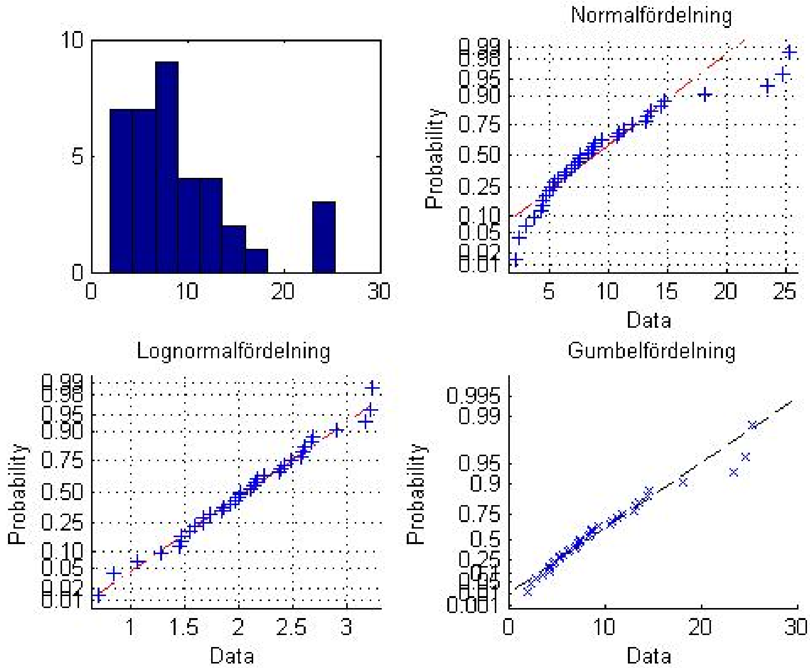
Matematisk statistik
Matematikcentrum
Lunds tekniska högskola

Kortfattade lösningar till tentamen i
FMSF75-Matematisk statistik för W,
2018-10-27

1. (a) Enligt figuren nedan verkar data passa bäst till en lognormalfördelning.



(b) X=vindhastighet; $\ln(X) \in N(\mu,\ \sigma,)$. Om $y_i = \ln(x_i)$, d.v.s. logaritmerade vinhastigheter skattas modellens parametrar med $\mu^*_{obs} = \frac{1}{37}\sum_{i=1}^{37} y_i = 2.0591$ och $\sigma^*_{obs} = \sqrt{\frac{1}{37-1}\sum_{i=1}^{37}(y_i-\bar{y})^2} = 0.6180$.
$P(4 \leq X \leq 22) = P(\ln(4) \leq \ln(X) \leq \ln(22)) = \Phi(\frac{\ln(22)-2.0591}{0.6180}) - \Phi(\frac{\ln(4)-2.0591}{0.6180}) = 0.9525 - 0.1381 = 0.8144$.

(c) P(ingen energi kan utvinnas)$= 1 - 0.8144 = 0.1856$. Y =antal dagar av de 10 då ingen energi kan utvinnas; $Y \in Bin(10,\ 0.1856)$
P(ingen av de 10 dagarna kan energi utvinnas)=P$(Y = 10) = 0.1856^{10} \cdot (1-0.1856)^0 = 0.1856^{10} = 4.85 \cdot 10^{-8}$.

2. (a) X=antal olyckor per månad. Det gäller att $\mathrm{P}(X = 3) = 1 - (0.25 + 0.35 + 0.25) = 0.15$.
$P(X \geq 2) = P(X = 2) + P(X = 3) = 0.25 + 0.15 = 0.40$

(b) $E(X) = \sum_{x=1}^{3} x \cdot p(x) = 0 \cdot 0.25 + 1 \cdot 0.35 + 2 \cdot 0.25 + 3 \cdot 0.15 = 1.3$

(c) $V(X) = \sum_x (x-1.3)^2 \cdot p(x) = (0-1.3)^2 \cdot 0.25 + (1-1.3)^2 \cdot 0.35 + (2-1.3)^2 \cdot 0.25 + (3-1.3)^2 \cdot 0.15 = 1.01$
$D(X) = \sqrt{1.01} = 1.005$

