

Industriell matematik och statistik, LMA136 2013/14

7 Mars 2014

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Disposition

Föreläsning 11

Repetition

Repetition

Beroendestrukturer

Beroendestrukturer

Regression

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Duggor

Dugga 1

Dugga 2

Dugga 3

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

► Konfidsensintervall

- Statistika Z (eller T) har fördelning F (Z en funktion av $\hat{\theta}$ och θ)
- $q_{1-\alpha/2}$ är $(1 - \alpha/2)$ -kvantil för fördelningen F .
 $P(-q_{1-\alpha/2} < Z < q_{1-\alpha/2}) = 1 - \alpha$
- Lös ut den sökta parametern ur olikheterna
- $(1-\alpha)$ -konfidsensintervallet för θ defineras av det resulterande intervallet.

► $(1-\alpha)$ -konfidsensintervall för väntevärdesparameter ges av

$$(\bar{x} - q_{1-\alpha/2}S, \bar{x} + q_{1-\alpha/2}S)$$

då $S = \sigma/\sqrt{n}$ om standardavvikelsen känd och
 $S = s/\sqrt{n}$ då σ okänd.

Konfidensintervall och hypotestest

- ▶ Hypotestest; Objektivt dra slutsats om data
 - ▶ Nollhypotes, H_0 , samt mothypotes, H_1 .
 - ▶ Signifikansnivå $\alpha = P(\text{Förkasta } H_0 \text{ sann} | H_0 \text{ sann})$
 - ▶ Testvariabel (statistikan) Z , T , med tillhörande kvantiler q för förkastelseområdet.
 - ▶ Beräkna observerade testvariabeln Z_{Obs} , T_{Obs}
 - ▶ Förkasta H_0 (och acceptera H_1) om observerade testvariabeln inom förkastelseområdet, annars acceptera H_0 .

Testets p-värde (beror av mothypotesen)

Mothypotes	p-värde
$H_1 : \mu \neq \mu_0$	$P(Z > Z_{Obs} H_0 \text{ sann})$
$H_1 : \mu > \mu_0$	$P(Z > Z_{Obs} H_0 \text{ sann})$
$H_1 : \mu < \mu_0$	$P(Z < Z_{Obs} H_0 \text{ sann})$

- ▶ Hypotestest med $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$
 Testregel: förkasta H_0 på signifikansnivå α om $(1-\alpha)$ -konfidensintervallet inte innehåller μ_0 .

Repetition

Beroendestruktur

Regression

Duggor

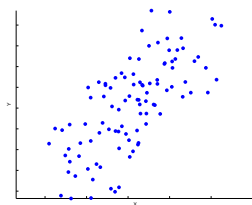
Dugga 1

Dugga 2

Dugga 3

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

- ▶ Nederbörd - Vattenskador
- ▶ Trafikhastighet - Personskador
- ▶ Bokslut -Aktiekurs



Figur: Bivariat data

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Beroenden

Föreläsning 11

Repetition

Beroendestruktur

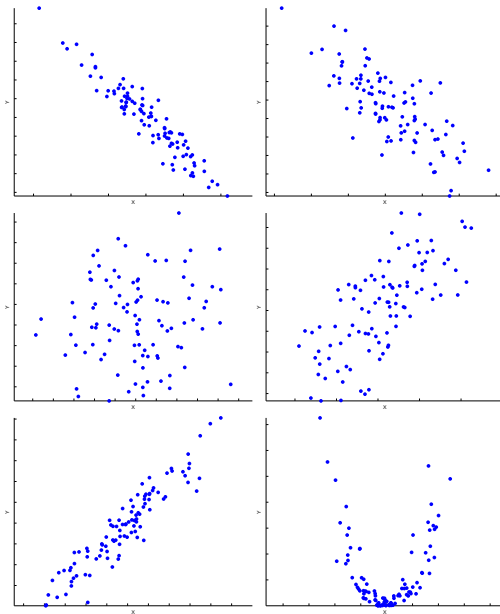
Regression

Duggor

Dugga 1

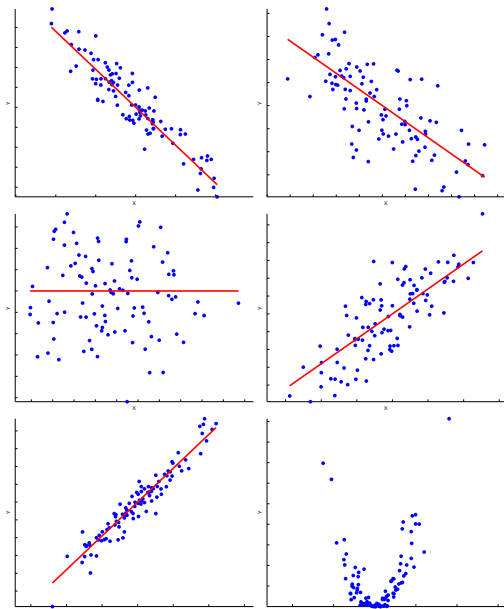
Dugga 2

Dugga 3



Linjärt beroende

$$y = bx + a$$



Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Linjär regression

- ▶ finna a, b som bäst beskriver relationen $y = bx + a$
- ▶ Skatta a, b
- ▶ låt $\hat{y}(x) = bx + a$

