Teoretická odvození pro zkoušku z předmětu KPST

1. Nezávislost náhodných jevů $P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A)P(B)$

$$P(A) = P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A)P(B)$$

2. Bayesova věta $P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|\overline{A})P(\overline{A})}$

$$\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \text{ a zároveň } P(B|A) = \frac{P(A \cap B)}{P(A)} \Rightarrow \\
P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\
P(B) &= P(B|A)P(A) + P(B|\overline{A})P(\overline{A}) \Rightarrow \\
P(A|B) &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})}
\end{aligned}$$

s využitím věty o úplné pravděpodobnosti

3. Vlastnosti střední hodnoty $\mathsf{E}(a + bX) = a + b\mathsf{E}X$

$$\begin{aligned}
\mathsf{E}(a + bX) &= \sum_{i=1}^{k}(a + bX_i)p_i = \sum_{i=1}^{k}(ap_i + bX_ip_i) = \sum_{i=1}^{k} ap_i + \sum_{i=1}^{k} bX_ip_i = \\
&= a\sum_{i=1}^{k} p_i + b\sum_{i=1}^{k} X_ip_i = a + b(E)X
\end{aligned}$$

4. Vlastnosti rozptylu $\mathsf{Var}(a + bX) = b^2\mathsf{Var}X$

$$\begin{aligned}
\mathsf{Var}(a + bX) &= \sum_{i=1}^{2}(a + bX_i - (E)(a + bX))^2 p_i = \sum_{i=1}^{2}(a + bX_i - (a + b(E)X))^2 p_i = \\
&= \sum_{i=1}^{2}(b(X_i - (E)X))^2 p_i = b^2\sum_{i=1}^{2}(X_i - (E)X)^2 p_i = b^2\mathsf{Var}X
\end{aligned}$$

5. Dva tvary rozptylu $\mathsf{Var}X = \mathsf{E}(X_i - \mathsf{E}X)^2 = \mathsf{E}X^2 - (\mathsf{E}X)^2$

$$\begin{aligned}
\sum_{i=1}^{k}(X_i - \mathsf{E}X)^2 p_i &= \sum_{i=1}^{k}(X_i^2 - 2X_i\mathsf{E}X + (\mathsf{E}X)^2)p_i = \sum_{i=1}^{k} X_i^2 p_i - 2\mathsf{E}X\sum_{i=1}^{k} X_ip_i + (\mathsf{E}X)^2\sum_{i=1}^{k} p_i = \\
&= \mathsf{E}X^2 - 2(\mathsf{E}X)^2 + (\mathsf{E}X)^2 = \mathsf{E}X^2 - (\mathsf{E}X)^2
\end{aligned}$$

6. Vlastnosti výběrové šikmosti $\mathsf{Skew}X = \frac{1}{n}\sum_{i=1}^{n} Z_i^3$, kde $z_i = \frac{X_i - \overline{X}}{\mathsf{sd}X}$ počítá se přes Z-skóry a pro ně platí

$$\begin{aligned}
ZX_i &= \frac{X_i - \overline{X}}{\mathsf{sd}X} \\
\text{pro } Y_i &= a + bX \text{ dostávám } ZY_i = \frac{Y_i - \overline{Y}}{\mathsf{sd}Y} \\
ZY_i &= \frac{a + bX_i - \overline{a + bX}}{\mathsf{sd}(a + bX)} = \frac{a + bX_i - a + b\overline{X}}{b\mathsf{sd}X} = \frac{b(X_i - \overline{X})}{b\mathsf{sd}X} = ZX_i
\end{aligned}$$

Z-skóry tedy vychází stejně pro proměnnou X i Y a tedy pro ně musí vycházet stejně i šikmost

7. Střední hodnota binomického rozdělení $Bi(n,p)$ s parametry $n = 3, p = 1/2$ je $\mathsf{E}X = np$

$$\begin{aligned}
\mathsf{E}X &= \sum_{k=0}^{n} kP(X = k) = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\
\mathsf{E}X &= 0\binom{3}{0}\left(\frac{1}{2}\right)^0\left(1 - \left(\frac{1}{2}\right)\right)^{3-0} + 1\binom{3}{1}\left(\frac{1}{2}\right)^1\left(1 - \left(\frac{1}{2}\right)\right)^{3-1} + \\
&+ 2\binom{3}{2}\left(\frac{1}{2}\right)^2\left(1 - \left(\frac{1}{2}\right)\right)^{3-2} + 3\binom{3}{3}\left(\frac{1}{2}\right)^3\left(1 - \left(\frac{1}{2}\right)\right)^{3-3} = \\
&= 0 + 3\left(\frac{1}{2}\right)^3 + 2 \times 3\left(\frac{1}{2}\right)^3 + 3\left(\frac{1}{2}\right)^3 = \frac{3+6+3}{8} = \frac{3}{2} = np
\end{aligned}$$

