



Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1912/SF1914/SF1915/SF1916 SANNOLIKHETSTEORI OCH STATISTIK,
ONSDAG 18 DECEMBER 2024 KL 8.00–13.00.

Examinator SF1914/SF1915/SF1916: Mykola Shykula, 08-790 66 44

Examinator SF1912: Celia Garcia Pareja, 072-872 34 06

Tillåtna hjälpmedel: Formel- och tabellsamling i Matematisk statistik (utdelas vid tentamen), miniräknare.

Tentamen består av två delar, benämnda del I och del II. Del I består av flerval uppgifterna 1–12, där i varje uppgift *endast ett svarsalternativ är korrekt*. På Del I ska endast svar i form av det korrekta svarsalternativet anges. Svaren ska anges på svarsblanketten (utdelas vid tentamen). Studenter som är godkända på kontrollskrivningen behöver ej besvara uppgift 1–3, utan får tillgodoräkna sig dessa tre uppgifter (i svarsblanketten anges ordet "Bonus"). Studenter som är godkända på den andra datorlaborationen behöver ej besvara uppgift 12, utan får tillgodoräkna sig denna uppgift (i svarsblanketten anges ordet "Bonus"). Bonusen gäller för den här omtentamen i december 2024 endast. Gränsen för godkänt är 9 poäng. Möjlighet att komplettera ges för tentander med 8 poäng.

Del II består av uppgifterna 13–16 och varje korrekt lösning ger 10 poäng. På denna del ska resonemang och uträkningar vara så utförliga och väl motiverade att de är lätta att följa och förstå. Införda beteckningar ska förklaras och definieras samt numeriska svar ska anges med minst tre värdesiffrors noggrannhet. Del II rättas bara för studenter som är godkända på eller får komplettera del I, och poäng på del II krävs för högre betyg än E. Studenter som är godkända på den andra datorlaborationen får dessutom tre bonuspoäng på del II. Bonusen gäller för den här omtentamen i december 2024 endast.

Tentamen kommer att vara rättad inom tre arbetsveckor från skrivningstillfället och kommer att finnas tillgänglig på studentexpeditionen minst sju veckor efter skrivningstillfället.

Del I

Uppgift 1

Röda Korset klassificerar den amerikanska befolkningen efter deras blodgrupp enligt fördelningen nedan:

Blodgrupp	A	B	AB	O
Sannolikhet	0.40	0.45	0.04	a

där a är en konstant. En läkare undersöker en grupp patienter som inte är släkt på något sätt, det vill säga vi kan anta att de är oberoende av varandra. Om två patienter från denna grupp väljs slumpmässigt, vad är då sannolikheten att den ena är från blodgruppen A och den andra är från blodgruppen O ?

- A: 0.29
- B: 0.51
- C: 0.44
- D: 0.088

Uppgift 2

En stokastisk variabel X har fördelningsfunktionen

$$F_X(x) = \begin{cases} 0 & , \quad x < 0, \\ \frac{x}{2} & , \quad 0 \leq x \leq 2, \\ 1 & , \quad x > 2. \end{cases}$$

Bestäm väntevärdet $E(e^X - 1)$.

- A: 1.19
- B: 2.19
- C: 3.19
- D: 5.39

Uppgift 3

Låt X och Y vara två oberoende stokastiska variabler sådana att $X \in Po(2.1)$ och $Y \in Po(2.5)$. Beräkna sannolikheten $P(X + Y \geq 2)$.

- A: ca 0.944
- B: ca 0.837
- C: ca 0.163
- D: ca 0.056

Uppgift 4

Låt X_1 och X_2 vara två oberoende och lika fördelade stokastiska variabler sådana att $X_1 \in Exp(1)$ och $X_2 \in Exp(1)$. Låt $Z = \max(X_1, X_2)$. Beräkna sannolikheten $P(Z \leq 1)$.

