Något om binomialfördelningen

Binomialfördelningen är en av de mest vanliga statistiska modellerna när man diskuterar proportioner t.ex. felkvoter. Fördelningen innehåller två parametrar n och p där n är stickprovets storlek och p är felkvoten.
(Notera att felkvot får representera varje typ av kvot (andel), t.ex. andelen människor med en viss sjukdom, andelen bygglovsansökningar som inte är kompletta vid inlämning, etc.)

Teoretiskt medelvärde (μ) och standardavvikelse (σ) i en diskret fördelning beräknas med följande två formler:

$$\mu = \sum p_i x_i \qquad\qquad \sigma = \sqrt{\sum p_i x_i^2 - \mu^2}$$

Efter en smula enklare matematisk behandling erhåller följande resultat:

		μ	σ
1	Felkvot:	$\mu = p$	$\sigma = \sqrt{\dfrac{p(1-p)}{n}}$
2	*X* (antal felaktiga)	$\mu = np$	$\sigma = \sqrt{np(1-p)}$

Kommentar

1 ’Felkvot’

Det är inte förvånande att teoretiskt medelvärde blir just p. Resultatet uttrycks som ett tal i intervallet [0, 1] och om man vill ha det i procent måste det multipliceras med 100.
Det blir naturligtvis en variation mellan beräknar (skattningar) av ’p’ från flera stickprov. Formeln för denna variation (sigma) finns i tabellen till höger. Det framgår t.ex. att om sökas i nämnaren så minskar variationen. Detta är ett av flera sätt att minska sigma.

2 ’Antal felaktiga’

Om man noterar antal felaktiga blir uttrycken för teoretiskt medelvärde och sigma enligt formlerna i tabellen.
Ex.: om processens $p = 0.021$ och stickprov $n = 100$ blir teoretiskt medelvärde 2.1 felaktiga per stickprov.

Anm. Formlerna oven är egentligen ganska lätta att härleda med enkla matematiska metoder. Det tar bar några minuter…

Ett exempel

Diagrammet visar en binomialfördelning med n = 80 och p = 0.02. Varje stapel anger sannolikheten att få exakt X felaktiga i stickprovet.

De tre röda staplarna är 0.198, 0.324, 0.261 och summan av dessa svarar på frågan ’Vad är sannolikheten att få 2 eller färre fel i stickprovet?.

(Kanske 2 är en övre gräns av något slag.)