

Matematisk statistik för B, K, N, BME och Kemister

Föreläsning 8

Johan Lindström

20 september 2017

Mozquizto

Repetition

- Hypotesprövning
- Styrkefunktion
- Diskret data

Hypotesprövning för diskret data

- Exempel: Poisson & Binomial
- Metoder för hypotesprövning

Styrkefunktion för diskret data

- Exempel

Mozquizto

Repetition

- Hypotesprövning
- Styrkefunktion
- Diskret data

Hypotesprövning för diskret data

- Exempel: Poisson & Binomial
- Metoder för hypotesprövning

Styrkefunktion för diskret data

- Exempel

Mozquizto — Resultat

	Test 1		Test 2	
	Alla	> 0 rätt	Alla	> 0 rätt
Försök	412	333	528	452
Försök/student	2.69	2.18	3.45	2.95
Godkända tester	37.1%	46.0%	28.0%	33.7%

Mozquizto — Exempel

Alice kastar en tärning 11 gånger och får 1 poäng varje gång hon får en trea. Ville däremot kastar 13 gånger och får 1 poäng varje gång han får en sexa. Beräkna sannolikheten att deras totala poängsumma understiger 5. Svara med minst tre decimalers noggrannhet.

X_A : Antal poäng Alice får

X_V : Antal poäng Ville får

Mozquizto — Exempel

En forskare mäter högsta utandningsflöde hos 20-åriga friska män med en peakflowmeter. Han antar att kapaciteten följer en normalfördelning med väntevärde $\mu = 545$ l/min och standardavvikelse $\sigma = 14$ l/min.

- Mellan vilka värden finns, enligt modellen, i stort sett samtliga maximala utandningsflödena hos 20-åriga friska män?
 - ▶ Mellan 713 och 713 l/min
 - ▶ Mellan 503 och 587 l/min
 - ▶ Mellan 531 och 559 l/min
- Vilket högsta utandningsflöde överstigs av 90% av männen?

X : utandningsflöde hos en godtycklig man.

Mozquizto — Exempel

Tiden det tar att betjäna en kund vid station A är en stokastisk variabel med väntevärde 5.4 minuter och standardavvikelse 4 minuter. Vid station B tar det i genomsnitt 3.7 minuter att betjäna en kund och standardavvikelsen är 1.75.

- b) Beräkna sannolikheten att det går snabbare att betjäna 50 kunder vid A än 80 kunder vid B.

A_i : Tid för kund i , station A

B_i : Tid för kund i , station B

$$S_A = \sum_{i=1}^{50} A_i : \text{Tid för 50 kunder, station A}$$

$$S_B = \sum_{i=1}^{80} B_i : \text{Tid för 80 kunder, station B}$$

Mozquizto

Repetition

Hypotesprövning

Styrkefunktion

Diskret data

Hypotesprövning för diskret data

Exempel: Poisson & Binomial

Metoder för hypotesprövning

Styrkefunktion för diskret data

Exempel

Hypotesprövning (Kap. 9.1.3 & 9.2.3 & 9.3.3)

H_0 förkastas om observationerna, ϑ^* , avviker för mycket från nollhypotesen ϑ_0 .

Testa **nollhypotesen** $H_0: \vartheta = \vartheta_0$

mot **mothypotesen** (tex) $H_1: \vartheta \neq \vartheta_0$

på nivån α , **felrisken** α ges av

$$\alpha = P(H_0 \text{ förkastas trots att den är sann})$$

Olika metoder för att utföra hypotestest

- 1. Direktmetoden eller P-värde**
 - ▶ Antag att H_0 är sann
 - ▶ Räkna ut **P-värdet** $p = P(\text{Få det vi fått eller värre})$
 - ▶ Om $p < \alpha$ förkastas H_0
- 2. Konfidensmetoden.** Gör ett konfidensintervall med konfidensgraden $1 - \alpha$ och förkasta H_0 på nivån α om intervallet ej täcker ϑ_0 . Intervallen skall, beroende på H_1 , vara

