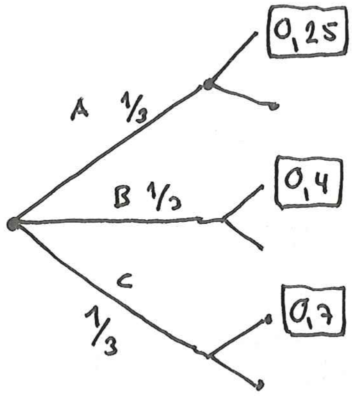
# TENTA 11-01-11

1. a) Vi kan använda träddiagram:




Sannolikheten för punktering blir:

$$\underline{P(P)} = \frac{1}{3}\cdot 0{,}25 + \frac{1}{3}\cdot 0{,}4 + \frac{1}{3}\cdot 0{,}7 =$$

$$= \frac{1}{3}(0{,}25 + 0{,}4 + 0{,}7) = \frac{1{,}35}{3} = \boxed{0{,}45}$$

Alternativt kan vi använda <u>Lagen om total sannolikhet</u>:

Vi har $P(P|A)$ = (sannolikheten för punktering om vi väljer väg A) = 0,25 och motsvarande för de andra, så Lagen ger:

$$P(P) = P(P|A)\cdot P(A) + P(P|B)\cdot P(B) + P(P|C)\cdot P(C) =$$

$$= 0{,}25\cdot\frac{1}{3} + 0{,}4\cdot\frac{1}{3} + 0{,}7\cdot\frac{1}{3} = \underline{0{,}45}.$$

b) Vi vill beräkna $P(C|P)$ och använder <u>Bayes sats</u> som ger:

$$P(C|P) = \frac{P(C)\cdot P(P|C)}{P(A)P(P|A) + P(B)\cdot P(P|B) + P(C)P(P|C)} = \frac{\frac{1}{3}\cdot 0{,}7}{0{,}45} =$$

$$= \frac{0{,}7}{1{,}35} = 0{,}5185 \approx \boxed{0{,}52}$$

(2) Låt $X$ vara slumpvariabeln som beskriver antalet av de 99 som får punktering. Eftersom det handlar om antal och två (oberoende) alternativ: punktering eller inte så antar vi att $X$ är binomialfördelad:

$X \sim Bin(99, p) = \boxed{Bin(99, 0{,}45)}$ enligt 1. Denna åter finns inte i tabellen, men eftersom $np(1-p) = 99 \cdot 0{,}45 \cdot 0{,}55 = 24{,}5$ är så stort (större än 5, se approximations-schemat i formelsamlingen) så kan vi approximera med en normalfördelning som har samma väntevärde och varians som $Bin(99, 0{,}45)$.

För $Bin(n,p)$ gäller Väntevärde: $n \cdot p$ Varians: $np(1-p)$ så i vårt fall blir talen: $99 \cdot 0{,}45$ resp. $99 \cdot 0{,}45 \cdot 0{,}55$ eller $44{,}55$ och $24{,}5$ och vi har:

$$Bin(99, 0{,}45) \approx N(44{,}55,\ 24{,}5) = \boxed{N(44{,}55,\ 4{,}95^2)}$$

Sannolikheten att fler än hälften får punktering blir:

$P(X \geq 50) = 1 - P(X \leq 49)$ och enligt formel (sid 105 i boken; eller föreläsning 8) med $\Phi$ fördelningsfunktionen för $N(0,1)$ är detta:

$$1 - \Phi\left(\frac{49 - 44{,}55}{4{,}95}\right) = 1 - \Phi\left(\frac{4{,}45}{4{,}95}\right) = 1 - \Phi(0{,}8990) =$$

$$= \left(\begin{array}{c}\text{Tabell och}\\ \text{interpolation}\end{array}\right) 1 - 0{,}8156 = 0{,}1844 \approx \boxed{18\%}$$

③ Låt $X_i \sim N(30, 8^2)$ vara tiden för att lösa sannolikhets-problem och $Y_i \sim N(40, 5^2)$ tiden för statistik problem $(i = 1,2,3,4)$

Då gäller $(X_1 + X_2 + X_3 + X_4) \sim N(4 \cdot 30, 4 \cdot 8^2)$ och

$$Y = (Y_1 + Y_2 + Y_3 + Y_4) \sim N(4 \cdot 40, 4 \cdot 5^2) \quad \text{(sid 153)}$$

och den totala tiden beskrivs av slumpvariabeln

$$Z = X + Y \sim N(4 \cdot 30 + 4 \cdot 40,\ 4 \cdot 8^2 + 4 \cdot 5^2) = \boxed{N(280, 18{,}87^2)}$$

Sannolikheten att Doris skall lösa problemen inom 5 timmar (eller 300 minuter) blir då:

$$P(Z \leq 300) = (\text{samma formel som i ②}) = \Phi\left(\frac{300-280}{18{,}87}\right) =$$

$$= \Phi\left(\frac{20}{18{,}87}\right) = \Phi(1{,}059) = (\text{Tabell}) = \boxed{0{,}855} \approx 86\,\%$$

---

④ För $Po(\lambda)$ vet vi att väntevärde är $\lambda$ som också är variansen. I vårt fall är $\lambda = 300$.

