Projekt MVE490 Del 2

Det är tillåtet att sammarbeta, men alla lösningar **skall** lämnas in individuellt.

Sista inlämningsdag är 18de oktober på föreläsningen. Det är ok att lämna in elektroniskt genom att maila till simonssi 'at' chalmers.se. OBS! Alla elektroniskt inlämnade uppgifter måste vara skrivna i en texteditor (word, latex, etc). Inscannad handskriven text rättas ej!!!

Glöm inte att motivera era svar! Svar utan motivering ger 0 poäng! Glöm ej heller att skriva namn och personnummer.

Denna deluppgift kan ge högst 2.5 bonuspoäng till tentan med följande gränser:

Poäng	Bonus
< 8	0
≥ 8	0.5
≥ 14	1
≥ 20	1.5
≥ 26	2.0
≥ 32	2.5

1. Låt Y vara en normalfördelad slumpvariabel med väntevärde -1 och varians 4 och låt $X_1, X_2, \ldots, X_{200}$ vara ett stickprov från samma fördelning som Y.

   (a) Beräkna $P(Y > 5)$. (3p)

   Lösning: Eftersom $Y \sim N(-1, 4)$ så vet vi att

   $$Z = \frac{Y - (-1)}{\sqrt{4}} = \frac{Y + 1}{2}$$

   är $N(0, 1)$-fördelad. Vi får då att

   $$\begin{aligned} P(Y > 5) &= P\left(\frac{Y+1}{2} > \frac{5+1}{2}\right) \\ &= P(Z > 3) \\ &= 1 - P(Z < 3) \\ &= 1 - P(Z < 0) - P(0 < Z < 3) \\ &= 1 - 0.5 - P(0 < Z < 3) \\ &= 0.5 - P(0 < Z < 3). \end{aligned}$$

   Vi tittar i tabell IV och ser att $P(0 < Z < 3) = 0.4987$ så vi får då att $P(Y > 5) = 0.5 - 0.4987 = 0.0013$.

   (b) Vad har $\bar{X} = \frac{1}{n} \sum X_k$ för fördelning? (3p)

   Lösning: Eftersom stickprovet kommer från en $N(-1, 4)$-fördelning så vet vi att även $\bar{X}$ är normalfördelat med väntevärde -1 men med varians $4/200 = 0.02$.

OBS! Eftersom stickprovet kommer från en normalfördelning så behöver vi *inte* använda oss av centrala gränsvärdessatsen!

(c) Beräkna $P(\bar{X} > -0.8)$. (3p)

Lösning: Eftersom $\bar{X} \sim N(-1, 0.02)$ så har vi att

$$Z = \frac{\bar{X} + 1}{\sqrt{0.02}}$$

är $N(0,1)$-fördelad. Vi får då att

$$\begin{aligned} P(\bar{X} > -0.8) &= P\left(\frac{\bar{X}+1}{\sqrt{0.02}} > \frac{-0.8+1}{\sqrt{0.02}}\right) \\ &= P(Z > 1.41) \\ &= 0.5 - P(0 < Z < 1.41). \end{aligned}$$

Vi tittar i tabell IV igen och ser att $P(0 < Z < 1.41) = 0.4207$ så vi får då att $P(\bar{X} > -0.8) = 0.5 - 0.4207 = 0.0793$.

(d) Låt $T = 3X_1 - X_2 - 5X_{34} + 2X_{57}$. Beräkna $E(T)$. (3p)

Lösning: Här använder vi räknereglerna $E(X + Y) = E(X) + E(Y)$ och $E(aX) = aE(X)$, samt att $E(X_k) = -1$ för alla k. Vi får att

$$\begin{aligned} E(T) &= E(3X_1 - X_2 - 5X_{34} + 2X_{57}) \\ &= 3E(X_1) - E(X_2) - 5E(X_{34}) + 2E(X_{57}) \\ &= 3 \times (-1) - (-1) - 5 \times (-1) + 2 \times (-1) \\ &= -3 + 1 + 5 - 2 \\ &= 1. \end{aligned}$$

Alltså har vi att $E(T) = 1$.

2. Du och din kompis Mika jobbar extra på ett IKEA-lager. Det har kommit in en del klagomål från kunder om att en viss sorts glödlampa inte fungerar. Därför får ni i uppdrag att undersöka proportionen p av lamporna på lagret som inte fungerar. Det är dock bråttom så ni hinner inte testa alla lampor. Istället väljer ni ut tvåhundra lampor slumpmässigt och testar hur många som inte fungerar. Mika tar sedan fram ett konfidensintervall för p med hjälp av formlerna ni lärt er i denna kurs: $[0.0172; 0.0628]$.

