Tentamen i Dataanalys och statistik för I den 28 okt 2015

Tentamen består av åtta uppgifter om totalt 50 poäng. Det krävs minst 20 poäng för betyg 3, minst 30 poäng för 4 och minst 40 för 5.
Examinator: Ulla Blomqvist
Hjälpmedel: Chalmersgodkänd miniräknare, Matematisk statistik av Ulla Dahlbom och Håkan Blomqvists formelsamling. **Boken eller formelsamlingen får inte innehålla egna anteckningar.**
Jour: Besöker tentamen c:a kl 10.00

Lycka till!

Uppgift 1: Anta att man har två händelser A och B. För dessa gäller att P(A) = 0.05, P(B) = 0.10 och $P(A^C \cup B^C) = 0.97$. Beräkna P(B | A).

(6 poäng)

Uppgift 2: Anta att antal akutfall på ett sjukhus mellan kl 00.00 och 02.00 en lördagsnatt är poissonfördelat med $\lambda = 3.5$.

a) Vad är sannolikheten att det under en slumpmässigt vald lördagsnatt inte kommer något akutfall till sjukhuset mellan dessa tider?
b) Vad är sannolikheten att det går kortare tid mellan två akutfall än en halvtimme?
c) Anta att man studerar fyra lördagsnätter mellan kl 00.00 och 02.00. Vad är sannolikheten att det vid minst tre av tillfällena kommer exakt ett akutfall under den första timmen?

(8 poäng)

Uppgift 3: Den stokastiska variabeln ξ har följande frekvensfunktion:

$$f(x) = \begin{cases} Cx & 0 \le x \le 1 \\ 2 - x & 1 < x \le 2 \\ 0 & \text{annars} \end{cases}$$

a) Beräkna konstanten C.
b) Beräkna variansen för ξ.
c) Beräkna medianen för ξ.

(6 poäng)

Uppgift 4: Anta att restiden till arbetet för en viss person kan beskrivas av en normalfördelning med $\mu = 45$ min och $\sigma = 10$ min. Under en 2-veckors period gör denne person 10 resor till arbetet. Vad är sannolikheten att han vid högst ett tillfälle får en restid som överstiger 55 min?

(6 poäng)

Uppgift 5: Ett föremål vägs en gång på en våg A och en gång på våg B. Kalla utslagen för x_A och x_B. Motsvarande stokastiska variabler ξ_A och ξ_B har båda väntevärdet lika med det vägdas vikt, men ξ_A har bara hälften så stor standardavvikelse som ξ_B. Vilken av nedanstående förslag är bäst om man vill skatta den sanna vikten?

a) $\frac{2x_A}{4} + \frac{x_B}{4}$ b) x_A

c) $\frac{4x_A}{5} + \frac{x_B}{5}$ d) $\frac{2x_A}{3} + \frac{x_B}{3}$

(6 poäng)

Uppgift 6: Två olika institut tänker göra var sin opinionsundersökning angående göteborgarnas (över 18 år) inställning till byggandet av Västlänken. Det ena institutet tänker fråga 1500 personer medan det andra anser att det räcker med 1000 personer. Anta att 74% av göteborgarna över 18 år är emot byggandet av Västlänken. Vad är sannolikheten att resultaten i de båda undersökningarna skiljer sig åt med mer än 5%?

(6 poäng)

Uppgift 7: Vid tillverkning av exponeringsmätare till kameror har man under en längre tid vid slutkontrollen funnit att 7% av mätarna måste justeras innan de kan säljas. Man ändrar därför vissa moment i tillverkningen och finner att av 500 slumpmässigt utvalda mätare behöver 25 justeras. Formulera lämplig nollhypotes och mothypotes för att kontrollera om ändringen har medfört någon förbättring. Genomför testet på 5%:s signifikansnivå.

(6 poäng)

Uppgift 8: Ett visst försök kan lyckas eller misslyckas. Man vill undersöka om antalet lyckade försök i en serie om 5 är binomialfördelat. Man gör därför 200 oberoende serier med 5 försök i varje serie och noterar antal lyckade försök i varje serie.

