UPPSALA UNIVERSITET
Matematiska institutionen
Niclas Petersson

Statistik IP1
Byggteknik
2006-02-22

# Lösningsskisser till Inlämningsuppgift 2

1.a) Andelen kvinnor som är längre än medellängden för män $\Leftrightarrow$ Sannolikheten att en slumpmässigt vald kvinna är längre än medellängden för män.

$$\begin{aligned}\mathbb{P}(X > 179.0) &= 1 - \mathbb{P}(X \leq 179.0) = 1 - \Phi(\frac{179.0 - 165.5}{6.15}) = 1 - \Phi(2.20) = \\ &= 1 - 0.9861 = 0.0139 = \mathbf{0.014}\end{aligned}$$

1.b) Låt $Z = X - Y$ = *längdskillnad på en slumpmässigt vald kvinna och en slumpmässigt vald man*. Då är $Z \in N(165.5 - 179.0, 37.8 + 46.9) = N(-13.5, 84.7)$.

$$\begin{aligned}\mathbb{P}(Z > 0) &= 1 - \mathbb{P}(Z \leq 0) = 1 - \Phi(\frac{0 - (-13.5)}{\sqrt{84.7}}) = 1 - \Phi(1.47) = \\ &= 1 - 0.9292 = 0.0708 = \mathbf{0.071}\end{aligned}$$

1.c) Beräkna först sannolikheten att *en* slumpmässigt vald kvinna är 161.5 – 174.5 cm lång.

$$\begin{aligned}\mathbb{P}(161.5 < X < 174.5) &= \mathbb{P}(X < 174.5) - \mathbb{P}(X < 161.5) = \Phi(\frac{174.5 - 165.5}{6.15}) \\ &- \Phi(\frac{161.5 - 165.5}{6.15}) = \Phi(1.46) - \Phi(-0.65) = \Phi(1.46) \\ &- 1 + \Phi(0.65) = 0.9279 - 1 + 0.7422 = 0.6701\end{aligned}$$

Låt $U$ = *antalet kvinnor som är* $161.5 - 174.5$ *cm långa (av 6 slumpmässigt utvalda kvinnor)*. Då är $U \in Bin(6, 0.67)$.

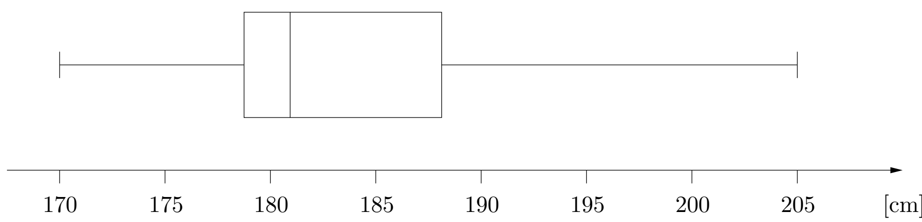
$$\mathbb{P}(U = 6) = \binom{6}{6} 0.67^6 \cdot (1 - 0.67)^0 = 0.67^6 = 0.09054 = \mathbf{0.091}$$

1.d) Låt $\bar{X} = \frac{1}{6}(X_1 + X_2 + ... + X_6)$ där $X_1$, $X_2$,... är oberoende stokastiska variabler med fördelningen $N(165.5, 37.8)$. Då är $\bar{X} \in N(165.5, \frac{37.8}{6}) = N(165.5, 6.30)$.

$$\begin{aligned}\mathbb{P}(161.5 < \bar{X} < 174.5) &= \mathbb{P}(\bar{X} < 174.5) - \mathbb{P}(\bar{X} < 161.5) = \Phi(\frac{174.5 - 165.5}{\sqrt{6.30}}) \\ &- \Phi(\frac{161.5 - 165.5}{\sqrt{6.30}}) = \Phi(3.59) - \Phi(-1.59) = \Phi(3.59) \\ &- 1 + \Phi(1.59) = 0.9998 - 1 + 0.9441 = 0.9439 = \mathbf{0.94}\end{aligned}$$

