Matematiska vetenskaper
Chalmers tekniska högskola

Tentamen: 2017-08-25 kl $14^{00} - 18^{00}$
Matematisk statistik för K, TMA073 & TMA074

Lärare och Jour: David Bolin, telefon 772 53 75.

Tillåtna hjälpmedel: Formelsamling och valfri miniräknare med tömda minnen.

Korrekt, väl motiverad lösning ger poängen som är indikerad i parentes vid vardera uppgift. Totalt kan man få 40 poäng och betygsgränserna för betyg 3, 4 och 5 är 16, 24 och 32 poäng.

---

1. Ett bostadsbolag gör en undersökning bland sina hyresgäster och finner att 35% av hushållen äger ett piano. Fem procent av alla hushåll har någon gång klagat till bolagets kundtjänst på grund av grannar som spelat för hög musik. Av alla klagomål kommer en tiondel från personer som äger ett piano. Låt A beteckna händelsen att ett hushåll äger ett piano, och låt B beteckna händelsen att ett hushåll klagar så sina grannar. Beräkna sannolikheten för att ett hushåll

   (a) äger ett piano och har klagat; (1p)

   (b) varken äger ett piano eller har klagat; (2p)

   (c) har klagat givet att det äger ett piano. (2p)

2. Antalet hårddiskar som går sönder per år i en stor serverhall antas vara Poissonfördelat med parameter $\mu = 2.5$. I serverhallens budget har man avsatt pengar till att byta ut tre hårddiskar per år.

   (a) Ge en konkret tolkning av parametern μ. (1p)

   (b) Vad är sannolikheten för att pengarna räcker till att att byta alla hårddiskar som går sönder under ett år? (2p)

   (c) När nio månader av året har gått så har två hårddiskar redan gått sönder, vad är sannolikheten att budgeten räcker året ut? (2p)

3. Antag att vikten hos lärare på Chalmers varierar enligt en normalfördelning med väntevärde 75kg och standardavvikelse 10kg.

   (a) Mellan vilka vikter, symmetriskt kring 75kg, har 95% av alla lärare sina vikter? (2p)

   (b) I hissen i matematiska vetenskapers höga hus står det att dess maximala last är 1000kg eller 13 personer. Antag att 13 lärare tar hissen, vad är sannolikheten att de tillsammans väger mer än 1000kg? (3p)

4. Ett läkemedelsföretag undersöker om en ändrad dosering kan minska bieffekterna av ett läkemedel. Med vanlig dosering får cirka en av tio personer någon slags biverkning. I ett försök med 200 personer rapporterar 14 personer att de fick biverkningar om den alternativa doseringen användes.

   (a) Testa på nivå 1% om det finns en minskad risk för biverkningar. Kom ihåg att ange vad som testas och vilka antaganden som görs. (4p)

   (b) Vid planering av försök som detta är det ofta viktigt att försäkra sig om att testet får tillräcklig styrka, förklara detta begrepp. (1p)

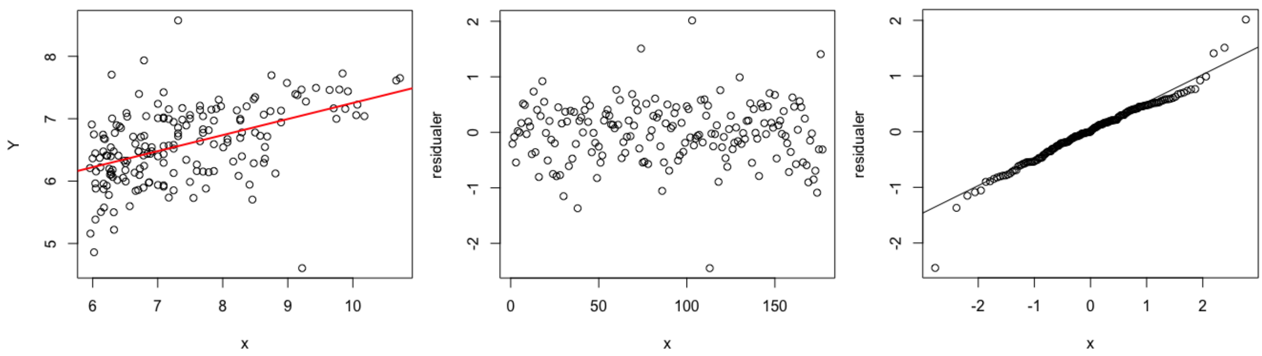

Figure 1: Resultat för regressionsmodellen i uppgift 5. Till vänster ses datan och den anpassade regressionlinjen, i mitten ses en residualplot och till höger ses en nomralfördelningsplot.