Antal lyckade försök	0	1	2	3	4	5
Antal serier	5	34	60	65	29	7

Genomför ett test på 5%:s signifikansnivå och avgör om resultatet motsäger att antalet lyckade försök är binomialfördelat.

(6 poäng)

Lösningar till Dataanalys och statistik 201510028

Uppgift 1: $P(A) = 0.05$ $P(B) = 0.10$ $P(A^C \cup B^C) = 0.97$.

Använd de Morgans sats $P(A \cap B) = 1 - P(A^C \cup B^C) = 1 - 0.97 = 0.03$

	A	**A^C**	
B	0.03	0.07	**0.10**
B^C	0.02	0.88	**0.90**
	0.05	**0.95**	**1.00**

$$P(B \mid A). = \frac{P(A \cap B)}{P(A)} = \frac{0.03}{0.05} = 0.60$$

Uppgift 2: ξ = antal akutfall $\xi = Po(\lambda = 3.5 \text{ akutfall/2 tim})$

a) $P(\xi = 0) = e^{-3.5} \cdot \dfrac{3.5^0}{0!} \approx 0.0302$

b) η = tiden mellan två akutfall
Räkna om λ till antal akutfall/ tim. $\lambda = 1.75$ akutfall/ tim
$\eta = Exp(\lambda = 1.75 \text{ akutfall/tim})$
$P(\xi < 0.5) = 1 - e^{-1.75 \cdot 0.5} \approx 0.583$

c) ζ = antal tillfällen det kommer exakt ett akutfall under den första timmen
$\zeta = Bin(n, p) = Bin(4, p)$ där p beräknas enligt

$$p = P(\xi = 1) = e^{-1.75} \cdot \frac{1.75^1}{1!} = 0.304$$

$$P(\zeta \geq 3) = P(\zeta = 3) + P(\zeta = 4) = \binom{4}{3} 0.304^3 \cdot (1-0.304)^1 +$$

$$+ \binom{4}{4} 0.304^4 \cdot (1-0.304)^0 = 0.0867$$

Uppgift 3:

$$f(x) = \begin{cases} Cx & 0 \leq x \leq 1 \\ 2-x & 1 < x \leq 2 \\ 0 & \text{annars} \end{cases}$$

Fortsättning uppgift 3 på nästa sida

Fortsättning uppgift 3

a) Eftersom f(x) är en frekvensfunktion så gäller att

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \quad \Rightarrow \quad \int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 Cx\,dx + \int_1^2 (2-x)\,dx =$$

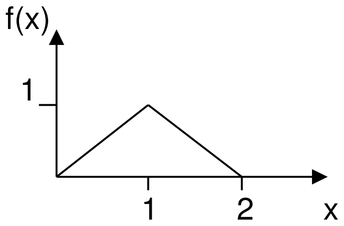
$$= \left[C\frac{x^2}{2}\right]_0^1 + \left[2x - \frac{x^2}{2}\right]_1^2 = 0.5\,C + 0.5 = 1 \quad \Rightarrow \quad C = 1$$

b) $\mathrm{Var}\,(\xi) = \int_{-\infty}^{\infty} x^2 f(x)\,dx - [E(\xi)]^2$ E (ξ) beräknas först

$$E\,(\xi) = \int_{-\infty}^{\infty} x\,f(x)\,dx = \int_0^1 x^2\,dx + \int_1^2 x(2-x)\,dx = \left[\frac{x^3}{3}\right]_0^1 + \left[\frac{2x^2}{2} - \frac{x^3}{3}\right]_1^2 = 1$$

$$\mathrm{Var}\,(\xi) = \int_0^1 x^3\,dx + \int_1^2 x^2(2-x)\,dx - 1^2 = \left[\frac{x^4}{4}\right]_0^1 + \left[\frac{2x^3}{3} - \frac{x^4}{4}\right]_1^2 - 1 = \frac{1}{6}$$

c) $P(\xi < Md) = 0.5$



Av figuren framgår att medianen är 1 men beräkningar genomförs för att få detta bekräftat