- A: 0.6321
- B: 0.3679
- C: 0.3996
- D: 0.6004

Uppgift 5

Låt 20 personer utföra mätningar (40 mätningar var) på en $N(\mu, \sigma)$ -fördelad stokastisk variabel, där σ är okänd. Därefter gör de var och en ett 95% konfidensintervall för μ . Vilket är det troligaste värdet (dvs det förväntade värdet) på det antal intervall som kommer att missa μ ?

A: 10

B: 5

C: 2

D: 1

Uppgift 6

Peter ska veckohandla i en stor affär och göra ett inköp på 48 artiklar. Antag att alla belopp avrundas nedåt vart och ett till hela kronor. Avrundningsfelen $X_1, X_2, \dots, X_{48}$ antas vara oberoende stokastiska variabler som samtliga är likformig fördelade $U(0, 1)$. Peters totala avrundningsfel betecknas $X = X_1 + X_2 + \dots + X_{48}$. Bestäm den approximativa sannolikheten $P(X > 20)$.

A: 0.023

B: 0.159

C: 0.841

D: 0.977

Uppgift 7

En s.v. X har täthetsfunktionen

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Beräkna maximum-likelihood-skattningen av λ då $(x_1, x_2, x_3, x_4) = (1.23, 3.32, 1.98, 2.12)$ är 4 oberoende observationer av X .

A: 4.624

B: 8.163

C: 0.462

D: 0.816

Uppgift 8

En elev vid en stor högskola i U.S.A. vill uppskatta andelen studenter som är registrerade väljare. Hen undersöker 500 studenter och finner att 300 är registrerade väljare. Beräkna ett approximativt 90% konfidensintervall för den verkliga andelen p studenter som är registrerade väljare.

- A: (0.557, 0.643)
- B: (0.564, 0.636)
- C: (0.599, 0.601)
- D: (0.578, 0.621)

Uppgift 9

Antag att nollhypotesen för ett högersidigt hypotestest är

$$H_0 : \text{andelen vuxna med jobb är } 88\%.$$

Vilket av följande refererar till signifikansnivån α när vi utför detta test?

- A: Sannolikheten att inte förkasta nollhypotesen att andelen vuxna som har jobb är 88% när andelen verkligen är mer än 88%.
- B: Sannolikheten att förkasta nollhypotesen att andelen vuxna som har jobb är 88% när andelen verkligen är mer än 88%.
- C: Sannolikheten att inte förkasta nollhypotesen att andelen vuxna som har jobb är 88% när andelen verkligen är 88%.
- D: Sannolikheten att förkasta nollhypotesen att andelen vuxna som har jobb är 88% när andelen verkligen är 88%.

Uppgift 10

Låt $x_1, \dots, x_n$ vara ett slumpmässigt stickprov från $N(\mu, \sigma)$. Vi skall pröva hypotesen $H_0 : \mu = 212$ mot $H_1 : \mu < 212$ på risknivån 10%. Antag att σ är okänd. Om $n = 30$, $\bar{x} = 211.2$ och $s = 2.2$, vilket av följande är sant?

- A: Testvariabeln $t_{obs} = -1.992$, kritiskt område är $(-\infty, -1.31)$ och därmed förkastas hypotesen H_0 .
- B: Testvariabeln $t_{obs} = -1.992$, kritiskt område är $(-\infty, -1.31)$ och därmed förkastas hypotesen H_0 inte.
- C: Testvariabeln $t_{obs} = -1.992$, kritiskt område är $(-\infty, -1.70) \cup (1.70, \infty)$ och därmed förkastas hypotesen H_0 .
- D: Testvariabeln $t_{obs} = -1.992$, kritiskt område är $(-\infty, -1.70) \cup (1.70, \infty)$ och därmed förkastas hypotesen H_0 inte.

