1. (a) Den totala vikten är $Z = X + Y$, som är Normalfördelad eftersom det är en summa av två (beroende) Normalfördelade variabler. Väntevärdet är $E(Z) = E(X) + E(Y) = 13$, och variansen ges av $V(Z) = V(X) + V(Y) + 2C(X,Y) = 2^2 + 1^2 + 2 = 7$, så att sammanfattningsvis $Z \in N(13, \sqrt{7})$. (3p)

   (b) Sannolikheten för överviktigt bagage är $P(Z > 20) = 1 - P(Z \leq 20) = 1 - \Phi\left(\frac{20-13}{\sqrt{7}}\right) \approx 1 - \Phi(2.6458) \approx 0.004075$. (Tabell: mellan 0.00402 och 0.00415) (3p)

   (c) Risken att minst en av de 100 passagerarna har överviktigt bagage ges av $P(\text{minst ett } Z_i \text{ är} > 20) = 1 - P(\text{alla } Z_i \text{ är} \leq 20) = 1 - P(\text{ett givet } Z_i \text{ är} \leq 20)^{100} = 1 - P(Z \leq 20)^{100} = 1 - (1 - 0.004075)^{100} \approx 0.3352$. (Tabell: mellan 0.33 och 0.34) (4p)

2. (a) Den sökta sannolikheten är (5p)

$$P(X > 40) = \int_{40}^{\infty} f_X(x)\mathrm{d}x = \int_{40}^{\infty} \frac{5}{(5+x)^2}\mathrm{d}x = \left[\frac{-5}{5+x}\right]_{40}^{\infty} = 0 + \frac{5}{45} = 1/9 \approx 0.1111.$$

   (b) Antalet "lyckade" försök att överskrida gränsen under ett år är $Y \in \mathrm{Bin}(365, p)$, där $p = 1/9$. Eftersom $365 \cdot 1/9 \cdot 8/9 \approx 36.05 > 10$ kan vi tillåta oss att normalapproximera binomialfördelningen med $Y \approx \mathrm{N}(365/9, \sqrt{36.0494})$, och får den sökta sannolikheten genom (5p)

$$P(Y \leq 50) = P(Y \leq 50.5) \approx \Phi(\frac{50.5 - 40.5556}{6.004}) \approx \Phi(1.6563) \approx 0.9512.$$

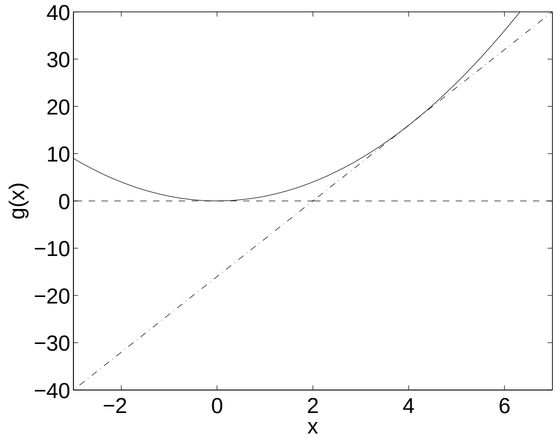
   (Exakt räkning ger $P(Y \leq 50) \approx 0.9477$.)

3. Antag att $X \in N(\mu, 2)$.

   (a) Eftersom $V(X) = E(X^2) - E(X)^2$ så får vi att $E(X^2) = V(X) + E(X)^2 = 2^2 + \mu^2 = 4 + \mu^2$. (2p)

   (b) Sätt $g(x) = x^2$, och beräkna derivatan, $g'(x) = 2x$. Gauss-approximation ger nu att $E(X^2) \approx g(E(X)) = \mu^2$. För $\mu = 0$ och 4 ger Gauss-appr. $E(X^2) \approx 0$ och 16, att jämföra med de exakta värdena 4 och 20. Gauss-appr. underskattar alltså här tydligt det sökta väntevärdet, men det relativa felet är mindre då $\mu = 4$. (4p)

