; O

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

	A	B	AB	O	Total
1	176	41	19	164	400
	192,00	37,14	16,57	154,29	
	1,333	0,401	0,356	0,612	
2	112	16	6	66	200
	96,00	18,57	8,29	77,14	
	2,667	0,356	0,631	1,610	
3	48	8	4	40	100
	48,00	9,29	4,14	38,57	
	0,000	0,178	0,005	0,053	
Total	336	65	29	270	700

Chi-Sq = 8,200; DF = 6; P-Value = 0,224
1 cells with expected counts less than 5.

Fra Gunnar Løvås: Statistikk for universiteter og høgskoler. Universitetsforlaget.

Simpson's paradox

Våre statistiske konklusjoner må alltid ledsages av sunn fornuft og god kunnskap om det fenomenet vi studerer. I noen tilfeller kan vi faktisk få helt feil bilde dersom vi utelater en viktig variabel, slik at den riktige sammenhengen er motsatt av den vi har oppdaget. Dette kalles Simpsons paradoks og illustreres med det følgende eksemplet. Slike feilkonklusjoner skyldes ofte at data fra ulike kilder (tabeller) er slått sammen i en tabell. Les eksemplet, og husk budskapet neste gang du uttaler deg på bakgrunn av statistiske undersøkelser.

Eksempel 258 *Det hadde vært interessant å sende ut et spørreskjema til bilførere for å kartlegge hvor mange som har fått en bilskade de siste årene. Resultatet av undersøkelsen kunne blitt slik:*

	Bilskade	*Ikke bilskade*	*Total*
Mann	233	323	556
Kvinne	87	194	281
Total	320	517	837

Her er det en signifikant forskjell på kjønnenes evne til å kjøre skadefritt. Andelen $233/556 = 0.42$ *av mennene var involvert i en skade, mens kvinnenes andel bare var* $87/281 = 0.31$. *Betyr dette at kvinner er dyktigere sjåfører enn menn?*

Løsning: Ved å undersøke spørreskjemaene nærmere kunne vi foreta en ekstra kategorisering, avhengig av hvor stor bil personene kjører. Da kunne resultatene presenteres slik:

	Store biler			*Små biler*		
	Skade	*Ikke skade*	*Total*	*Skade*	*Ikke skade*	*Total*
Mann	150	35	185	82	288	370
Kvinne	16	2	18	72	192	264
Total	166	37	203	154	480	634

Nå er bildet snudd på hodet. For store biler var 88 % av kvinnene mot 81 % av mennene involvert i en skade. For små biler var 27 % av kvinnene mot 22 % av mennene involvert. Kvinnene har høyest skadeandel både for store og små biler!