Kvadratiske avståndet

$$Q = \sum_{i=1}^n (y_i - \hat{y}(x_i))^2 = \sum_{i=1}^n (y_i - (bx_i + a))^2$$

$$\frac{dQ}{da} = -2 \sum_{i=1}^n (y_i - bx_i - a) = 0$$

$$\frac{dQ}{db} = -2 \sum_{i=1}^n (y_i - bx_i - a)x_i = 0$$

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Forts. linjär regression

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Om

- ▶ $\bar{x^2} = \frac{1}{n} \sum_{i=1}^n x_i^2$
- ▶ $\bar{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i$

$$\begin{aligned} b\bar{x} + a &= \bar{y}, \\ b\bar{x^2} + a\bar{x} &= \bar{xy} \end{aligned}$$

dvs

- ▶ $a = \bar{y} - b\bar{x}$
- ▶ $b = \frac{\bar{xy} - \bar{y}\bar{x}}{\bar{x^2} - \bar{x}^2}$

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Mått på hur starkt det linjärt beroendet

$$r = \frac{\bar{x}\bar{y} - \bar{\bar{x}}\bar{\bar{y}}}{\sqrt{(\bar{x}^2 - \bar{\bar{x}}^2)(\bar{y}^2 - \bar{\bar{y}}^2)}}$$

Förklaringsgrad $R^2 = r^2$

Multivariat linjär regression

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Data $\{(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_n, y_n, z_n)\}$

Ansätt funktion $\hat{z}(x, y) = bx + cy + a$

Kvadratiske avståndet;

$$Q = \sum_{i=1}^n (z_i - \hat{z}(x_i, y_i))^2 = \sum_{i=1}^n (z_i - (bx_i + cy_i + a))^2$$

Minsta kvadratiske avståndet;

$\frac{dQ}{da} = 0, \frac{dQ}{db} = 0, \frac{dQ}{dc} = 0$ lös ekvationssystemet

Generell minsta kvadrat-regression

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Ansätt funktion en passande funktion t.ex.;

► $\hat{y}(x) = bx^2 + a$

► $\hat{y}(x) = e^{bx} + a$

Minsta kvadratiske avståndet;

$$Q = \sum_{i=1}^n (y_i - \hat{y}(x_i))^2$$

$\frac{dQ}{da} = 0, \frac{dQ}{db} = 0$ lös ekvationssystemet

Antag att vi har två händelser A och B , där $P(A) = 0.48$, $P(B) = 0.42$ och $P(A \cup B) = 0.55$.

- ▶ a) $P(A^c)$
- ▶ b) $P(A \cap B)$
- ▶ c) $P(A \cup B)$
- ▶ d) $P(A \cap B^c)$
- ▶ e) $P(A^c \cup B^c)$
- ▶ f) $P(A^c \cap B^c)$
- ▶ g) Är händelserna A och B oberoende?

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Ett brev levereras på 1, 2, 3 eller 4 dagar med respektive sannolikhet 0.28, 0.31, 0.24, 0.17.

- ▶ a) Vad är risken att det tar mer än en dag att leverera brevet?
- ▶ b) Vad är den förväntade leveranstiden?
- ▶ c) Vad är variansen på leveranstiden?

En elev, David, går i klass 9A. David har 24 klasskamrater. Elevernas födelsedagar antas vara oberoende samt likformigt fördelade över året.

- ▶ a) Matilda går i samma klass, vad är chansen att Matilda fyller år samma dag som David?
- ▶ b) Kim går också i samma klass, vad är chansen att ingen av de tre; David, Matilda och Kim har födelsedag samma dag?
- ▶ c) Vad är chansen att ingen i klassen har födelsedag på samma dag?

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

På en telefonkundtjänst beräknas betjäningstiden X modelleras som en kontinuerlig stokastisk variabel med frekvensfunktion $f(x) = \frac{c}{8}x^{c-1}$ om x ligger i intervallet $(0,2)$ minuter noll om x ligger utanför detta intervall.

