**UPPSALA UNIVERSITET**
**Matematiska institutionen**
Fredrik Jonsson
jonsson@math.uu.se

MATEMATISK STATISTIK
Tentamen för NVP kemi/geologi
Matematik och statistik NV1

# Lösningar till tentamen 070111

1. (a) $P(\text{någon T-8 av tre visar } 8) = 1 - P(\text{alla tre visar } \leq 7) = 1 - (7/8)^3 \approx 0.33$.
   Oscars påstående ger alltså en för hög uppskattning av den aktuella sannolikheten.
   Det är ej korrekt att resonera:

   $$P(\text{tärning 1, tärning 2 eller tärning 3 visar 8}) = 3 \cdot P(\text{tärning 1 visar 8}) = 3 \cdot 1/8,$$

   ty händelserna $\{\text{tärning } i \text{ visar } 8\ \}$, $i = 1, 2, 3$ är inte oförenliga. (Man kan få två eller tre åttor.)

   (b) Inför beteckningar $X_1$, $X_2$ och $X_3$ på de slumpvariabler som svarar mot de tre tärningarna, $Y = X_1 + X_2 + X_3$. Då gäller att

   $$\begin{aligned} P(Y = 17 \text{ eller } 18) &= P(Y = 17) + P(Y = 18) \\ &= P(\text{en av tärningarna visar 5 och övriga två 6}) + (1/6)^3 \\ &= 3 \cdot (1/6)^3 + (1/6)^3 \approx 0.0185. \end{aligned}$$

   (c) Vi jämför $E(Y)$ med väntevärdet för en T-20.
   $E(Y) = E(X_1 + X_2 + X_3) = 3E(X_1) = 3 \cdot 1/6(1 + 2 + 3 + 4 + 5 + 6) = 10.5$.
   Detta sammanfaller med väntevärdet för en T-20, ty $\frac{1}{20}\sum_{i=1}^{20} i = 10.5$.
   De båda procedurerna ger alltså i genomsnitt lika höga värden.

2. (a) Låt $X$ =antalet defekta bland 12, då gäller att $X$ är $Bin(12, 0.15)$.
   $P(\text{minst 10 felfria}) = P(\text{högst 2 defekta}) = P(X \leq 2) =$[från tabell]$= 0.7358 \approx 0.74$.

   (b) Låt $Y$ =antalet defekta bland $20 \cdot 12 = 240$, då gäller att $Y$ är $Bin(240, 0.15)$.
   $Y$ är approx. $N(36, 30.6)$, ty $np(1 - p) > 10$.

   $$\begin{aligned} P(Y \leq 30) &\approx \Phi\left(\frac{30 + 1/2 - 36}{\sqrt{30.6}}\right) = \Phi(-0.994) \approx \Phi(-0.99) = 1 - \Phi(0.99) \\ &\approx 1 - 0.8389 = 0.1611 \approx 0.16. \end{aligned}$$

   (c) Låt $U$ =antalet defekta bland $24 \cdot 10 = 240$, då gäller att $U$ är $Bin(240, 0.005)$.
   $U$ är approx. $Po(1.2)$, ty $240 > 10$ och $0.005 < 0.1$.
   $P(U > 2) = 1 - P(U \leq 2) \approx 1 - 0.8795 = 0.1205 \approx 0.12$.
   Det går också att beräkna $P(U \leq 2)$ exakt med hjälp av lämplig sannolikhetsfunktion till $Bin(240, 0.005)$.

3. (a) Alla positiva heltalsvärden och 0; $P(X = 0) = p$.

   (b) $P(X = k) = p \cdot (1 - p)^k$.

(c) $P$(Isac misslyckas alla tre ggr) $= 1/2 \cdot 1/2 \cdot 1/2 = 1/8$. Sannolikheten att han får med sig poängen är alltså $1 - 1/8 = 7/8$.