   (c) Gauss-appr. ger nu $V(X^2) \approx g'(E(X))^2 V(X) = (2\mu)^2 \cdot 2^2 = 16\mu^2$, dvs $V(X^2) \approx 0$ och 256. Det är här uppenbart att Gauss-approximationen är olämplig då $\mu = 0$, eftersom $E(X^2)$ är ett mått på hur mycket X i (kvadrat-) medel avviker från 0. Taylor-utvecklingen av $g(x)$ kring μ, som Gauss-approximationen bygger på, ignorerar helt variationen i X och avbildar alla x-värden på 0 när $\mu = 0$, men bibehåller en del variation när $\mu = 4$. Figuren visar Taylor-utvecklingen av $g(x)$ i de två fallen (heldragen= $g(x)$, streckad=fallet $\mu = 0$, streck-prickad=fallet $\mu = 4$). Vi ser då också att om variansen varit större hade vi inte ens kunnat vänta oss en bra approximation ens för fallet $\mu = 4$. (4p)



4. (a) Det är rimligt att anta att de olika observationerna i ett stickprov är oberoende av varandra, och att tiderna har ändlig varians. Eftersom de är dragna slumpmässigt ur en viss population kan vi också se dem som observationer av en gemensam fördelning. Förutsättningarna för CGS är därför uppfyllda, och alltså blir stickprovsmedelvärdet $\bar{x}$ ungefär normalfördelat, med något väntevärde $E(X_i)$ och standardavvikelse σ_x. Medelfelet för skattningen av väntevärdet ges av $s_x/\sqrt{n}$, $n = 20$, så att det sökta konfidensintervallet ges av (5p)

$$
\begin{aligned}
I_{E(X_i)} &: \bar{x} \pm t_{0.025}(n-1) \cdot s_x/\sqrt{n} = 27.10 \pm t_{0.025}(19) \cdot 10.5526/\sqrt{20} \\
&\approx [\text{t-kvantil } = 2.0930] \approx [22.1612,\ 32.0388] \approx [22.16,\ 32.04] \\
&\approx [\text{t-kvantil från tabell } = 2.09] \approx [22.178,\ 32.032]
\end{aligned}
$$

(b) Skillnaden i medelvärde före och efter är ca 2, vilket riskerar att överskuggas av variationen mellan individer. Det är däremot rimligt att anta en modell där varje försöksperson har sitt eget personliga väntevärde, och att förändringen ska mätas relativt dessa. Situationen är alltså en vanlig ”stickprov i par”, där vi ska testa om det gemensamma väntevärdet för Z_i är 0 eller inte: $H_0 : E(Z) = 0$, $H_1 : E(Z) \neq 0$. (10p)

Med en likadan motivering som i (a) kan vi normalapproximera skattningen av väntevärdet, $\bar{z}$, och vi kan testa hypoteserna med konfidensmetoden: Ett 95% konfidensintervall för $E(Z)$ ges av

$$
\begin{aligned}
I_{E(Z)} &: \bar{z} \pm t_{0.025}(n-1) \cdot s_z/\sqrt{n} = -1.750 \pm t_{0.025}(19) \cdot 3.8781/\sqrt{20} \\
&\approx [\text{t-kvantil } = 2.0930] \approx [-3.560,\ 0.0650] \\
&\approx [\text{t-kvantil från tabell } = 2.09] \approx [-3.5624,\ 0.0624]
\end{aligned}
$$

Eftersom 0 ligger i intervallet kan vi inte förkasta H_0 (på nivån 0.05).

