1. Det är skillnaderna man ska se på och göra ett parat t-test. Som vi gjort i kursen använder vi ett parat t-test Till det behöver vi medelvärde och standardavvikelse av skillnaderna 0.8, 0.6, 0.3, -0.1, 1.1, -0.2, 0.3, 0.5, 0.5, 0.3.

   Ett 95 procents konfidensintervall ges av

$$\bar{x} \pm t_{0.025}(n-1)\frac{s}{\sqrt{n}}, \text{ där} \tag{1}$$

$$\begin{aligned} \bar{x} &= 0.410 \\ \bar{s} &= 0.387 \\ n &= 10 \\ t_{0.025}(9) &= 2.26 (\text{ur tabell}) \end{aligned}$$

   Insatt i (1) blir detta (0.13,0.69). Detta ger svaret på (a).

   Svaret på (b) är att vi *kan* dra slutsatsen att dett finns en skillnad mellan materialen, och eftersom skillnaden är positiv betyder detta att B slits mer än A.

   Svaret på (c) är att det dels finns ett beroende mellan skorna, vilket gör antagandet om oberoende stickprov felaktigt, dels (viktigare) kommer variationen i slitage *mellan* pojkar att dominera skillnaden mellan fötterna *inom* varje pojke, som är det vi är intresserade av.

2. Detta är ett klassiskt fall för satsen om total sannolikhet och Bayes' sats. Bilda händelserna:

   - A_1 : Diktsamlingen är usel
   - A_2 : Diktsamlingen är medioker
   - A_3 : Diktsamlingen är strålande och epokgörande.
   - B : Diktsamlingen får strålande recensioner

   Det är ofta bra (om än inte nödvändigt) att rita upp problemet som nedan. (Som vanligt ej skalenligt.)

   (a) Ur texten får vi följande förhållanden:

$$\begin{aligned} \mathbb{P}(B|A_1) = 0.25 \quad &; \quad \mathbb{P}(A_1) = 0.30 \\ \mathbb{P}(B|A_2) = 0.30 \quad &; \quad \mathbb{P}(A_2) = 0.68 \\ \mathbb{P}(B|A_3) = 0.80 \quad &; \quad \mathbb{P}(A_3) = 0.02 \end{aligned}$$

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2) + \mathbb{P}(B|A_3)\mathbb{P}(A_3) \\ &= 0.25 \cdot 0.30 + 0.30 \cdot 0.68 + 0.80 \cdot 0.02 \\ &= 0.295. \end{aligned}$$

   (b) Här behöver vi Bayes' sats:

$$\begin{aligned} \mathbb{P}(A_3|B) &= \frac{\mathbb{P}(B|A_3)}{\mathbb{P}(B)}\mathbb{P}(A_3) \\ &= \frac{0.80}{0.30} \cdot 0.02 \\ &= 0.054. \end{aligned}$$

3. Definiera X och Y som antalet våldsbrott som rapporteras före respektive efter projekten

$$X \sim Bin(200, p_1);\ Y \sim Bin(200, p_2)$$

Vi ska alltså göra ett konfidensintervall för

$$\theta = p_2 - p_1.$$

θ skattas som

$$\begin{aligned}
\theta^* &= p_2^* - p_1^*, \text{ där} \\
p_2^* &= \frac{Y}{200} \text{ och} \\
p_1^* &= \frac{X}{200} \\
\mathbb{V}(\theta^*) &= \frac{p_2(1-p_2)}{200} + \frac{p_1(1-p_1)}{200} \\
d(\theta^*) &= \sqrt{\frac{p_2^*(1-p_2^*)}{200} + \frac{p_1^*(1-p_1^*)}{200}} = 0.03157
\end{aligned}$$

Med $p_1^* = \frac{20}{200} = 0.100$ och $p_2^* = \frac{25}{200} = 0.125$ har vi $200 \cdot 0.1(1-0.1) = 18.000 > 10$ och $200 \cdot 0.125(1-0.125) = 21.875 > 10$, så normalapproximation är tillåten.