Test	$H_1: \vartheta < \vartheta_0$	$H_1: \vartheta \neq \vartheta_0$	$H_1: \vartheta > \vartheta_0$
Intervall:	uppåt begr	tvåsidigt	nedåt begr
- 3. Testkvantitet $T(X)$ och kritiskt område C** Förkasta H_0 om testkvantiteten hamnar i det kritiska området.
 C och T skall väljas så att

$$\alpha = P(T(X) \in C) = P(\text{"Förkasta } H_0 \text{ om } H_0 \text{ är sann"})$$

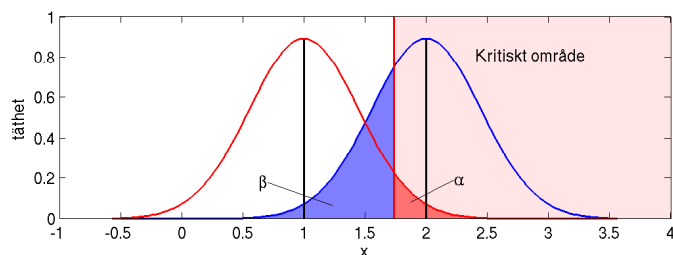
Styrkefunktion (Kap. 7.6)

Användas för att avgöra hur bra testet skiljer H_0 från H_1 .

$$h(\vartheta) = P(\text{"Förkasta } H_0 \text{ om } \vartheta \text{ är rätt värde"})$$

Typ 1 fel: $\alpha = P(H_0 \text{ förkastas om } H_0 \text{ sann})$

Typ 2 fel: $\beta = P(H_0 \text{ förkastas ej om } H_0 \text{ ej sann})$



Konfidensintervall (Kap. 9.1.2 & 9.2.2 & 9.3.2)

Ett **konfidensintervall** för en parameter ϑ täcker rätt värde på ϑ med sannolikheten $1 - \alpha$.

$1 - \alpha$ kallas **konfidensgrad**. Vanliga värden är 0.95, 0.99 och 0.999.

Normalfördelad skattning, $\vartheta^* \in N(\vartheta, V(\vartheta^*))$

$D(\vartheta^*)$ **känd**: $I_\vartheta = \vartheta^* \pm \lambda_{\alpha/2} D(\vartheta^*)$

$D(\vartheta^*)$ **okänd**: $I_\vartheta = \vartheta^* \pm t_{\alpha/2}(f) d(\vartheta^*)$

Normalapproximation, $\vartheta^* \in N(\vartheta, V(\vartheta^*))$ (Ex: CGS)

$D(\vartheta^*)$ **känd**: $I_\vartheta = \vartheta^* \pm \lambda_{\alpha/2} D(\vartheta^*)$

$D(\vartheta^*)$ **okänd**: $I_\vartheta = \vartheta^* \pm \lambda_{\alpha/2} d(\vartheta^*)$ (alltid λ -kvantil)

Mozquizto

Repetition

Hypotesprövning
Styrkefunktion
Diskret data

Hypotesprövning för diskret data

Exempel: Poisson & Binomial
Metoder för hypotesprövning

Styrkefunktion för diskret data

Exempel

Exempel — Dreamliner

NTSB Interim Factual Report (March 7, 2013)

Boeing also determined that the probability that a battery could vent was **once in every 10 million flight hours**. As of January 16, 2013, the in-service 787 fleet had accumulated **less than 52 000 flight hours**, and during this period **two events** involving smoke emission from a 787 battery had occurred . . .



Antag att antalet fel per flygtimme är oberoende poisson.

- ▶ Vad är fördelningen för antalet fel under 52 000 flygtimmar?
- ▶ Undersök, på nivån $\alpha = 0.001$, om två fel är oväntat många?

Exempel — Blir allvarliga stormar vanligare?

Under de 10 åren 1975–1984 förekom **53** orkaner i nord Atlanten, av dessa var **18 (33.96%)** allvarlig, d.v.s. styrka 3+. Motsvarande siffor för de senaste 10 åren, 2007–2016, är **65** orkaner varav **28 (43.04%)** allvarliga.

Undersök, på signifikansnivå $\alpha = 0.05$, om andelen allvarliga orkaner har ökat.

Hypotestest – Vilken metod?

