STOCKHOLMS UNIVERSITET
MATEMATISKA INSTITUTIONEN
Avd. Matematisk statistik

MT1011
25 maj 2020

Lösningar till tentamen i Sannolikhetslära och statistik för lärare 25 maj 2020

Examinator: Jan-Olov Persson

Uppgift 1

Frekvenstabellen som motsvarar histogrammet är

Smältpunkt	71.6-71.8	71.8-72.0	72.0-72.2	72.2-72.4	72.4-72.6
Frekvens	2	12	18	9	1

och det sammanlagda antalet mätningar är $n = 2 + 12 + 18 + 9 + 1 = 42$

a) Med $n = 42$ är medianen medelvärdet av observation 21 och 22 för det storleksordnade datamaterialet. Ur tabellen framgår det att medianen ligger i intervallet 72.0-72.2.

b) Den nedre gränsen för medelvärdet blir

$$(2 \cdot 71.6 + 12 \cdot 71.8 + 18 \cdot 72.0 + 9 \cdot 72.2 + 1 \cdot 72.4)/42 = 71.98$$

Den övre gränsen för medelvärdet blir $71.98 + 0.2 = 72.18$.

Uppgift 2

Om vi ser problemet ur ett sannolikhetsperspektiv och låter A vara händelsen att bära på virus vid provtillfället och B händelsen att ha antikroppar vid provtillfället så gäller följande:
$P(A) = 0.07$, $P(B) = 0.10$ och $P(A \cap B) = 0.024 = 2.4\%$

a) Andelen som någon gång ha varit smittad (vid provtillfället eller tidigare) kan beräknas som $P(A \cup B)$ med formeln $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Med insatta värden får vi $P(A \cup B) = 0.07 + 0.10 - 0.024 = 0.146 = 14.6\%$.

b) Vi söker $P(B|A^*)$ och får **Rita Venndiagram?**

$$P(B|A^*) = \frac{P(A^* \cap B)}{P(A^*)} = \frac{P(B) - P(A \cap B)}{1 - P(A)} = \frac{0.10 - 0.024)}{1 - 0.07} = 0.082 = 8.2\%$$

Uppgift 3

a) Sannolikheten för att norrsken *inte* inträffar under en bestämd natt är $1 - 0.3 = 0.7$. Sannolikheten för att norrsken inte inträffar under någon av de sju nätterna blir då $0.7^7 = 0.082$ (oberoende skulle antas).

b) Sannolikheten för att norrsken endast inträffar natt 1 och natt 7 är $0.3 \cdot 0.7^5 \cdot 0.3 = 0.015$

c) Låt X vara antalet nätter med norrsken. $X \sim Bin(n = 7, p = 0.3)$ och

$$P(X = 3) = \binom{7}{3} \cdot 0.3^3 \cdot 0.7^4 = 0.227$$

Uppgift 4

Om X är regnmängden ett visst år gäller det att $X \sim N(\mu = 1100, \sigma = 210)$.

a) $P(X < 900) = P(\frac{X-1100}{210} < \frac{900-1100}{210}) = P(Z < -0.952)$ där $Z \sim N(0,1)$. Med normalfördelningstabellen får vi:
$P(Z < -0.952) = \phi(-0.95) = 1 - \phi(0.952) \approx 0.17$.

b) Vi söker $X_{0.95}$. Enligt normalfördelningstabellen är $Z_{0.95} = 1.645$ och med sambandet

$$Z = \frac{X - \mu}{\sigma} \Leftrightarrow X = \sigma Z + \mu$$

får man $X_{0.95} = 210 \cdot 1.645 + 1100 \approx 1445$ mm.

c) $P(Y > 600) = P(100 + 0.4X > 600) = P(X > \frac{600-100}{0.4}) = P(X > 1250) = 1 - P(Z < \frac{1250-1100}{210}) = 1 - P(Z < 0.714) \approx 0.24$.