- ▶ a) Vad är konstanten c för att $f(x)$ är en frekvensfunktion?
- ▶ b) Vad är den förväntade betjäningstiden?
- ▶ c) Vad är variansen av betjäningstiden?
- ▶ d) Om inbördes betjäningstider är oberoende och det ett år inkommer 10005 samtal till kundtjänsten, vad är sannolikheten att den totala betjäningstiden överstiger 250 timmar (15000 minuter)?

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Q2

Maja spelar Blackjack, hon spelar alltid med insatsen 100kr/spel och är ganska duktig. Hon beräknar att hon har en vinstchans på $p=0.47$ i varje enskild spelomgång. Vid vinst utbetalas 2 ggr insatsen.

- ▶ a) Beräkna förväntade antal spelomgångar som Maja vinner om hon deltar i 10 spel.
- ▶ b) Beräkna Majas förväntade resultat i kr efter 10 spelomgångar (Kom ihåg att det kostar 100kr/spel att delta).
- ▶ c) För att öka intresset för Black Jack på casinot introduceras en bonus på 50kr för spelare som deltar i 7 efterföljande spel.
Beräkna Majas förväntade resultat i kr om hon deltar i 7 spel och erhåller bonusen.
- ▶ d) Vad är det minsta antal spel som casinot kan erbjuda Maja en bonus om 50kr för att delta i och samtidigt ha en positiv förväntad vinst? (dvs. Maja har ett negativt förväntat resultat)

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Svenska kvinnors kroppslängd kan anses vara normalfördelad med väntevärde 166.6 cm och standardavvikelse 6.6 cm.

- ▶ a) Hur stor är sannolikheten att en slumpvis utvald svensk kvinna är mellan 160 och 173.2 cm lång?
- ▶ b) Hur stor är sannolikheten att en slumpvis utvald svensk kvinna är mellan 153.4 och 179.8 cm lång?
- ▶ c) Hur stor är sannolikheten att en slumpvis utvald svensk kvinna är kortare än 166.6 cm?

Vid en opinionsundersökning uppskattas stödet för Piratpartiet. I ett stickprov om 4505 människor har 165 uttryckt sitt stöd för partiet. Beräkna konfidensintervall på nivåerna 95% respektive 99% för andelen som stödjer partiet.

- ▶ a) Beräkna nedre intervallgräns för 95%-intervallet.
- ▶ b) Beräkna nedre intervallgräns för 99%-intervallet.
- ▶ c) Sverige har en småpartispärr på 4% vid riksdagsval som försvårar för små partier att få mandat i allmänna val. Erhåller partiet mindre än 4% röster ges de inget mandat alls.

Baserat på undersökningen, vad är sannolikheten att Piratpartiet inte får något mandat?

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Medelsträckan svenska pendlare kör i veckan är 27 km. I ett slumpmässigt urval om 20 kvinnor från Norrbotten beräknades deras medelpendlarsträcka till 28 km och stickprovsvariansen (s^2) 16 km². Man vill testa om medelsträckan kvinnor i Norrbotten pendlar till jobber skiljer sig från medelsträckan för alla pendlare. Sträckorna kan antas vara approximativt normalfördelade.

- ▶ a) Man testar på 0.5% signifikansnivå, nollhypotesen är H_0 : Equation 250 = 27 km. Vad är mothypotesen?
- ▶ b) Vad är testvariabeln t för detta test?
- ▶ c) Vad är p-värdet för testet?
- ▶ d) Förkasta eller acceptera H_0 ?

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3

Q3

X är rektangelfördelad mellan 0 och okänd parameter b . Eftersom rektangelfördelningen är symmetrisk kring väntevärdet innebär det att medianen för X är lika med väntevärdet för X .

$\hat{\mu}_1 = 1/3(X_1 + X_2 + X_3)$ och $\hat{\mu}_2 = c \cdot m(X_1, X_2, X_3)$ då $m(\dots)$ är en funktion som väljer ut det mittersta värdet av datapunkterna, är två skattare av medianen.

- ▶ a) Är $\hat{\mu}_1$ en väntevärdesriktig skattare av medianen för X ?
- ▶ b) Man har funnit att tätheten $f(x)$ till den stokastiska variabeln $m(X_1, X_2, X_3)$ är $f(x) = 6x(b - x)/b^3$ för x mellan 0 och b och 0 annars.
För vilket c är $\hat{\mu}_2$ en väntevärdesriktig skattning av medianen?
- ▶ Om c är den konstant som gör $\hat{\mu}_2$ till en väntevärdesriktig skattning av medianen för X , vilken av skattarna $\hat{\mu}_1$ och $\hat{\mu}_2$ är effektivast?

Repetition

Beroendestruktur

Regression

Duggor

Dugga 1

Dugga 2

Dugga 3