8. Interval spolehlivosti pro průměr vychází z $\overline{X} \sim N(\mu, \sigma^2/n)$

$$\begin{aligned}
Z &= \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim N(0,1) \\
\text{Chceme} & \quad P(\mu \in CI) = P(LL \leq \mu \leq UL) = 1 - \alpha \\
\text{Víme} & \quad P(q(\alpha/2) \leq Z \leq q(1-\alpha/2)) = 1 - \alpha;\ \text{a } q(\alpha/2) = -q(1-\alpha/2) \\
P(-q(1-\alpha/2) &\leq \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \leq q(1-\alpha/2)) = \\
&= P\left(-q(1-\alpha/2)\sigma/\sqrt{n} - \overline{X} \leq -\mu \leq q(1-\alpha/2)\sigma/\sqrt{n} - \overline{X}\right) = \\
&= P\left(-q(1-\alpha/2)\sigma/\sqrt{n} - \overline{X} \leq -\mu \leq q(1-\alpha/2)\sigma/\sqrt{n} - \overline{X}\right) = \\
&= P\left(\overline{X} + q(1-\alpha/2)\sigma/\sqrt{n} \geq \overline{X} - \mu \geq q(1-\alpha/2)\sigma/\sqrt{n}\right) \\
LL &= \overline{X} - \mu \geq q(1-\alpha/2)\sigma/\sqrt{n} \\
UL &= \overline{X} + \mu \geq q(1-\alpha/2)\sigma/\sqrt{n}
\end{aligned}$$

kde $q(1-\alpha/2)$ je kvantil standardního normálního rozdělení

9. Spojitost mezi jednovýběrovým t-testem a intervalem spolehlivosti pro průměr: testová statistika pro jednovýběrový t-test je $T = \frac{\overline{X}-\mu_0}{\mathsf{sd}X}\sqrt{n}$, tato statistika má za platnosti nulové hypotézy t-rozdělení o $n-1$ stupních volnosti, nulovou hypotézu (při oboustranné alternativě) zamítáme, pokud $|T| \geq qt_{n-1}(1-\alpha/2)$, kde $qt_{n-1}(1-\alpha/2)$ je kvantil t-rozdělení o $n-1$ stupních volnosti, meze intervalu spolehlivosti pro průměr při neznámém rozptylu jsou

$$\begin{aligned} LL &= \overline{X} - \mu \geq qt_{n-1}(1-\alpha/2)\mathsf{sd}X/\sqrt{n} \\ UL &= \overline{X} + \mu \geq qt_{n-1}(1-\alpha/2)\mathsf{sd}X/\sqrt{n} \end{aligned}$$

z testové statistiky lze odvodit, že nulovou hypotézu zamítáme, když

$$\begin{aligned} T &= \frac{\overline{X}-\mu_0}{\mathsf{sd}X}\sqrt{n} \geq qt_{n-1}(1-\alpha/2) \Rightarrow \overline{X} - \mu_0 \geq qt_{n-1}(1-\alpha/2)\frac{\mathsf{sd}X}{\sqrt{n}} \Rightarrow \\ \mu_0 &\leq \overline{X} - qt_{n-1}(1-\alpha/2)\frac{\mathsf{sd}X}{\sqrt{n}} \\ &\text{nebo} \\ -T &= \frac{\mu_0-\overline{X}}{\mathsf{sd}X}\sqrt{n} \geq qt_{n-1}(1-\alpha/2) \Rightarrow \mu_0 - \overline{X} \geq qt_{n-1}(1-\alpha/2)\frac{\mathsf{sd}X}{\sqrt{n}} \Rightarrow \\ \mu_0 &\geq \overline{X} + qt_{n-1}(1-\alpha/2)\frac{\mathsf{sd}X}{\sqrt{n}} \end{aligned}$$