För att restaurangen skall gå med vinst måste $k$ vara större än det värde då det går precis jämnt upp:

$16k + 1000 = 20k$ ger $\underline{k = 250}$.

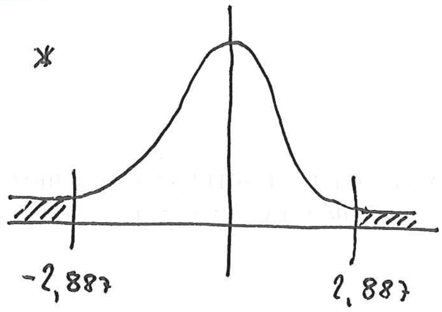
Vidare gäller $Po(\lambda) \approx N(\lambda, \lambda)$ om $\lambda > 15$ (se schemat) så i vårt fall är det OK.

Det betyder att om $X$ beskriver antalet sålda hamburgare så gäller $X \sim N(300, 17{,}32^2)$ och

$$P(\bar{X} > 250) = 1 - P(\bar{X} \leq 250) = (\text{som ovan!}) =$$

$$= 1 - \Phi\left(\frac{250-300}{17,32}\right) = 1 - \Phi\left(-\frac{50}{17,32}\right) = 1 - \Phi(-2,887)$$

$$\overset{*}{=} 1 - (1 - \Phi(2,887)) = \Phi(2,887) = \boxed{0,9981} \approx 99,8\ \%$$




(5.) Det handlar här om konstruktion av konfidensintervall i fallet två oberoende stickprov med samma obekanta varians. Referensvariabeln är då (se tabell)

$$R_\mu = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{där} \quad S_p^2 = \frac{S_{xx} + S_{yy}}{n_1 + n_2 - 2}$$ och vi måste

räkna ut alla storheter (utom $\lambda_1 - \lambda_2$) med hjälp av de givna siffrorna:

$$\bar{X} = \frac{690 + 630 + 672 + 657}{4} = \underline{662,25}$$

$$\bar{Y} = \frac{598 + 572 + 634 + 658 + 607}{5} = \underline{613,8}$$

$$S_{xx} = \sum_{i=1}^{4} (x_i - \bar{x})^2 = 27,75^2 + 32,25^2 + 9,75^2 + 5,25^2 = \underline{1932,25}$$

$$S_{yy} = \sum_{i=1}^{5} (y_i - \bar{y})^2 = \ldots = \underline{4.405.16}$$

Enl. ovan är då $S_p^2 = \dfrac{1932{,}25 + 4405{,}16}{4+5-2} =$

$= \dfrac{6337{,}05}{7} = 905{,}29$ och $\underline{S_p = 30{,}09}$ så

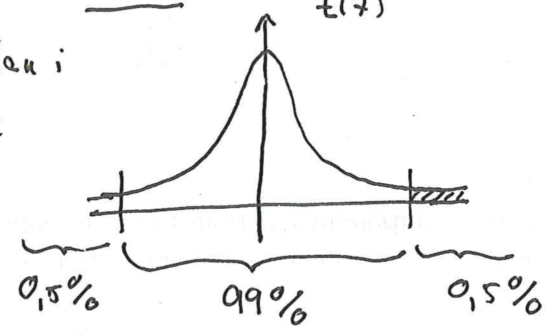
$S_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}} = S_p\sqrt{\dfrac{1}{4}+\dfrac{1}{5}} = S_p \cdot \sqrt{\dfrac{9}{20}} =$

$= 30{,}09 \cdot \dfrac{3}{\sqrt{20}} = \dfrac{90{,}29}{\sqrt{20}} = \underline{20{,}19}$

Sammanfattningsvis blir referensvariabeln:

$$R_\mu = \frac{(662{,}25 - 613{,}8) - (\mu_1 - \mu_2)}{20{,}19} = \boxed{\frac{48{,}45 - (\mu_1-\mu_2)}{20{,}19}}$$