(a) Hur många lampor av de ni testade fungerade inte? (4p)

Lösning: Mitten av ett konfidensintervall för en proportion p är proportionen i stickprovet, dvs $\hat{p}$. Mitten på intervallet $[0.0172; 0.0628]$ är 0.04, alltså måste vi har att $\hat{p} = 0.04$. Men vi räknar ut $\hat{p}$ med formeln

$$\hat{p} = \frac{\text{antalet observationer i stickprovet med den sökta egenskapen}}{\text{stickprovsstorleken}}$$

vilket i detta fall blir

$$\hat{p} = \frac{\text{antalet testade lampor som inte fungerar}}{200}$$

så vi har att

$$\text{antalet testade lampor som inte fungerar} = 200 \times 0.04 = 8.$$

(b) Vilken signifikansnivå har Mika använt? (4p)

Lösning: Formeln för ett konfidensintervall för en proportion är

$$\ell = \hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad u = \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Vi vet att $u = 0.0628$, att $n = 200$ och att $\hat{p} = 0.04$ så vi har då att

$$z_{\frac{\alpha}{2}} = \frac{u - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} = \frac{0.0628 - 0.04}{\sqrt{\frac{0.04 \times 0.96}{200}}} = 1.645.$$

Vi tittar i tabell IV (eller i tabell 2 på sida 316 i boken) och ser att detta motsvarar en signifikansnivå på $\alpha = 0.1$.

Man kan även använda formeln för ℓ istället för formeln för u och ta fram α på liknande sätt.

(c) Er chef är inte helt nöjd. Hen tycker att konfidensintervallet är för stort och beordrar er att ta fram ett konfidensintervall vars längd är 0.03. Om vi antar att er punktskattning $\hat{p}$ förblir samma och att ni använder samma signifikansnivå som tidigare, hur många lampor måste ni testa om detta ska vara uppfyllt? (4p)

Lösning Om intervallets längd ska vara 0.03 så måste vi ha att

$$z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.015.$$

Eftersom vi känner till både $z_{\frac{\alpha}{2}}$ och $\hat{p}$ så kan vi lösa ut n ur denna formel. Vi får då

$$n = \frac{z^2_{\frac{\alpha}{2}} \hat{p}(1-\hat{p})}{0.015^2} = \frac{1.645^2 \times 0.04 \times 0.96}{0.015^2} = 461.83 \approx 462.$$

Om chefen ska få sin önskan uppfylld måste ni alltså testa minst 462 lampor.

3. Stina jobbar på ett företag som producerar müsli. En av deras sorter innehåller mycket russin. Det är meningen att varje paket av denna sort ska innehålla ca 200 russin men det har inkommit klagomål på att den innehåller för många russin. Stina får därför i uppgift att testa om müslipaketen i genomsnitt innehåller fler än 200 russin. Om så är fallet måste de göra något åt problemet. Slumpmässigt väljer Stina ut 40 paket och räknar hur många russin det är i varje paket. Hon hittade i genomsnitt 201.5 russin i paketen. Vi kan anta att antalet russin i ett paket är normalfördelat och att variansen är $\sigma^2 = 5$.

(a) Ställ upp hypoteserna för att utföra testet. (1p)

Lösning: Låt μ vara det riktiga väntevärdet som vi är intresserade av. Eftersom vi bara är intresserade av avvikeler åt ena hållet så blir testet ensidigt. Hypoteserna blir

$$H_0 : \mu \leq 200$$

$$H_1 : \mu > 200$$

(b) Vilken teststatistika bör Stina använda? Vad har den för fördelning? (3p)

Lösning: Eftersom vi känner till variansen så använder vi oss av teststatistikan

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Eftersom stickprovet kan antas komma från en normalfördelning så är $\bar{X}$ normalfördelad och Z är då $N(0, 1)$-fördelad.

Återigen behöver vi inte använda oss av centrala gränsvärdessatsen eftersom vi kan anta att stickprovet kommer från en normalfördelning

(c) Bör Stina förkasta nollhypotesen på signifikansnivå $\alpha = 0.01$? (4p)

Lösning: Vi har att $\mu_0 = 200$, $\bar{x} = 201.5$, $\sigma = \sqrt{5}$ och $n = 40$. Vår observerade teststatistika blir alltså

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{201.5 - 200}{\sqrt{5}/\sqrt{40}} \approx 4.24.$$

Vi tittar i normalfördelningstabellen och ser att det kritiska värdet är $z_\alpha = 2.33$. Eftersom vår observerade teststatistika är större än (ligger längre ifrån 0) det kritiska värdet så förkastar vi H_0 på signifikansnivå $\alpha = 0.01$.

(d) Antalet russin i paketen är egentligen inte normalfördelat. Varför inte? Ge endast ett argument. (2p)

Lösning: Det mest uppenbara argumentet är att antalet russin i ett paket är diskret eftersom det bara kan vara ett heltal. Det kan t ex inte vara 200.3 russin i ett paket. Normalfördelningen är en kontinuerlig fördelning så antalet russin i ett paket kan inte vara normalfördelat.

Ett annat argument är att en normalfördelad slumpvariabel kan anta negativa värden och att vi inte kan ha negativt antal russin i ett paket.

(e) Medelvärdet av stickprovet är större än 200. En av cheferna säger att på grund av detta så kan vi direkt dra slutsatsen att det i genomsnitt är för många russin i paketen generellt sett. Förklara varför det inte är rimligt att resonera på detta sätt. (2p)

Lösning: Det viktiga här är att påpeka att stickprovet är just ett stickprov och inte nödväntigtvis (även om vi hoppas på det) representativt för hela populationen.