TILASTOTIETEEN VALINTAKOE 2006
RATKAISUT

1.1 Mikä on todennäköisyys, että kahden harhattoman nopan heitossa silmälukujen summa on ainakin 5? (1p) — a) b) c) d)

a) $\frac{1}{6}$ b) $\frac{1}{2}$ c) $\frac{2}{3}$ d) $\frac{5}{6}$ — ☐ ☐ ☐ ☒

$P(X_1 + X_2 \geq 5) = 1 - P(X_1 + X_2 < 5) = 1 - (\frac{1}{36} + \frac{2}{36} + \frac{3}{36}) = 1 - \frac{1}{6} = \frac{5}{6}$.

1.2 Oletetaan, että A ja B ovat riippumattomia tapahtumia ja että $P(A) = 0.1$ ja $P(B) = 0.7$. Tällöin $P(A \cup B)$ on (1p)

a) 0.07 b) 0.60 c) 0.70 d) 0.73 — ☐ ☐ ☐ ☒

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B)$
$= 0.1 + 0.7 - 0.1 \cdot 0.7 = 0.8 - 0.07 = 0.73.$

1.3 Keskihajonnan neliötä kutsutaan (1p)

a) variaatiokertoimeksi b) varianssiksi
c) kovarianssiksi d) vinoudeksi — ☐ ☒ ☐ ☐

1.4 Mitä tapahtuu perusjoukon odotusarvon luottamusvälille, kun otoskokoa kasvatetaan? (1p)

a) se kasvaa b) se ei muutu
c) se lyhenee d) se ei ole määritelty — ☐ ☐ ☒ ☐

1.5 Nominaaliasteikolle sopiva keskiluku on (1p)

a) geometrinen keskiarvo b) moodi
c) aritmeettinen keskiarvo d) mediaani — ☐ ☒ ☐ ☐

1.6 Testin tulos ei ole tilastollisesti merkitsevä. Tämä tarkoittaa sitä, että testisuureen havaittu arvo on (1p)

a) hyväksymisalueella b) hylkäämisalueella
c) luottamusvälin ulkopuolella d) vaihteluvälillä — ☒ ☐ ☐ ☐

1.7 Laatikossa on x punaista ja y sinistä palloa. Laatikosta nostetaan kaksi palloa umpimähkään ilman takaisinpanoa. Mikä on todennäköisyys, että jälkimmäiseksi nostettu pallo on punainen? (1p)

a) $\frac{x^2+xy-x}{(x+y)^2}$ b) $\frac{x}{x+y}$ c) $\frac{y}{x+y}$ d) $\frac{xy}{(x+y)^2}$ ☐ ☒ ☐ ☐

$P(2.pun.) = P(1.sin., 2.pun.) + P(1.pun., 2.pun.)$
$= \frac{y}{x+y} \cdot \frac{x}{x+y-1} + \frac{x}{x+y} \cdot \frac{x-1}{x+y} = \frac{x}{x+y}.$

1.8 Jatkoa edelliseen. Mikä on todennäköisyys, että ensimmäinen pallo oli punainen, kun jälkimmäiseksi nostettu pallo on punainen? $(1p)$

a) $\frac{x-1}{x+y-1}$ b) $\frac{y}{x+y-1}$ c) $\frac{x}{x+y}$ d) 1 ☒ ☐ ☐ ☐

$P(1.pun.|2.pun.) = \frac{P(1.pun.,2.pun.)}{P(2.pun.)} = \frac{\frac{x}{x+y} \cdot \frac{x-1}{x+y-1}}{\frac{x}{x+y}} = \frac{x-1}{x+y-1}.$

