1. (a) Vi har

$$\mathsf{E}(X) = \sum_{k=0}^{5} kf(k) = 0 \cdot 0.1 + 1 \cdot 0.05 + 2 \cdot 0.1 + 3 \cdot 0.25 + 4 \cdot 0.2 + 5 \cdot 0.3 = 3.3,$$
$$\mathsf{E}(X^2) = \sum_{k=0}^{5} k^2 f(k) = 0^2 \cdot 0.1 + 1^2 \cdot 0.05 + 2^2 \cdot 0.1 + 3^2 \cdot 0.25 + 4^2 \cdot 0.2 + 5^2 \cdot 0.3 = 13.4.$$

Alltså är $\mathsf{V}(X) = \mathsf{E}(X^2) - \mathsf{E}(X)^2 = 13.4 - 3.3^2 = 2.51$.

(b) Låt X_i beteckna antalet poäng student i får. Vi har då $\bar{X} = \frac{1}{45}\sum_{i=1}^{45} X_i$ och

$$\mathsf{E}(\bar{X}) = \frac{1}{45}\sum_{i=1}^{45} \mathsf{E}(X_i) = 3.3,$$
$$\mathsf{V}(\bar{X}) = \frac{1}{45^2}\sum_{i=1}^{45} \mathsf{V}(X_i) = 2.51/45 = 0.0558.$$

CGS ger alltså att $\bar{X}$ är approximativt $\mathsf{N}(3.3, 0.0558)$-fördelad.

(c) Vi söker

$$\mathsf{P}(\bar{X} > 3) = 1 - \mathsf{P}(\bar{X} \leq 3) = 1 - \Phi\left(\frac{3 - 3.3}{\sqrt{0.0558}}\right)$$
$$= 1 - \Phi(-1.2703) = 0.898.$$

2. (a) Medianen hos fördelningen för X är det tal M så att $\mathsf{P}(X \leq M) = \mathsf{P}(X \geq M) = 0.5$.

(b) Om vi vill utföra hypotestestet

$$H_0 : M = M_0$$
$$H_1 : M < M_0$$

kan vi definiera Q_+ som antalet observerade värden större än M_0. Om H_0 är sann är $Q_+ \sim \text{Bin}(n, 0.5)$. Vi kan alltså beräkna p-värdet för testet som $\mathsf{P}(X < Q_+)$ där $X \sim \text{Bin}(n, 0.5)$. Vi förkastar H_0 om denna sannolikhet är mindre än den valda signifikansnivån.

(c) Vi vill nu utföra testet med $M_0 = 0$, och räknar då antalet observationer som är positiva, $Q_+ = 3$. p-värdet ges alltså av

$$\mathsf{P}(X \leq 3) = \mathsf{P}(X = 0) + \mathsf{P}(X = 1) + \mathsf{P}(X = 2) + \mathsf{P}(X = 3)$$
$$= \binom{10}{0}0.5^0 0.5^1 0 + \binom{10}{1}0.5^1 0.5^9 + \binom{10}{2}0.5^2 0.5^8 + \binom{10}{3}0.5^3 0.5^7$$
$$= 0.5^{10}(1 + 10 + 45 + 120) = 0.1719$$

Eftersom p-värdet är större än α kan vi inte förkasta H_0. Studien har alltså inte visat att dieten minskar medianvikten.

