Tentamen i: **Finansiell statistik**

Examinator: Mikael Möller

Hjälpmedel: Miniräknare (med, i förekommande fall, tomt programmerbart minne).
Utdelad formelsamling samt *Tabeller över statistiska fördelningar.*

Tentamensdag: 030211

Skrivningstid: 5 timmar

Återlämningsdag: Fredag 21:e Februari kl. $13^{00} - 13^{30}$ i B705 plan 7.
Därefter kan skrivningarna hämtas på expeditionen plan 7 i B-huset.

Betygskrav: För godkänt krävs preliminärt minst 30 poäng varav minst 15 poäng på tal 4, 5 och 6.
För väl godkänt krävs preliminärt minst 50 poäng varav minst 20 poäng på tal 4, 5 och 6.

Lösningarna skall vara **väl motiverade** och lätta att följa. Saknas motivering görs avdrag. Om antaganden och analys ej klart framgår kan det bli fullt avdrag.
Om mer än en lösning inlämnas på en uppgift rättas den sämsta lösningen.
Endast ett tal:s lösning per pappersark.
Tabellsamlingen får inte innehålla några anteckningar.

Observera: De studenter **som tidigare ansökt** om på statistiska institutionen **och beviljats** rätt att tillgodoräkna sig del av kursen får välja mellan två alternativ. Valet skall meddelas skrivvakten innan tentamen påbörjas.
Alternativen är att

a. göra hela tentan. Skrivtiden är då 5 timmar. Märk kuvertet "5 timmar".

b. endats göra uppgifterna under rubriken *Tillämpad statistik* (tal 4, 5 och 6). Skrivtiden är då 2.5 timmar. I detta fall är maximal poäng 30 och för godkänt krävs minst 15 poäng. *Betyget väl godkänt delas inte ut i detta fall.* Märk kuvertet ”2.5 timmar”.

Lycka till!

Sannolikhetslära och inferens

Lösning 1: Vi har följande modell

$$\begin{aligned} X &= \text{ett hushålls dagliga vattenförbrukning} \\ \mu &= 100 \\ \sigma &= 12 \\ X &\in N(\mu, \sigma) \end{aligned}$$

a.

$$
\begin{aligned}
p\left(X > 80\right) &= p\left(\frac{X-100}{12} > \frac{80-100}{12}\right) \\
&= p(z > -1.67) \\
&= p(z \leq 1.67) \\
&= 0.9525
\end{aligned}
$$

b. Låt a beteckna den högsta förbrukningen för att få rabatt.

$$
\begin{aligned}
p\left(X \leq a\right) &= 0.166 \\
p\left(\frac{X-100}{12} \leq \frac{a-100}{12}\right) &= 0.166 \\
p\left(z \leq \frac{a-100}{12}\right) &= 0.166 \\
\frac{a-100}{12} &= -0.97
\end{aligned}
$$

varav det följer att

$$a = 100 - 0.97 \cdot 12 = 88.36$$

Lösning 2: Vi har följande modell (3p)

$$
\begin{aligned}
X_{Si} &= \begin{cases} 1 & \text{stadsbo } i \text{ röstar för} \\ 0 & \text{annars} \end{cases} \qquad i = 1, 2, \ldots, 500 \\
X_{Li} &= \begin{cases} 1 & \text{landsbo } i \text{ röstar för} \\ 0 & \text{annars} \end{cases} \qquad i = 1, 2, \ldots, 500
\end{aligned}
$$

där det gäller att

$$X_S = \sum_{i=1}^{500} X_{Si} \in Bin(500, p_S) \quad \text{och} \quad X_L = \sum_{i=1}^{500} X_{Li} \in Bin(500, p_L)$$

Eftersom

$$
\begin{aligned}
500 \cdot \frac{250}{500} &= 250,\ 500 \cdot \left(1 - \frac{250}{500}\right) = 250 \\
500 \cdot \frac{400}{500} &= 400,\ 500 \cdot \left(1 - \frac{400}{500}\right) = 100
\end{aligned}
$$

gäller att centrala gränsvärdessatsen kan användas.
Ett 95 procentigt symmetriskt konfidensintervall för $p_S - p_L$ när standardavvikelsen är okänd kan skrivas (7p)

$$
\begin{aligned}
\bar{x}_S - \bar{x}_L \pm z\sqrt{\frac{\bar{x}_S(1-\bar{x}_S)}{500} + \frac{\bar{x}_L(1-\bar{x}_L)}{500}} &= 0.5 - 0.8 \pm 1.96\sqrt{\frac{0.5 \cdot 0.5}{500} + \frac{0.8 \cdot 0.2}{500}} \\
&= -0.3 \pm 0.06 \\
&= (-0.36, -0.24)
\end{aligned}
$$

Lösning 3: H_0: typ av resa är oberoende av resebyrå H_1: icke H_0
Sätt upp tabell inkluderande förväntade värden. Beräkna

$$\chi^2 = \frac{(2345 - 2090)^2}{2090} + \cdots + \frac{(4310 - 3929)^2}{3929} \approx 319$$