1.9 Oletetaan, että muuttujat X_i ovat korreloimattomia, $E(X_i) = 1$ ja $\mathrm{Var}(X_i) = 2$, $i = 1, 2, 3$. Tällöin $E\left[(X_1 - 2X_2 + X_3)^2\right]$ on $(1p)$

a) 0 b) 8 c) 12 d) 16 ☐ ☐ ☒ ☐

$E(X_1 - 2X_2 + X_3) = 0$, joten $E\left[(X_1 - 2X_2 + X_3)^2\right]$
$= Var(X_1 - 2X_2 + X_3) = Var(X_1) + 2^2 Var(X_2) + Var(X_3)$
$= 2 + 4 \cdot 2 + 2 = 12.$

1.10 Jatkoa edelliseen. Jos muuttujat X_i, ovat korreloituneita siten, että $\mathrm{Cov}(X_i, X_j) = 1$ kaikilla $i \neq j$, niin verrattuna 1.9:ssä saatuun arvoon $E\left[(X_1 - 2X_2 + X_3)^2\right]$ $(1p)$

a) suurenee b) ei ole määriteltävissä
c) pysyy samana d) pienenee ☐ ☐ ☐ ☒

$Var(X_1 - 2X_2 + X_3)$
$= Var(X_1) + 2^2 Var(X_2) + Var(X_3)$
$+2Cov(X_1, -2X_2) + 2Cov(X_1, X_3) + 2Cov(-2X_2, X_3)$
$= 2 + 4 \cdot 2 + 2 - 4 \cdot 1 + 2 \cdot 1 - 4 \cdot 1 = 12 - 6 = 6.$

1.11 Satunnaismuuttuja X noudattaa binomijakaumaa parametrein 6 ja 0.3. Tällöin $(1p)$

a) $X \in \{1, 2, 3, 4, 5, 6\}$ b) $E(X) = 0.3$
c) $E(X) = 1.8$ d) $Var(X) = 1.8$ ☐ ☐ ☒ ☐

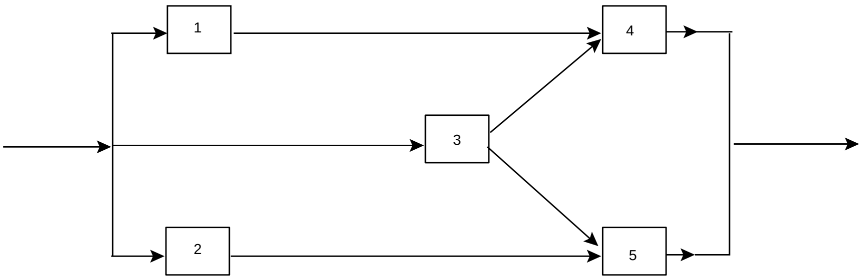
1.12 Kuinka monta erilaista osajoukkoa on n erilaista alkiota sisältävällä perusjoukolla, kun $n > 2$? $(1p)$

a) n b) n^2 c) 2^n d) n^n ☐ ☐ ☒ ☐

Osajoukon alkiot voidaan valita $\underbrace{2 \cdot \cdots \cdot 2}_{n \text{ kappaletta}} = 2^n$.

2. a) (3p) Veden pumppausjärjestelmä rakentuu viiden pumppauslaitoksen varaan. Alla kuva.

Oletetaan, että yksittäinen laitos toimii jonakin ajankohtana todennäköisyydellä p. Laitosten toiminta on toisistaan riippumatonta. Määritä todennäköisyys, että kuluttaja saa vettä.



Merkitään $A_i = \{\text{pumppausasema } i\}$, $i = 1, 2, \ldots, 5$, $P(A_i) = p$ ja $B = \{\text{kuluttaja saa vettä}\}$. Todennäkösyys, että kuluttaja saa vettä on

