Kurskod: TAMS65

Provkod: TEN1

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Tentamen tisdagen den 9 mars 2010 kl 14-18.

Hjälpmedel: Formelsamling i matematisk statistik utgiven av matematiska institutionen samt miniräknare med tömda minnen.
Betygsgränser: 8-11 poäng ger betyg 3, 11.5-14.5 ger betyg 4 och 15-18 poäng ger betyg 5.
Resultatet meddelas via LADOK.

1. En modell för dagsavkastningen (D) hos en aktie är att den är log-normalfördelad, dvs. att

$$U = \ln D \sim N(\mu, \sigma).$$

För logaritmen av dagsavkastningen, $u_i = \ln d_i$, $i = 1 \ldots, 58$, hos Ericsson B de senaste tre månaderna är medelvärdet $\bar{u} = 8.424 \cdot 10^{-4}$ och stickprovsstandardavvikelsen $s_u = 0.0117$.

a) Bilda I_μ, dvs. ett konfidensintervall för μ, med konfidensgraden 95%. (1p)

b) Observationerna har följande klassindelning.

Klassindelning	Antal u_i
$u_i < -0.02$	1
$-0.02 \leq u_i < -0.01$	9
$-0.01 \leq u_i < 0$	18
$0 \leq u_i < 0.01$	18
$0.01 \leq u_i < 0.02$	8
$0.02 \leq u_i$	4

Undersök med lämpligt χ^2-test (nivå 0.05) om det verkar rimligt att anta att logaritmen av avkastningarna är normalfördelade. (2p)

2. Ett företag erbjuder olika former av logistiska lösningar som involverar transport och distribution. De vill nu hitta den effektivaste av tre olika transportmetoder och mäter därför tiden ”dörr till dörr”, dvs. tiden det tar från avsändare till mottagare, för ett antal jämförbara och oberoende transporter. Resultat (i timmar):

Kurskod: TAMS65

Lösningar

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Lösningar till tentamen tisdagen den 9 mars 2010 kl 14-18.

1. a) Skatta μ med $\hat{\mu} = \bar{u} = 8.424 \cdot 10^{-4}$ och σ med $s_u = 0.0117$. Eftersom $U_i \sim N(\mu, \sigma)$ så gäller att $\bar{u}$ är en observation från

$$\bar{U} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Hjälpvariabel ges av

$$\frac{\bar{U} - \mu}{S_u/\sqrt{n}} \sim t(n-1)$$

och

$$\mathrm{P}\left(-a \le \frac{\bar{U} - \mu}{S_u/\sqrt{n}} \le a\right) = 0.95$$

där $n = 58$ och a fås ur en $t(57)$-tabell, $a \approx 2.00$. Vidare är

$$\mathrm{P}\left(\bar{U} - a\frac{S_u}{\sqrt{n}} \le \mu \le \bar{U} + a\frac{S_u}{\sqrt{n}}\right) = 0.95$$

och konfidensintervallet

$$I_\mu = \left(\bar{u} \mp a\frac{s_u}{\sqrt{n}}\right) = \left(8.424 \cdot 10^{-4} \mp 2.00\frac{0.0117}{\sqrt{58}}\right) = (-0.0022, 0.0039)\,.$$

b) Skatta μ med $\hat{\mu} = \bar{u} = 8.424 \cdot 10^{-4}$ och σ med $s_u = 0.0117$. Beräkna sannolikheterna för varje intervall enligt

$$\begin{aligned}
p_1 &= \mathrm{P}\,(U < -0.02) = \Phi\left(\frac{-0.02 - 8.424 \cdot 10^{-4}}{0.0117}\right)\\
&= 1 - \Phi(1.78) = 1 - 0.9625 = 0.0375,\\
p_2 &= \mathrm{P}\,(-0.02 \le U < -0.01) = \Phi(-0.93) - \Phi(-1.78)\\
&= \Phi(1.78) - \Phi(0.93) = 0.9625 - 0.8238 = 0.1387,\\
&\vdots\\
&\text{osv.}
\end{aligned}$$

Observationerna har följande klassindelning.