Nu gäller att

$$\chi^2_{0.05}((2-1)(3-1)) = 5.99 << 319$$

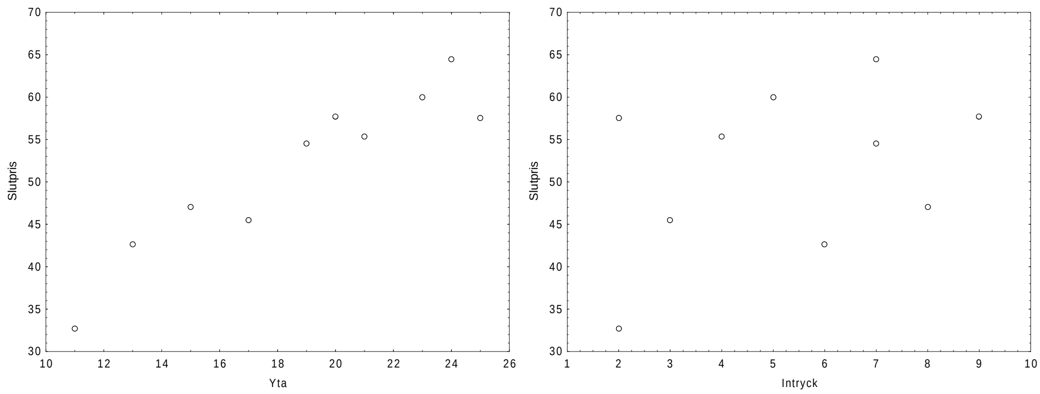
Därför finns belägg för att resebyråerna har olika profil på 5 % nivån.

Tillämpad statistik

Lösning 4: Vi börjar med att ange den fullständiga variansanalystabellen

Source	SS	df	MS	F
Regression	819,3280	2	409,6640	350,8665
Error	8,1730	7	1,1676	
Total	827,5010	9		

a. (1)



(2) Följande blir de fyra stegen

Steg 1 H_0: $\beta_1 = \beta_2 = 0$ H_1: $\neg H_0$

Steg 2 Som testvariabel tar vi

$$F = \frac{\text{Regression}/2}{\text{Error}/7}$$

där $F \in F(2, 7)$.

Steg 3 Beslutsregeln erhålls ur ekvationen

$$\begin{aligned} \alpha &= 0.05 = P(\text{förkasta } H_0 \text{ givet } H_0 \text{ sann}) \\ &= P(F > a) \\ a &= F_{0.05}(2, 7) = 4.74 \end{aligned}$$

Steg 4 Eftersom $F_{obs} = 350 > 4.74$ förkastas nollhypotesen.

Linjär regression är signifikant.

b. Den linjära modellen kan skrivas

$$y = 9.782271 + 1.870935x_1 + 1.278141x_2$$

varför prognosen blir

$$y = 9.782271 + 1.870935 \times 18 + 1.278141 \times 5 = 49.8$$

d v s 4 980 000 kronor.

c. Båda variablerna har observerade t-värden (24.56147 och 8.85137) som överstiger t-tabellens värde $t(7) = 1.895$ d v s båda variablerna är signifikanta.

d. Den justerade förklaringsgraden erhålls till

$$\bar{R} = \sqrt{1 - \frac{8.1730/7}{827.5010/9}} \approx 0.994.$$

Lösning 5: Följande gäller

a. för väntevärdet

$$\begin{aligned} \mu_t &= E(y_t) = E(y_{t-1}) + E(a_t) = \mu_{t-1} + 0 \\ &= \mu_{t-1}. \end{aligned}$$

Följ iterationen ända ut

$$\mu_t = \mu_{t-1} = \mu_{t-2} = \cdots = \mu_1 = \mu_0 = E(y_0) = 0.$$

b. för variansen erhålls på grund av oberoendet

$$\begin{aligned} V(y_t) &= V(y_{t-1} + a_t) = V(y_{t-1}) + V(a_t) \\ &= V(y_{t-1}) + \sigma^2. \end{aligned}$$

Följ iterationen ända ut

$$\begin{aligned} V(y_t) &= V(y_{t-1}) + \sigma^2 = V(y_{t-2}) + \sigma^2 + \sigma^2 = V(y_{t-3}) + 3\sigma^2 \\ &= V(y_0) + t\sigma^2 = t\sigma^2. \end{aligned}$$

c. för kovariansen gäller om $t > s$

$$\begin{aligned} C(y_t, y_s) &= C(y_s + a_{s+1} + \cdots + a_t, y_s) \\ &= C(y_s, y_s) = V(y_s) = s\sigma^2 \end{aligned}$$

och om $t < s$ erhålls

$$\begin{aligned} C(y_t, y_s) &= C(y_t, y_t + a_{t+1} + \cdots + a_s) \\ &= C(y_t, y_t) = V(y_t) = t\sigma^2. \end{aligned}$$

Detta kan sammanfattas i

$$C(y_t, y_s) = \min(t, s)\sigma^2.$$

d. Villkoret som kräver att kovariansen endast beror av tidsdifferensen är ej uppfyllt.

Lösning 6: Sätt

$$\begin{aligned} A &= \text{BBA går upp} \\ E &= \text{Vännen prognosticerar att den går upp} \end{aligned}$$

och vi söker

$$P\big(A \mid \complement E\big) = \frac{P\big(\complement E \mid A\big)\, P(A)}{P\big(\complement E \mid A\big)\, P(A) + P\left(\complement E \mid \complement A\right) P(\complement A)}.$$

Följande värden är givna

$$\begin{aligned} P\big(\complement A\big) &= 0.2 \\ P(E \mid A) &= 0.7 \\ P\big(\complement E \mid \complement A\big) &= 0.4 \end{aligned}$$

och med deras hjälp erhålls

$$\begin{aligned} P(A) &= 1 - 0.2 \\ &= 0.8 \\ P\big(\complement E \mid A\big) &= 1 - P(E \mid A) \\ &= 0.3 \end{aligned}$$

varför sökt sannolikhet blir

$$P\big(A \mid \complement E\big) = \frac{0.3 \times 0.8}{0.3 \times 0.8 + 0.4 \times 0.2} = 0.75.$$