UPPSALA UNIVERSITET
Matematiska institutionen
Fredrik Jonsson

PROV I MATEMATISK STATISTIK
Sannolikhet och Statistik, W m.fl.
5:e juni 2008

Skrivtid: 8-13. Maxpoäng 40. För betygen 3,4 resp. 5 krävs 18, 25 resp. 32 poäng, inklusive ev. bonuspoäng. Tillåtna hjälpmedel: Formelsamling med tabeller samt miniräknare. Lösningarna skall vara väl motiverade och försedda med förklarande text. Om du inte kan få till en fullständig lösning, försök då att ange i ord och med resultat hur du tänkt och hur långt du kommit.

1. Två vanliga, sex-sidiga tärningar kastas. Beteckna resultaten X_1 respektive X_2 och inför den stokastiska variabeln $Y = X_1 + X_2$ (siffersumman). Bestäm

   (2p) $P(Y \geq 8)$

   (3p) $P(Y \geq 8 \mid X_1 = 5)$

2. En person singlar slant ett antal gånger och för statistik över det inträffade. Låt n beteckna antalet upprepningar. Antag samma chans för krona respektive klave.

   (2p) Om $n = 10$, hur stor är chansen att få krona lika många gånger som klave?

   (2p) Om $n = 20$, hur stor är chansen att få bara krona eller bara klave?

   (2p) Om $n = 100$, hur stor är chansen att få krona lika många gånger som klave?

3. Effektiviteten hos två luftfuktare A och B jämfördes genom att de placerades i två likvärdiga rum med samma ursprungsfuktighet. Efter en timme mätte man luftfuktigheten (μ_1 respektive μ_2) i vardera rum. Eftersom mätutrustningen inte var den mest tillförlitliga valde man att göra 7 mätningar i vardera rum. Antag att det vid varje mätning tillkommer ett mätfel $\varepsilon \in N(0, \sigma)$. Man erhöll:

A :	70.8	70.6	70.4	70.9	70.3	70.3	70.7
B :	70.2	69.9	70.2	70.5	69.8	70.1	70.3

   (4p) Bilda ett 99% konfidensintervall för $\mu_1 - \mu_2$.

   (2p) Hur tolkar du föregående resultat?

4. På kontrollavdelningen vid en kvarn gör man regelbundet mätningar av halten kadmium i vetemjöl. Mätningarna är behäftade med normalfördelade mätfel så att uppmätt halt X i ett parti med mjöl anses vara normalfördelad med väntevärde μ = den genomsnittliga halten i partiet, och med standardavvikelse $\sigma = 10.0$ (i viss enhet). Det är angeläget (ur hälsosynpunkt) att genomsnittshalten understiger 40.0.

   (2p) Vad är sannolikheten att ett uppmätt värde understiger 40.0 när genomsnittshalten är $\mu = 45.0$?

   (2p) Antag att vi gör 3 bestämningar av halten i ett och samma parti. Vad är sannolikheten att alla tre mätvärdena understiger 40.0 när $\mu = 45.0$?

   (2p) Antag att vi gör n bestämningar i ett parti med $\mu = 42.0$. Hur stort måste n vara för att sannolikheten att alla n mätvärden understiger 40.0 ska vara högst 0.01?

5. (5p) Vid en enorm tomatodling i Holland förekommer det att plantor angrips av bladmögel. Av 225 undersökta plantor befanns 70 vara angripna. Bilda ett 95% konfidensintervall för p = andel angripna plantor. Ange slumpmässig modell och eventuella approximationer.

6. På en arbetsplats har man problem med en kopiator. Man är intresserad av $p =$*andel misslyckade kopieringar*. Man höll uppsikt under 90 kopieringstillfällen och det visade sig uppkomma papperstrassel vid 8 av dessa. Betrakta kopieringstillfällen som oberoende slumpmässiga försök under likartade förhållanden.

   (1p) Vilken är den rimliga skattningen av p?

   Analysera skattningsmetodiken genom att besvara följande frågor (med skattning avser vi nu en stokastisk variabel).

   (2p) Vad är väntevärdet och variansen av skattningen?

   (1p) Är skattningen *väntevärdesriktig*?

   (2p) Vad händer med variansen av skattningen om man ökar antalet kontrollerade kopieringstillfällen? Hur kan man tolka detta beteende?

7. Uppenbarligen är det så, att ju fler bilar som befinner sig på en väg, desto långsammare flyter trafiken. För en trafikplanerare kan det i konkreta fall vara önskvärt att se närmare på detta samband.

   Följande är data från en vägsträcka i USA år 1956. Vid 24 tillfällen har $x =$*trafiktäthet* (enhet: bilar/km) och $y =$*genomsnittlig hastighet hos bilar* (enhet: km/h) uppmätts.

x	y	x	y	x	y	x	y	x	y	x	y
12.7	62.4	87.8	12.4	81.1	13.8	18.3	51.2	18.6	46.3	66.3	18.3
17.0	50.7	81.3	13.4	62.8	17.9	19.1	50.9	66.0	16.9	61.7	18.0
66.0	17.1	75.6	13.7	77.0	15.8	16.5	54.7	60.3	19.8	66.6	16.6
50.0	25.9	66.2	17.9	89.6	12.6	22.2	46.5	56.0	21.2	67.8	18.3

   Det kan anses naturligt att se y som en funktion av x, men att viss variation inträder vid mättillfällen, pga. mätfel, individuell bilkörning och andra faktorer (jmf. Fig. 1). I Fig. 2 plottas y-data mot logaritmerat x-data (här kallat för u-data).

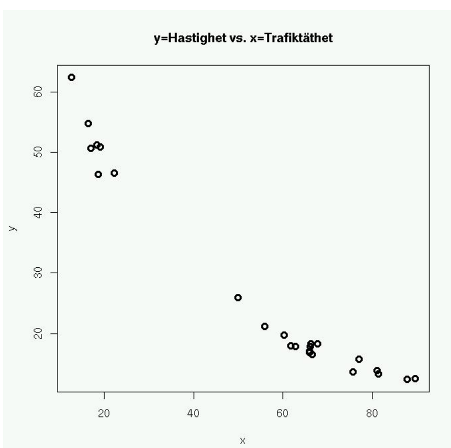

Fig. 1

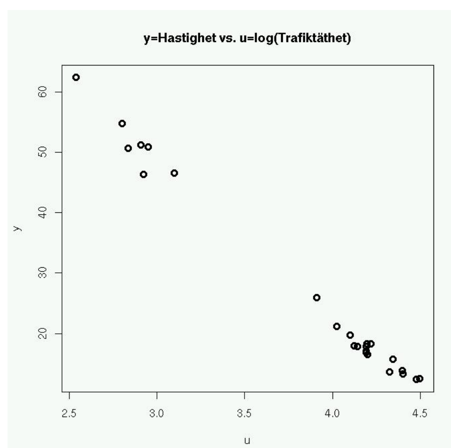

Fig. 2

   (2p) Upprätta, genom sedvanlig regressionsanalys och utifrån de numeriska uppgifterna nedan, ett lineärt samband mellan u och y (y som funktion av u).

$$\bar{u} = 3.83, \quad \bar{y} = 27.18, \quad \sum uy = 2250.74, \quad \sum u^2 = 362.27, \quad \sum y^2 = 24098.88$$

   (2p) Komplettera det förra resultatet med konfidensintervall för lutningskoefficient respektive intercept för den anpassade linjen.

   (2p) Vilken statistisk modell ligger bakom konfidensintervallen i föregående uppgift? Hur modelleras därmed sambandet mellan y och x?