Klassindelning					Antal $u_i = N_i$	p_i	$np_i = 58p_i$
		u_i	$<$	-0.02	1	0.0375	2.175
-0.02	$\leq$	u_i	$<$	-0.01	9	0.1387	8.045
-0.01	$\leq$	u_i	$<$	0	18	0.2959	17.162
0	$\leq$	u_i	$<$	0.01	18	0.3102	17.992
0.01	$\leq$	u_i	$<$	0.02	8	0.1661	9.6334
0.02	$\leq$	u_i			4	0.0516	2.993

Förväntade antalet i klass nummer ett ($u_i < -0.02$) och sex ($0.02 \leq u_i$) är för litet (≤ 5), så slå ihop klasser.

Klassindelning					Antal $u_i = N_i$	p_i	$np_i = 58p_i$
		u_i	$<$	-0.01	10	0.1762	10.220
-0.01	$\leq$	u_i	$<$	0	18	0.2959	17.162
0	$\leq$	u_i	$<$	0.01	18	0.3102	17.992
0.01	$\leq$	u_i			12	0.2177	12.627

Teststorhet

$$T = \sum_{i=1}^{4} \frac{(N_i - np_i)^2}{np_i} = \cdots \approx 0.077.$$

Förkasta hypotesen om normalfördelning om $T > c$ där c fås ur en $\chi^2(\underbrace{4-1-2}_{=1})$-tabell, dvs. $c = 3.84$. Förkasta **inte** hypotesen om normalförelning.

2. a) $H_0 : \sigma_i = \sigma_j$ för $i \neq j$ mot $H_1 : \sigma_i \neq \sigma_j$ på nivån 0.05.

Teststorhet $v = s_i^2/s_j^2$ är en observation från den s.v.

$$V = \frac{S_i^2}{S_j^2} \sim F(n_i - 1, n_j - 1) \quad \text{under } H_0.$$

Förkasta H_0 om $v < a$ eller $v > b$, där a och b fås ur en F-tabell.

Ex. $i = 1, j = 2$ $\quad v = s_1^2/s_2^2 = 0.936.$

$$\begin{aligned}
0.025 &= \mathrm{P}(V < a \text{ om } V \sim F(10-1, 8-1)) = \mathrm{P}\left(\frac{1}{a} < \frac{1}{V} \text{ om } \frac{1}{V} \sim F(7,9)\right) \\
&= 1 - \mathrm{P}\left(\frac{1}{V} \leq \frac{1}{a} \text{ om } \frac{1}{V} \sim F(7,9)\right) \\
0.975 &= \mathrm{P}\left(\frac{1}{V} \leq \frac{1}{a} \text{ om } \frac{1}{V} \sim F(7,9)\right)
\end{aligned}$$

$F(7,9)$-tabell ger $1/a = 4.20$ dvs. $a = 1/4.20 = 0.24$.

$$0.025 = \mathrm{P}(V > b \text{ om } V \sim F(9,7))$$

$F(9,7)$-tabell ger $b = 4.82$. Alltså, $a = 0.24 < v = 0.936 < 4.82 = b$ och förkasta **inte** H_0. På samma sätt och samma slutsats för de två andra testen.

b) Skatta σ^2 med

$$s^2 = \frac{9s_1^2 + 7s_2^2 + 8s_3^2}{9+7+8} = 3.4648.$$

Frihetsgrader $9+7+8 = 24$.

Hjälpvariabel

$$\frac{\bar{X}_i - \bar{X}_j - (\mu_i - \mu_j)}{S\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \sim t(24)$$

som ger konfidensintervallen

$$I_{\mu_i - \mu_j} = \left(\bar{x}_i - \bar{x}_j \mp t_{0.99}(24) s \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}\right),$$

där $t_{0.99}(24) = 2.49$. De tre konfidensintervallen blir nu

$$\begin{aligned} I_{\mu_1-\mu_2} &= \left(15.00 - 18.8475 \mp 2.49\sqrt{3.4648}\sqrt{\frac{1}{10} + \frac{1}{8}}\right) \\ &= \underline{(-6.05, -1.65)} < 0, \\ I_{\mu_1-\mu_3} &= \underline{(0.49, 4.75)} > 0, \\ I_{\mu_2-\mu_3} &= \underline{(4.22, 8.72)} > 0. \end{aligned}$$

Alltså, det är troligt att $\mu_3 < \mu_1 < \mu_2$ dvs. metod 3 är effektivast.