2.a) Från bilaga 2 inses/beräknas följande:

| | | | | | |
|---|---|---|---|---|---|
| minsta värde: | 170 | median: | 181 | undre kvartil: | 179 |
| största värde: | 205 | variationsbredd: | 35 | övre kvartil: | 188 |



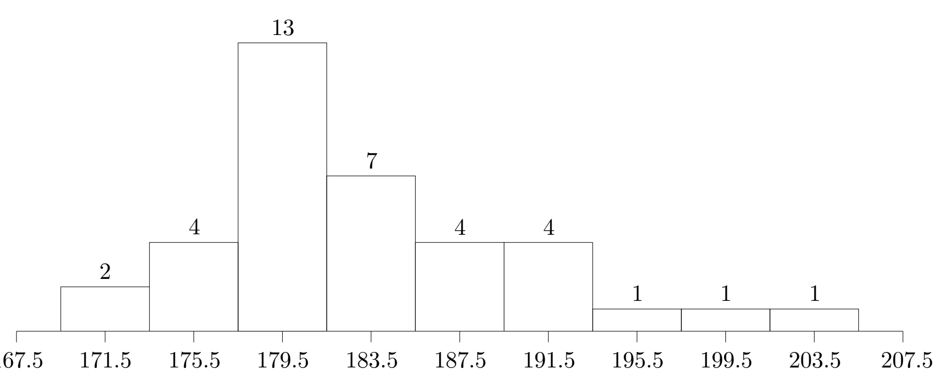

Figur 1: Lådagram över mäns längder.

2.b) Klassindelningen måste (åtminstone) innehålla intervallet $(169.5, 205.5)$. Då $205.5 - 169.5 = 36$ finns det 2 väldigt naturliga alternativ.

**Alternativ 1:**
Välj 9 st klasser med klassbredd 4 cm. Med 37 mätvärden kommer det ge ett snitt på ca 4 st mätvärden per klass. Detta val ger som synes både snygga klassgränser och ett lagom antal klasser.

| Klassgränser | Frekvens $f_i$ | Rel. frek. (i %) | Klassmitt $y_i$ | $f_i y_i$ | $f_i y_i^2$ |
|---|---|---|---|---|---|
| 169.5-173.5 | 2 | 5.4 | 171.5 | 343 | 58824.5 |
| 173.5-177.5 | 4 | 10.8 | 175.5 | 702 | 123201 |
| 177.5-181.5 | 13 | 35.1 | 179.5 | 2333.5 | 418862.25 |
| 181.5-185.5 | 7 | 18.9 | 183.5 | 1284.5 | 235705.75 |
| 185.5-189.5 | 4 | 10.8 | 187.5 | 750 | 140625 |
| 189.5-193.5 | 4 | 10.8 | 191.5 | 766 | 146689 |
| 193.5-197.5 | 1 | 2.7 | 195.5 | 195.5 | 38220.25 |
| 197.5-201.5 | 1 | 2.7 | 199.5 | 199.5 | 39800.25 |
| 201.5-205.5 | 1 | 2.7 | 203.5 | 203.5 | 41412.25 |
| | 37 | 99.9 | | 6777.5 | 1243340.25 |

Med klassindelningen ovan får man följande histogram.

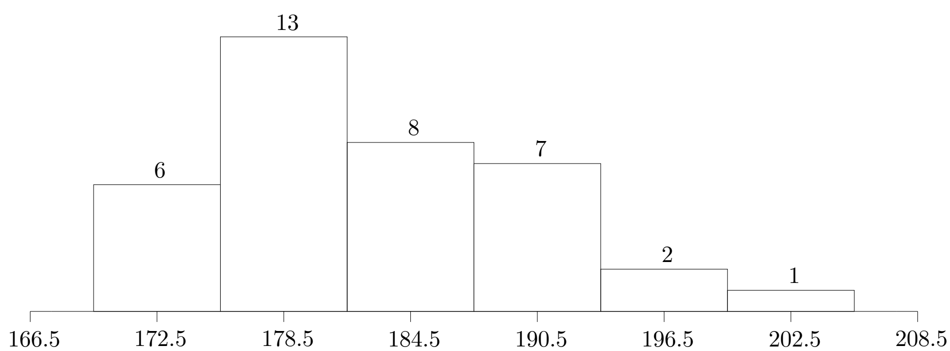



Figur 2: Histogram över mäns längder.