(d) Om X_i= antal olyckor under månad i är antal olyckor på fem år $\sum_{i=1}^{60} X_i$.
Enligt centrala gränsvärdessatsen (många, oberoende och samma fördelning) gäller $\sum_{i=1}^{60} X_i \underset{\sim}{\in} N(60 \cdot 1.3,\ \sqrt{60}\sqrt{1.01})$
$P(\sum_{i=1}^{60} X_i \geq 100) = 1 - P(\sum_{i=1}^{60} X_i \leq 99) \approx 1 - \Phi(\frac{99-60\cdot 1.3}{\sqrt{60 \cdot 1.01}}) = 0.0035$.

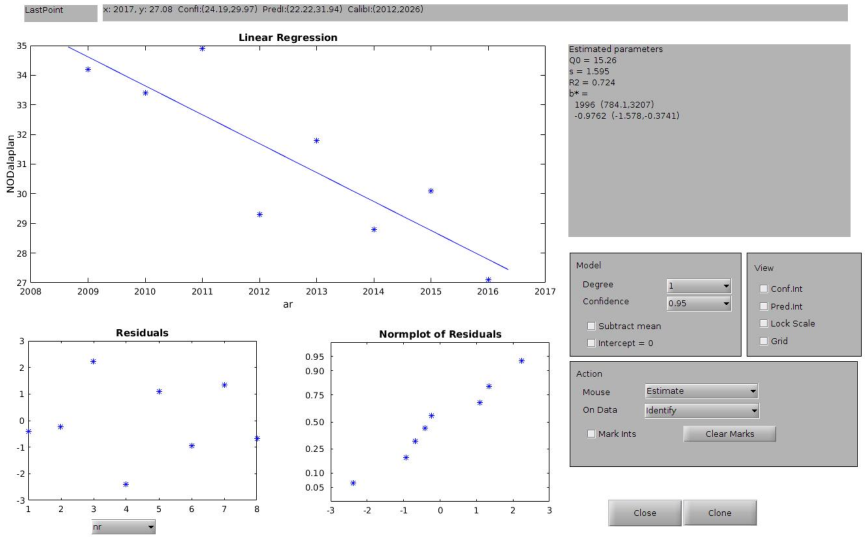
3. (a) X=antal gräsbränder under årets juli; $X \in Po(\lambda)$.
Intressanta hypoteser: $H_0 : \lambda \leq 10$; $H_1 : \lambda > 10$.
Med direktmetoden fås P-värde=$P(X \geq 19 \text{ om } X \in Po(10)) =$
$1 - P(X \leq 18 \text{ om } X \in Po(10)) = 0.0072$.
Eftersom P-värde=$0.0072 < 0.01$ (standardgräns) förkastas H_0 på nivå 0.01. Mer precist kan H_0 förkastas på nivå 0.0072.
Ja, denna typ av bränder tycks ha ökat detta år.

   (b) Y=antalet allvarliga bränder i år av de 19 observerade;
$Y \in Bin(19, p)$, där p=P(en gräsbrand är allvarlig).
Intressanta hypoteser: $H_0 : p \geq 0.4$; $H_1 : p < 0.4$. Med direktmetoden fås P-värde=$P(Y \leq 4 \text{ om } Y \in Bin(19, 0.4)) = 0.0696$.
Eftersom P-värde=$0.0696 > 0.05$ (standardgräns) kan H_0 ej förkastas på nivå 0.05.
Nej, de observerade data tyder inte på att andelen allvarliga tillbud minskat.