Man kan även komma fram till svaret genom att låta $X$ beteckna slumpvariabeln som anger antalet misslyckade försök innan han lyckas (upprepning i det oändliga!). Då gäller att $X \in Ge(1/2)$. Med hjälp av (b) fås då att

$$\begin{aligned} P(\text{Isac lyckas}) &= P(X \leq 2) = p(0) + p(1) + p(2) \\ &= 1/2 + 1/2 \cdot 1/2 + 1/2 \cdot 1/2 \cdot 1/2 = 4/8 + 2/8 + 1/8 = 7/8. \end{aligned}$$

4.

$$\text{Hanar: } x_1, \ldots, x_5 \text{ stickprov från } N(\mu_1, \sigma^2), \ \sum x_i = 111.32, \ \sum x_i^2 = 2479.5.$$
$$\text{Honor: } y_1, \ldots, y_5 \text{ stickprov från } N(\mu_2, \sigma^2), \ \sum y_i = 110.24, \ \sum y_i^2 = 2431.1.$$

(a) 95% intervall för $\mu_1 - \mu_2$: $\bar{x} - \bar{y} \pm t_{0.025,8} \cdot s\sqrt{1/5 + 1/5}$.
$s^2 = \dfrac{4 \cdot 0.27938 + 4 \cdot 0.12847}{8} \approx 0.204$, ger:
$22.264 - 22.048 \pm 2.306\sqrt{0.204}\sqrt{1/5 + 1/5}$,
vilket slutligen ger det ungefärliga intervallet (-0.44, 0.88).

(b) Ingen skillnad påvisad, eftersom 0 tillhör intervallet och således är ett rimligt värde på $\mu_1 - \mu_2$.

(c) Vi behöver bestämma $n$ så att

$$t_{0.025,2n-2} \cdot s\sqrt{1/n + 1/n} \approx 1/2 \cdot t_{0.025,8} \cdot s\sqrt{1/5 + 1/5}.$$

Om t-värdet varit oförändrat hade vi fått $n = 20$, men kvantilen beror av $n$, så det räcker med $n = 16$ (c:a).

5. $H_0$: oberoende, $H_1$: ej oberoende.
Observationer med tillhörande $\hat{e}_{i,j}$-värde:

| | *För vapenlicens* | *Mot vapenlicens* | *Summa* |
|---|---|---|---|
| *För dödsstraff* | 784 | 311 | 1095 |
| | 799.5 | 295.5 | 1095 |
| *Mot dödsstraff* | 236 | 66 | 302 |
| | 220.5 | 81.5 | 302 |

$$\chi^2 = \frac{(784 - 799.5)^2}{799.5} + \frac{(311 - 295.5)^2}{295.5} + \frac{(236 - 220.5)^2}{220.5} + \frac{(66 - 81.5)^2}{81.5} \approx 5.15$$

$\chi^2 \in \chi^2(1)$ under $H_0$.
$\chi^2_{0.05}(1) \approx 3.84$, $\chi^2 > 3.84$.

Dvs. $H_0$ förkastas, dvs. oberoende kan inte antas. Betraktar man data tycks det vara så att personer som är mot dödsstraff i högre grad är för vapenlicens (jämfört med personer som är för dödsstraff), samt att personer som är för vapenlicens i högre grad är mot dödsstraff (än de som är mot vapenlicens). Testet uttalar sig om att det är osannolikt att denna fördelning enbart beror på slump.

6. (a) [Stolpdiagram och/eller lådagram.] $B$ har vunnit 27 av 50 gånger, men å andra sidan kan man avläsa från diagramen att när $A$ vinner så gör han det med större marginal. Det finns olika tänkbara tolkningar, t.ex. att endera spelaren är ojämn i sitt spel. $A$ kanske tar större risker, när det går bra är han betydligt bättre än $B$ men i vissa fall klantar han till det så att $B$ nätt och jämnt kan vinna. Det är också en tolkningsfråga vad som ska värdesättas, att vinna oftast eller att totalt sätt samla ihop mest poäng.

   (b) Ett stolpdiagram visar frekvenser av observationer (y-axel) i respektive kategori (x-axel). Ett lådagram är ett sätt att grafiskt åskådliggöra minsta/största värde, median och övre/nedre kvartil i ett numeriskt datamaterial.

   Båda ger information om spridning och läge vad gäller datamaterialet, den förra relaterar till aktuella kategorier medan den senare delar in datamängden i fyra delar med avseende på storleksordning och med i princip lika många observationer i varje del.

   Den senare ger mer kortfattad information om datamaterialets utbredning och "tyngdpunkt". Den är också enklare att ta fram vid stora datamängder och många kategorier. Om någon eller några kategorier är överrepresenterade i förhållande till andra, eller om det föreligger någon slags skevhet eller symmetri i utbredningen i $x$-led upptäcks detta bäst med ett stolpdiagram.