Uppgift 11

En forskare vill testa om de fem sätten (att dricka koffeinhaltiga drycker, ta en tupplur, gå en promenad, äta ett sockerrikt mellanmål, annat) som människor använder för att bekämpa dåsigheit mitt på dagen är lika fördelade mellan kontorsanställda. Ett slumpmässigt urval av 60 kontorsanställda väljs ut och följande data erhålls.

Metod	dryck	tupplur	promenad	mellanmål	annat
Antal	19	16	12	8	5

Vi utför ett χ^2 -goodness-of-fit test för att avgöra om alla de fem metoderna är lika vanliga för att bekämpa dåsigheit mitt på dagen, där nollhypotesen är H_0 : alla metoder är lika vanliga. Givet data ovan blir teststorheten $Q = 10.83$, och därmed är slutsatsen av testet:

- A: Nollhypotesen kan förkastas på risknivån 5%, men inte på risknivån 1%.
- B: Nollhypotesen kan förkastas på risknivån 1%, men inte på risknivån 5%.
- C: Nollhypotesen kan förkastas både på risknivån 1% och på risknivån 5%.
- D: Nollhypotesen kan varken förkastas på risknivån 1% eller på risknivån 5%.

Uppgift 12

Vid enkel linjär regression gäller följande (linjära) samband:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

där man observerat paren (x_i, y_i) och ε_i betecknar de slumpmässiga felen, $i = 1, \dots, n$. Man önskar testa hypotesen $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$ och får testets P -värde till 0.027. Vilket av följande påståenden är sant?

- A: Vi förkastar nollhypotesen H_0 och kan därmed inte förkasta att y_i :na linjärt beror av x_i :na på risknivån 5%.
- B: Vi förkastar inte nollhypotesen H_0 och kan därmed inte förkasta att y_i :na linjärt beror av x_i :na på risknivån 5%.
- C: Vi förkastar nollhypotesen H_0 och kan därmed förkasta att y_i :na linjärt beror av x_i :na på risknivån 5%.
- D: Vi förkastar inte nollhypotesen H_0 och kan därmed förkasta att y_i :na linjärt beror av x_i :na på risknivån 5%.

Var god vänd!

Del II

Uppgift 13

Låt X vara en diskret stokastisk variabel sådan att $P(X = 1) = P(X = -1) = 0.25$ och $P(X = 0) = 0.5$. Vidare, låt Y vara en annan diskret stokastisk variabel sådan att $Y = 0$ när $X \neq 0$ med $P(Y = 0) = 0.5$, och $Y = 1$ eller $Y = -1$ när $X = 0$ med $P(Y = 1) = P(Y = -1) = 0.25$.

a) Är X och Y okorrelerade? Motivera ditt svar. (7 p)

b) Är X och Y oberoende? Motivera ditt svar. (3 p)

Uppgift 14

Av intresse är ett samband mellan snitttiden för att få en fast anställning (i år) bland arbetare som har en kandidatexamen eller högre och de som inte har det. En forskare tog två slumpmässiga och oberoende urval med följande resultat: $n_1 = 18$ arbetare med kandidatexamen eller högre fick en fast anställning efter $\bar{x}_1 = 5.21$ år (i snitt) och $n_2 = 24$ arbetare utan examen fick en fast anställning efter $\bar{x}_2 = 5.00$ år (i snitt). Stickprovs standardavvikelserna för båda grupperna blev $s_1 = s_2 = 0.4$. Vi antar att tiden för att få en fast anställning följer en normalfördelning i båda grupperna, dvs $N(\mu_1, \sigma)$ för dem med kandidatexamen eller högre och $N(\mu_2, \sigma)$ för dem utan examen, där σ är okänd men lika i de båda grupperna.