5. Man vill undersöka hur månadslönen hos ett aktiebolags VD hänger samman med företagets marknadsvärde och tar därför reda på dessa uppgifter hos 177 aktiebolag. Man ansätter följande regressionsmodell för hur lön och marknadsvärde hänger samman:

$$Y_i = \beta_0 + \beta_1 x_i + e_i,$$

där Y betecknar logaritmerad lön (enhet 1000\$), x betecknar logaritmen av företagets marknadsvärde (enhet miljoner \$) och e är mätfel. Följande storheter kan beräknas från datan

$$S_{xx} = 226.0944 \qquad S_{yy} = 64.64622 \qquad S_{xy} = 58.21094 \qquad \bar{x} = 7.39941 \qquad \bar{y} = 6.58284$$

Resultatet från en minsta-kvadrat anpassning av modellen kan ses i Figur 1.

(a) Vilka antaganden behövs för att enkel linjär regression ska kunna användas? Verkar dessa vara uppfyllda i det här fallet? (2p)

(b) Skatta β_0 och β_1. (1p)

(c) Enligt modellen, vad är den förväntade logaritmerade lönen hos en VD om företagets marknadsvärde är 8 miljarder? Beräkna ett 95% prediktionsintervall för lönen. (3p)

(d) Vad är skillnaden mellan ett prediktionsintevall för $Y(x)$ och ett konfidensintervall för $\mu_Y(x)$ i en regressionsmodell? Ange en kortfattad tolkning av båda intervallen. (2p)

(e) Ett företags vd har en lön på 5 i den logaritmerade skalan. Vad är är det förväntade marknadsvärdet på hans bolag enligt regressionmodellen? Ser du några problem med att göra denna prediktion? I så fall vad? (2p)

6. Anna och Anders ska mäta halten fosfor, p, i en sjö. Anna gör n oberoende mätningar, $x_1, \ldots, x_n$, med ett instrument som inte har något systematiskt fel och vars mätresultat har varians σ_1^2. Hennes skattning av vikten är $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Anders beräknar en skattning $\bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i$ baserat på m andra oberoende mätningar, $y_1, \ldots, y_m$, tagna med ett annat instrument som också saknar systematiskt fel och vars mätresultat har varians σ_2^2. De vill nu kombinera sina resultat för att skapa en bättre skattning av halten, $p^* = \alpha\bar{x} + \beta\bar{y}$ där α och β är två konstanter.

(a) Beräkna $\mathsf{E}(p^*)$ och $\mathsf{V}(p^*)$ (2p)

(b) Ange de villkor på α och β som gör p^* till en väntevärdesriktig skattare av p. (1p)

(c) Hitta de värden på α och β som ger den mest effektiva (alltså har minst varians) väntevärdesriktiga skattaren. (3p)

(d) Antag att $\sigma_1 = \sigma_2 = 10$, $m = 5$ och $n = 10$. Vad har den mest effektiva väntevärdesriktiga skattaren för varians i detta fall? (1p)

Eftersom det är tidskrävande att ta sig till sjön och samla in mätningar tänker Anders att han istället för att göra egna mätningar kan använda sitt instrument på de prov som Anna har samlat in. Han får då forsformätningar $\tilde{y}_1, \ldots, \tilde{y}_n$ som är korrelerade med Annas mätningar så att $\mathsf{C}(\tilde{y}_i, x_j) = \rho$ om $i = j$ och $\mathsf{C}(\tilde{y}_i, x_j) = 0$ om $i \neq j$. Han skapar en skattning $\tilde{y} = \frac{1}{n}\sum_{i=1}^{n} \tilde{y}_i$ och de kombinerar sina resultat till en skattning $\tilde{p}^* = \gamma\bar{x} + (1-\gamma)\tilde{y}$ där γ är en konstant.

(e) Beräkna $\mathsf{V}(\tilde{p}^*)$ och hitta det värde på γ som gör $\tilde{p}^*$ så effektiv som möjligt. (3p)

Lycka till!