4. (a) X=dioxinhalt i ekologiska ägg 2010-2015; $X \in N(\mu_F, \sigma_F)$
Y=dioxinhalt i ekologiska ägg 2017; $Y \in N(\mu_E, \sigma_E)$
Gör ett 95 % intervall för $\mu_F - \mu_E$:
$I_{\mu_F - \mu_E} = (1.41 - 0.90 \pm 1.96\sqrt{\frac{0.66^2}{28} + \frac{0.19^2}{13}}) = (0.51 \pm 1.96 \cdot 0.1354) = (0.2446, 0.7754)$
I intervallet ovan är inte de två variansskattningarna sammanpoolade eftersom kvoten mellan de två variansskattningarna är stor: $\frac{0.66^2}{0.19^2} \approx 12$.
OM en sammanpoolning görs är den gemensamma skattade standardavvikelsen
$s_p = \sqrt{\frac{27 \cdot 0.66^2 + 12 \cdot 0.19^2}{27+12}} = 0.5592$ och intervallet
$I_{\mu_F - \mu_E} = (1.41 - 0.90 \pm 1.96 s_p\sqrt{\frac{1}{28} + \frac{1}{13}}) = (0.51 \pm 1.96 \cdot 0.1877) = (0.1422, 0.8778)$
I intervallen ovan används lambda-kvantil eftersom antalet observationer är så stort att den är approximativt lika stor som eventuella t-kvantiler.

   (b) Z=dioxinhalt i konventionella ägg 2010-2016; $Z \in N(\mu_K, \sigma)$
Y=dioxinhalt i ekologiska ägg 2017; $Y \in N(\mu_E, \sigma)$
$H_0 : \mu_E - \mu_K \leq 0$; $H_1 : \mu_E - \mu_K > 0$; Gör ett nedåt begränsat, 95 % intervall för $\mu_E - \mu_K$:
Eftersom de två skattningarna av σ är ungerfär lika stora görs en gemensam skattning: $s_p = \sqrt{\frac{39 \cdot 0.125^2 + 12 \cdot 0.19^2}{39+12}} = 0.143$ och intervallet är
$I_{\mu_E - \mu_K} = (0.90 - 0.46 - t_{0.05}(51) s_p\sqrt{\frac{1}{40} + \frac{1}{13}}, \infty) = (0.44 - 1.68 \cdot s_p\sqrt{\frac{1}{40} + \frac{1}{13}}, \infty) = (0.44 - 0.0765, \infty) = (0.364, \infty)$
Eftersom intervallet inte täcker över 0 förkastas H_0 och ja, ekologiska ägg har högre dioxinhalt jämfört med konventionella ägg.

5. (a) Med y_t=årsmedelvärde av NO_2 år t ansätts modellen $y_t = \alpha + \beta \cdot t + \varepsilon_t$ där ε_t är oberoende och $N(0, \sigma)$. Modellen verkar rimlig eftersom residualerna tycks oberoende, visar inget mönster och är normalfördelade.

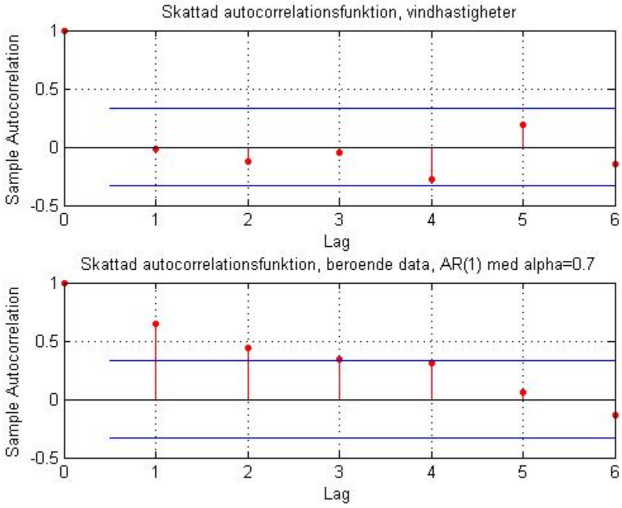


Intervallet för β är (-1.578, -0.3741), vilket inte täcker över 0. Därför förkastas hypotesen $H_0 : \beta = 0$ och en signifikant (nedåtgående) trend i data är visad.