$$
\begin{aligned}
P(B) &= P[(A_1 \cap A_4) \cup (A_3 \cap A_4) \cup (A_3 \cap A_5) \cup (A_2 \cap A_5)] \\
&= P(A_1 \cap A_4) + P(A_3 \cap A_4) + P(A_3 \cap A_5) + P(A_2 \cap A_5) \\
&- P(A_1 \cap A_4 \cap A_3 \cap A_4) - P(A_1 \cap A_4 \cap A_3 \cap A_5) - P(A_1 \cap A_4 \cap A_2 \cap A_5) \\
&- P(A_3 \cap A_4 \cap A_3 \cap A_5) - P(A_3 \cap A_4 \cap A_3 \cap A_5) - P(A_3 \cap A_4 \cap A_2 \cap A_5) \\
&+ P(A_1 \cap A_4 \cap A_3 \cap A_4 \cap A_3 \cap A_5) + P(A_1 \cap A_4 \cap A_3 \cap A_4 \cap A_2 \cap A_5) \\
&+ P(A_1 \cap A_4 \cap A_3 \cap A_5 \cap A_2 \cap A_5) + P(A_3 \cap A_4 \cap A_3 \cap A_5 \cap A_2 \cap A_5) \\
&- P(A_1 \cap A_4 \cap A_3 \cap A_4 \cap A_3 \cap A_5 \cap A_2 \cap A_5) \\
&= p^2 + p^2 + p^2 + p^2 - p^3 - p^4 - p^4 - p^3 - p^3 - p^4 + p^4 + p^5 + p^5 + p^4 - p^5 \\
&= 4p^2 - 3p^3 - p^4 - p^5
\end{aligned}
$$

2. b) (3p) Kierrellessään kaupoissa henkilö A unohtaa päähineensä kauppaan todennäköisyydellä $\frac{1}{4}$. Eräänä päivänä henkilö A oli käynyt neljässä kaupassa, ja kotiin tultuaan huomasi, että oli kadottanut päähineensä. Mikä on todennäköisyys, että päähine on kolmannessa kaupassa? (3p)

Merkitään $A = \{\text{henkilö unohtaa päähineensä}\}$ ja $B = \{\text{kauppa } i\}$, $i = 1, 2, 3, 4$. Tällön $P(A) = \frac{1}{4} + \frac{3}{4} \cdot \frac{1}{4} + \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} + \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{175}{256}$ ja $P(B_3|A) = \frac{\frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{4}}{\frac{175}{256}} = \frac{36}{175}$.

3. (6p) Kolme akvaarioharrastajaa vertaili saman kalalajin pituuksia kolmessa eri akvaariossa A, B ja C. Mitatut tulokset ovat alla.

Akvaario A:	7	5	5	8	7	10	6	11	6
Akvaario B:	8	11	7	10	9	9	12	8	10
Akvaario C:	5	7	4	8	8	6	7	6	7

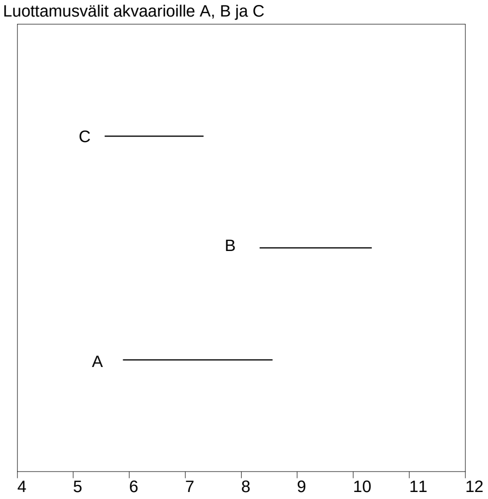
Oletetaan, että hajonnat ovat tiedossa ja ne likimain ovat akvaarioille A 2, B $\frac{3}{2}$ ja C $\frac{4}{3}$. Määritä 95 %:n luottamusväliä muodostettaessa käytettävä kriittinen arvo, kun oletamme, että voimme soveltaa normaalijakaumaa. Laske 95 %:n luottamusvälit akvaarioiden A, B ja C kalojen pituuksien odotusarvoille ja vertaa saamiasi tuloksia. (Laskuissa kriittisen arvon voi pyöristää lähimpään kokonaislukuun.) Piirrä myös kuva.

