Facit Tenta TMS145 HT 2006

1 $a = 2, b = 2$ så $f_X(x) = abx^{a-1}(1-x^a)^{b-1} = 4x - 4x^3, \;\; 0 \le x \le 1.$

(a)

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx = 4\int_0^1 x^2 - x^4 \, dx = 4\left(\left[\frac{x^3}{3}\right]_0^1 - \left[\frac{x^5}{5}\right]_0^1\right)$$
$$= 4(1/3 - 1/5) = 8/15 \approx 0.53$$

(b)

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx - \left(\frac{8}{15}\right)^2$$
$$= 4\int_0^1 x^3 - x^5 \, dx - \left(\frac{8}{15}\right)^2 = 4(1/4 - 1/6) - \left(\frac{8}{15}\right)^2 \approx 0.0489$$

(c) Vi söker maxpunkter genom att derivera och sätta derivatan till noll.

$$f_X'(x) = 4 - 12x^2 = 0 \Rightarrow x^2 = \frac{1}{3} \Rightarrow x_1 = \frac{1}{\sqrt{3}}, \;\; x_2 = -\frac{1}{\sqrt{3}}$$

Eftersom $0 \le x \le 1$ så är $x_1 = 1/\sqrt{3}$ enda lösningen och en enkel skiss av funktionen visar att detta är ett maximum.

2 Låt X_i vara antalet fel på sidan $i = 1, \ldots, 796$.

(a) Vi vet att $X_i \sim \mathrm{Po}(2)$ så enligt centrala gränsvärdessatsen (CGS) är

$$X = \sum_{i=1}^{796} X_i \sim N(796 \times 2, 796 \times 2) \text{ approximativt}$$

ty $\mathbb{E}[X_i] = \lambda = 2$ och $Var[X_i] = \lambda = 2$ för alla X_i. Eftersom $796 \times 2 = 1592$ får vi då

$$\begin{aligned}
\mathrm{P}(X > 1650) &= 1 - \mathrm{P}(X < 1650) \\
&= 1 - \mathrm{P}\left(\frac{X - 1592}{\sqrt{1592}} < \frac{1650 - 1592}{\sqrt{1592}}\right) \\
&= 1 - \Phi(1.454) = 1 - 0.927 = 0.0731
\end{aligned}$$

(b) Enligt CGS är $X \sim N(796\lambda, 796\lambda)$ approximativt. Vi söker λ så att $\mathrm{P}(X \leq 1000) = 0.90$. Detta kan skrivas som

$$\begin{aligned}
\mathrm{P}(X \leq 1000) &= \mathrm{P}\left(\frac{X - 796\lambda}{\sqrt{X - 796\lambda}} \leq \frac{1000 - 796\lambda}{\sqrt{796\lambda}}\right) \\
&= \Phi\left(\frac{1000 - 796\lambda}{\sqrt{796\lambda}}\right) = 0.90.
\end{aligned}$$

Tabell ger oss att

$$\frac{1000 - 796\lambda}{\sqrt{796\lambda}} = 1.28,$$

vilket kan genom att sätta $t^2 = \lambda$ skrivas som andragradsekvationen

$$t^2 + 0.0454t - 1.256 = 0.$$

Denna ekvation har lösningarna

$$\begin{aligned}
t_1 &= -0.0227 + 1.1257 = 1.09835 \\
t_2 &= -0.0227 - 1.1257 = -1.14379
\end{aligned}$$

där endast t_1 är intressant vilket ger oss att $\lambda = t^2 = 1.206$.