3. a) $H_0 : \beta_2 = 0$ mot $H_1 : \beta_2 \neq 0$ på nivån 0.05.

Teststorhet

$$v = \frac{(Q_{res}^{(1)} - Q_{res}^{(2)})/p}{Q_{res}^{(2)}/(n-(k+p)-1)} = \frac{(52214508 - 19028699)/1}{19028699/76} = 132.5,$$

som är en observation från den s.v.

$$V = \frac{(Q_{res}^{(1)} - Q_{res}^{(2)})/1}{Q_{res}^{(2)}/76} \sim F(1,76) \quad \text{under } H_0.$$

Förkasta H_0 om $v > c$ där c fås ur en $F(1,76)$-tabell, $c \approx 3.99$.

$v = 132.5 > 3.99 = c$, <u>Förkasta H_0. Den kvadratiska modellen är bättre!</u>

b) Bilda I_{β_3}. Skatta β_3 med $\hat{\beta}_3 = -278$ och σ med $s = 500.377$.

Eftersom den s.v. $\hat{\beta}_3 \sim N(\beta_3, \sigma\sqrt{h_{33}})$ bilda hjälpvariabeln

$$\frac{\hat{\beta}_3 - \beta_3}{S\sqrt{h_{33}}} \sim t(76).$$

Konfidensintervallet blir då

$$I_{\beta_3} = \big(\hat{\beta}_3 \mp \underbrace{t_{0.975}(76)}_{\approx 1.99} \underbrace{s\sqrt{h_{33}}}_{=d(\hat{\beta}_3)=111.9}\big) = (-278 \mp 1.99 \cdot 111.9)$$
$$= \underline{(-500.9, -55.3)} < 0.$$

Ja, det verkar finnas skillnader mellan män och kvinnor (kvinnor får färre cancerdiagnoser).

c) Prediktionsintervallet ges av $\underline{I_{Y_0} = (2361.2, 4394.9) \text{ för män och 63-67 år}}$ (åldern är ett intervall med 65 år som mittpunkt).

4. a) Likelihood-funktionen ges av

$$L(\lambda) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \frac{\lambda(\lambda y_i)^{v-1} e^{-\lambda y_i}}{\Gamma(v)}$$

och log-likelihood-funktionen är

$$l(\lambda) = \ln L(\lambda) = \sum_{i=1}^{n} \ln(\lambda(\lambda y_i)^{v-1} e^{-\lambda y_i}) - \ln \Gamma(v)$$
$$= \sum_{i=1}^{n} (\ln \lambda + (v-1)(\ln \lambda + \ln y_i) - \lambda y_i) - \ln \Gamma(v)$$
$$= nv \ln \lambda - \lambda \sum_{i=1}^{n} y_i + konst.$$

$$\frac{dl}{d\lambda} = \frac{nv}{\lambda} - \sum_{i=1}^{n} y_i$$

$$\frac{dl}{d\lambda} = 0 \Leftrightarrow \lambda = \underline{\frac{nv}{\sum_{i=1}^{n} y_i} = \hat{\lambda}}.$$

Maximum ty

$$\frac{d^2 l}{d\lambda^2} = -\frac{nv}{\lambda^2} < 0.$$

b) Visa att $\mathrm{E}\left(\frac{1}{\hat{\lambda}}\right) = \frac{1}{\lambda}$.

$$\mathrm{E}\left(\frac{1}{\hat{\lambda}}\right) = \mathrm{E}\left(\frac{\sum_{i=1}^{n} Y_i}{nv}\right) = \frac{\sum_{i=1}^{n} \mathrm{E}\, Y_i}{nv} = \frac{\sum_{i=1}^{n} \frac{v}{\lambda}}{nv} = \frac{n\frac{v}{\lambda}}{nv} = \frac{1}{\lambda} \quad ok!$$