tedy když μ_0 je mimo meze intervalu spolehlivosti

10. Rozklad součtu čtverců u analýzy rozptylu, tj. rozklad celkové variability na variabilitu

vysvětlenou a nevysvětlenou

$$
\begin{aligned}
SST &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_{..})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_{i.}+\overline{X}_{i.}\overline{X}_{..})^2 = \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_{i.})^2 + 2\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_{i.})(\overline{X}_{i.}-\overline{X}_{..}) + \sum_{i=1}^{k} n_i(\overline{X}_{i.}-\overline{X}_{..})^2 = \\
&= \sum_{i=1}^{k} n_i(\overline{X}_{i.}-\overline{X}_{..})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_{i.})^2 = \\
&= SSA + SSe \\
& \quad \text{jelikož platí, že} \\
\sum_{i=1}^{k}\sum_{j=1}^{n_i} & (X_{ij}-\overline{X}_{i.})(\overline{X}_{i.}-\overline{X}_{..}) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}\overline{X}_{i.} - X_{ij}\overline{X}_{..} - \overline{X}_{i.}^2 + \overline{X}_{i.}\overline{X}_{..}) = \\
&= \sum_{i=1}^{k}\overline{X}_{i.}\sum_{j=1}^{n_i}X_{ij} - \overline{X}_{..}\sum_{i=1}^{k}\sum_{j=1}^{n_i}X_{ij} - \sum_{i=1}^{k}\overline{X}_{i.}^2\sum_{j=1}^{n_i}1 + \overline{X}_{..}\sum_{i=1}^{k}\overline{X}_{i.}\sum_{j=1}^{n_i}1 = \\
&= \sum_{i=1}^{k}\overline{X}_{i.}n_i\overline{X}_{i.} - \overline{X}_{..}n\overline{X}_{..} - \sum_{i=1}^{k}n_i\overline{X}_{i.}^2 + \overline{X}_{..}\sum_{i=1}^{k}n_i\overline{X}_{i.} = \\
&= \sum_{i=1}^{k}n_i\overline{X}_{i.}^2 - n\overline{X}_{..}^2 - \sum_{i=1}^{k}n_i\overline{X}_{i.}^2 + n\overline{X}_{..}^2 = 0
\end{aligned}
$$

11. odvození regresních koeficientů v jednoduché lineární regresi
    metoda nejmenších čtverců, hledáme min $\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2$

$$
\begin{aligned}
\min_{b0,b1}\sum_{i=1}^{n} & (Y_i-(b_0+b_1X_i))^2 \\
& \text{derivace podle } b_0 \\
\frac{\partial}{\partial b_0}\sum_{i=1}^{n} & (Y_i-(b_0+b_1X_i))^2 = \sum_{i=1}^{n}-2(Y_i-b_0-b_1X_i) \\
& \text{derivace podle } b_1 \\
\frac{\partial}{\partial b_1}\sum_{i=1}^{n} & (Y_i-(b_0+b_1X_i))^2 = \sum_{i=1}^{n}-2(Y_i-b_0-b_1X_i)X_i
\end{aligned}
$$

položíme rovno nule

$$
\begin{aligned}
\sum_{i=1}^{n} -2(Y_i - b_0 - b_1 X_i)) &= 0 \\
2n\overline{Y} - 2nb_0 - 2b1 \sum_{i=1}^{n} X_i &= 0 \\
b_0 &= \frac{2n\overline{Y} - 2nb_1\overline{X}}{2n} \\
b_0 &= \overline{Y} - b_1\overline{X} \\
\sum_{i=1}^{n} -2(Y_i - b_0 - b_1 X_i)X_i &= 0 \\
-2\sum_{i=1}^{n} Y_i X_i + 2b_0 \sum_{i=1}^{n} X_i + 2b_1 \sum_{i=1}^{n} X_i^2 &= 0 \\
\text{dosadím za } b_0 & \\
-2\sum_{i=1}^{n} Y_i X_i + 2(\overline{Y} - b_1\overline{X}) \sum_{i=1}^{n} X_i + 2b_1 \sum_{i=1}^{n} X_i^2 &= 0 \\
-2\sum_{i=1}^{n} Y_i X_i + 2\overline{Y} \sum_{i=1}^{n} X_i - 2b_1\overline{X} \sum_{i=1}^{n} X_i + 2b_1 \sum_{i=1}^{n} X_i^2 &= 0 \\
\sum_{i=1}^{n} X_i Y_i - n\overline{XY} &= b_1(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2) \\
b_1 &= \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}
\end{aligned}
$$

12. odvoyení očekávané četnosti v χ^2-testu nezávislosti z definice nezávislosti $P(A \cap B) = P(A)P(B)$

$$
\begin{aligned}
P(A_1) &= \frac{\sum_{j=1}^{c} X_{1j}}{n} = \frac{n_{1.}}{n} \\
P(B_1) &= \frac{\sum_{i=1}^{r} X_{i1}}{n} = \frac{n_{.1}}{n} \\
& \text{očkávaná četnost 1. políčka tabulky (se souřadnicemi)1, 1} \\
& \frac{n_{1.}}{n}\frac{n_{.1}}{n} n = \frac{n_{1.} n_{1.}}{n}
\end{aligned}
$$