a) Bestäm ett tvåsidigt 99% konfidensintervall för skillnaden $\mu_1 - \mu_2$ baserat på dessa data, dvs bestäm $I_{\mu_1 - \mu_2}(0.99)$. Tolka resultatet. Finns det en signifikant skillnad? (5 p)

b) Ange det ungefärliga P -värdet för ett hypotestest för att testa om det finns en skillnad mellan väntevärden μ_1 och μ_2 . Svaret kan även ges i form av ett intervall, dvs $P_1 \leq P\text{-värdet} \leq P_2$. Tolka resultatet. (5 p)

Uppgift 15

Man vill undersöka om följande data kommer från en binomialfördelning med parametrar $n = 3$ och $p = 0.4$, dvs H_0 : data kommer från $Bin(3, 0.4)$ mot H_1 : data kommer inte från $Bin(3, 0.4)$.

	0	1	2	3	Totalt
antal observationer	15	54	28	3	100

Beräkna först förväntat antal observationer under hypotesen H_0 och sedan pröva hypotesen H_0 mot H_1 med ett χ^2 -test med teststorheten Q . Kan man förkasta H_0 på risknivån 5%? Tolka resultatet. (10 p)

Uppgift 16

En läkare vill undersöka om svenska kvinnor har större underarmsmått på den dominanta sidan (till exempel höger sida för högerhänta) eller om måttet på höger och vänster sida i genomsnitt är detsamma. Hen har tillgång till en undersökning på måfå utvalda 10 kvinnor där man inte noterat de faktiska mätvärdena, utan endast angivit med + när personens dominanta mått var större och - annars. Resultaten sammanfattas i följande tabell:

Kvinna nr	1	2	3	4	5	6	7	8	9	10
Mätning	+	+	-	+	+	-	-	+	+	+

Läkaren förutsätter att mätvärdena för olika kvinnor är oberoende och väljer att använda teststorheten $\xi = \text{antal plustecken}$ för att testa hypoteserna

H_0 : ingen genomsnittlig skillnad,

H_1 : genomsnittlig skillnad finns.

Hen väljer beslutsregeln "Förkasta H_0 om antalet plustecken ξ är högst 1 eller minst 9".

Vilken signifikansnivå har den beslutsregeln? (10 p)

Lycka till!



Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1912/SF1914/SF1915/SF1916
SANNOLIKHETSTEORI OCH STATISTIK,
ONSDAG 18 DECEMBER 2024 KL 8.00–13.00.

Del I, Svar.

1. D
2. B
3. A
4. C
5. D
6. D
7. C
8. B
9. D
10. A
11. A
12. A

Del I, Lösningsförslag.**Uppgift 1**

Först, gäller att $0.40 + 0.45 + 0.04 + a = 1$. Därför, $a = P(O) = 0.11$.

Vidare, för två oberoende patienter $P1$ och $P2$ valda på måfå, vi söker:

$$\begin{aligned} P\left(\left(\{P1 = A\} \cap \{P2 = O\}\right) \cup \left(\{P1 = O\} \cap \{P2 = A\}\right)\right) &= |\text{disjunkta}| = \\ &= P(\{P1 = A\} \cap \{P2 = O\}) + P(\{P1 = O\} \cap \{P2 = A\}) = |\text{oberoende}| = \\ &= P(P1 = A)P(P2 = O) + P(P1 = O)P(P2 = A) = 0.40 \cdot 0.11 + 0.11 \cdot 0.40 = 0.088 \end{aligned}$$

Svaret är D.

Uppgift 2

Först ser vi att s.v. variabel X har täthetsfunktionen

$$f_X(x) = F'_X(x) = \begin{cases} 0 & , \quad x < 0 \text{ eller } x > 2, \\ \frac{1}{2} & , \quad 0 \leq x \leq 2. \end{cases}$$

Därmed, $E(e^X - 1) = E(e^X) - 1 =$

$$= \int_0^2 e^x (1/2) dx - 1 = \frac{e^2 - e^0}{2} - 1 \simeq 3.19 - 1 = 2.19$$

Svaret är B.