Vi vet att denna är t-fördelad, $\sim \underline{t(7)}$ och eftersom vi skall ha konfidensgrad $\underline{99\%}$ uppsöker vi $0{,}5\% = 0{,}005$ kvantilen i tabellen (för $f=7$) och får att den är $\underline{3{,}50}$. Detta medför:




$$P\left(-3{,}5 < \frac{48{,}45-(\mu_1-\mu_2)}{20{,}19} < 3{,}5\right) = 0{,}99 \iff$$

$$P\left(-3{,}5 \cdot 20{,}19 < 48{,}45 - (\mu_1-\mu_2) < 3{,}5 \cdot 20{,}19\right) = 0{,}99$$

$$P\left(-70{,}66 < 48{,}45 - (\mu_1-\mu_2) < 70{,}66\right) = 0{,}99$$

Så vi har:

$$P(-70{,}66 < (\mu_1-\mu_2) - 48{,}45 < +70{,}66) = 0{,}99$$ eller

$$P(48{,}45 - 70{,}66 < \mu_1-\mu_2 < 48{,}45+70{,}66) = 0{,}99$$

och konfidensintervallet kan skrivas:

$\boxed{(48{,}45 \pm 70{,}66)}$ eller $\boxed{(-22{,}2\ ,\ 119{,}11)}$

---

(6) Om $X$ är antalet hårdkokta ägg så har vi: (antal och två möjligheter!) $X \sim Bin(60, p)$, $p$ okänd.

Om vi skattar $p$ med $p^* = 35/60$ blir $n \cdot p^*(1-p^*) =$

$= 60 \cdot \frac{35}{60} \cdot \frac{25}{60} = 14{,}59$ som är större än 5 så vi

kan approximera med normalfördelning. Vi har alltså

$Bin(60, p^*) \approx N(60 \cdot p^*, 60 \cdot p^*(1-p^*))$ och vi har

$$X \sim N(np^*, np^*(1-p^*)) \Rightarrow P = \frac{X}{n} \sim N\Big(p^*, \overbrace{\frac{p^*(1-p^*)}{n}}^{\delta^2}\Big)$$

referensvariabel

$$R_p = \frac{p^* - p}{\delta} = \frac{7/12 - p}{\sqrt{\frac{\frac{35}{60}\cdot\frac{25}{60}}{60}}} = \frac{7/12 - p}{0{,}063}$$

Så vi har

$$P\left(-1{,}96 < \frac{7/12 - p}{0{,}063} < 1{,}96\right) = 0{,}95$$

dvs. $P\left(\frac{7}{12} - 0{,}063 \cdot 1{,}96 < p < \frac{7}{12} + 0{,}063 \cdot 1{,}96\right) = 0{,}95$

eller $P(0{,}583 - 0{,}123 < p < 0{,}583 + 0{,}123) = 0{,}95$

Så $I_p = \boxed{(0{,}58 \pm 0{,}12)} = \boxed{(0{,}46\,,\, 0{,}70)}$

b) Bredden av intervallet är (för 95% konfidens)

$$l = 2 \cdot \lambda_{0{,}025} \cdot \sqrt{\frac{p^*(1-p^*)}{n}}$$

och om vi vill ha $l$ mindre än eller lika med $0{,}1$ så ser vi att

$$2 \cdot \lambda_{0{,}025} \cdot \sqrt{\frac{p^*(1-p^*)}{n}} < 0{,}1 \iff$$

$$\sqrt{\frac{p^*(1-p^*)}{n}} < \frac{0{,}1}{2\lambda_{0{,}025}} \iff \frac{p^*(1-p^*)}{n} < \left(\frac{0{,}1}{2\lambda_{0{,}025}}\right)^2$$

$$\iff n > \frac{p^*(1-p^*) \cdot 4 \cdot \lambda_{0{,}025}^2}{0{,}01} =$$

$$= 400\, p^*(1-p^*) \cdot \lambda_{0{,}025}^2 = 400 \cdot \frac{7}{12} \cdot \frac{5}{12} \cdot 1{,}96^2$$

$$= 373{,}49$$

Svar: Man måste koka $\boxed{374 \text{ ägg}}$.

(7). I detta problem handlar det om Stickprov i par. (sid. 342 i boken).

Låt $X_i$ vara den slumpvariabel som beskriver skillnaden i sträckor ("efter"-"före") i km.

Vi antar att $X_i \sim N(\Delta, \sigma^2)$, ($\sigma^2$ okänd).

och att de är oberoende. Vi vill hitta ett 95% konfidensintervall och se om $\Delta = 0$ ligger i detta eller ej!

Genom att bilda skillnader för vi stickprovet.