3. (a) Vi börjar med att beräkna medelvärdena för varje grupp samt det totala medelvärdet:

$$\bar{y}_1 = 1.816, \quad \bar{y}_2 = 1.02, \quad \bar{y}_3 = 0.5, \quad \bar{y} = 1.2217.$$

Vi kan nu beräkna

$$SS_A = 5(\bar{y}_1 - \bar{y})^2 + 4(\bar{y}_2 - \bar{y}) + 3(\bar{y}_3 - \bar{y}) = 3.4912,$$
$$SS_E = \sum_{ij}(y_{ij} - \bar{y}_i)^2 = \sum_j (y_{1j} - \bar{y}_1)^2 + \sum_j (y_{2j} - \bar{y}_2)^2 + \sum_j (y_{3j} - \bar{y}_3)^2 = 1.3169.$$

Summerar vi dessa får vi $SS_{Tot} = 4.8082$. Vi har nu allt som behövs för att fylla i värdena i variansanalystabellen:

Variation	Kvadratsumma	Frihetsgrader	Medelkvadratsumma	Teststorhet
Faktor A	3.4912	2	1.74562	11.93
Residual	1.3169	9	0.14632	
Total	4.8082	11		

(b) Vi ska jämföra värdet på teststorheten med det kritiska värdet $F_{0.05}(2, 9) = 4.2565$. Eftersom teststorheten är större än detta värde kan vi förkasta nollhypotesen att faktorn inte påverkar halten. Alltså har vi en signifikant variation i HG halterna mellan de olika sjöarna.

4. (a) Eftersom variablerna är oberoende ges den simultana täthetsfunktionen av

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) = \frac{1}{5}e^{-x/5}\frac{1}{10}e^{-y/10} = \frac{1}{50}e^{-(x/5+y/10)}$$

för $x, y > 0$.Vi har $\mathsf{E}(Z) = \mathsf{E}(X) + \mathsf{E}(Y) = 5 + 10 = 15$.

(b) Vi beräknar

$$\begin{aligned}
\mathsf{P}(X + Y < 20) &= \int_0^{20}\int_0^{20-x} \frac{1}{50}e^{-(x/5+y/10)}dydx = \frac{1}{50}\int_0^{20} e^{-x/5}\int_0^{20-x} e^{-y/10}dydx \\
&= \frac{1}{50}\int_0 20e^{-x/5}\left[-10e^{-y/10}\right]_0^{20-x} dx = \frac{1}{5}\int_0^{20} -e^{-2}e^{-x/10} + e^{-x/5}dx \\
&= \frac{1}{5}\left[10e^{-2}e^{-x/10} - 5e^{-x/5}\right]_0^{20} = 1 + e^{-4} - 2e^{-2}.
\end{aligned}$$

Den sökta sannolikheten är $\mathsf{P}(X + Y > 20) = 1 - \mathsf{P}(X + Y < 20) = 2e^{-2} - e^{-4} \approx 0.2524$.

5. (a) Skattningen av parametrarna fås genom att beräkna

$$\begin{pmatrix}\beta_0^* \\ \beta_1^* \\ \beta_2^*\end{pmatrix} = (X^TX)^{-1}X^TY = \begin{pmatrix} 3.3503 & -0.4406 & -0.1982 \\ -0.4406 & 0.0754 & -0.0135 \\ -0.1982 & -0.0135 & 0.1106 \end{pmatrix}\begin{pmatrix}534.28 \\ 3394.4 \\ 1438.8\end{pmatrix} = \begin{pmatrix}9.2555 \\ 1.1102 \\ 7.4126\end{pmatrix}.$$

(b) Vi beräknar först $Q_0 = Y^TY - \beta^TX^*Y = 216.32$ och $s = \sqrt{Q_0/(15-3)} = 1.8505$. Vi behöver också $t_{0.025}(12) = 2.1788$ och kan nu beräkna

$$I_{\beta_1} = [\beta_1^* - t_{0.025}(12)s\sqrt{0.0754}, \beta_1^* + t_{0.025}(12)s\sqrt{0.0754}] = [-1.43, 3.6504]$$
$$I_{\beta_2} = [\beta_2^* - t_{0.025}(12)s\sqrt{0.1106}, \beta_2^* + t_{0.025}(12)s\sqrt{0.1106}] = [4.3361, 10.489]$$

Eftersom $0 \in I_{\beta_1}$ kan vi inte förkasta nollhypotesen $H_0 : \beta_1 = 0$. Däremot kan vi förkasta $H_0 : \beta_2 = 0$ eftersom $0 \notin I_{\beta_2}$.