Uppgift 3

$X + Y \in Po(4.6)$, ty X och Y är två *oberoende diskreta* s.v. sådana att $X \in Po(2.1)$ och $Y \in Po(2.5)$. Därmed, $P(X + Y \geq 2) =$

$$= 1 - P(X + Y \leq 1) = |\text{Table 5 för } Po(4.6)| = 1 - 0.05629 \simeq 0.94371 \simeq 0.944$$

Svaret är A.

Uppgift 4

Vi har: $P(Z \leq 1) = P(\max(X_1, X_2) \leq 1) = P(X_1 \leq 1, X_2 \leq 1) = |\text{oberoendet}| =$
 $= P(X_1 \leq 1)P(X_2 \leq 1) = |X_1, X_2 \in Exp(1)| = \left(\int_0^1 e^{-x} dx\right) \left(\int_0^1 e^{-x} dx\right) = (1 - e^{-1})^2 =$
 $= (0.63212)^2 \simeq 0.399576 \simeq 0.3996$

Svaret är C.

Uppgift 5

Sannolikheten är lika med 0.05 för att den sanna parametern μ missar ett 95% k.i. för μ . Dvs antal X av dom som missar (av 20 st sådana oberoende k.i.) är en diskret stokastisk variabel sådan att $X \in \text{Bin}(n, p)$, där $n = 20$, $p = 0.05$, och det troligaste (förväntade) värdet på det antal intervall som kommer att missa μ är

$$E(X) = np = 20(0.05) = 1$$

Svaret alltså är D.

Uppgift 6

Eftersom $X_i \in U(0, 1)$, har vi: $\mu = E(X_i) = 1/2$, $\sigma = D(X_i) = \sqrt{1/12}$, $i = 1, \dots, 48$. Enligt CGS (ty $n = 48 > 40$),

$$X = \sum_{i=1}^n X_i \stackrel{Approx}{\in} N(n\mu, \sigma\sqrt{n}) = N(48(0.5), \sqrt{1/12}\sqrt{48}) = N(24, 2).$$

Därför,

$$\begin{aligned} P(X > 20) &= 1 - P(X \leq 20) = 1 - \Phi\left(\frac{20 - 24}{2}\right) = \\ &= 1 - \Phi(-2) = |\text{symmetri}| = \Phi(2) = |\text{Table 1}| = 0.97725 \end{aligned}$$

Svaret är D.

Uppgift 7

ML-skattningen. Täthetsfunktionen för $X_i \in \text{Exp}(\lambda)$ är

$$f_{X_i}(x_i; \lambda) = \lambda e^{-\lambda x_i} \text{ för } x_i \geq 0,$$

där $i = 1, \dots, 4$. Därför, log-likelihooden blir

$$\log(L(x_1, x_2, x_3, x_4; \lambda)) = \log\left(\prod_{i=1}^4 \lambda e^{-\lambda x_i}\right) = \log(\lambda^4 e^{-\lambda \sum_{i=1}^4 x_i}).$$

Eftersom vi observerat $(x_1, x_2, x_3, x_4) = (1.23, 3.32, 1.98, 2.12)$, gäller i så fall att

$$\log(L(1.23, 3.32, 1.98, 2.12; \lambda)) = \log(\lambda^4 e^{-8.65\lambda}) = 4\log(\lambda) - 8.65\lambda.$$

Deriveringen map λ och = 0 ger: $\lambda = \frac{4}{8.65} = 0.462$.

Svaret är C.

Uppgift 8

Vi har $x = 300$ och $n = 500$, som ger $p_{obs}^* = \frac{x}{n} = 300/500 = 0.600$, och $1 - p_{obs}^* = 1 - 0.600 = 0.400$.

Eftersom det är 90%-k.i., då blir $\alpha = 1 - 0.90 = 0.10$. Och, för den tvåsidiga k.i. används det $\alpha/2 = 0.05$.