5. a) Teststorhet

$$z = \frac{\bar{x} - 54}{10/\sqrt{n}}$$

som är en observation från den s.v.

$$Z = \frac{\bar{X} - 54}{10/\sqrt{n}} \sim N(0,1) \quad \text{under } H_0.$$

Förkasta H_0 om $z > a = 2.33$ där $a = 2.33$ fås från en $N(0,1)$-tabell.

b)

$$\begin{aligned}
0.90 = h(\mu = 66) &= \mathrm{P}\left(\text{Förkasta } H_0 \text{ om } \mu = 66 \text{ är det sanna värdet}\right) \\
&= \mathrm{P}\left(\frac{\bar{X} - 54}{10/\sqrt{n}} > 2.33 \text{ om } \bar{X} \sim N(66, 10/\sqrt{n})\right) \\
&= \mathrm{P}\left(\frac{\bar{X} - 66}{10/\sqrt{n}} > 2.33 + \frac{54 - 66}{10/\sqrt{n}} \text{ om } \bar{X} \sim N(66, 10/\sqrt{n})\right) \\
&= 1 - \Phi\left(2.33 + \frac{54 - 66}{10/\sqrt{n}}\right) = \Phi\left(-\left(2.33 + \frac{54 - 66}{10/\sqrt{n}}\right)\right)
\end{aligned}$$

Tabell ger

$$1.28 = -\left(2.33 + \frac{54 - 66}{10/\sqrt{n}}\right)$$

dvs.

$$n = \left(\frac{10(1.28 + 2.33)}{66 - 54}\right)^2 = 9.$$

6. Beräkningar ger

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix},$$

$$(\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{n \sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} = \begin{pmatrix} h_{00} & h_{01} \\ h_{10} & h_{11} \end{pmatrix}$$

och

$$\mathbf{X}^T\boldsymbol{y} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}.$$

Detta ger

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{y} = \frac{1}{n\sum x_i^2 - \left(\sum x_i\right)^2}\begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix}\begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

$$= \frac{1}{n\sum x_i^2 - \left(\sum x_i\right)^2}\begin{pmatrix} \left(\sum x_i^2\right)\left(\sum y_i\right) - \left(\sum x_i\right)\left(\sum x_i y_i\right) \\ -\left(\sum x_i\right)\left(\sum y_i\right) + n\sum x_i y_i \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$$

och man ser att

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n}\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{\sum_{i=1}^n x_i^2 - \frac{1}{n}\left(\sum_{i=1}^n x_i\right)^2}.$$

Vidare har man att

$$\begin{aligned}
\hat{\beta}_0 &= \frac{\left(\sum x_i^2\right)\bar{y} - \bar{x}\left(\sum x_i y_i\right)}{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2} \\
&= \frac{\left(\sum x_i^2\right)\bar{y} - \frac{1}{n}\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right) - \bar{x}\left(\left(\sum x_i y_i\right) - \frac{1}{n}\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)\right)}{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2} \\
&= \bar{y} - \hat{\beta}_1\bar{x}.
\end{aligned}$$

b) Från a) ser man att $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1\bar{x}$ och

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} h_{00} & h_{01} \\ h_{10} & h_{11} \end{pmatrix}$$

där

$$h_{11} = \frac{1}{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2}.$$

Vidare är

$$\begin{aligned}
Q_{regr} &= \sum(\hat{y}_i - \bar{y})^2 = \sum(\hat{\beta}_0 + \hat{\beta}_1 x_i - (\hat{\beta}_0 + \hat{\beta}_1\bar{x}))^2 = \hat{\beta}_1^2\sum(x_i - \bar{x})^2 \\
&= \hat{\beta}_1^2\sum\left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right) = \hat{\beta}_1^2\Big(\sum x_i^2 - 2\bar{x}\underbrace{\sum x_i}_{=n\bar{x}} + n\bar{x}^2\Big) \\
&= \hat{\beta}_1^2\left(\sum x_i^2 - n\bar{x}^2\right) = \hat{\beta}_1^2\left(\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2\right) = \frac{\hat{\beta}_1^2}{h_{11}}
\end{aligned}$$

och $s^2 = Q_{res}/(n-2)$. Teststoheten v kan nu skrivas som

$$v = \frac{Q_{regr}}{Q_{res}/(n-2)} = \frac{\hat{\beta}_1^2}{h_{11}}\frac{1}{s^2} = \left(\frac{\hat{\beta}_1}{s\sqrt{h_{11}}}\right)^2 = t^2 \quad ok!$$