3 (a) Låt B vara händelsen att $X = 8$. Vi söker då $\mathrm{P}(A|B)$. Från texten i talet har vi följande

$$\begin{aligned}
\mathrm{P}(B|A) &= \begin{pmatrix} 10 \\ 8 \end{pmatrix} 0.8^8 0.2^2 = 0.302 \\
\mathrm{P}(B|A^c) &= \begin{pmatrix} 10 \\ 8 \end{pmatrix} 0.7^8 0.3^2 = 0.0.0014
\end{aligned}$$

ty X är binomialfördelad i båda fallen (i första fallet med parametrar $n = 10, p = 0.8$ och i andra fallet med parametrar $n = 10, p = 0.3$). Bayes sats ger nu att

$$\mathrm{P}(A|B) = \mathrm{P}(B|A)\frac{\mathrm{P}(A)}{\mathrm{P}(B)}$$

där allt är känt förutom $\mathrm{P}(B)$ vilken vi kan beräkna med hjälp av satsen om total sannolikhet,

$$\mathrm{P}(B) = \mathrm{P}(B|A)\,\mathrm{P}(A) + \mathrm{P}(B|A^c)\,\mathrm{P}(A^c) = 0.302 \times 0.5 + 0.0014 \times 0.5 = 0.1517.$$

Vi får

$$\mathrm{P}(A|B) = 0.302\frac{0.5}{0.1517} = 0.995.$$

(b) Enligt definitionen av väntevärde och med satsen om total sannolikhet får vi

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=-\infty}^{\infty} x\,\mathrm{P}_X(x) = \sum_{x=-\infty}^{\infty} x\,\mathrm{P}(X = x) \\
&= \sum_{x=-\infty}^{\infty} x\,(\mathrm{P}(X = x|A)\,\mathrm{P}(A) + \mathrm{P}(X = x|A^c)\,\mathrm{P}(A^c)) \\
&= 0.5\left(\sum_{x=-\infty}^{\infty} x\,\mathrm{P}(X = x|A) + \sum_{x=-\infty}^{\infty} x\,\mathrm{P}(X = x|A^c)\right) \\
&= 0.5\,(\mathbb{E}[X|A] + \mathbb{E}[X|A^c]) = 0.5(10 \times 0.8 + 10 \times 0.3) = 5.5.
\end{aligned}$$

Observera att vi har har använt att väntevärdet för en binomialfördelad variabel med parametrar n och p är np.

4 (a) Vi vill undersöka hur vår förklaringsvariabel x, föroreningstalet, påverkar y, renheten hos syret, under destilleringprocessen. Det mest informativa är ett konfidensintervall för parametern β_1, men det är också möjligt att utföra antingen ett t-test eller ett F-test för att pröva om $\beta_1 = 0$.

Ett konfidensintervall för β_1 är

$$I_{\beta_1} = (\hat{\beta}_1 \mp a \cdot s\sqrt{h_{11}})$$

Från utskriften fås att $\hat{\beta}_1 = -2.901$ och $s\sqrt{h_{11}} = 0.3056$. Från t-tabell med 13 frihetsgrader fås att $a = 3.012$. Detta ger att $I_{\beta_1} = (-3.82, -1.98)$.

Ett t-test ger teststatistikan

$$u = \frac{\hat{\beta}_1}{s\sqrt{h_{11}}} = -9.49$$

Vi kan förkasta nollhypotesen $\beta_1 = 0$ eftersom den kritiska gränsen är 3.012.

Ett F-test ger teststatistikan

$$v = \frac{SS_R/1}{SS_E/13} = \frac{16.4908/1}{2.3786/13} = 90.13$$

Vi kan förkasta nollhypotesen $\beta_1 = 0$ eftersom den kritiska gränsen är 9.07.

(b) Konfidensintervallet i (a) inte innehåller 0:an, vilket innebär att föroreningstalet har betydelse som förklaringsvariabel för renheten hos syret.

(c) Förklaringsgraden R^2 är högst i modellen med urbaniseringsgraden x_2 som förklarande variabel (värde 0.5367) så den är ett förnuftigt val.

5 (a) Vi testar hypotesen

$$H_0 : \sigma_{kont} = \sigma_{IHD} \quad \text{mot} \quad H_1 : \sigma_{kont} \neq \sigma_{IHD}$$

med hjälp av ett test eller konfidensintervall.