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$. |
|---|---|---|---|---|---|---|---|---|
| 0,2 | 0,6 | -0,5 | 2,1 | 1,7 | 1,3 | -0,4 | 3,5 | 1,4 |

$n = 9$

och räknar ur detta ut,

$$\begin{cases} \bar{x} = \frac{1}{9}\sum_{i=1}^{n} x_i = \underline{1,1} \\ s^2 = \frac{1}{8}\sum_{i=1}^{9}(x_i - \bar{x})^2 = 1,64 \Rightarrow s = \sqrt{1,64} = \underline{1,28} \end{cases}$$

Vi kan då ställa upp en referensvariabel (se formelsamligen 2.2.1. ett stickprov, $\sigma$ okänd.)

$$R_{\bar{X}} = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$$ dvs. i vårt fall:

$$R_{\bar{X}} = \frac{1{,}1-\Delta}{1{,}28/\sqrt{9}} \sim t(8). \qquad (1)$$

Tabellen för t-fördelningen ger ($f=8$, $\alpha = 0{,}025$)

$t_{0{,}025}(8) = 2{,}31$ och (1) ger

$$P\left(-2{,}31 < \frac{1{,}1-\Delta}{1{,}28/3} < 2{,}31\right) = 0{,}95 \iff$$

$$P\left(-2{,}31 \cdot \frac{1{,}28}{3} < 1{,}1-\Delta < \underbrace{2{,}31 \cdot \frac{1{,}28}{3}}\right) = 0{,}95 \iff$$

$$P\left(1{,}1-0{,}9856 < \Delta < 0{,}9856\right) = 0{,}95$$ så

intervallet är $\boxed{(0{,}114 < \Delta < 2{,}08)}$ och eftersom

$\Delta = 0$ INTE ligger i intervallet är det, med 95% konfidens, en signifikant skillnad. □

(8). a) Vi låter förvaringstiden $x_i$ (i dagar) vara den bestämda; icke-slumpvisa variabeln och $y_i$ den slumpvariabel och modellen är:

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad \text{med } \varepsilon_i \sim N(0, \sigma^2)$$

För att punktskatta $\alpha$ och $\beta$ beräknar vi (se formelsamlingen): $\bar{x} = 4{,}167$ $\quad \bar{y} = 75{,}167$

$$S_{xx} = \sum_{i=1}^{6} (x_i - \bar{x})^2 = \left( \sum_{i=1}^{6} x_i^2 - 6\bar{x}^2 \,! \right)^{*} = 74{,}83$$

$$S_{xy} = \sum_{i=1}^{6} (x_i - \bar{x})(y_i - \bar{y}) = \left( \sum_{i=1}^{n} x_i y_i - 6\bar{x}\cdot\bar{y}^{**} \right) = -456{,}17$$

$$S_{yy} = \sum_{i=1}^{6} (y_i - \bar{y})^2 \left( = \sum_{i=1}^{n} y_i^2 - 6\bar{y}^2 \right)^{*} = 3064{,}8.$$

Då får vi: $\beta^* = \dfrac{S_{xy}}{S_{xx}} = \boxed{-6{,}036}$ och

$$\alpha^* = \bar{y} - \beta^* \bar{x} = 75{,}17 + 6{,}036 \cdot 4{,}17 = \boxed{100{,}56}$$

b) Förklaringsgraden är $R^2 = \dfrac{S_{xy}^2}{S_{xx} S_{yy}} = \left( = 1 - \dfrac{Q_0}{S_{yy}} \right) \Rightarrow$

$\boxed{R = 0{,}907}$ så vi har en bra förklaringsgrad!

c) Den förväntade minskningen av C-vitaminhalten per dag är precis <u>linjens lutning</u>, $\beta$.

Vi vet att $\beta \sim N(\beta, \sigma^2/\sqrt{S_{xx}})$ (se formelsaml.)

och $\sigma^2$ skattas med $s_r^2 = \dfrac{Q_0}{n-2} = \dfrac{S_{yy} - S_{xy}^2/S_{xx}}{6-2}$

---

* Se formel under 2.1 Allmänt i Formelsamlingen!

** Se formel under 2.3 Regressionsanalys.

Så $S_r^2 = 71{,}037 \Rightarrow S_r = 8{,}43$.

Vi får referensvariabeln för $\beta$ (Formelsamling).

$$R_\beta = \frac{\beta^* - \beta}{S_r/\sqrt{S_{xx}}} = \frac{-6{,}036 - \beta}{8{,}43/\sqrt{79{,}83}} \sim t_{0{,}025}(4)$$

Vilket medför: (tabell för $t_{0{,}025}(4)$)

$$P\left(-2{,}7764 < \frac{-6{,}036 - \beta}{0{,}975} < 2{,}7764\right) = 0{,}95$$

så

$$P(-6{,}036 - 2{,}7764 < \beta < -6{,}036 + 2{,}7764) = 0{,}95$$

eller $\boxed{I_\beta = (-8{,}8\,,\ 3{,}3)}$