Det kritiska värdet, från Tabell 2: $z_{\alpha/2} = z_{0.05} = 1.645$, som ger följande felmarginal:

$$m = z_{\alpha/2} \cdot \sqrt{\frac{p_{obs}^*(1 - p_{obs}^*)}{n}} = 1.645 \cdot \sqrt{\frac{0.60 \cdot 0.40}{500}} = 0.036$$

Därmed,

$$I_p(0.9) = (p_{obs}^* \pm m) = (0.6 \pm 0.036) = (0.564, 0.636).$$

Svaret är B.

Uppgift 9

Signifikansnivån α (eller Typ I fel) – det är sannolikheten att förkasta nollhypotesen H_0 givet att nollhypotesen H_0 är sann. I det här fallet: Sannolikheten att förkasta nollhypotesen att andelen vuxna som har jobb är 88% när andelen verkligen är 88%.

Svaret är D.

Uppgift 10

Signifikansnivån $\alpha = 0.1$. Teststorheten (under H_0) är:

$$t_{obs} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{211.2 - 212}{\frac{2.2}{\sqrt{30}}} = -1.9917 \approx -1.992$$

Kritiskt område för vänster-sidigt test ($H_1 : \mu < 212$) är

$$(-\infty, -t_{\alpha}(n-1)) = (-\infty, -t_{0.1}(29)) = (-\infty, -1.31),$$

enligt Tabell 3. Slutligen, eftersom $-1.992 \in (-\infty, -1.31)$ kan vi förkasta hypotesen H_0 .

Svaret är A.

Uppgift 11

Vi har:

$$H_0 : p_1 = p_2 = p_3 = p_4 = p_5 = 0.2 \text{ vs } H_1 : \text{ Någon } p_i \text{ är inte } 0.2,$$

med teststorheten Q som är approximativt $\chi^2_\alpha(r-1)$ fördelad, $r-1 = 5-1 = 4$. Därför är det kritiska värdet för signifikansnivån $\alpha = 0.05$ är enl Tabell 4: $\chi^2_{0.05}(4) = 9.49$. Och, det kritiska värdet för signifikansnivån $\alpha = 0.01$ är enl Tabell 4: $\chi^2_{0.01}(4) = 13.3$.

Eftersom $9.49 = \chi^2_{0.05}(4) < Q = 10.83 < \chi^2_{0.01}(4) = 13.3$, kan alltså därför nollhypotesen förkastas på risknivån 5%, men inte på risknivån 1%.

Svaret är A.

Uppgift 12

Eftersom P -värdet är lika med $0.027 < 0.05$, så kan vi förkasta $H_0 : \beta_1 = 0$ med förmån till $H_1 : \beta_1 \neq 0$ på risknivån 5%.

Att förkasta nollhypotesen H_0 innebär att vi påstår att $\beta_1 \neq 0$ på risknivån 5%. Vilket i sin tur innebär att det inte kan förkastas att y_i :na linjärt beror av x_i :na på risknivån 5%, dvs att $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $i = 1, \dots, n$.

Svaret är A.

Del II, Lösningsförslag.**Uppgift 13**

- a) S.v. X och Y är okorrelerade om $C(X, Y) = E(XY) - E(X)E(Y) = 0$. Vi har:

$$E(X) = \sum_k k p_X(k) = -1 \cdot 0.25 + 0 \cdot 0.5 + 1 \cdot 0.25 = 0,$$

och, likaså,

$$E(Y) = \sum_j j p_Y(j) = -1 \cdot 0.25 + 0 \cdot 0.5 + 1 \cdot 0.25 = 0.$$

Dessutom, $E(XY) = 0$ ty $Y = 0$ när $X \neq 0$ och vice versa. Därmed,

$$C(X, Y) = E(XY) - E(X)E(Y) = 0 - (0)(0) = 0,$$

dvs X och Y är okorrelerade.