Ett annat alternativ är att ställa upp teststorheten $T = |(\bar{z} - 0)/(s_z/\sqrt{n})|$ och förkasta H_0 om $T > t_{0.025}(n-1)$. Här får vi $T = 2.0181 < 2.0930 = t_{0.025}(19)$ och kan alltså inte heller här förkasta H_0. I själva verket är detta test ekvivalent med konfidensintervallmetoden i denna problemformulering. (Tabell: $t_{0.025}(19) = 2.09$)

(c) Att ett z_i är negativt betyder att det tog längst tid att montera byggsatsen första gången. Här vet vi egentligen inte vilken fördelning Z har, men hoppas att den kan approximeras med en normalfördelning. Den sökta sannolikheten är alltså (5p)

$$
P(Z \leq -5) \approx \Phi\left(\frac{-5 - (-1.750)}{3.8781}\right) \approx \Phi(-0.8380) \approx 0.2010.
$$

Tabell: $1 - \Phi(0.8380) \approx 1 - 0.7995 = 0.2005$

5. (a) Skattningarna av α och β ges enligt formelsamlingen av (8p)

$$
\begin{aligned}
\beta^* &= \frac{S_{xy}}{S_{xx}} = \frac{535281}{798444} \approx 0.6704 \\
\alpha^* &= \bar{y} - \beta^* \bar{x} = 695.6 - \beta^* \cdot 762.7 \approx 184.3
\end{aligned}
$$

För att få fram medelfelen (skattningar av skattningarnas standardavvikelse) behöver vi en skattning av σ, och eftersom frihetsgraderna är $n - 2 = 28$ tar vi $s = \sqrt{Q_0/(n-2)} = \sqrt{44993/28} \approx 40.0861$. Medelfelen blir

$$
\begin{aligned}
d(\beta^*) &= \frac{s}{\sqrt{S_{xx}}} \approx 0.0449 \\
d(\alpha^*) &= s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} \approx 34.9896
\end{aligned}
$$

(b) Enklaste sättet att konstruera ett konfidensintervall för $1/\beta$ med rätt konfidensgrad är att utgå från ett konfidensintervall för β, (12p)

$$I_\beta = [a, b] = \beta^* \pm t_{0.025}(n-2)\, d(\beta^*)$$
$$\approx [\text{t-kvantil från tabell =2.05}] \approx 0.6704 \pm 2.05 \cdot 0.0449 \approx [0.5784,\ 0.7624].$$

Eftersom $[a, b]$ täcker β med sannolikhet 0.95 och $a, b > 0$ kommer intervallet $[1/b, 1/a] \approx [1.3116,\ 1.7290]$ att täcka $1/\beta$ med sannolikhet 0.95:

$$\{a < \beta < b\} \Leftrightarrow \{a < \beta \text{ och } \beta < b\} \Leftrightarrow \{1/\beta < 1/a \text{ och } 1/b < 1/\beta\} \Leftrightarrow \{1/b < 1/\beta < 1/a\}$$

En omständligare och dessutom oprecis metod är att använda Gauss-approximation. Problemet med detta är att $1/\beta^*$ inte är normalfördelad, vilket gör att vi inte har någon garanti för att konfidensgraden blir tillräckligt stor. Om vi ändå utför approximationen får vi $E(1/\beta^*) \approx 1/\beta$ och $V(1/\beta^*) \approx V(\beta^*)/\beta^4$, så att $D(1/\beta^*) = D(\beta^*)/\beta^2$, som kan skattas med $d(1/\beta^*) = d(\beta^*)/(\beta^*)^2 = 0.0998$. Om nu $1/\beta^*$ vore normalfördelad skulle ett 95% konfidensintervall baserat på normalfördelningskvantil ($d(1/\beta^*)$ är inte χ^2-fördelad) ges av $1/\beta^* \pm 1.96 \cdot d(\beta^*)/(\beta^*)^2 \approx [1.2960,\ 1.6973]$. Detta intervall är lite smalare än det exakta vi räknade ut tidigare, men har nästan samma läge, vilket tyder på att $1/\beta^*$ är tillräckligt normalfördelad för att tillåta approximationen. Det hade dock varit svårt att avgöra analytiskt utan att räkna ut det exakta intervallet. Att intervallet är smalare tyder också på att konfidensgraden är mindre än 95%, men det går inte att avgöra säkert utan mer analys.