(c) Skattningarna ges av $\beta_2^* = S_{xy}/S_{xx} = 7.526$ och $\beta_0 = \bar{y} - \beta_1^*\bar{x}_2 = 16.332$.

(d) Vi beräknar först $Q_0 = S_{yy} - S_{xy}^2/S_{xx} = 41.092$ och $s = \sqrt{Q_0/(15-2)} = 1.7779$. Dessutom är $t_{0.025}(13) = 2.1604$, och vi kan nu beräkna

$$I_{\beta_2} = [\beta_2^* - t_{0.025}(13)s/\sqrt{S_{xx}}, \beta_2^* + t_{0.025}(13)s/\sqrt{S_{xx}}] = [6.2627, 8.7893].$$

Eftersom detta intervall inte innehåller 0 kan vi förkasta $H_0 : \beta_2 = 0$.

(e) Förklaringsgraden R^2 definieras som andelen av variationen i datan som förklaras av regressionslinjen. För Modell 2 kan vi beräkna denna som

$$R^2 = \frac{\beta_2^* S_{xy}}{S_{yy}} = 0.92723.$$

(f) Eftersom Modell 1 endast har en något högre förklaringsgrad och eftersom parametern β_2 inte är signifikant i den modellen verkar Modell 2 vara mer lämplig för datan.

6. (a) Fördelningsfunktionen ges av

$$F(x) = \int_0^x f(y)dy = \int_0^x a(1+y)^{-(a+1)}dy = [-(1+y)^{-a}]_0^x = 1-(1+x)^{-a}$$

(b) $\mathsf{P}(X < 0.1) = F(0.1) = 1 - (1+0.1)^{-3} = 0.2487.$

(c) Vi söker

$$\begin{aligned} \mathsf{P}(X < 0.2 | X > 0.1) &= \frac{\mathsf{P}(0.1 < X < 0.2)}{\mathsf{P}(X > 0.1)} = \frac{F(0.2) - F(0.1)}{1 - F(0.1)} \\ &= \frac{1 - (1+0.2)^{-3} - 1 + (1+0.1)^{-3}}{(1+0.1)^{-3}} \\ &= \frac{(1+0.1)^{-3} - (1+0.2)^{-3}}{(1+0.1)^{-3}} = 1 - \frac{1.2^{-3}}{1.1^{-3}} \\ &= 0.2297 \end{aligned}$$

Vi ser att risken att gå i konkurs under det andra året alltså är något mindre än risken att gå i konkurs under det första året, givet att företaget överlevde det första året. Så fördelningen verkar alltså fånga effekten att risken för konkurs minskar med tiden.

(d) Vätevärdet av X kan beräknas med hjälp av partialintegration:

$$\begin{aligned} \mathsf{E}(X) &= \int_0^\infty xa(1+x)^{-(a+1)}dx = [-x(1+x)^{-a}]_0^\infty + \int_0^\infty (1+x)^{-a}dx \\ &= \left[\frac{1}{1-a}(1+x)^{1-a}\right]_0^\infty = -\frac{1}{1-a} = \frac{1}{a-1}. \end{aligned}$$

(e) Log-likelihood funktionen ges av

$$\ell(a) = \sum_{i=1}^n \log\left(a(1+x_i)^{-(a+1)}\right) = \sum_{i=1}^n \log(a) - (a+1)\log(1+x_i) = n\log(a) - \sum_{i=1}^n (a+1)\log(1+x_i).$$

Derivatan med avseende på a är

$$\frac{\partial \ell}{\partial a} = \frac{n}{a} - \sum_{i=1}^n \log(1+x_i).$$

Vi sätter derivatan lika med noll och löser ut a, vilket ger

$$a_{ML}^* = \frac{n}{\sum_{i=1}^n \log(1+x_i)}.$$