Alltså är ett 95 procents approximativt konfidensintervall för $\theta = p_2 - p_1$

$$p_2^* - p_1^* \pm z_{0.025} \cdot d(p_2^* - p_1^*) = 0.125 - 0.1 \pm 1.960 \cdot 0.032,$$

vilket blir $(-0.037, 0.087)$

4. (a) Ur bilden ser man att de katolska byarna är $n_x = 26$ stycken och de ickekatolska $n_y = 16$ I medel ligger de katolska byarna något i stil med 80, så $\bar{y} = 80.55$ är det enda rimliga. På samma sätt ses att $\bar{x} = 66.59$. Av de återstående värdena måste de ickekatolska byarna stå för den mindre standardavvikelsen, eftersom den fördelningen är toppigare och därmed smalare: $s_x = 6.99$. Alltså återstår $s_y = 8.23$

   (b) Vi antar oberoende observationer och samma varians i de två grupperna. Då skattas den gemensamma standardavvikelsen:

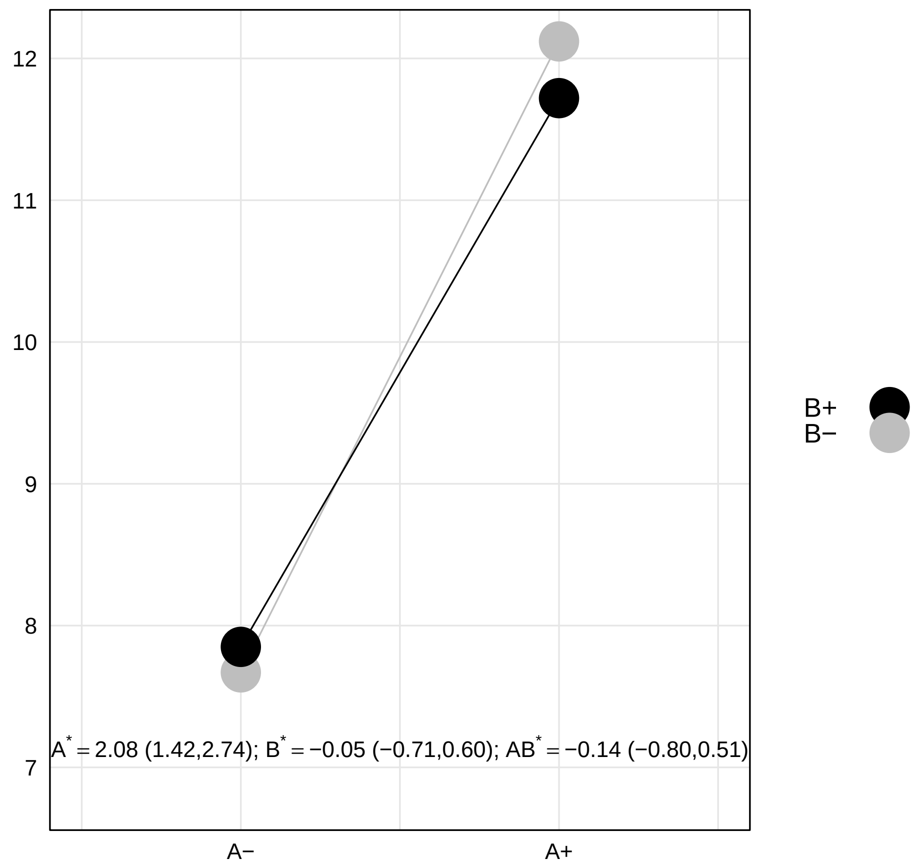
$$s = \sqrt{\frac{(n_x-1)s_x^2 + (n_y-1)s_y^2}{n_x+n_y-2}} = 7.479.$$

Ett 95 procents konfidensintervall ges av

$$\bar{y} - \bar{x} \pm \cdot s \cdot t_{0.025}\underbrace{(n_x+n_y-2)}_{40}\sqrt{\frac{1}{n_x} + \frac{1}{n_y}} =$$

$$80.55 - 66.59 \pm \cdot 7.479 \cdot 2.02\sqrt{\frac{1}{26} + \frac{1}{16}} = (9.2, 18.8).$$

Detta konfidensintervall täcker inte värden 0, så *kan* säga att det skiljer i fertilitet mellan katolska och icke-katolska byar.