- b) X och Y är uppenbarligen *inte* oberoende (värdet som X tar påverkar direkt värdet som Y kan ta). T.ex., om $X = 0$ vet vi att antingen $Y = -1$ eller $Y = 1$. En sådan resonemang räcker för (3p) i denna deluppgift.

Dock *för att formellt visa att X och Y inte är oberoende*: Först ska man beräkna den simultana sannolikhetsfunktionen $p_{X,Y}(k, j) = P(X = k, Y = j)$, där $k, j \in \{-1, 0, 1\}$, dvs ta fram 3x3-tabell med de simultana sannolikheterna. Och sedan, välja ut ett par (k, j) som uppfyller

$$P(X = k) \neq P(X = k|Y = j) \stackrel{\text{Def}}{=} \frac{P(X = k, Y = j)}{P(Y = j)}.$$

Uppgift 14

- a) 99% tväsidig k.i. (dvs $\alpha = 0.01$) för $\mu_1 - \mu_2$, när σ är okänd:

$$\begin{aligned} I_{\mu_1 - \mu_2}(0.99) &= (\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}(n_1 + n_2 - 2)s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}) = \\ &= (5.21 - 5.00) \pm 2.70 \cdot 0.4\sqrt{\frac{1}{18} + \frac{1}{24}} = (0.21 \pm 0.337) = (-0.127, 0.547). \end{aligned}$$

Eftersom k.i. $I_{\mu_1 - \mu_2}(0.99) = (-0.127, 0.547)$ överlappar 0, kan vi dra slutsatsen att det inte finns tillräckligt med stöd i data (dvs, i de erhållna stickproven) för att påstå på risknivå 1% att det finns en skillnad i snitttiden för att få en fast anställning bland arbetare som har en kandidatexamen och de som inte har det.

b) Beräkna nu P -värdet för ett hypotestest som testar följande noll- och alternativhypoteser:

$$H_0 : \mu_1 = \mu_2,$$

$$H_1 : \mu_1 \neq \mu_2.$$

Vi har P -värde $= 2\left(P(T > |t_{obs}| \mid H_0 \text{ är sann})\right)$, och

$$t_{obs} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = |\mu_1 - \mu_2 = 0, \text{ under } H_0| = \frac{0.21 - 0}{0.4\sqrt{\frac{1}{18} + \frac{1}{24}}} \simeq 1.684,$$

där testvariabel, under H_0 , har följande fördelning:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \in t(n_1 + n_2 - 2),$$

ty vi har normalfördelningsantagande i de både angivna grupperna.

Vidare, för $t(n_1 + n_2 - 2) = t(40)$ -fördelning, ger Tabell 3:

$$P(T > 1.684) \approx 0.05,$$

dvs ganska nära 0.05 (mer exakt, $0.025 < P(T > 1.684) < 0.05$).

Därmed, P -värdet $= 2(P(T > 1.684)) \approx 0.10$, eller mer exakt: $0.05 < P\text{-värdet} < 0.1$

Det erhållna P -värdet (som är > 0.05) är klart *större* än testets signifikansnivå $\alpha = 1 - 0.99 = 0.01$, vilket innebär att vi inte kan förkasta H_0 på 1% signifikansnivån. Vilket i sin tur stöder slutsatsen från a), dvs att det inte finns tillräckligt med bevis i data för att stödja en skillnad mellan de två grupperna (av snitttiden för att få en fast anställning).

Uppgift 15

Man vill undersöka om det givna datamaterialet kommer från en binomialfördelning med parametrar $n = 3$ och $p = 0.4$, dvs H_0 : data kommer från $Bin(3, 0.4)$ mot H_1 : data kommer inte från $Bin(3, 0.4)$. Signifikansnivån (eller risknivån) är $\alpha = 0.05$.