- ▶ Normalfördelad skattning.
 - σ känd: Vilken som helst.
 - σ okänd: Direktmetoden kräver t -fördelningens fördelningsfunktion.
- ▶ Fördelning där $\mu^* = \bar{X} \in N(\mu, V(\mu^*)) \dots$ enl. CGS.
 - ▶ Vilken som helst
- ▶ Bin, Po, ... där $D(\vartheta^*)$ innehåller ϑ .
 - Direktmetoden **Går alltid** att använda, ibland med normalapproximation.
 - Testkvantitet Kräver normalt **normalapproximation**.

Vid **styrkefunktion** är det naturligt att **utgå från testkvantitet**.

Testkvantiter

Antag att vi vill testa $H_0: \vartheta = \vartheta_0$.

Model	Skattning	$T(X)$	$D(\vartheta^*)/d(\vartheta^*)$	kvantil	
$X_i \in N(\mu, \sigma^2)$	σ känd	$\mu^* = \bar{X}$	$\frac{\mu^* - \mu_0}{D(\mu^*)}$	$\frac{\sigma}{\sqrt{n}}$	λ
	σ okänd		$\frac{\mu^* - \mu_0}{d(\mu^*)}$	$\frac{s}{\sqrt{n}}$	$t(f)$
$X \in \text{Bin}(n, p)$	$p^* = \frac{X}{n}$	$\frac{p^* - p_0}{D_0(p^*)}$	$\sqrt{\frac{p_0(1-p_0)}{n}}$	λ	
$X_i \in \text{Po}(\mu)$	$\mu^* = \bar{X}$	$\frac{\mu^* - \mu_0}{D_0(\mu^*)}$	$\sqrt{\frac{\mu_0}{n}}$	λ	

Notera:

1. Standardavvikelse/medelfel **räknas under H_0** .
2. Bin och Po fallet kräver **normalapproximation**.
3. α -kvantil om **ensidigt**, $\alpha/2$ -kvantil om **tvåsidigt**.

Mozquizto

Repetition

Hypotesprövning
Styrkefunktion
Diskret data

Hypotesprövning för diskret data

Exempel: Poisson & Binomial
Metoder för hypotesprövning

Styrkefunktion för diskret data

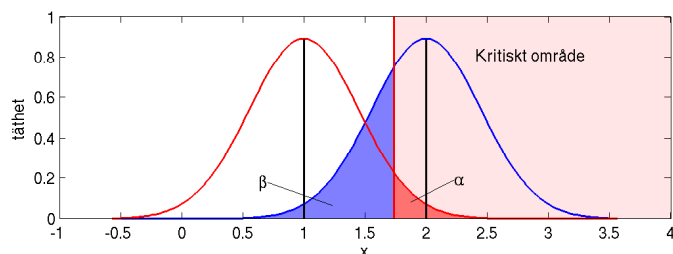
Exempel

Styrkefunktion

(Kap. 7.6)

Användas för att avgöra hur bra testet skiljer H_0 från H_1 .

$$h(\vartheta) = P(\text{"Förkasta } H_0 \text{ om } \vartheta \text{ är rätt värde"})$$

Typ 1 fel: $\alpha = P(H_0 \text{ förkastas om } H_0 \text{ sann})$ Typ 2 fel: $\beta = P(H_0 \text{ förkastas ej om } H_0 \text{ ej sann})$ 

Exempel — Styrkefunktion för diskret data

Antag att antalet orkaner per år i nord Atlanten kan modelleras som en Poissonfördelning med väntevärde $\mu = 5.5$ orkaner/år (baserat på historisk data). Under de senaste 5 åren, 2012–2016, noterades 29 orkaner i nord Atlanten.

1. Vilken fördelning kan antalet orkaner under 5 år antas följa?
2. Testa på nivån 5% om de 29 orkanerna under 5 år tyder på ett ökat antal stormar?
3. Under hur många år måste vi räkna orkaner för att, med 99% sannolikhet, upptäcka en 50% ökning i förväntat antalet orkaner per år (d.v.s en ökning från $\mu = 5.5$ till $\mu = 5.5 \cdot 1.5 = 8.25$)?