5. Effekterna skattas:

$$\begin{aligned} A^* &= \frac{1}{2}\left(\frac{11.72+12.12}{2} - \frac{7.85+7.67}{2}\right) = 2.080 \\ B^* &= \frac{1}{2}\left(\frac{11.72+7.85}{2} - \frac{12.12+7.67}{2}\right) = -0.055 \\ (AB)^* &= \frac{1}{2}\left(\frac{11.72-7.85}{2} - \frac{12.12-7.67}{2}\right) = -0.145 \end{aligned}$$

Den gemensamma standardavvikelsen skattas:

$$s = \sqrt{\frac{s_1^2 + s_2^2 + s_3^2 + s_2^4}{4}} = 1.206$$

Ett 95 procents konfidensintervall för effekterna är

$$\text{Effektskattning} \pm t_{0.025}(12)\frac{s}{\sqrt{4n}} = \text{Effektskattning} \pm 0.657$$

Alltså är de effektskattningar som till beloppet är större än 0.657 signifikanta. De andra inte. Sålunda är huvudeffekten av A starkt signifikant på 0.05-nivån medan huvudeffekten a B inte är det.

Slutsatsen är att kläderna har en tydlig effekt. Att språket skulle ha effekt finns det inga belägg för. Det finns inget signifikant samspel, vilket betyder att effekten av kläder är densamma oavsett språket.

6. Teoriuppgifter

(a)

$$1 = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 1 - \mathbb{P}(A \cap B)$$

Ur detta ser vi att

$$\mathbb{P}(A \cap B) = 0$$

Detta tolkas som att en av Jimmie och Janne alltid går på fest men att de aldrig går på samma fest.

Lägg märke till att A och B *inte* är oberoende.

(b)

$$\begin{aligned} -8X \sim N(-8, (-8)^2 \cdot 2) &= N(-8, 128) \\ X - 2Y \sim N(-1, (1 + 2^2) \cdot 2) &= N(-1, 10) \end{aligned}$$

(c) Lägg först märke till att

$$\mathbb{P}(X = 0) = 1 - \mathbb{P}(X = 1 \cdot 10^6) + \mathbb{P}(X = 5 \cdot 10^5) + \mathbb{P}(X = 2.5 \cdot 10^5) = 0.91.$$

$$\mathbb{E}(X) = 0 \cdot \mathbb{P}(X = 0) + 1 \cdot 10^6 \cdot \mathbb{P}(X = 1 \cdot 10^6) + 5 \cdot 10^5 \cdot \mathbb{P}(X = 5 \cdot 10^5) + 2.5 \cdot 10^5 \cdot \mathbb{P}(X = 2.5 \cdot 10^5) =$$

$$0 \cdot 0.91 + 1000000 \cdot 0.01 + 500000 \cdot 0.03 + 250000 \cdot 0.05 = 37500$$

Det går alltså inte att ta mindre än 37500 för försäkringen om den ska löna sig.

(d)
- $\mathbb{E}(X) = \int_1^{91} x f(x) dx = 43.6$
- Medianen är lösningen till $\int_1^x f(t) dt = 0.5$, alltså 52.33.
- För en kontinumerlig fördelning gäller alltid $\mathbb{P}(X = x) = 0$, även för $x = 35$.
- $\mathbb{P}(30 \leq X < 40) = \int_{30}^{40} f(x) dx = 0.0770$.

(e) I den första planen förekommer varje metod precis två gånger för varje barn, vilket gör det till att föredra i jämförelse med att fullständigt randomisera, eftersom variationen mellan barnen i så fall riskerar att överskugga den intressanta källan till variation, den mellan metoderna.