Vi har för s.v. $X \in Bin(3, 0.4)$:

$$P(X = 0) = \binom{3}{0} 0.4^0 0.6^3 = 0.6^3 \simeq 0.216,$$

$$P(X = 1) = \binom{3}{1} 0.4^1 0.6^2 = 3(0.4^1) 0.6^2 \simeq 0.432,$$

$$P(X = 2) = \binom{3}{2} 0.4^2 0.6^1 = 3(0.4^2) 0.6^1 \simeq 0.288,$$

$$P(X = 3) = \binom{3}{3} 0.4^3 0.6^0 = 0.4^3 \simeq 0.064$$

Därmed, *förväntat antal obs.*, E_i , under H_0 är angivet (tillsammans med det ursprungliga datat, O_i) i tabellen nedan:

	0	1	2	3	Totalt
antal obs., O_i	15	54	28	3	100
förväntat antal obs., E_i	21.6	43.2	28.8	6.4	100

där $21.6 = 100 \cdot P(X = 0)$, $43.2 = 100 \cdot P(X = 1)$, $28.8 = 100 \cdot P(X = 2)$, $6.4 = 100 \cdot P(X = 3)$ (observera att alla de fyra förväntade antal är större än 5, vilket måste uppfyllas för att χ^2 -testet ska vara tillförlitligt).

Teststorheten Q under H_0 beräknas till

$$Q_{obs} = \sum_i \frac{(O_i - E_i)^2}{E_i} = \frac{(15 - 21.6)^2}{21.6} + \frac{(54 - 43.2)^2}{43.2} + \frac{(28 - 28.8)^2}{28.8} + \frac{(3 - 6.4)^2}{6.4} \simeq$$

$$\simeq 2.02 + 2.7 + 0.022 + 1.81 \simeq 6.55$$

Från Tabell 4, är det kritiska värdet lika med

$$\chi^2_{\alpha}(r-1) = \chi^2_{0.05}(3) = 7.81,$$

och därmed **kan nollhypotesen H_0 ej förkastas** på risknivån 5%, ty $Q_{obs} \simeq 6.55 < 7.81 = \chi^2_{\alpha}(r-1)$, där $r = 4$ är antal grupper i datamaterialet.

Tolkning: På risknivån 5% kan vi inte förkasta nollhypotesen H_0 , dvs vi kan inte påstå att det angivna datamaterialet inte kan tänka sig komma från en $Bin(3, 0.4)$ -fördening på 5% signifikansnivån.

Uppgift 16

Om H_0 är sann så är sannolikheten lika med 0.5 att en slumpvis vald kvinna har större mått på dominant sida. Eftesom de 10 måtten i stickprovet antas vara oberoende så följer att $\xi \in Bin(10, 0.5)$ då H_0 är sann. Signifikansnivån är då:

$$\begin{aligned} \alpha &= P(H_0 \text{ förkastas} \mid H_0 \text{ sann}) \\ &= P(\xi \leq 1 \text{ eller } \xi \geq 9 \mid \xi \in Bin(10, 0.5)) \\ &= P(\xi \leq 1 \mid \xi \in Bin(10, 0.5)) + P(\xi \geq 9 \mid \xi \in Bin(10, 0.5)) \\ &= P(\xi \leq 1 \mid \xi \in Bin(10, 0.5)) + (1 - P(\xi \leq 8 \mid \xi \in Bin(10, 0.5))) \\ &\approx 0.01074 + (1 - 0.98926) = 0.02148, \end{aligned}$$

enligt Tabell 6 (för $Bin(10, 0.5)$).

Alternativt, utan tabell och med $0.5^k 0.5^{10-k} = 0.5^{10}$:

$$\begin{aligned} \alpha &= P(\xi \leq 1 \text{ eller } \xi \geq 9 \mid \xi \in Bin(10, 0.5)) \\ &= \left(\binom{10}{0} + \binom{10}{1} + \binom{10}{9} + \binom{10}{10} \right) 0.5^{10} \approx 0.0214844. \end{aligned}$$

Svar: ca 0.0215