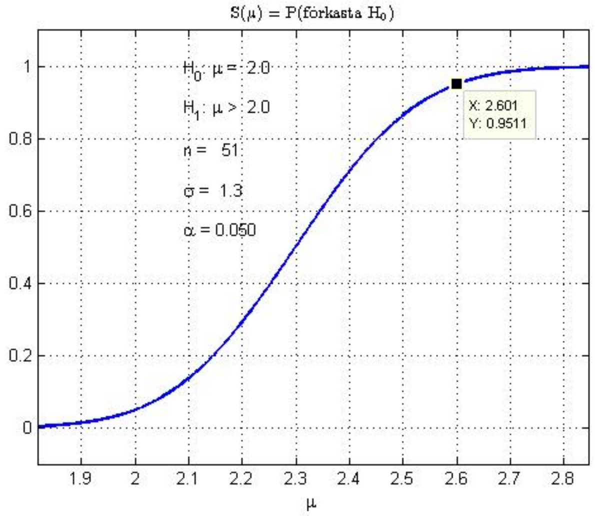
(b) Ett 95 % prediktionsintervall för Y_t vid år 2017 uppskattas från rutinen `reggui` till (22, 32) medan ett 99 % prediktionsintervall uppskattas till (19, 34) och ett 99.9 % prediktionsintervall uppskattas till (15, 39).
Sannolikheten att Y_{2017} överstiger MKN-normen på 40 är väldigt liten, understiger 0.0005 eftersom det 99.9 % prediktionsintervallet ej täcker över 40.
Sannolikheten att Y_{2017} överstiger ÖUT-tröskeln på 32 är ungefär 0.025 eftersom det 95 % prediktionsintervallets gräns är ungefär 32.
Sannolikheten att Y_{2017} överstiger NUT-tröskeln på 26 är svårt att bedöma utifrån enbart `reggui` men kommer att överstiga 0.5 eftersom mitten av prediktionsintervallet överstiger denna tröskel.

(c) Z=skillnaden i årsmedelvärden av NO_2 mellan Dalaplan och Rådhuset under perioden 2009-2016; $Z \in N(\Delta, \sigma)$.
Från data fås: $\Delta^* = \bar{z} = 15.1250$ och $\sigma^* = s_z = \sqrt{\frac{1}{8-1}\sum_{i=1}^{8}(z_i - \bar{z})^2} = 1.7645$.
Ett 95 % konfidensintervall för Δ ges av $I_\Delta = (\bar{z} \pm t_{0.025}(8-1) \cdot \frac{s_z}{\sqrt{8}}) = (13.650,\ 16.600)$.

6. (a) Övre plotten i bilden nedan visar den skattade autokorrelationsfunktionen för vindhastigheterna. Eftersom samtliga skattade korrelationskoefficienter ligger innanför bandet dras slutsatsen att data är oberoende.
I den undre plotten ges ett exempel på hur den skattade autokorrelationsfunktionen kan se ut då det finns ett positivt beroende i data.



(b) Genom att pröva sig fram med olika värden på n i rutinen styrkefkn ses att villkoret är uppfyllt om $n = 51$.



(c) X=kvinnans vikt; Y=pojkens vikt. Totalvikt på bron är $X+19+5+Y+7.5 = X+Y+31.5$.
$E(X+Y+31.5) = E(X)+E(Y)+31.5 = 62+35+31.5 = 128.5$ (kg)
$V(X+Y+31.5) = V(X)+V(Y) = 3^2+2^2 = 13$
$D(X+Y+31.5) = \sqrt{13} = 3.6056$

(d) Eftersom intervallet för α täcker 0 är parametern α ej signifikant skilt från 0. Modellen $y_i = \beta x_i + \varepsilon_i$ är alltså att föredra. Däremot är β signifikant skilt från 0 och termen med denna parameter bör ej tas bort från modellen.

(e) $\overline{X} = \frac{1}{4}\sum_{i=1}^{4} X_i \in N(\mu,\ \frac{0.3}{\sqrt{4}})$.
$P(|\overline{X}-\mu| > 0.2) = 1 - P(-0.2 \le \overline{X}-\mu \le 0.2) = 1 - (\Phi(\frac{0.2}{\frac{0.3}{2}}) - \Phi(\frac{-0.2}{\frac{0.3}{2}})) = 1 - (\Phi(\frac{0.4}{0.3}) - \Phi(\frac{-0.4}{0.3})) = 1 - (0.9088 - 0.0912) = 0.1824$.