$P(\xi \le 1) = \int_0^1 x\,dx = 0.5$ Md = medianen = 1

Uppgift 4: : ξ = restiden ξ = N (μ, σ) = N(45, 10) min

$$P(\xi > 55) = 1 - P\left(Z < \frac{55-45}{10}\right) = 1 - \Phi(1) \approx 1 - 0.8413 = 0.1587$$

η = antal resor där restiden överstiger 55 min
η = Bin(n, p) = Bin(10, 0.1587)

$$P(\eta \le 1) = P(\eta = 0) + P(\eta = 1) = \binom{10}{0} 0.1587^0 \cdot (1-0.1587)^{10} +$$

$$+ \binom{10}{1} 0.1587^1 \cdot (1-0.1587)^9 = 0.5127$$

Uppgift 5: $S(\xi_A) = \dfrac{\sigma}{2}$ och $S(\xi_B) = \sigma \Rightarrow Var(\xi_A) = \dfrac{\sigma^2}{4}$ och $Var(\xi_B) = \sigma^2$

Vi söker den sammanvägningen där summan av vikterna = 1 och variansen är minst

a) $\dfrac{2x_A}{4} + \dfrac{x_B}{4}$ Summan av vikterna: $\dfrac{2}{4} + \dfrac{1}{4} = \dfrac{3}{4}$

b) x_A Summan av vikterna: 1

Variansen: $Var(\xi_A) = \dfrac{\sigma^2}{4} = 0.25\sigma^2$

c) $\dfrac{4x_A}{5} + \dfrac{x_B}{5}$ Summan av vikterna: $\dfrac{4}{5} + \dfrac{1}{5} = 1$

Variansen: $Var(\dfrac{4\xi_A}{5} + \dfrac{\xi_B}{5}) = \dfrac{16}{25} Var(\xi_A) + \dfrac{1}{25} Var(\xi_B) =$

$= \dfrac{16}{25} \cdot \dfrac{\sigma^2}{4} + \dfrac{1}{25}\sigma^2 = 0.2\,\sigma^2$

d) $\dfrac{2x_A}{3} + \dfrac{x_B}{3}$ Summan av vikterna: $\dfrac{2}{3} + \dfrac{1}{3} = 1$

Variansen: $Var(\dfrac{2\xi_A}{3} + \dfrac{\xi_B}{3}) = \dfrac{4}{9} Var(\xi_A) + \dfrac{1}{9} Var(\xi_B) =$

$= \dfrac{4}{9} \cdot \dfrac{\sigma^2}{4} + \dfrac{1}{9}\sigma^2 = 0.22\,\sigma^2$

Välj alternativ c) dvs $\dfrac{4x_A}{5} + \dfrac{x_B}{5}$

Uppgift 6: ξ_1 = antalet som är emot i undersökning 1 $\xi_1 = Bin(1500, 0.74)$
ξ_2 = antalet som är emot i undersökning 1 $\xi_2 = Bin(1000,0.74)$

$\hat{p}_x = \dfrac{\xi_x}{n_x}$ är skattningen av p i undersökning x där $E(\hat{p}_x) = p$ och $Var(\hat{p}_x) = \dfrac{p(1-p)}{n_x}$

$\eta = \hat{p}_A - \hat{p}_B$ alt $\eta = \hat{p}_B - \hat{p}_A$ är skillnaden mellan undersökningsresultaten

$E(\eta) = E(\hat{p}_A - \hat{p}_B) = E(\hat{p}_A) - E(\hat{p}_B) = p - p = 0$

$Var(\eta) = Var(\hat{p}_A - \hat{p}_B) = Var(\hat{p}_A) + Var(\hat{p}_B) = \dfrac{0.74 \cdot 0.26}{1500} + \dfrac{0.74 \cdot 0.26}{1000} =$

$= \dfrac{5 \cdot (0.74 \cdot 0.26)}{3000}$

Eftersom n_1 och n_2 båda är större än 100 $\Rightarrow$ använd normalapproximation

Fortsättning uppgift 6 på nästa sida

Fortsättning uppgift 6

P(skillnaden är större än 5%) = $P(|\eta| > 0.05) = P(\eta < -0.05) + P(\eta > 0.05) =$

$$2\,(1 - P(\eta < 0.05)) = 2\,(1 - P(Z < \frac{0.05-0}{\sqrt{\frac{5\cdot 0.74\cdot 0.26}{3000}}}) = 2\,(1-\Phi(2.79)) = 0.0052$$