**Alternativ 2:**
Välj 6 st klasser med klassbredd 6 cm. Med 37 mätvärden kommer det ge ett snitt på ca 6 st mätvärden per klass. Detta val ger också snygga klassgränser och ett lagom antal klasser. (Vad som är lagom är ofta en fråga om tycke och smak!)

| Klassgränser | Frekvens $f_i$ | Rel. frek. (i %) | Klassmitt $y_i$ | $f_i y_i$ | $f_i y_i^2$ |
|---|---|---|---|---|---|
| 169.5-175.5 | 6 | 16.2 | 172.5 | 1035 | 178537.5 |
| 175.5-181.5 | 13 | 35.1 | 178.5 | 2320.5 | 414209.25 |
| 181.5-187.5 | 8 | 21.6 | 184.5 | 1476 | 272322 |
| 187.5-193.5 | 7 | 18.9 | 190.5 | 1333.5 | 254031.75 |
| 193.5-199.5 | 2 | 5.4 | 196.5 | 393 | 77224.5 |
| 199.5-205.5 | 1 | 2.7 | 202.5 | 202.5 | 41006.25 |
| | 37 | 99.9 | | 6760.5 | 1237331.25 |

Med klassindelningen ovan får man följande histogram.




Figur 3: Histogram över mäns längder.

**Anm:** Andra indelningar är förstås möjliga. Antalet klasser bör dock vara 6-9 st för att histogrammet ska ha något informationsvärde. Med det smått utvidgade intervallet $(167.5, 207.5)$ får man t.ex. $207.5 - 167.5 = 40$, vilket också bäddar för ett snyggt histogram (8 st klasser med klassbredd 5 cm).

2.c) Varken lådagram eller histogram stödjer hypotesen att en slumpmässigt vald mans längd kan ses som en observation från en normalfördelning. Figurerna är inte symmetriska och histogrammets utseende avviker påtagligt från en normalfördelad täthetsfunktion. Mätvärdena är dock ganska få i sammanhanget så man bör inte förkasta möjligheten att normalfördelningen trots allt kan vara en bra modell utan att först göra en djupare analys.

2.d) Skattat väntevärde och standardavvikelse utifrån explicita mätvärden:

$$\bar{y} = \frac{1}{37}\sum_{i=1}^{37} y_i = \mathbf{182.8}, \qquad s = \sqrt{\frac{1}{37-1}\sum_{i=1}^{37}(y_i - \bar{y})} = \mathbf{7.23}$$

Skattat väntevärde och standardavvikelse utifrån klassindelat material:

**Alternativ 1:**

$$\bar{y} = \frac{6777.5}{37} = \mathbf{183.2}, \qquad s = \sqrt{\frac{1}{37-1}(1243340.25 - \frac{1}{37}6777.5^2)} = \mathbf{7.20}$$

**Alternativ 2:**

$$\bar{y} = \frac{6760.5}{37} = \mathbf{182.7}, \qquad s = \sqrt{\frac{1}{37-1}(1237331.25 - \frac{1}{37}6760.5^2)} = \mathbf{7.60}$$

En enskild slumpmässigt vald mans längd kan ses som en observation från en okänd fördelning med väntevärde $\mu$, och standardavvikelse $\sigma$. Varje enskild mans längd anses vara en observation från samma (okända) fördelning samt vara oberoende av övriga mäns längder. Centrala gränsvärdessatsen (CGS) kan tillämpas ty antalet stokastiska variabler är 37 st (vilket är $\geq 20$). Med $\bar{Y} = \frac{1}{37}(Y_1 + Y_2 + ... + Y_{37})$ får vi:

$$\bar{Y} \tilde{\in} N(\mu, \frac{\sigma^2}{37}) \approx N(\bar{x}, \frac{s^2}{37}) = \mathbf{N(182.8, 1.40)}$$

när vi approximerar $\mu$ och $\sigma$ med våra skattade värden $\bar{y}$ och $s$.