Data Display

```
Row     x1     x2     x3
  1  14,13  17,61  11,82
  2  11,67  19,52  10,42
  3  15,25  16,76  12,56
  4  15,58  22,58   9,46
  5  12,71  17,71  13,36
  6  17,38  18,24  15,08
  7  17,38  20,24  10,69
  8  14,92  18,12  13,63
  9  15,65         14,38
 10  15,35
```

```
Mean of x1 = 15,00
Mean of x2 = 18,8475
Mean of x3 = 12,3778
```

```
Standard deviation of x1, s1 = 1,80644
Standard deviation of x2, s2 = 1,86717
Standard deviation of x3, s3 = 1,91647
```

Modell: Vi har tre oberoende stickprov från $N(\mu_i, \sigma_i)$.

a) Är det rimligt att anta lika standardavvikelser, dvs. kan vi anta $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$? Motivera ditt svar med lämpliga konfidensintervall eller test var och ett på nivån 0.05. Det räcker att du genomför ett konfidensintervall eller test men det ska tydligt framgå hur du drar dina slutsatser. (1p)

b) Antag att $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$. Kan man med någon säkerhet hävda att någon transportmetod är effektivare än de andra. Motivera ditt svar med lämpliga konfidensintervall vart och ett med konfidensgraden 98%. (2p)

3. På Socialstyrelsens hemsida kan man hitta statistik om antalet ny cancerfall i Sverige under åren 2005 till 2008.

Ålder	2008		2007		2006		2005	
	Män	Kvinnor	Män	Kvinnor	Män	Kvinnor	Män	Kvinnor
40-44	349	860	340	849	335	823	318	706
45-49	561	1152	495	1193	594	1140	549	1183
50-54	1059	1567	1097	1634	1084	1566	1099	1566
55-59	2073	2109	2148	2229	2335	2374	2372	2433
60-64	3881	3198	3897	3258	3959	3187	3819	2942
65-69	4173	3275	4202	2970	4077	2805	4142	2808
70-74	4112	2957	4026	2893	4119	2857	4179	2793
75-79	3863	2947	3924	3030	4098	2993	4198	3005
80-84	3181	2804	3238	2722	3455	2805	3585	2873
85-	2549	2746	2419	2793	2468	2851	2358	2714

Data har analyserats enligt två olika modeller.

Modell 1: $Y = \beta_0' + \beta_1' x + \beta_2' z + \varepsilon'$,

Modell 2: $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 z + \varepsilon$, där ε-variablerna antas vara normalfördelade, x är mitten av varje åldersintervall ($x = 42, 47, 52, \ldots, 87$) och

$$z = \begin{cases} 0 & \text{för män,} \\ 1 & \text{för kvinnor.} \end{cases}$$

Datautskrifter från Minitab finns nedan.

a) Beskriver modell 2 data bättre än modell 1? Motivera ditt svar med ett lämpligt test på nivån 0.05. Både nollhypotes och mothypotes ska anges. Slutsats av testet ska tydligt framgå. (1.5p)

b) I modell 2, gör z nytta som förklaringsvariabel, dvs. kan vi skilja på män och kvinnor? Motivera ditt svar med ett lämpligt konfidensintervall eller test på nivån 0.05. (1p)

c) I Minitab utskriften för modell 2 finns också ett prediktionsintervall. Skriv upp intervallet och ange för vilken ålder och kön som det är beräknat för. (0.5p)