Luottamusväli odotusarvolle on $\overline{x} \pm z_c \frac{s}{\sqrt{n}}$, jossa z_c on määritettävä normaalijakauman taulukosta. 95 %:n luottamusvälillä $z_c = 1.96$ (ja pyöristettynä lähimpään kokonaislukuun $z_c = 2$). Keskiarvot akvaarioille A, B ja C ovat $\overline{x}_A = \frac{65}{9}$, $\overline{x}_B = \frac{84}{9}$ ja $\overline{x}_C = \frac{58}{9}$. Lasketaan luottamusvälit kolmelle akvaariolle

Akvaario A: $\frac{65}{9} \pm 2 \cdot \frac{2}{\sqrt{9}}$ $\qquad (\frac{53}{9} \leq \mu \leq \frac{77}{9})$

Akvaario B: $\frac{84}{9} \pm 2 \cdot \frac{\frac{3}{2}}{\sqrt{9}}$ $\qquad (\frac{75}{9} \leq \mu \leq \frac{93}{9})$

Akvaario C: $\frac{58}{9} \pm 2 \cdot \frac{\frac{4}{3}}{\sqrt{9}}$ $\qquad (\frac{50}{9} \leq \mu \leq \frac{66}{9})$



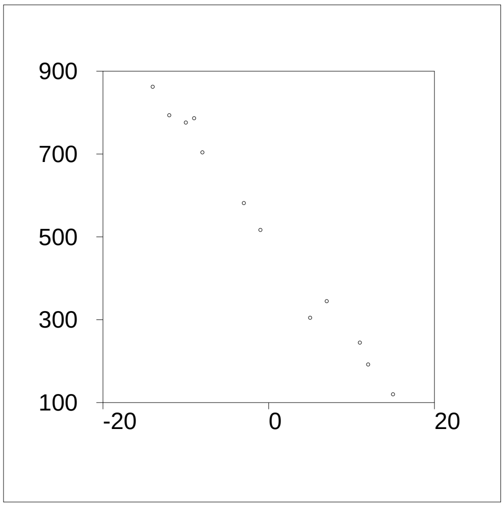
Akvaarion B kalojen pituuksien odotusarvojen voidaan tulkita poikkeavimmiksi kahdesta muusta tarkastellusta akvaariosta.

4. (6p) Allaolevassa taulukossa on esitetty aineisto, jossa x kuvaa kuukausittaista keskilämpötilaa Celsiusasteina ja y kuvaa kuukauden aikana lämmitykseen käytettyjen puupellettien määrää kiloina.

$x, ^\circ C$	-14	-12	-9	-3	5	11	15	12	7	-1	-8	-10
y, kg	863	794	787	582	305	245	120	192	345	517	704	776

a) Piirrä aineiston hajontakuvio. Tulkitse kuvaa sanallisesti ja arvioi erityisesti korrelaatiokertoimen suuruutta.

Korrelaatiokerroin on negatiivinen ja itseisarvoltaan lähellä ykköstä ($r \approx -0.99$).
b) Millä mitta-asteikoilla korrelaatiokerroin on sopiva riippuvuutta kuvaava suure?
Pearsonin korrelaatiokerroin soveltuu intervalli-ja suhde-asteikollisille muuttujille.
c) Määrittele käsitteet nollahypoteesi, vastahypoteesi, kaksisuuntainen testi ja yksisuuntainen testi.