Teststorheten

$$v = s^2_{IHD}/s^2_{kont} = 0.979^2/0.347^2 \approx 7.96$$

Från F-tabell med frihetsgrader ν_1=15 (konservativt val) och ν_2=6 fås den kritiska gränsen 3.94 (exakt värde med 17 resp 6 frihetsgrader är 3.908). Den nedre gränsen fås från F-tabell med frihetsgrader 6 och 17 och blir 1/2.7=0.37. H_0 förkastas. Ett 90%-igt konfidensintervall för $\sigma_{IHD}/\sigma_{kont}$ blir (2.037, 21.48).

(b) Hypoteserna som prövas är

$$H_0 : \mu_{IHD} = \mu_{kont} \quad \text{mot} \quad H_1 : \mu_{IHD} > \mu_{kont}$$

Varianserna kan inte antas vara lika, enligt testet i (a), så Welchs method måste användas för att jämföra väntevärdena. Teststorheten

$$u = \frac{\bar{x}_{IHD} - \bar{x}_{kont}}{\sqrt{\frac{s^2_{IHD}}{18} + \frac{s^2_{kont}}{7}}} \approx 7.19$$

Den kritiska gränsen är 2.508 från t-tabell med 22 frihetsgrader så H_0 förkastas. Ett konfidensintervall för att testa hypoteserna är

$$I_{\mu_{IHD}-\mu_{kont}} = (2.57, \infty)$$

Det verkar troligt att IHD-patienterna har ett högre innehåll av endotelceller i blodet.

6 (a) Likelihood-funktionen blir, med $m = 250$:

$$L(p) = \prod_{i=1}^{m} \left[\binom{5}{x_i} \cdot p^{x_i} \cdot (1-p)^{5-x_i} \right] = \prod_{i=1}^{m} \binom{5}{x_i} \cdot p^{\sum_{i=1}^m x_i} \cdot (1-p)^{\sum_{i=1}^m (n-x_i)}$$

Detta ger att

$$l(p) = \ln(\prod_{i=1}^{m} \binom{5}{x_i}) + \sum_{i=1}^{m} x_i \cdot \ln(p) + \sum_{i=1}^{m} (n - x_i) \cdot \ln(1-p)$$

Genom att derivera och sätta derivatan lika med noll får man att

$$\hat{p} = \frac{\sum_{i=1}^m x_i}{5m}$$

vilket är ett maximum. Sätter man in värdena fås att $\hat{p} = 664/(250 \cdot 5) = 0.5312$

(b) Ett χ^2-test görs för att testa om binomialfördelnings-antagandet är korrekt. I uppgiften anges att $X \sim Bin(5, p)$ och därmed finns

6 kategorier, för vilka vi behöver beräkna sannolikheter.

$$
\begin{aligned}
p_0 &= P(X=0) = \binom{5}{0} p^0 (1-p)^{5-0} = (1-p)^5 \\
p_1 &= P(X=1) = \binom{5}{1} p^1 (1-p)^{5-1} = 5 \cdot p \cdot (1-p)^4 \\
&\vdots \\
p_5 &= P(X=5) = \binom{5}{5} p^5 (1-p)^{5-5} = p^5
\end{aligned}
$$

Genom att sätta in det skattade värdet på p från (a) fås att $(\hat{p}_0, \ldots, \hat{p}_5) = (0.022643, 0.12829, 0.29072, 0.32942, 0.18663, 0.042295)$. En frekvenstabell ser ut som följer

Kategori	0	1	2	3	4	5
Förv. frekvens ($m\hat{p}_i$)	5.66	32.07	72.68	82.35	46.66	10.57
Obs. frekvens (N_i)	25	65	33	21	60	46

Teststorheten är, med $m = 250$ som tidigare, $Q = \sum_{i=0}^{5} \frac{(N_i - m\hat{p}_i)^2}{m\hat{p}_i} \approx 290$.

Den kritiska gränsen är 9.488 från χ^2-tabell med (6-1-1)=4 frihetagrader. H_0 förkastas. Antagandet om binomialfördelning är mycket troligen falskt.