Modell 1

Regression Analysis: y versus x; z

```
The regression equation is
y = - 1176 + 58,8 x - 278 z

Predictor      Coef  SE Coef
Constant    -1176,1    433,5
x            58,828    6,411
z            -278,0    184,1

S = 823,475   R-Sq = 52,9%   R-Sq(adj) = 51,7%

Analysis of Variance
Source          DF         SS        MS
Regression       2   58647549  29323774
Residual Error  77   52214508    678110
Total           79  110862057
```

Modell 2

Regression Analysis: y versus x; x2; z**

```
The regression equation is
y = - 15195 + 516 x - 3,55 x**2 - 278 z

Predictor      Coef  SE Coef
Constant     -15195     1246
x            516,19    39,92
x**2        -3,5455   0,3080
z            -278,0    111,9

S = 500,377   R-Sq = 82,8%   R-Sq(adj) = 82,2%

Analysis of Variance
Source          DF         SS        MS
Regression       3   91833358  30611119
Residual Error  76   19028699    250378
Total           79  110862057

Predicted Values for New Observations
New
Obs     Fit  SE Fit       95% CI              95% PI
  1  3378,0   101,4  (3176,0; 3580,0)  (2361,2; 4394,9)

Values of Predictors for New Observations
New
Obs     x  x**2         z
  1  65,0  4225  0,000000
```

4. Avståndet (X) mellan mutationer på en DNA-sträng kan antas vara exponentialfördelade med väntevärde $1/\lambda$.

   Vid en undersökning av n oberoende DNA-strängar fann man att avsåndet mellan v mutationer på sträng i var y_i, ($i = 1, \ldots, n$). Vi vet att summan av oberoende exponentialvariabler är gammafördelad, dvs. vi har att y_i är en observation från

$$Y_i = \sum_{j=1}^{v} X_j \sim \Gamma(v, \lambda)\,, \text{ där } i = 1, \ldots, n.$$

   Täthetsfunktionen för $Y_i \sim \Gamma(v, \lambda)$ ges av

$$f(y) = \frac{\lambda(\lambda y)^{v-1} e^{-\lambda y}}{\Gamma(v)}, \text{ för } y > 0$$

   och där $\Gamma(v) = (v-1)!$.

   a) Skatta λ med maximum-likelihood-metoden. (2p)

   b) Visa att $1/\hat{\lambda}$ är en väntevärdesriktig skattning av $1/\lambda$.

   Tips: Utnyttja att $\mathrm{E}(Y) = v/\lambda$. (1p)

5. Ett slumpmässigt stickprov med n observationer ska tas från en normalfördelning $N(\mu, \sigma)$ med $\sigma = 10$ för att testa nollhypotesen

$$H_0 : \mu = 54 \quad \text{mot} \quad H_1 : \mu > 54,$$

på nivån $\alpha = 0.01$.

a) Skriv upp teststorheten och ange för vilka värden på teststorheten som nollhypotesen ska förkastas. (1p)

b) Bestäm stickprovsstorleken n som krävs så att styrkan för att förkasta H_0 när $\mu = 66$ är 0.90. (2p)

6. Låt $y_1, \ldots, y_n$ vara oberoende observationer från de s.v. $Y_1, \ldots, Y_n$, där $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma)$ och $x_1, \ldots, x_n$ är fixa tal. Vi har alltså modellen

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \text{ är obs. av } \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}}_{=\mathbf{X}} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}}_{=\boldsymbol{\beta}} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

där $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$.

a) Utgå från MK-skattningen $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{y}$ och visa att

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n}\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{\sum_{i=1}^n x_i^2 - \frac{1}{n}\left(\sum_{i=1}^n x_i\right)^2}, \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}. \end{aligned}$$

(1.5p)

b) För att testa hypotesen

$$H_0 : \beta_1 = 0 \quad \text{mot} \quad H_1 : \beta_1 \neq 0,$$

kan vi använda två olika teststorheter. En som är t-fördelad

$$t = \frac{\hat{\beta}_1}{s\sqrt{h_{11}}}$$

och en som är F-fördelad

$$v = \frac{Q_{regr}/1}{Q_{res}/(n-2)}.$$

Visa att $t^2 = v$, dvs. dessa två teststorheter är ekvivalenta. (1.5p)