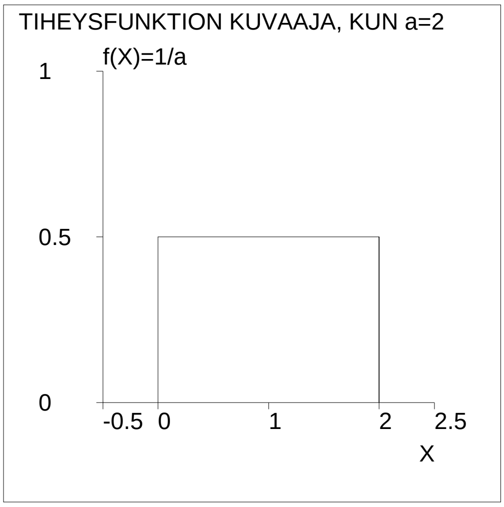


Nollahypoteesi on testattavan perusjoukon tilaa koskeva oletus. Tilastollinen testi johdetaan olettaen, että nollahypoteesin mukainen väittämä on voimassa perusjoukossa. *Vastahypoteesi* on oletus, joka pätee testattavalle perusjoukolle, mikäli nollahypoteesi on epätosi. Täten vastahypoteesin sisältämä väittämä sisältyy aina nollahypoteesin väittämän komplementtiin. Vastahypoteesi määrittää tilastollisen testin hylkäämisalueen sijainnin.
Kaksisuuntaisessa testissä vastahypoteesi H_1 on muotoiltu siten, että testin hylkäämisalueet sijaitsevat testisuureen jakauman molemmilla hännillä. Esimerkiksi keskiarvoja testattaessa nollahypoteesi voisi olla $H_0 : \mu_1 = \mu_2$ ja vastahypoteesina $H_1 : \mu_1 \neq \mu_2$.
Yksisuuntaisessa testissä hylkäämisalue sijaitseen jommalla kummalla jakauman hännällä. Keskiarvoja testattaessa yksisuuntainen vastahypoteesi voisi olla $H_1 : \mu_1 < \mu_2$ tai $H_1 : \mu_1 > \mu_2$.
d) Teoreettisen korrelaatiokertoimen ρ poikkeavuutta nollasta voidaan testata approksimoimalla testisuureen jakaumaa t-jakaumalla. Muotoile tämän kaksisuuntaisen testin nollahypoteesi H_0 ja vastahypoteesi H_1 parametrein.
$H_0 : \rho = 0$
$H_1 : \rho \neq 0$
e) Eräässä otoksessa, jossa vapausasteluku oli 20, havaittiin d)-kohdan mukaisen testin testisuureen arvoksi 3,1. Mikä on testin johtopäätös 1 %:n merkitsevyystasolla?
T-jakauman taulukosta saadaan $t_{0.005}(20) \approx 2.8 < 3.1$, joten H_0 hylätään 1%:n merkitsevyystasolla.

5. (6p) Olkoon $a > 0$ kiinteä reaaliluku. Satunnaismuuttuja X on tasaisesti jakautunut välillä $(0, a)$, jos sen tiheysfunktio on

$$f_X(x) = \begin{cases} \frac{1}{a}, & \text{kun } x \in (0, a) \\ 0, & \text{muulloin.} \end{cases}$$

Piirrä tiheysfunktion kuvaaja.



Määritä tasaisen jakauman kertymäfunktio.

Olkoon $x \in (0, a)$, niin $F_X(x) = \int\limits_0^x \frac{1}{a} du = \frac{x}{a}$. Täten $F_X(x) = \begin{cases} 0, & \text{kun } x \leq 0 \\ \frac{x}{a}, & \text{kun } x \in (0, a) \\ 1, & \text{kun } x \geq a. \end{cases}$

Määritä $P(X \leq ab)$, kun $0 < b < 1$. $P(X \leq ab) = F_X(ab) = \frac{ab}{a} = b$.

Määritä tasaisen jakauman mediaani. $P(X \leq Md) = 0.5$, joten edellisen perusteella $Md = 0.5a = \frac{a}{2}$.

Miten määritellään jatkuvan satunnaismuuttujan odotusarvo? $\mathbb{E}(X) = \int\limits_{-\infty}^{\infty} x \cdot f_X(x) dx$.

Määritä tasaisen jakauman odotusarvo. $\mathbb{E}(X) = \int\limits_0^a x \frac{1}{a} dx = \frac{1}{a} \cdot \frac{1}{2}(a^2 - 0^2) = \frac{a}{2}$.