Uppgift 5

Sannolikheten för en sexa i första kastet är $1/6$ vilket ger vinst 20 kr (återbetalt minus insats). Sannolikheten för första sexa i andra kastet är $5/6 \cdot 1/6 = 5/36$ och vinsten blir då 10 kr. Sannolikheten för första sexa i tredje kastet är $5/6 \cdot 5/6 \cdot 1/6 = 25/216$ och ger vinst 0 kr. Sannolikheten för ingen sexa på tre kast är $5/6 \cdot 5/6 \cdot 5/6 = 125/216$ och då blir vinsten -10 kr (förlust 10 kr). Tabellen nedan visar sannolikhetsfunktionen för vinsten, X, vid ett spel. (man kan notera att antal kast till första sexa beskrivs av en för-första-gången-fördelad slumpvariapel med parameter $p = 1/6$).

Antal kast till sexa	1	2	3	> 3
x (kr)	20	10	0	-10
P(X=x)	1/6	5/36	25/216	125/216

a)

$E(X) = \sum xP(X = x) = 20\cdot1/6+10\cdot5/36+0\cdot25/216-10\cdot125/216 \approx -1.065$

I genomsnitt förlorar man alltså drygt 1 krona per spel.

b) Låt $Y = \sum_{i=1}^{100} X_i$ vara den sammanlagda vinsten vid 100 spel. Enligt centrala gränsvärdessatsen gäller approximativt att $Y \sim N(\mu, \sigma)$ där $\mu = 100E(X_i)$ och $\sigma = D(X_i)\sqrt{100} = 10D(X_i)$. Väntevärdet $E(X_i)$ beräknades i uppgift a och vi behöver beräkna $D(X_i)$, vilket görs med formeln (utan index i):

$$D(X) = \sqrt{\sum (x - E(X))^2 P(X = x)}$$

Med $E(X) = -1.065$ och x-värden och sannolikheter enligt tabellen får man $D(X) = 11.72$. Approximativt gäller det således att

$$Y \sim N(\mu = -106.5, \sigma = 117.2)$$

och sannolikheten för en sammanlagd vinsten större än 50 kr blir

$$P(Y > 50) = 1 - \phi(\frac{50 - -106.5}{117.2}) = 1 - \phi(1.335) \approx 0.09 = 9\%$$

Uppgift 6

Ett 95% konfidensintervall för μ_1 beräknas med formeln

$$\bar{x}_1 \pm t_{0.025}(n-1)s_1/\sqrt{n}$$

Med konfidensintervall (4.69 , 5.55) , $n_1 = 5$ och $t_{0.025}(4) = 2.776$ får man följande ekvationssystem:

$\bar{x} - 2.776 \cdot s_1/\sqrt{5} = 4.69$
$\bar{x} + 2.776 \cdot s_1/\sqrt{5} = 5.55$

Med lösningar $\bar{x}_1 = (4.69 + 5.55)/2 = 5.12$ och $s_1 = \frac{(5.55-4.69)\sqrt{5}}{2 \cdot 2.776} = 0.346$

b) Ett hypotestest av $H0 : \mu_1 = \mu_2$ mot $H1 : \mu_1 \neq \mu_2$ på signifikansnivå 5% kan göras med hjälp av ett 95% konfidensintervall för $\mu_1 - \mu_2$ som beräknas med formeln:

$$\bar{x}_1 - \bar{x}_2 \pm t_{0.025}(n_1 + n_2 - 2)s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

där $s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$.

Medelvärde och standardavvikelse för stickprov 2 beräknas som i uppgift a och ger med $n_2 = 10$ och $t_{0.025}(9) = 2.262$, $\bar{x}_2 = 5.58$ och $s_2 = 0.280$.

Med $t_{0.025}(13) = 2.160$ och $s_p = 0.302$ får man ett 95% konfidensintervall för $\mu_1 - \mu_2$ till -0.46 ± 0.36 eller $(-0.82, -0.10)$. Eftersom noll inte ingår i intervallet kan H_0 förkastas vilket betyder att X_1 har ett signifikant lägre väntevärde än X_2, på nivån 5%.