6. (a) Eftersom $X \in \text{Po}(\theta)$ och $Y \in \text{Po}(4\theta)$ så blir (8p)

$$E(\theta_1^*) = E\left(\frac{1}{2}\left(\frac{X}{1} + \frac{Y}{4}\right)\right) = \frac{1}{2}\left(\frac{E(X)}{1} + \frac{E(Y)}{4}\right) = \frac{1}{2}\left(\frac{\theta}{1} + \frac{4\theta}{4}\right) = \theta$$
$$E(\theta_2^*) = E\left(\frac{X+Y}{5}\right) = \frac{E(X)+E(Y)}{5} = \frac{\theta + 4\theta}{5} = \theta$$
$$E(\theta_3^*) = E\left(\frac{X+4Y}{17}\right) = \frac{E(X)+4E(Y)}{17} = \frac{\theta + 4\cdot 4\theta}{17} = \theta$$

och eftersom de är oberoende blir

$$V(\theta_1^*) = \frac{V(X)}{2^2} + \frac{V(Y)}{8^2} = \frac{\theta}{2^2} + \frac{4\theta}{8^2} = \theta\left(\frac{1}{4} + \frac{1}{16}\right) = \frac{5\theta}{16} = 0.3125\,\theta$$
$$V(\theta_2^*) = \frac{V(X)+V(Y)}{5^2} = \frac{\theta+4\theta}{25} = \frac{\theta}{5} = 0.2\,\theta$$
$$V(\theta_3^*) = \frac{V(X)+4^2V(Y)}{17^2} = \frac{\theta + 4^2\cdot 4\theta}{17^2} = \frac{65\theta}{289} \approx 0.225\,\theta$$

Alla tre skattningarna är alltså väntevärdesriktiga, och eftersom $V(\theta_1^*) > V(\theta_3^*) > V(\theta_2^*)$, så är θ_2^* den effektivaste skattningen.

(b) Vi väljer att använda θ_2^* eftersom det är den effektivaste skattningen av de tre givna, och får $\theta_2^* = \frac{x+y}{5} = \frac{22+73}{5} = \frac{95}{5} = 19$. Eftersom $D(\theta^*) = \sqrt{\theta/5}$ enligt (a) får vi medelfelet som $d(\theta^*) = \sqrt{\theta_2^*/5} = \sqrt{3.8} \approx 1.9494$. Eftersom $\theta_2^* = 19 > 15$ tror vi att det är ok att använda normalapproximation, och får ett intervall med approximativ konfidensgrad 95% genom $I_\theta : \theta_2^* \pm \lambda_{0.025}\, d(\theta_2^*) = 19 \pm 1.96 \cdot 1.9494 = [15.1792,\ 22.8208]$. (Notera att intervallets undre gräns är så pass låg att vår normalapproximation kan vara lite riskabel.) (8p)

(c) Sätt $Z = X + Y$. Då är $Z \in \text{Po}(5\theta)$. Med $a = 5$ och $\mu = \theta$ får vi då genast att ML-skattningen av θ baserat på en observation z av Z ges av $z/5$, vilket matchas av $\theta_2^* = (x+y)/5$. (4p)

För fullständighets skull bör vi kontrollera att detta verkligen är ML-skattningen baserat på de två observationerna; man skulle ju kunna tänka sig att de enskilda observationerna innehöll mer information om θ än vad summan av dem gör. Den simultana sannolikhetsfunktionen ges av

$$p_{X,Y}(x,y) = e^{-\theta}\frac{\theta^x}{x!} \cdot e^{-4\theta}\frac{(4\theta)^y}{y!} = e^{-5\theta}\frac{\theta^{x+y}4^y}{x!\,y!} = \frac{4^y}{x!\,y!}\exp\left(-5\theta + (x+y)\ln\theta\right)$$

Läget för maximum m.a.p. θ beror endast av $x+y$, och vi är klara.