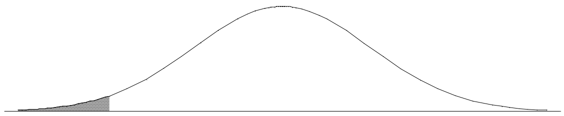
Uppgift 7:

Steg 1: H_0: $p \geq 0.07$

H_1: $p < 0.07$

$n \cdot p\,(1-p) = 0.07 \cdot 0.93 \cdot 500 > 10 \Rightarrow$ använd normalapproximation

Steg 2: $\alpha = 5\%$



-1.645

Steg 3: Välj testvariabeln $Z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$

Steg 4: Urval gav $\hat{p} = \dfrac{25}{500} = 0.05$

$$Z = \frac{0.05 - 0.07}{\sqrt{\frac{0.07\cdot 0.93}{500}}} \approx -1.75$$

Steg 5: H_0 förkastas. Denna undersökning tyder på att färre behöver justeras vilket tyder på att förändringen har gjort att kvaliteten har blivit bättre.

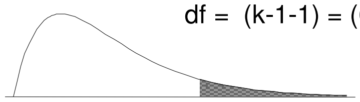
Uppgift 8:

Steg 1: H_0: binomialfördelat

H_1: <u>inte</u> binomialfördelat

df = (k-1-1) = (6-1-1) = 4.

Steg 2: $\alpha = 0.05$



χ^C

χ^2-tabellen ger att det kritiska värdet är 9.488

Steg 3: Välj testvariabeln $\chi^2 = \sum_{i=1}^{6} \dfrac{(O_i - E_i)^2}{E_i}$

Fortsättning uppgift 8 på nästa sida

Fortsättning uppgift 8

Steg 4: Vi beräknar en skattning av p från de observerade värdena.

$$\hat{p} = \frac{1}{5 \cdot 200}(0 \cdot 5 + 1 \cdot 34 + 2 \cdot 60 + 3 \cdot 65 + 4 \cdot 29 + 5 \cdot 7) = 0.5$$

Vi börjar med att beräkna sannolikheterna i en binomialfördelning med p = 0.5. Därefter beräknar vi de förväntade värdena, E_i.

$\xi = x$	$P(\xi = x)$	Förväntat antal
0	$\binom{5}{0} 0.5^0 \cdot 0.5^5 = 0.03125$	$200 \cdot 0.03125 = 6.25$
1	$\binom{5}{1} 0.5^1 \cdot 0.5^4 = 0.15625$	$200 \cdot 0.15625 = 31.25$
2	$\binom{5}{2} 0.5^2 \cdot 0.5^3 = 0.3125$	$200 \cdot 0.3125 = 62.5$
3	$\binom{5}{3} 0.5^3 \cdot 0.5^2 = 0.3125$	$200 \cdot 0.3125 = 62.5$
4	$\binom{5}{4} 0.5^4 \cdot 0.5^1 = 0.15625$	$200 \cdot 0.15625 = 31.25$
5	$\binom{5}{5} 0.5^5 \cdot 0.5^0 = 0.03125$	$200 \cdot 0.03125 = 6.25$

Vi ställer nu upp de observerade och de förväntade värdena i en tabell.

	0	1	2	3	4	5	Totalt
Obs antal, O_i	5	34	60	65	29	7	200
Förväntat antal, E_i	6.25	31.25	62.5	62.5	31.25	6.25	200

$$\chi^2 = \frac{(5-6.25)^2}{6.25} + \frac{(34-31.25)^2}{31.25} + \frac{(60-62.5)^2}{62.5} + \frac{(65-62.5)^2}{62.5} +$$

$$+ \frac{(29-31.25)^2}{31.25} + \frac{(7-6.25)^2}{6.25} = 0.944 < 9.49$$

Värdet hamnar i acceptansområdet.

Steg 5: H_0 kan inte förkastas. Undersökningen motsäger inte hypotesen att antalet lyckade försök skulle kunna vara binomialfördelat